Abstract

This is the typset sources for the expl3 programming environment; see the matching interface3 PDF for the API reference manual. The expl3 modules set up a naming scheme for \LaTeX{} commands, which allow the \LaTeX{} programmer to systematically name functions and variables, and specify the argument types of functions.

The \TeX{} and \epsilon-\TeX{} primitives are all given a new name according to these conventions. However, in the main direct use of the primitives is not required or encouraged: the expl3 modules define an independent low-level \LaTeX{}3 programming language.

The expl3 modules are designed to be loaded on top of \LaTeX{}2e. With an up-to-date \LaTeX{}2e kernel, this material is loaded as part of the format. The fundamental programming code can also be loaded with other \TeX{} formats, subject to restrictions on the full range of functionality.

*E-mail: latex-team@latex-project.org
Contents

I Introduction 1

1 Introduction to expl3 and this document 2
  1.1 Naming functions and variables 2
    1.1.1 Scratch variables 5
    1.1.2 Terminological inexactitude 5
  1.2 Documentation conventions 5
  1.3 Formal language conventions which apply generally 7
  1.4 \TeX concepts not supported by \LaTeXeX 7

II Bootstrapping 8

2 The l3bootstrap module: Bootstrap code 9
  2.1 Using the \LaTeXeX modules 9

3 The l3names module: Namespace for primitives 11
  3.1 Setting up the \LaTeXeX programming language 11

III Programming Flow 12

4 The l3basics module: Basic definitions 13
  4.1 No operation functions 13
  4.2 Grouping material 13
  4.3 Control sequences and functions 14
    4.3.1 Defining functions 14
    4.3.2 Defining new functions using parameter text 15
    4.3.3 Defining new functions using the signature 17
    4.3.4 Copying control sequences 19
    4.3.5 Deleting control sequences 20
    4.3.6 Showing control sequences 20
    4.3.7 Converting to and from control sequences 21
  4.4 Analysing control sequences 22
  4.5 Using or removing tokens and arguments 23
    4.5.1 Selecting tokens from delimited arguments 26
  4.6 Predicates and conditionals 27
    4.6.1 Tests on control sequences 28
    4.6.2 Primitive conditionals 28
  4.7 Starting a paragraph 30
  4.8 Debugging support 30
10 The \texttt{l3sys} module: System/runtime functions \hfill 74
10.1 The name of the job \hfill 74
10.2 Date and time \hfill 74
10.3 Engine \hfill 75
10.4 Output format \hfill 76
10.5 Platform \hfill 76
10.6 Random numbers \hfill 76
10.7 Access to the shell \hfill 77
10.8 System queries \hfill 78
10.9 Loading configuration data \hfill 79
10.9.1 Final settings \hfill 79

11 The \texttt{l3msg} module: Messages \hfill 80
11.1 Creating new messages \hfill 80
11.2 Customizable information for message modules \hfill 81
11.3 Contextual information for messages \hfill 82
11.4 Issuing messages \hfill 83
11.4.1 Messages for showing material \hfill 87
11.4.2 Expandable error messages \hfill 87
11.5 Redirecting messages \hfill 88

12 The \texttt{l3file} module: File and I/O operations \hfill 90
12.1 Input–output stream management \hfill 90
12.1.1 Reading from files \hfill 92
12.1.2 Reading from the terminal \hfill 96
12.1.3 Writing to files \hfill 96
12.1.4 Wrapping lines in output \hfill 98
12.1.5 Constant input–output streams, and variables \hfill 99
12.1.6 Primitive conditionals \hfill 99
12.2 File operations \hfill 99
12.2.1 Basic file operations \hfill 99
12.2.2 Information about files and file contents \hfill 100
12.2.3 Accessing file contents \hfill 103

13 The \texttt{l3luatex} module: Lua\TeX{}-specific functions \hfill 105
13.1 Breaking out to \LaTeX{} \hfill 105
13.2 Lua interfaces \hfill 106

14 The \texttt{l3legacy} module: Interfaces to legacy concepts \hfill 108

IV Data types \hfill 109
15 The l3tl module: Token lists
- 15.1 Creating and initialising token list variables
- 15.2 Adding data to token list variables
- 15.3 Token list conditionals
  - 15.3.1 Testing the first token
- 15.4 Working with token lists as a whole
  - 15.4.1 Using token lists
  - 15.4.2 Counting and reversing token lists
  - 15.4.3 Viewing token lists
- 15.5 Manipulating items in token lists
  - 15.5.1 Mapping over token lists
  - 15.5.2 Head and tail of token lists
  - 15.5.3 Items and ranges in token lists
  - 15.5.4 Sorting token lists
- 15.6 Manipulating tokens in token lists
  - 15.6.1 Replacing tokens
  - 15.6.2 Reassigning category codes
- 15.7 Constant token lists
- 15.8 Scratch token lists

16 The l3tl-build module: Piecewise tl constructions
- 16.1 Constructing ⟨tl var⟩ by accumulation

17 The l3str module: Strings
- 17.1 Creating and initialising string variables
- 17.2 Adding data to string variables
- 17.3 String conditionals
- 17.4 Mapping over strings
- 17.5 Working with the content of strings
- 17.6 Modifying string variables
- 17.7 String manipulation
- 17.8 Viewing strings
- 17.9 Constant strings
- 17.10 Scratch strings

18 The l3str-convert module: String encoding conversions
- 18.1 Encoding and escaping schemes
- 18.2 Conversion functions
- 18.3 Conversion by expansion (for PDF contexts)
- 18.4 Possibilities, and things to do

19 The l3quark module: Quarks and scan marks
- 19.1 Quarks
- 19.2 Defining quarks
- 19.3 Quark tests
- 19.4 Recursion
  - 19.4.1 An example of recursion with quarks
- 19.5 Scan marks
24 The \texttt{13token} module: Token manipulation 195
24.1 Creating character tokens ........................................... 196
24.2 Manipulating and interrogating character tokens ................. 197
24.3 Generic tokens ......................................................................... 200
24.4 Converting tokens ................................................................. 201
24.5 Token conditionals ............................................................... 201
24.6 Peeking ahead at the next token ........................................... 205
24.7 Description of all possible tokens ......................................... 210
25 The \texttt{13prop} module: Property lists 213
25.1 Creating and initialising property lists ................................. 214
25.2 Adding and updating property list entries ............................. 216
25.3 Recovering values from property lists ................................... 217
25.4 Modifying property lists ....................................................... 218
25.5 Property list conditionals ..................................................... 218
25.6 Recovering values from property lists with branching .......... 219
25.7 Mapping over property lists .................................................. 220
25.8 Viewing property lists ......................................................... 221
25.9 Scratch property lists .......................................................... 222
25.10 Constants .............................................................................. 222
26 The \texttt{13skip} module: Dimensions and skips 223
26.1 Creating and initialising \texttt{dim} variables ............................... 223
26.2 Setting \texttt{dim} variables ........................................................ 224
26.3 Utilities for dimension calculations ...................................... 224
26.4 Dimension expression conditionals ..................................... 225
26.5 Dimension expression loops ................................................. 227
26.6 Dimension step functions ...................................................... 228
26.7 Using \texttt{dim} expressions and variables ............................... 229
26.8 Viewing \texttt{dim} variables ....................................................... 231
26.9 Constant dimensions ............................................................ 232
26.10 Scratch dimensions ............................................................... 232
26.11 Creating and initialising \texttt{skip} variables ............................. 232
26.12 Setting \texttt{skip} variables ....................................................... 233
26.13 Skip expression conditionals ................................................. 234
26.14 Using \texttt{skip} expressions and variables ............................. 234
26.15 Viewing \texttt{skip} variables ...................................................... 234
26.16 Constant skips ..................................................................... 235
26.17 Scratch skips ....................................................................... 235
26.18 Inserting skips into the output ............................................. 235
26.19 Creating and initialising \texttt{muskip} variables ........................ 236
26.20 Setting \texttt{muskip} variables .................................................. 236
26.21 Using \texttt{muskip} expressions and variables ........................ 237
26.22 Viewing \texttt{muskip} variables ................................................ 237
26.23 Constant muskips ................................................................. 238
26.24 Scratch muskips ................................................................. 238
26.25 Primitive conditional .......................................................... 238
27 The \texttt{l3keys} module: Key–value interfaces \hspace{1cm} 239
   27.1 Creating keys \hspace{1cm} 240
   27.2 Sub-dividing keys \hspace{1cm} 245
   27.3 Choice and multiple choice keys \hspace{1cm} 246
   27.4 Key usage scope \hspace{1cm} 248
   27.5 Setting keys \hspace{1cm} 248
   27.6 Handling of unknown keys \hspace{1cm} 249
   27.7 Selective key setting \hspace{1cm} 250
   27.8 Digesting keys \hspace{1cm} 251
   27.9 Utility functions for keys \hspace{1cm} 251
   27.10 Low-level interface for parsing key–val lists \hspace{1cm} 252

28 The \texttt{l3intarray} module: Fast global integer arrays \hspace{1cm} 255
   28.1 Creating and initialising integer array variables \hspace{1cm} 255
   28.2 Adding data to integer arrays \hspace{1cm} 256
   28.3 Counting entries in integer arrays \hspace{1cm} 256
   28.4 Using a single entry \hspace{1cm} 256
   28.5 Integer array conditional \hspace{1cm} 256
   28.6 Viewing integer arrays \hspace{1cm} 256
   28.7 Implementation notes \hspace{1cm} 257

29 The \texttt{l3fp} module: Floating points \hspace{1cm} 258
   29.1 Creating and initialising floating point variables \hspace{1cm} 260
   29.2 Setting floating point variables \hspace{1cm} 260
   29.3 Using floating points \hspace{1cm} 261
   29.4 Floating point conditionals \hspace{1cm} 262
   29.5 Floating point expression loops \hspace{1cm} 264
   29.6 Symbolic expressions \hspace{1cm} 266
   29.7 User-defined functions \hspace{1cm} 268
   29.8 Some useful constants, and scratch variables \hspace{1cm} 269
   29.9 Scratch variables \hspace{1cm} 269
   29.10 Floating point exceptions \hspace{1cm} 270
   29.11 Viewing floating points \hspace{1cm} 271
   29.12 Floating point expressions \hspace{1cm} 271
      \hspace{1cm} 29.12.1 Input of floating point numbers \hspace{1cm} 271
      \hspace{1cm} 29.12.2 Precedence of operators \hspace{1cm} 272
      \hspace{1cm} 29.12.3 Operations \hspace{1cm} 273
   29.13 Disclaimer and roadmap \hspace{1cm} 280

30 The \texttt{l3fparray} module: Fast global floating point arrays \hspace{1cm} 283
   30.1 Creating and initialising floating point array variables \hspace{1cm} 283
   30.2 Adding data to floating point arrays \hspace{1cm} 283
   30.3 Counting entries in floating point arrays \hspace{1cm} 284
   30.4 Using a single entry \hspace{1cm} 284
   30.5 Floating point array conditional \hspace{1cm} 284

31 The \texttt{l3bitset} module: Bitsets \hspace{1cm} 285
   31.1 Creating bitsets \hspace{1cm} 286
   31.2 Setting and unsetting bits \hspace{1cm} 287
   31.3 Using bitsets \hspace{1cm} 287
32 The \texttt{ctab} module: Category code tables  \hspace{1cm} 289
  32.1 Creating and initialising category code tables \hspace{1cm} 289
  32.2 Using category code tables \hspace{1cm} 290
  32.3 Category code table conditionals \hspace{1cm} 290
  32.4 Constant and scratch category code tables \hspace{1cm} 290

V Text manipulation  \hspace{1cm} 292
33 The \texttt{unicode} module: Unicode support functions  \hspace{1cm} 293

34 The \texttt{text} module: Text processing  \hspace{1cm} 296
  34.1 Expanding text \hspace{1cm} 296
  34.2 Case changing \hspace{1cm} 297
  34.3 Removing formatting from text \hspace{1cm} 299
  34.4 Control variables \hspace{1cm} 299
  34.5 Mapping to graphemes \hspace{1cm} 300

VI Typesetting  \hspace{1cm} 301
35 The \texttt{box} module: Boxes  \hspace{1cm} 302
  35.1 Creating and initialising boxes \hspace{1cm} 302
  35.2 Using boxes \hspace{1cm} 303
  35.3 Measuring and setting box dimensions \hspace{1cm} 304
  35.4 Box conditionals \hspace{1cm} 305
  35.5 The last box inserted \hspace{1cm} 305
  35.6 Constant boxes \hspace{1cm} 305
  35.7 Scratch boxes \hspace{1cm} 305
  35.8 Viewing box contents \hspace{1cm} 306
  35.9 Boxes and color \hspace{1cm} 306
  35.10 Horizontal mode boxes \hspace{1cm} 306
  35.11 Vertical mode boxes \hspace{1cm} 307
  35.12 Using boxes efficiently \hspace{1cm} 309
  35.13 Affine transformations \hspace{1cm} 310
  35.14 Viewing part of a box \hspace{1cm} 313
  35.15 Primitive box conditionals \hspace{1cm} 314

36 The \texttt{coffins} module: Coffin code layer  \hspace{1cm} 315
  36.1 Creating and initialising coffins \hspace{1cm} 315
  36.2 Setting coffin content and poles \hspace{1cm} 316
  36.3 Coffin affine transformations \hspace{1cm} 317
  36.4 Joining and using coffins \hspace{1cm} 318
  36.5 Measuring coffins \hspace{1cm} 318
  36.6 Coffin diagnostics \hspace{1cm} 319
  36.7 Constants and variables \hspace{1cm} 320
## 46 l3regex implementation

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>46.1 Plan of attack</td>
<td>470</td>
</tr>
<tr>
<td>46.2 Helpers</td>
<td></td>
</tr>
<tr>
<td>46.2.1 Constants and variables</td>
<td>474</td>
</tr>
<tr>
<td>46.2.2 Testing characters</td>
<td>474</td>
</tr>
<tr>
<td>46.2.3 Internal auxiliaries</td>
<td>475</td>
</tr>
<tr>
<td>46.2.4 Character property tests</td>
<td>478</td>
</tr>
<tr>
<td>46.2.5 Simple character escape</td>
<td>480</td>
</tr>
<tr>
<td>46.3 Compiling</td>
<td></td>
</tr>
<tr>
<td>46.3.1 Variables used when compiling</td>
<td>487</td>
</tr>
<tr>
<td>46.3.2 Generic helpers used when compiling</td>
<td>488</td>
</tr>
<tr>
<td>46.3.3 Mode</td>
<td>489</td>
</tr>
<tr>
<td>46.3.4 Framework</td>
<td>491</td>
</tr>
<tr>
<td>46.3.5 Quantifiers</td>
<td>494</td>
</tr>
<tr>
<td>46.3.6 Raw characters</td>
<td>497</td>
</tr>
<tr>
<td>46.3.7 Character properties</td>
<td>499</td>
</tr>
<tr>
<td>46.3.8 Anchoring and simple assertions</td>
<td>500</td>
</tr>
<tr>
<td>46.3.9 Character classes</td>
<td>500</td>
</tr>
<tr>
<td>46.3.10 Groups and alternations</td>
<td>504</td>
</tr>
<tr>
<td>46.3.11 Catcodes and csnames</td>
<td>506</td>
</tr>
<tr>
<td>46.3.12 Raw token lists with \u</td>
<td>510</td>
</tr>
<tr>
<td>46.3.13 Other</td>
<td>514</td>
</tr>
<tr>
<td>46.3.14 Showing regexes</td>
<td>514</td>
</tr>
<tr>
<td>46.4 Building</td>
<td></td>
</tr>
<tr>
<td>46.4.1 Variables used while building</td>
<td>521</td>
</tr>
<tr>
<td>46.4.2 Framework</td>
<td>522</td>
</tr>
<tr>
<td>46.4.3 Helpers for building an NFA</td>
<td>525</td>
</tr>
<tr>
<td>46.4.4 Building classes</td>
<td>526</td>
</tr>
<tr>
<td>46.4.5 Building groups</td>
<td>528</td>
</tr>
<tr>
<td>46.4.6 Others</td>
<td>532</td>
</tr>
<tr>
<td>46.5 Matching</td>
<td></td>
</tr>
<tr>
<td>46.5.1 Variables used when matching</td>
<td>534</td>
</tr>
<tr>
<td>46.5.2 Matching: framework</td>
<td>537</td>
</tr>
<tr>
<td>46.5.3 Using states of the NFA</td>
<td>540</td>
</tr>
<tr>
<td>46.5.4 Actions when matching</td>
<td>541</td>
</tr>
<tr>
<td>46.6 Replacement</td>
<td></td>
</tr>
<tr>
<td>46.6.1 Variables and helpers used in replacement</td>
<td>543</td>
</tr>
<tr>
<td>46.6.2 Query and brace balance</td>
<td>543</td>
</tr>
<tr>
<td>46.6.3 Framework</td>
<td>545</td>
</tr>
<tr>
<td>46.6.4 Submatches</td>
<td>546</td>
</tr>
<tr>
<td>46.6.5 Csnames in replacement</td>
<td>549</td>
</tr>
<tr>
<td>46.6.6 Characters in replacement</td>
<td>551</td>
</tr>
<tr>
<td>46.6.7 An error</td>
<td>552</td>
</tr>
<tr>
<td>46.7 User functions</td>
<td></td>
</tr>
<tr>
<td>46.7.1 Variables and helpers for user functions</td>
<td>556</td>
</tr>
<tr>
<td>46.7.2 Matching</td>
<td>561</td>
</tr>
<tr>
<td>46.7.3 Extracting submatches</td>
<td>562</td>
</tr>
<tr>
<td>46.7.4 Replacement</td>
<td>567</td>
</tr>
<tr>
<td>46.7.5 Peeking ahead</td>
<td>570</td>
</tr>
<tr>
<td>46.8 Messages</td>
<td></td>
</tr>
<tr>
<td>46.8.1 Variables and helpers for user functions</td>
<td>576</td>
</tr>
<tr>
<td>46.8.2 Matching</td>
<td>571</td>
</tr>
<tr>
<td>46.8.3 Extracting submatches</td>
<td>572</td>
</tr>
<tr>
<td>46.8.4 Replacement</td>
<td>576</td>
</tr>
</tbody>
</table>
46.9 Code for tracing ............................................. 582

47 \texttt{l3prg} implementation .............................. 584
47.1 Primitive conditionals .................................... 584
47.2 Defining a set of conditional functions ................. 584
47.3 The boolean data type .................................... 584
47.4 Internal auxiliaries ...................................... 586
47.5 Boolean expressions ................................. 588
47.6 Logical loops ............................................. 592
47.7 Producing multiple copies ............................ 594
47.8 Detecting \TeX{}'s mode .................................. 595
47.9 Internal programming functions ....................... 596

48 \texttt{l3sys} implementation .............................. 598
48.1 Kernel code ............................................. 598
  48.1.1 Detecting the engine ............................ 598
  48.1.2 Platform ........................................ 601
  48.1.3 Configurations .................................. 601
  48.1.4 Access to the shell ............................... 604
48.2 Dynamic (every job) code ............................. 606
  48.2.1 The name of the job ............................. 606
  48.2.2 Time and date .................................. 607
  48.2.3 Random numbers ................................. 608
  48.2.4 Access to the shell ............................... 609
48.3 System queries .......................................... 609
  48.3.1 Held over from \texttt{l3file} ..................... 611
48.4 Last-minute code ....................................... 611
  48.4.1 Detecting the output ............................. 611
  48.4.2 Configurations ................................. 612

49 \texttt{l3msg} implementation ............................ 614
49.1 Internal auxiliaries ................................. 614
49.2 Creating messages ..................................... 614
49.3 Messages: support functions and text ................ 616
49.4 Showing messages: low level mechanism .............. 617
49.5 Displaying messages ................................. 619
49.6 Kernel-specific functions ........................... 628
49.7 Internal messages ..................................... 629
49.8 Expandable errors ..................................... 636
49.9 Message formatting .................................... 637
50 l3file implementation

50.1 Input operations ............................................ 638
  50.1.1 Variables and constants .................................. 638
  50.1.2 Stream management ....................................... 639
  50.1.3 Reading input ........................................... 642
50.2 Output operations ........................................... 645
  50.2.1 Variables and constants .................................. 645
  50.2.2 Internal auxiliaries ..................................... 646
50.3 Stream management ........................................... 647
  50.3.1 Deferred writing .......................................... 649
  50.3.2 Immediate writing ........................................ 650
  50.3.3 Special characters for writing ............................ 651
  50.3.4 Hard-wrapping lines to a character count ............... 651
50.4 File operations .............................................. 660
  50.4.1 Internal auxiliaries ..................................... 662
50.5 GetIdInfo .................................................... 678
50.6 Checking the version of kernel dependencies ................. 679
50.7 Messages ..................................................... 681
50.8 Functions delayed from earlier modules ....................... 681

51 l3luatex implementation ................................. 683

51.1 Breaking out to Lua .......................................... 683
51.2 Messages ..................................................... 684
51.3 Lua functions for internal use ................................. 684
51.4 Preserving iniTeX Lua data for runs ......................... 690

52 l3legacy implementation ............................... 692

53 l3tl implementation ........................................ 694

53.1 Functions .................................................... 694
53.2 Constant token lists ......................................... 696
53.3 Adding to token list variables ................................. 696
53.4 Internal quarks and quark-query functions ..................... 699
53.5 Reassigning token list category codes ......................... 700
53.6 Modifying token list variables ................................. 703
53.7 Token list conditionals ....................................... 707
53.8 Mapping over token lists ..................................... 712
53.9 Using token lists ............................................ 714
53.10 Working with the contents of token lists ...................... 714
53.11 The first token from a token list ............................ 717
53.12 Token by token changes ...................................... 722
53.13 Using a single item ......................................... 724
53.14 Viewing token lists ......................................... 727
53.15 Internal scan marks .......................................... 729
53.16 Scratch token lists ......................................... 729

54 l3tl-build implementation ............................ 730
### 55 l3str implementation

<table>
<thead>
<tr>
<th>Section</th>
</tr>
</thead>
<tbody>
<tr>
<td>55.1 Internal auxiliaries</td>
</tr>
<tr>
<td>55.2 Creating and setting string variables</td>
</tr>
<tr>
<td>55.3 Modifying string variables</td>
</tr>
<tr>
<td>55.4 String comparisons</td>
</tr>
<tr>
<td>55.5 Mapping over strings</td>
</tr>
<tr>
<td>55.6 Accessing specific characters in a string</td>
</tr>
<tr>
<td>55.7 Counting characters</td>
</tr>
<tr>
<td>55.8 The first character in a string</td>
</tr>
<tr>
<td>55.9 String manipulation</td>
</tr>
<tr>
<td>55.10 Viewing strings</td>
</tr>
</tbody>
</table>

### 56 l3str-convert implementation

<table>
<thead>
<tr>
<th>Section</th>
</tr>
</thead>
<tbody>
<tr>
<td>56.1 Helpers</td>
</tr>
<tr>
<td>56.1.1 Variables and constants</td>
</tr>
<tr>
<td>56.2 String conditionals</td>
</tr>
<tr>
<td>56.3 Conversions</td>
</tr>
<tr>
<td>56.3.1 Producing one byte or character</td>
</tr>
<tr>
<td>56.3.2 Mapping functions for conversions</td>
</tr>
<tr>
<td>56.3.3 Error-reporting during conversion</td>
</tr>
<tr>
<td>56.3.4 Framework for conversions</td>
</tr>
<tr>
<td>56.3.5 Byte unescape and escape</td>
</tr>
<tr>
<td>56.3.6 Native strings</td>
</tr>
<tr>
<td>56.3.7 clist</td>
</tr>
<tr>
<td>56.3.8 8-bit encodings</td>
</tr>
<tr>
<td>56.4 Messages</td>
</tr>
<tr>
<td>56.5 Escaping definitions</td>
</tr>
<tr>
<td>56.5.1 Unescape methods</td>
</tr>
<tr>
<td>56.5.2 Escape methods</td>
</tr>
<tr>
<td>56.6 Encoding definitions</td>
</tr>
<tr>
<td>56.6.1 UTF-8 support</td>
</tr>
<tr>
<td>56.6.2 UTF-16 support</td>
</tr>
<tr>
<td>56.6.3 UTF-32 support</td>
</tr>
<tr>
<td>56.7 PDF names and strings by expansion</td>
</tr>
<tr>
<td>56.7.1 iso 8859 support</td>
</tr>
</tbody>
</table>

### 57 l3quark implementation

<table>
<thead>
<tr>
<th>Section</th>
</tr>
</thead>
<tbody>
<tr>
<td>57.1 Quarks</td>
</tr>
<tr>
<td>57.2 Scan marks</td>
</tr>
</tbody>
</table>
58 13seq implementation
58.1 Allocation and initialisation ........................................ 819
58.2 Appending data to either end ..................................... 822
58.3 Modifying sequences .............................................. 823
58.4 Sequence conditionals ............................................ 827
58.5 Recovering data from sequences ................................. 829
58.6 Mapping over sequences ......................................... 833
58.7 Using sequences .................................................. 838
58.8 Sequence stacks ................................................... 839
58.9 Viewing sequences ............................................... 839
58.10 Scratch sequences .............................................. 840

59 13int implementation
59.1 Integer expressions .............................................. 842
59.2 Creating and initialising integers ................................. 844
59.3 Setting and incrementing integers ................................. 846
59.4 Using integers ..................................................... 847
59.5 Integer expression conditionals .................................. 847
59.6 Integer expression loops ......................................... 851
59.7 Integer step functions ............................................ 852
59.8 Formatting integers ............................................... 854
59.9 Converting from other formats to integers ..................... 860
59.10 Viewing integer .................................................. 862
59.11 Random integers ................................................. 863
59.12 Constant integers ............................................... 863
59.13 Scratch integers ................................................ 864
59.14 Integers for earlier modules .................................... 864

60 13flag implementation .............................................. 865
60.1 Protected flag commands ........................................ 865
60.2 Expandable flag commands ...................................... 866
60.3 Old n-type flag commands ...................................... 867

61 13clist implementation ............................................ 869
61.1 Removing spaces around items .................................. 870
61.2 Allocation and initialisation .................................... 871
61.3 Adding data to comma lists ..................................... 873
61.4 Comma lists as stacks .......................................... 874
61.5 Modifying comma lists .......................................... 876
61.6 Comma list conditionals ....................................... 879
61.7 Mapping over comma lists ..................................... 880
61.8 Using comma lists .............................................. 884
61.9 Using a single item ............................................. 886
61.10 Viewing comma lists .......................................... 888
61.11 Scratch comma lists .......................................... 889
## 65 `l3keys` implementation

- **Low-level interface** .................................................. 965
- **Constants and variables** ........................................... 972
- **The key defining mechanism** ...................................... 975
- **Turning properties into actions** ................................ 977
- **Creating key properties** ........................................... 984
- **Setting keys** .......................................................... 990
- **Utilities** .............................................................. 999
- **Messages** ............................................................. 1002

## 66 `l3intarray` implementation

- **Lua implementation** .................................................. 1004
- **Font dimension based implementation** .......................... 1010
- **Common parts** ....................................................... 1016

## 67 `l3fp` implementation

## 68 `l3fp-aux` implementation

- **Access to primitives** .............................................. 1018
- **Internal representation** ......................................... 1018
- **Using arguments and semicolons** ............................... 1019
- **Constants, and structure of floating points** ................ 1020
- **Overflow, underflow, and exact zero** .......................... 1023
- **Expanding after a floating point number** ..................... 1023
- **Other floating point types** ...................................... 1024
- **Packing digits** ...................................................... 1027
- **Decimate (dividing by a power of 10)** ........................ 1030
- **Functions for use within primitive conditional branches** 1032
- **Integer floating points** .......................................... 1033
- **Small integer floating points** .................................. 1034
- **Fast string comparison** .......................................... 1035
- **Name of a function from its `l3fp-parse` name** .......... 1035
- **Messages** .......................................................... 1035

## 69 `l3fp-traps` implementation

- **Flags** .............................................................. 1036
- **Traps** .............................................................. 1036
- **Errors** ............................................................ 1040
- **Messages** .......................................................... 1040

## 70 `l3fp-round` implementation

- **Rounding tools** .................................................... 1042
- **The `round` function** ........................................... 1046
## 71 l3fp-parse implementation

<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>71.1</td>
<td>Work plan</td>
<td>1051</td>
</tr>
<tr>
<td>71.1.1</td>
<td>Storing results</td>
<td>1052</td>
</tr>
<tr>
<td>71.1.2</td>
<td>Precedence and infix operators</td>
<td>1053</td>
</tr>
<tr>
<td>71.1.3</td>
<td>Prefix operators, parentheses, and functions</td>
<td>1056</td>
</tr>
<tr>
<td>71.1.4</td>
<td>Numbers and reading tokens one by one</td>
<td>1057</td>
</tr>
<tr>
<td>71.2</td>
<td>Main auxiliary functions</td>
<td>1059</td>
</tr>
<tr>
<td>71.3</td>
<td>Helpers</td>
<td>1060</td>
</tr>
<tr>
<td>71.4</td>
<td>Parsing one number</td>
<td>1061</td>
</tr>
<tr>
<td>71.4.1</td>
<td>Numbers: trimming leading zeros</td>
<td>1067</td>
</tr>
<tr>
<td>71.4.2</td>
<td>Number: small significand</td>
<td>1068</td>
</tr>
<tr>
<td>71.4.3</td>
<td>Number: large significand</td>
<td>1070</td>
</tr>
<tr>
<td>71.4.4</td>
<td>Number: beyond 16 digits, rounding</td>
<td>1072</td>
</tr>
<tr>
<td>71.4.5</td>
<td>Number: finding the exponent</td>
<td>1075</td>
</tr>
<tr>
<td>71.5</td>
<td>Constants, functions and prefix operators</td>
<td>1078</td>
</tr>
<tr>
<td>71.5.1</td>
<td>Prefix operators</td>
<td>1078</td>
</tr>
<tr>
<td>71.5.2</td>
<td>Constants</td>
<td>1081</td>
</tr>
<tr>
<td>71.5.3</td>
<td>Functions</td>
<td>1082</td>
</tr>
<tr>
<td>71.6</td>
<td>Main functions</td>
<td>1083</td>
</tr>
<tr>
<td>71.7</td>
<td>Infix operators</td>
<td>1085</td>
</tr>
<tr>
<td>71.7.1</td>
<td>Closing parentheses and commas</td>
<td>1086</td>
</tr>
<tr>
<td>71.7.2</td>
<td>Usual infix operators</td>
<td>1088</td>
</tr>
<tr>
<td>71.7.3</td>
<td>Juxtaposition</td>
<td>1089</td>
</tr>
<tr>
<td>71.7.4</td>
<td>Multi-character cases</td>
<td>1089</td>
</tr>
<tr>
<td>71.7.5</td>
<td>Ternary operator</td>
<td>1090</td>
</tr>
<tr>
<td>71.7.6</td>
<td>Comparisons</td>
<td>1090</td>
</tr>
<tr>
<td>71.8</td>
<td>Tools for functions</td>
<td>1092</td>
</tr>
<tr>
<td>71.9</td>
<td>Messages</td>
<td>1095</td>
</tr>
</tbody>
</table>

## 72 l3fp-assign implementation

<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>72.1</td>
<td>Assigning values</td>
<td>1096</td>
</tr>
<tr>
<td>72.2</td>
<td>Updating values</td>
<td>1097</td>
</tr>
<tr>
<td>72.3</td>
<td>Showing values</td>
<td>1097</td>
</tr>
<tr>
<td>72.4</td>
<td>Some useful constants and scratch variables</td>
<td>1099</td>
</tr>
</tbody>
</table>

## 73 l3fp-logic implementation

<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>73.1</td>
<td>Syntax of internal functions</td>
<td>1100</td>
</tr>
<tr>
<td>73.2</td>
<td>Tests</td>
<td>1100</td>
</tr>
<tr>
<td>73.3</td>
<td>Comparison</td>
<td>1101</td>
</tr>
<tr>
<td>73.4</td>
<td>Floating point expression loops</td>
<td>1104</td>
</tr>
<tr>
<td>73.5</td>
<td>Extrema</td>
<td>1107</td>
</tr>
<tr>
<td>73.6</td>
<td>Boolean operations</td>
<td>1109</td>
</tr>
<tr>
<td>73.7</td>
<td>Ternary operator</td>
<td>1110</td>
</tr>
</tbody>
</table>
## 74 l3fp-basics implementation

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>74.1 Addition and subtraction</td>
<td>1112</td>
</tr>
<tr>
<td>74.1.1 Sign, exponent, and special numbers</td>
<td>1113</td>
</tr>
<tr>
<td>74.1.2 Absolute addition</td>
<td>1115</td>
</tr>
<tr>
<td>74.1.3 Absolute subtraction</td>
<td>1117</td>
</tr>
<tr>
<td>74.2 Multiplication</td>
<td>1121</td>
</tr>
<tr>
<td>74.2.1 Signs, and special numbers</td>
<td>1121</td>
</tr>
<tr>
<td>74.2.2 Absolute multiplication</td>
<td>1123</td>
</tr>
<tr>
<td>74.3 Division</td>
<td>1125</td>
</tr>
<tr>
<td>74.3.1 Signs, and special numbers</td>
<td>1125</td>
</tr>
<tr>
<td>74.3.2 Work plan</td>
<td>1126</td>
</tr>
<tr>
<td>74.3.3 Implementing the significand division</td>
<td>1129</td>
</tr>
<tr>
<td>74.4 Square root</td>
<td>1134</td>
</tr>
<tr>
<td>74.5 About the sign and exponent</td>
<td>1141</td>
</tr>
<tr>
<td>74.6 Operations on tuples</td>
<td>1142</td>
</tr>
</tbody>
</table>

## 75 l3fp-extended implementation

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>75.1 Description of fixed point numbers</td>
<td>1144</td>
</tr>
<tr>
<td>75.2 Helpers for numbers with extended precision</td>
<td>1145</td>
</tr>
<tr>
<td>75.3 Multiplying a fixed point number by a short one</td>
<td>1146</td>
</tr>
<tr>
<td>75.4 Dividing a fixed point number by a small integer</td>
<td>1146</td>
</tr>
<tr>
<td>75.5 Adding and subtracting fixed points</td>
<td>1147</td>
</tr>
<tr>
<td>75.6 Multiplying fixed points</td>
<td>1148</td>
</tr>
<tr>
<td>75.7 Combining product and sum of fixed points</td>
<td>1149</td>
</tr>
<tr>
<td>75.8 Extended-precision floating point numbers</td>
<td>1152</td>
</tr>
<tr>
<td>75.9 Dividing extended-precision numbers</td>
<td>1154</td>
</tr>
<tr>
<td>75.10 Inverse square root of extended precision numbers</td>
<td>1158</td>
</tr>
<tr>
<td>75.11 Converting from fixed point to floating point</td>
<td>1160</td>
</tr>
</tbody>
</table>

## 76 l3fp-expo implementation

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>76.1 Logarithm</td>
<td>1162</td>
</tr>
<tr>
<td>76.1.1 Work plan</td>
<td>1162</td>
</tr>
<tr>
<td>76.1.2 Some constants</td>
<td>1163</td>
</tr>
<tr>
<td>76.1.3 Sign, exponent, and special numbers</td>
<td>1163</td>
</tr>
<tr>
<td>76.1.4 Absolute In</td>
<td>1163</td>
</tr>
<tr>
<td>76.2 Exponential</td>
<td>1171</td>
</tr>
<tr>
<td>76.2.1 Sign, exponent, and special numbers</td>
<td>1171</td>
</tr>
<tr>
<td>76.3 Power</td>
<td>1175</td>
</tr>
<tr>
<td>76.4 Factorial</td>
<td>1181</td>
</tr>
<tr>
<td>Section</td>
<td>Title</td>
</tr>
<tr>
<td>---------</td>
<td>-------</td>
</tr>
<tr>
<td>77</td>
<td>l3fp-trig implementation</td>
</tr>
<tr>
<td>77.1</td>
<td>Direct trigonometric functions</td>
</tr>
<tr>
<td>77.1.1</td>
<td>Filtering special cases</td>
</tr>
<tr>
<td>77.1.2</td>
<td>Distinguishing small and large arguments</td>
</tr>
<tr>
<td>77.1.3</td>
<td>Small arguments</td>
</tr>
<tr>
<td>77.1.4</td>
<td>Argument reduction in degrees</td>
</tr>
<tr>
<td>77.1.5</td>
<td>Argument reduction in radians</td>
</tr>
<tr>
<td>77.1.6</td>
<td>Computing the power series</td>
</tr>
<tr>
<td>77.2</td>
<td>Inverse trigonometric functions</td>
</tr>
<tr>
<td>77.2.1</td>
<td>Arctangent and arccotangent</td>
</tr>
<tr>
<td>77.2.2</td>
<td>Arcsine and arccosine</td>
</tr>
<tr>
<td>77.2.3</td>
<td>Arccosecant and arcsecant</td>
</tr>
<tr>
<td>78</td>
<td>l3fp-convert implementation</td>
</tr>
<tr>
<td>78.1</td>
<td>Dealing with tuples</td>
</tr>
<tr>
<td>78.2</td>
<td>Trimming trailing zeros</td>
</tr>
<tr>
<td>78.3</td>
<td>Scientific notation</td>
</tr>
<tr>
<td>78.4</td>
<td>Decimal representation</td>
</tr>
<tr>
<td>78.5</td>
<td>Token list representation</td>
</tr>
<tr>
<td>78.6</td>
<td>Formatting</td>
</tr>
<tr>
<td>78.7</td>
<td>Convert to dimension or integer</td>
</tr>
<tr>
<td>78.8</td>
<td>Convert from a dimension</td>
</tr>
<tr>
<td>78.9</td>
<td>Use and eval</td>
</tr>
<tr>
<td>78.10</td>
<td>Convert an array of floating points to a comma list</td>
</tr>
<tr>
<td>79</td>
<td>l3fp-random implementation</td>
</tr>
<tr>
<td>79.1</td>
<td>Engine support</td>
</tr>
<tr>
<td>79.2</td>
<td>Random floating point</td>
</tr>
<tr>
<td>79.3</td>
<td>Random integer</td>
</tr>
<tr>
<td>80</td>
<td>l3fp-types implementation</td>
</tr>
<tr>
<td>80.1</td>
<td>Support for types</td>
</tr>
<tr>
<td>80.2</td>
<td>Dispatch according to the type</td>
</tr>
<tr>
<td>81</td>
<td>l3fp-symbolic implementation</td>
</tr>
<tr>
<td>81.1</td>
<td>Misc</td>
</tr>
<tr>
<td>81.2</td>
<td>Building blocks for expressions</td>
</tr>
<tr>
<td>81.3</td>
<td>Expanding after a symbolic expression</td>
</tr>
<tr>
<td>81.4</td>
<td>Applying infix operators to expressions</td>
</tr>
<tr>
<td>81.5</td>
<td>Applying prefix functions to expressions</td>
</tr>
<tr>
<td>81.6</td>
<td>Conversions</td>
</tr>
<tr>
<td>81.7</td>
<td>Identifiers</td>
</tr>
<tr>
<td>81.8</td>
<td>Declaring variables and assigning values</td>
</tr>
<tr>
<td>81.9</td>
<td>Messages</td>
</tr>
<tr>
<td>81.10</td>
<td>Road-map</td>
</tr>
<tr>
<td>82</td>
<td>l3fp-functions implementation</td>
</tr>
<tr>
<td>82.1</td>
<td>Declaring functions</td>
</tr>
<tr>
<td>82.2</td>
<td>Defining functions by their expression</td>
</tr>
</tbody>
</table>
83 `l3fpararray implementation .............................................. 1246
  83.1 Allocating arrays ................................................. 1246
  83.2 Array items ..................................................... 1247

84 `l3bitset implementation ................................................. 1251
  84.1 Messages .......................................................... 1256

85 `l3ctab implementation .................................................. 1257
  85.1 Variables .......................................................... 1257
  85.2 Allocating category code tables ................................... 1258
  85.3 Saving category code tables ....................................... 1259
  85.4 Using category code tables ....................................... 1260
  85.5 Category code table conditionals .................................. 1265
  85.6 Constant category code tables .................................... 1267
  85.7 Messages .......................................................... 1268

86 `l3unicode implementation ................................................. 1270
  86.1 User functions ..................................................... 1270
  86.2 Data loader ......................................................... 1274

87 `l3text implementation ................................................... 1285
  87.1 Internal auxiliaries .............................................. 1285
  87.2 Utilities ............................................................ 1286
  87.3 Codepoint utilities ............................................... 1289
  87.4 Configuration variables ......................................... 1292
  87.5 Expansion to formatted text ..................................... 1293

88 `l3text-case implementation .............................................. 1302
  88.1 Case changing ..................................................... 1302

89 `l3text-map implementation .............................................. 1337
  89.1 Mapping to text ................................................... 1337

90 `l3text-purify implementation .......................................... 1345
  90.1 Purifying text ..................................................... 1345
  90.2 Accent and letter-like data for purifying text ................... 1351

91 `l3box implementation .................................................... 1358
  91.1 Support code ...................................................... 1358
  91.2 Creating and initialising boxes .................................. 1358
  91.3 Measuring and setting box dimensions ............................ 1359
  91.4 Using boxes ......................................................... 1360
  91.5 Box conditionals .................................................. 1361
  91.6 The last box inserted ............................................. 1361
  91.7 Constant boxes .................................................... 1361
  91.8 Scratch boxes ..................................................... 1362
  91.9 Viewing box contents ............................................. 1362
  91.10 Horizontal mode boxes .......................................... 1363
  91.11 Vertical mode boxes ............................................. 1365
  91.12 Affine transformations .......................................... 1368
  91.13 Viewing part of a box ........................................... 1377
Part I

Introduction
Chapter 1

Introduction to expl3 and this document

This document is intended to act as a comprehensive reference manual for the expl3 language. A general guide to the \LaTeX3 programming language is found in expl3.pdf.

1.1 Naming functions and variables

\LaTeX3 does not use \texttt{@} as a “letter” for defining internal macros. Instead, the symbols \texttt{_} and \texttt{:} are used in internal macro names to provide structure. The name of each function is divided into logical units using \texttt{_}, while \texttt{:} separates the name of the function from the argument specifier (“arg-spec”). This describes the arguments expected by the function. In most cases, each argument is represented by a single letter. The complete list of arg-spec letters for a function is referred to as the signature of the function.

Each function name starts with the module to which it belongs. Thus apart from a small number of very basic functions, all expl3 function names contain at least one underscore to divide the module name from the descriptive name of the function. For example, all functions concerned with comma lists are in module \texttt{clist} and begin \texttt{clist_}.

Every function must include an argument specifier. For functions which take no arguments, this will be blank and the function name will end \texttt{:}. Most functions take one or more arguments, and use the following argument specifiers:

\texttt{N} and \texttt{n} These mean \textit{no manipulation}, of a single token for \texttt{N} and of a set of tokens given in braces for \texttt{n}. Both pass the argument through exactly as given. Usually, if you use a single token for an \texttt{n} argument, all will be well.

\texttt{c} This means csname, and indicates that the argument will be turned into a csname before being used. So \texttt{\textbackslash foo:c \{ArgumentOne\}} will act in the same way as \texttt{\textbackslash foo:N \textbackslash ArgumentOne}. All macros that appear in the argument are expanded. An internal error will occur if the result of expansion inside a \texttt{c}-type argument is not a series of character tokens.

\texttt{V} and \texttt{v} These mean \textit{value of variable}. The V and v specifiers are used to get the content of a variable without needing to worry about the underlying \LaTeX structure containing the data. A V argument will be a single token (similar to \texttt{N}), for example
\texttt{\textbackslash foo:v \textbackslash MyVariable}: on the other hand, using \texttt{v} a csname is constructed first, and then the value is recovered, for example \texttt{\textbackslash foo:v \{MyVariable\}}.

\begin{itemize}
    \item This means \textit{expansion once}. In general, the \texttt{V} and \texttt{v} specifiers are favoured over \texttt{o} for recovering stored information. However, \texttt{o} is useful for correctly processing information with delimited arguments.

    \item The \texttt{x} specifier stands for \textit{exhaustive expansion}: every token in the argument is fully expanded until only unexpandable ones remain. The \TeX{} \texttt{\textbackslash def} primitive carries out this type of expansion. Functions which feature an \texttt{x}-type argument \textit{are not} expandable.

    \item The \texttt{e} specifier is in many respects identical to \texttt{x}, but uses \texttt{\textbackslash expanded} primitive. Parameter character (usually \texttt{#}) in the argument need not be doubled. Functions which feature an \texttt{e}-type argument may be expandable.

    \item The \texttt{f} specifier stands for \textit{full expansion}, and in contrast to \texttt{x} stops at the first non-expandable token (reading the argument from left to right) without trying to expand it. If this token is a (\texttt{space token}), it is gobbled, and thus won’t be part of the resulting argument. For example, when setting a token list variable (a macro used for storage), the sequence

\begin{verbatim}
\tl_set:Nn \l_mya_tl { A }
\tl_set:Nn \l_myb_tl { B }
\tl_set:Nf \l_mya_tl { \l_mya_tl \l_myb_tl }
\end{verbatim}

will leave \texttt{\l_mya_tl} with the content \texttt{A\l_myb_tl}, as \texttt{A} cannot be expanded and so terminates expansion before \texttt{\l_myb_tl} is considered.

\end{itemize}

\textbf{T and F} For logic tests, there are the branch specifiers \texttt{T} (\texttt{true}) and \texttt{F} (\texttt{false}). Both specifiers treat the input in the same way as \texttt{n} (no change), but make the logic much easier to see.

\begin{itemize}
    \item The letter \texttt{p} indicates \TeX{} \textit{parameters}. Normally this will be used for delimited functions as expl3 provides better methods for creating simple sequential arguments.

    \item Finally, there is the \texttt{w} specifier for \textit{weird} arguments. This covers everything else, but mainly applies to delimited values (where the argument must be terminated by some specified string).

\end{itemize}

\textbf{D} The \texttt{D} stands for \textit{Do not use}. All of the \TeX{} primitives are initially \texttt{\textbackslash let} to a \texttt{D} name, and some are then given a second name. These functions have no standardized syntax, they are engine dependent and their name can change without warning, thus their use is strongly discouraged in package code: programmers should instead use the interfaces documented in interface3.pdf.

Notice that the argument specifier describes how the argument is processed prior to being passed to the underlying function. For example, \texttt{\textbackslash foo:c} will take its argument, convert it to a control sequence and pass it to \texttt{\textbackslash foo:N}.

Variables are named in a similar manner to functions, but begin with a single letter to define the type of variable:

\begin{itemize}
    \item \texttt{c} Constant: global parameters whose value should not be changed.
\end{itemize}
g Parameters whose value should only be set globally.

l Parameters whose value should only be set locally.

Each variable name is then build up in a similar way to that of a function, typically starting with the module\(^1\) name and then a descriptive part. Variables end with a short identifier to show the variable type:

- **bitset** a set of bits (a string made up of a series of 0 and 1 tokens that are accessed by position).
- **clist** Comma separated list.
- **dim** “Rigid” lengths.
- **fp** Floating-point values;
- **int** Integer-valued count register.
- **muskip** “Rubber” lengths for use in mathematics.
- **skip** “Rubber” lengths.
- **str** String variables: contain character data.
- **tl** Token list variables: placeholder for a token list.

Applying V-type or v-type expansion to variables of one of the above types is supported, while it is not supported for the following variable types:

- **bool** Either true or false.
- **box** Box register.
- **coffin** A “box with handles” — a higher-level data type for carrying out box alignment operations.
- **flag** Non-negative integer that can be incremented expandably.
- **fparray** Fixed-size array of floating point values.
- **intarray** Fixed-size array of integers.
- **ior/iow** An input or output stream, for reading from or writing to, respectively.
- **prop** Property list: analogue of dictionary or associative arrays in other languages.
- **regex** Regular expression.
- **seq** “Sequence”: a data type used to implement lists (with access at both ends) and stacks.

\(^1\)The module names are not used in case of generic scratch registers defined in the data type modules, e.g., the int module contains some scratch variables called \l_tmpa_int, \l_tmpb_int, and so on. In such a case adding the module name up front to denote the module and in the back to indicate the type, as in \l_int_tmpa_int would be very unreadable.
1.1.1  Scratch variables

Modules focussed on variable usage typically provide four scratch variables, two local and two global, with names of the form \(\langle\text{scope}\rangle_{\text{tmpa}}_{\langle\text{type}\rangle}/\langle\text{scope}\rangle_{\text{tmpb}}_{\langle\text{type}\rangle}\). These are never used by the core code. The nature of \TeX{} grouping means that as with any other scratch variable, these should only be set and used with no intervening third-party code.

1.1.2  Terminological inexactitude

A word of warning. In this document, and others referring to the \texttt{expl3} programming modules, we often refer to “variables” and “functions” as if they were actual constructs from a real programming language. In truth, \TeX{} is a macro processor, and functions are simply macros that may or may not take arguments and expand to their replacement text. Many of the common variables are also macros, and if placed into the input stream will simply expand to their definition as well — a “function” with no arguments and a “token list variable” are almost the same.\footnote{\TeX{}nically, functions with no arguments are \texttt{\long} while token list variables are not.} On the other hand, some “variables” are actually registers that must be initialised and their values set and retrieved with specific functions.

The conventions of the \texttt{expl3} code are designed to clearly separate the ideas of “macros that contain data” and “macros that contain code”, and a consistent wrapper is applied to all forms of “data” whether they be macros or actually registers. This means that sometimes we will use phrases like “the function returns a value”, when actually we just mean “the macro expands to something”. Similarly, the term “execute” might be used in place of “expand” or it might refer to the more specific case of “processing in \TeX{}’s stomach” (if you are familiar with the \TeX{}book parlance).

If in doubt, please ask; chances are we’ve been hasty in writing certain definitions and need to be told to tighten up our terminology.

1.2  Documentation conventions

This document is typeset with the experimental \texttt{l3doc} class; several conventions are used to help describe the features of the code. A number of conventions are used here to make the documentation clearer.

Each group of related functions is given in a box. For a function with a “user” name, this might read:

\begin{lstlisting}[language=TeX]
\ExplSyntaxOn \ExplSyntaxOn \ExplSyntaxOff \ExplSyntaxOn ...
\ExplSyntaxOff\ExplSyntaxOn ...
\ExplSyntaxOff
\end{lstlisting}

The textual description of how the function works would appear here. The syntax of the function is shown in mono-spaced text to the right of the box. In this example, the function takes no arguments and so the name of the function is simply reprinted.

For programming functions, which use \_ and \, in their name there are a few additional conventions: If two related functions are given with identical names but different argument specifiers, these are termed variants of each other, and the latter functions are printed in grey to show this more clearly. They will carry out the same function but will take different types of argument:
When a number of variants are described, the arguments are usually illustrated only for the base function. Here, \texttt{sequence} indicates that \texttt{seq_new:N} expects the name of a sequence. From the argument specifier, \texttt{seq_new:c} also expects a sequence name, but as a name rather than as a control sequence. Each argument given in the illustration should be described in the following text.

**Fully expandable functions** Some functions are fully expandable, which allows them to be used within an \texttt{x}-type or \texttt{e}-type argument (in plain \TeX terms, inside an \texttt{edef} or \texttt{expanded}), as well as within an \texttt{f}-type argument. These fully expandable functions are indicated in the documentation by a star:

\begin{verbatim}
\seq_map_function:NN \seq_map_function:NN \star
\end{verbatim}

As with other functions, some text should follow which explains how the function works. Usually, only the star will indicate that the function is expandable. In this case, the function expects a \texttt{cs}, shorthand for a \texttt{control sequence}.

**Restricted expandable functions** A few functions are fully expandable but cannot be fully expanded within an \texttt{f}-type argument. In this case a hollow star is used to indicate this:

\begin{verbatim}
\seq_map_function:NN \star
\end{verbatim}

**Conditional functions** Conditional (if) functions are normally defined in three variants, with \texttt{T}, \texttt{F} and \texttt{TF} argument specifiers. This allows them to be used for different “true”/“false” branches, depending on which outcome the conditional is being used to test. To indicate this without repetition, this information is given in a shortened form:

\begin{verbatim}
\sys_if_engine_xetex:TF \star
\end{verbatim}

The underlining and italic of \texttt{TF} indicates that three functions are available:

- \texttt{sys_if_engine_xetex:T}
- \texttt{sys_if_engine_xetex:F}
- \texttt{sys_if_engine_xetex:TF}

Usually, the illustration will use the \texttt{TF} variant, and so both \texttt{true code} and \texttt{false code} will be shown. The two variant forms \texttt{T} and \texttt{F} take only \texttt{true code} and \texttt{false code}, respectively. Here, the star also shows that this function is expandable. With some minor exceptions, all conditional functions in the expl3 modules should be defined in this way.

Variables, constants and so on are described in a similar manner:

\begin{verbatim}
\l_tmpa_tl
\end{verbatim}

A short piece of text will describe the variable: there is no syntax illustration in this case. In some cases, the function is similar to one in \texttt{if\TeX\ }2\texttt{e} or plain \TeX. In these cases, the text will include an extra “\TeXhackers note” section:
The normal description text.

\textbf{TexHackers note:} Detail for the experienced \TeX{} or \LaTeX{} programmer. In this case, it would point out that this function is the \TeX{} primitive \texttt{\string}.

\textbf{Changes to behaviour} When new functions are added to expl3, the date of first inclusion is given in the documentation. Where the documented behaviour of a function changes after it is first introduced, the date of the update will also be given. This means that the programmer can be sure that any release of expl3 after the date given will contain the function of interest with expected behaviour as described. Note that changes to code internals, including bug fixes, are not recorded in this way unless they impact on the expected behaviour.

\section{Formal language conventions which apply generally}

As this is a formal reference guide for \LaTeX{} programming, the descriptions of functions are intended to be reasonably "complete". However, there is also a need to avoid repetition. Formal ideas which apply to general classes of function are therefore summarised here.

For tests which have a TF argument specification, the test if evaluated to give a logically TRUE or FALSE result. Depending on this result, either the \texttt{true code} or the \texttt{false code} will be left in the input stream. In the case where the test is expandable, and a predicate (_p) variant is available, the logical value determined by the test is left in the input stream: this will typically be part of a larger logical construct.

\section{\TeX{} concepts not supported by \LaTeX{}3}

The \TeX{} concept of an "\texttt{\outer}" macro is not supported at all by \LaTeX{}3. As such, the functions provided here may break when used on top of \LaTeX{}2\epsilon if \texttt{\outer} tokens are used in the arguments.
Part II

Bootstrapping
Chapter 2

The \l3bootstrap module

Bootstrap code

2.1 Using the \LaTeX3 modules

The modules documented in interface3 (and this file) are designed to be used on top of \LaTeX2e and are already pre-loaded since \LaTeX2e 2020-02-02. To support older formats, the $\texttt{\usepackage{expl3}}$ or $\texttt{\RequirePackage{expl3}}$ instructions are still available to load them all as one.

As the modules use a coding syntax different from standard \LaTeX2e it provides a few functions for setting it up.

\begin{quote}
\texttt{\ExplSyntaxOn} ⟨\textit{code}⟩ \texttt{\ExplSyntaxOff}
\end{quote}

\begin{itemize}
\item \texttt{\ExplSyntaxOn} function switches to a category code regime in which spaces and new lines are ignored, and in which the colon (:) and underscore (_\texttt{)}} are treated as “letters”, thus allowing access to the names of code functions and variables. Within this environment, - is used to input a space. The \texttt{\ExplSyntaxOff} reverts to the document category code regime.
\end{itemize}

\textbf{\LaTeXX hackers note:} Spaces introduced by - behave much in the same way as normal space characters in the standard category code regime: they are ignored after a control word or at the start of a line, and multiple consecutive - are equivalent to a single one. However, - is not ignored at the end of a line.

\begin{quote}
\texttt{\ProvidesExplPackage ⟨\textit{package}⟩ ⟨\textit{date}⟩ ⟨\textit{version}⟩ ⟨\textit{description}⟩}
\end{quote}

These functions act broadly in the same way as the corresponding \LaTeX2e kernel functions \texttt{\ProvidesPackage}, \texttt{\ProvidesClass} and \texttt{\ProvidesFile}. However, they also implicitly switch \texttt{\ExplSyntaxOn} for the remainder of the code with the file. At the end of the file, \texttt{\ExplSyntaxOff} will be called to reverse this. (This is the same concept as \LaTeX2e provides in turning on \texttt{\makeatletter} within package and class code.) The \texttt{\langle date\rangle} should be given in the format \texttt{\langle year\rangle/\langle month\rangle/\langle day\rangle} or in the ISO date format \texttt{\langle year\rangle-(\langle month\rangle)-\langle day\rangle}. If the \texttt{\langle version\rangle} is given then a leading \texttt{v} is optional: if given as a “pure” version string, a \texttt{v} will be prepended.

Updated: 2011-08-13

Updated: 2023-08-03
\GetIdInfo

\GetIdInfo $Id: ⟨SVN info field⟩ $ ⟨(description)⟩

Updated: 2012-06-04

Extracts all information from a SVN field. Spaces are not ignored in these fields. The information pieces are stored in separate control sequences with \ExplFileName for the part of the file name leading up to the period, \ExplFileDate for date, \ExplFileVersion for version and \ExplFileDescription for the description.

To summarize: Every single package using this syntax should identify itself using one of the above methods. Special care is taken so that every package or class file loaded with \RequirePackage or similar are loaded with usual \LaTeX category codes and the \LaTeX category code scheme is reloaded when needed afterwards. See implementation for details. If you use the \GetIdInfo command you can use the information when loading a package with

\ProvidesExplPackage{\ExplFileName}
{\ExplFileDate}{\ExplFileVersion}{\ExplFileDescription}
Chapter 3

The \texttt{l3names} module
Namespace for primitives

3.1 Setting up the \LaTeX{}3 programming language

This module is at the core of the \LaTeX{}3 programming language. It performs the following tasks:

- defines new names for all \TeX{} primitives;
- emulate required primitives not provided by default in Lu\TeX{};
- switches to the category code régime for programming;

This module is entirely dedicated to primitives (and emulations of these), which should not be used directly within \LaTeX{}3 code (outside of “kernel-level” code). As such, the primitives are not documented here: \textit{The \TeX{}book}, \textit{\TeX{} by Topic} and the manuals for pd\TeX{}, Xe\TeX{}, Lu\TeX{}, p\TeX{} and up\TeX{} should be consulted for details of the primitives. These are named \texttt{\textbackslash\tex\langle\textit{name}\rangle}:D, typically based on the primitive’s \textit{name} in pd\TeX{} and omitting a leading \texttt{pdf} when the primitive is not related to pdf output.
Part III
Programming Flow
Chapter 4

The l3basics module

Basic definitions

As the name suggests, this module holds some basic definitions which are needed by most or all other modules in this set.

Here we describe those functions that are used all over the place. By that, we mean functions dealing with the construction and testing of control sequences. Furthermore the basic parts of conditional processing are covered; conditional processing dealing with specific data types is described in the modules specific for the respective data types.

4.1 No operation functions

\prg_do_nothing: \prg_do_nothing:
An expandable function which does nothing at all: leaves nothing in the input stream after a single expansion.

\scan_stop: \scan_stop:
A non-expandable function which does nothing. Does not vanish on expansion but produces no typeset output.

4.2 Grouping material

\group_begin: \group_end:
\group_begin: \group_end:
These functions begin and end a group for definition purposes. Assignments are local to groups unless carried out in a global manner. (A small number of exceptions to this rule will be noted as necessary elsewhere in this document.) Each \group_begin: must be matched by a \group_end:, although this does not have to occur within the same function. Indeed, it is often necessary to start a group within one function and finish it within another, for example when seeking to use non-standard category codes.

\TeXhackers note: These are the \TeX primitives \begingroup and \endgroup.
\group_insert_after:N \group_insert_after:N (token)

Adds \textit{token} to the list of \textit{tokens} to be inserted when the current group level ends. The list of \textit{tokens} to be inserted is empty at the beginning of a group: multiple applications of \texttt{\group_insert_after:N} may be used to build the inserted list one \textit{token} at a time. The current group level may be closed by a \texttt{\group_end:} function or by a token with category code 2 (close-group), namely a \textbackslash if standard category codes apply.

\textbf{\texttt{\TeX}hackers note:} This is the \TeX{} primitive \texttt{\aftergroup}.  

\texttt{\group_show_list:} \texttt{\group_log_list:}

Display (to the terminal or log file) a list of the groups that are currently opened. This is intended for tracking down problems.

\textbf{\texttt{\TeX}hackers note:} This is a wrapper around the \texttt{\e-\TeX{}} primitive \texttt{\showgroups}.

\section{4.3 Control sequences and functions}

As \TeX{} is a macro language, creating new functions means creating macros. At point of use, a function is replaced by the replacement text ("code") in which each parameter in the code (#1, #2, etc.) is replaced the appropriate arguments absorbed by the function. In the following, \textit{code} is therefore used as a shorthand for "replacement text".

Functions which are not "protected" are fully expanded inside an e-type or x-type expansions. In contrast, "protected" functions are not expanded within e and x expansions.

\subsection{4.3.1 Defining functions}

Functions can be created with no requirement that they are declared first (in contrast to variables, which must always be declared). Declaring a function before setting up the code means that the name chosen is checked and an error raised if it is already in use. The name of a function can be checked at the point of definition using the \texttt{\cs_new...} functions: this is recommended for all functions which are defined for the first time.

There are three ways to define new functions. All classes define a function to expand to the substitution text. Within the substitution text the actual parameters are substituted for the formal parameters (#1, #2, ...).

\textbf{new} Create a new function with the \texttt{new} scope, such as \texttt{\cs_new:Npn}. The definition is global and results in an error if it is already defined.

\textbf{set} Create a new function with the \texttt{set} scope, such as \texttt{\cs_set:Npn}. The definition is restricted to the current \TeX{} group and does not result in an error if the function is already defined.

\textbf{gset} Create a new function with the \texttt{gset} scope, such as \texttt{\cs_gset:Npn}. The definition is global and does not result in an error if the function is already defined.
Within each set of scope there are different ways to define a function. The differences depend on restrictions on the actual parameters and the expandability of the resulting function.

**nopar** Create a new function with the `nopar` restriction, such as `\cs_set_nopar:Npn`. The parameter may not contain `\par` tokens.

**protected** Create a new function with the `protected` restriction, such as `\cs_set_protected:Npn`. The parameter may contain `\par` tokens but the function will not expand within an `e`-type or `x`-type expansion.

Finally, the functions in Subsections 4.3.2 and 4.3.3 are primarily meant to define *base functions* only. Base functions can only have the following argument specifiers:

- **N and n** No manipulation.
- **T and F** Functionally equivalent to `n` (you are actually encouraged to use the family of `\prg_new_conditional`: functions described in Section 9.1).
- **p and w** These are special cases.

The `\cs_new`: functions below (and friends) do not stop you from using other argument specifiers in your function names, but they do not handle expansion for you. You should define the base function and then use `\cs_generate_variant:Nn` to generate custom variants as described in Section 5.2.

### 4.3.2 Defining new functions using parameter text

```latex
\cs_new:Npn \langle function \rangle \langle parameters \rangle \{ \langle code \rangle \}
```

Creates *(function)* to expand to *(code)* as replacement text. Within the *(code)*, the *(parameters)* (#1, #2, etc.) will be replaced by those absorbed by the function. The definition is global and an error results if the *(function)* is already defined.

```latex
\cs_new_nopar:Npn \langle function \rangle \langle parameters \rangle \{ \langle code \rangle \}
```

Creates *(function)* to expand to *(code)* as replacement text. Within the *(code)*, the *(parameters)* (#1, #2, etc.) will be replaced by those absorbed by the function. When the *(function)* is used the *(parameters)* absorbed cannot contain `\par` tokens. The definition is global and an error results if the *(function)* is already defined.

```latex
\cs_new_protected:Npn \langle function \rangle \langle parameters \rangle \{ \langle code \rangle \}
```

Creates *(function)* to expand to *(code)* as replacement text. Within the *(code)*, the *(parameters)* (#1, #2, etc.) will be replaced by those absorbed by the function. The *(function)* will not expand within an `e`-type or `x`-type argument. The definition is global and an error results if the *(function)* is already defined.
\cs_new_protected_nopar:Npn \cs_new_protected_nopar:cpn \cs_new_protected_nopar:Npe \cs_new_protected_nopar:Npx \cs_new_protected_nopar:cpx

Creates \texttt{function} to expand to \texttt{code} as replacement text. Within the \texttt{code}, the \texttt{parameters} (\#1, \#2, etc.) will be replaced by those absorbed by the function. When the \texttt{function} is used the \texttt{parameters} absorbed cannot contain \texttt{par} tokens. The \texttt{function} will not expand within an e-type or x-type argument. The definition is global and an error results if the \texttt{function} is already defined.

\cs_set:Npn \cs_set:cpn \cs_set:Npe \cs_set:cpe \cs_set:Npx \cs_set:cpx

Sets \texttt{function} to expand to \texttt{code} as replacement text. Within the \texttt{code}, the \texttt{parameters} (\#1, \#2, etc.) will be replaced by those absorbed by the function. The assignment of a meaning to the \texttt{function} is restricted to the current TeX group level.

\cs_set_protected:Npn \cs_set_protected:cpn \cs_set_protected:Npe \cs_set_protected:cpe \cs_set_protected:Npx \cs_set_protected:cpx

Sets \texttt{function} to expand to \texttt{code} as replacement text. Within the \texttt{code}, the \texttt{parameters} (\#1, \#2, etc.) will be replaced by those absorbed by the function. The assignment of a meaning to the \texttt{function} is restricted to the current TeX group level. The \texttt{function} will not expand within an e-type or x-type argument.
4.3.3 Defining new functions using the signature

<table>
<thead>
<tr>
<th>\cs_new:Nn</th>
<th>\cs_new:nnn { (code) }</th>
</tr>
</thead>
<tbody>
<tr>
<td>Creates (function) to expand to (code) as replacement text. Within the (code), the \langle parameters \rangle (#1, #2, etc.) will be replaced by those absorbed by the function. The definition is global and an error results if the (function) is already defined.</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>\cs_new_nopar:Nn</th>
<th>\cs_new_nopar:nnn { (code) }</th>
</tr>
</thead>
<tbody>
<tr>
<td>Creates (function) to expand to (code) as replacement text. Within the (code), the \langle parameters \rangle (#1, #2, etc.) will be replaced by those absorbed by the function. When the (function) is used the \langle parameters \rangle absorbed cannot contain \textbackslash par tokens. The definition is global and an error results if the (function) is already defined.</td>
<td></td>
</tr>
</tbody>
</table>
\cs_new_protected:Nn \cs_new_protected:{(cn|Ne|ce)} \cs_set_protected:Nn \cs_set_protected:{(code)}

Creates \textit{(function)} to expand to \textit{(code)} as replacement text. Within the \textit{(code)}, the number of \textit{(parameters)} is detected automatically from the function signature. These \textit{(parameters)} (#1, #2, etc.) will be replaced by those absorbed by the function. The \textit{(function)} will not expand within an \texttt{e}-type or \texttt{x}-type argument. The definition is global and an error results if the \textit{(function)} is already defined.

\cs_new_protected_nopar:Nn \cs_new_protected_nopar:{(cn|Ne|ce)} \cs_set_nopar:Nn \cs_set_nopar:{(code)}

Creates \textit{(function)} to expand to \textit{(code)} as replacement text. Within the \textit{(code)}, the number of \textit{(parameters)} is detected automatically from the function signature. These \textit{(parameters)} (#1, #2, etc.) will be replaced by those absorbed by the function. When the \textit{(function)} is used the \textit{(parameters)} absorbed cannot contain \texttt{par} tokens. The \textit{(function)} will not expand within an \texttt{e}-type or \texttt{x}-type argument. The definition is global and an error results if the \textit{(function)} is already defined.

\cs_set:Nn \cs_set:{(cn|Ne|ce)} \cs_set:{(code)}

Sets \textit{(function)} to expand to \textit{(code)} as replacement text. Within the \textit{(code)}, the number of \textit{(parameters)} is detected automatically from the function signature. These \textit{(parameters)} (#1, #2, etc.) will be replaced by those absorbed by the function. The assignment of a meaning to the \textit{(function)} is restricted to the current \TeX{} group level.

\cs_set_nopar:Nn \cs_set_nopar:{(cn|Ne|ce)} \cs_set_nopar:{(code)}

Sets \textit{(function)} to expand to \textit{(code)} as replacement text. Within the \textit{(code)}, the number of \textit{(parameters)} is detected automatically from the function signature. These \textit{(parameters)} (#1, #2, etc.) will be replaced by those absorbed by the function. When the \textit{(function)} is used the \textit{(parameters)} absorbed cannot contain \texttt{par} tokens. The assignment of a meaning to the \textit{(function)} is restricted to the current \TeX{} group level.

\cs_set_protected:Nn \cs_set_protected:{(cn|Ne|ce)} \cs_set_protected:{(code)}

Sets \textit{(function)} to expand to \textit{(code)} as replacement text. Within the \textit{(code)}, the number of \textit{(parameters)} is detected automatically from the function signature. These \textit{(parameters)} (#1, #2, etc.) will be replaced by those absorbed by the function. The \textit{(function)} will not expand within an \texttt{e}-type or \texttt{x}-type argument. The assignment of a meaning to the \textit{(function)} is restricted to the current \TeX{} group level.

\cs_set_protected_nopar:Nn \cs_set_protected_nopar:{(cn|Ne|ce)} \cs_set_protected_nopar:{(code)}

Sets \textit{(function)} to expand to \textit{(code)} as replacement text. Within the \textit{(code)}, the number of \textit{(parameters)} is detected automatically from the function signature. These \textit{(parameters)} (#1, #2, etc.) will be replaced by those absorbed by the function. When the \textit{(function)} is used the \textit{(parameters)} absorbed cannot contain \texttt{par} tokens. The \textit{(function)} will not expand within an \texttt{e}-type or \texttt{x}-type argument. The assignment of a meaning to the \textit{(function)} is restricted to the current \TeX{} group level.
\cs_gset:Nn \cs_gset:(cn|Ne|ce)

Sets \emph{function} to expand to \emph{code} as replacement text. Within the \emph{code}, the number of \emph{parameters} is detected automatically from the function signature. These \emph{parameters} (#1, #2, etc.) will be replaced by those absorbed by the function. The assignment of a meaning to the \emph{function} is global.

\cs_gset_nopar:Nn \cs_gset_nopar:(cn|Ne|ce)

Sets \emph{function} to expand to \emph{code} as replacement text. Within the \emph{code}, the number of \emph{parameters} is detected automatically from the function signature. These \emph{parameters} (#1, #2, etc.) will be replaced by those absorbed by the function. When the \emph{function} is used the \emph{parameters} absorbed cannot contain \texttt{par} tokens. The assignment of a meaning to the \emph{function} is global.

\cs_gset_protected:Nn \cs_gset_protected:(cn|Ne|ce)

Sets \emph{function} to expand to \emph{code} as replacement text. Within the \emph{code}, the number of \emph{parameters} is detected automatically from the function signature. These \emph{parameters} (#1, #2, etc.) will be replaced by those absorbed by the function. The \emph{function} will not expand within an \texttt{e}-type or \texttt{x}-type argument. The assignment of a meaning to the \emph{function} is global.

\cs_gset_protected_nopar:Nn \cs_gset_protected_nopar:(cn|Ne|ce)

Sets \emph{function} to expand to \emph{code} as replacement text. Within the \emph{code}, the number of \emph{parameters} is detected automatically from the function signature. These \emph{parameters} (#1, #2, etc.) will be replaced by those absorbed by the function. When the \emph{function} is used the \emph{parameters} absorbed cannot contain \texttt{par} tokens. The \emph{function} will not expand within an \texttt{e}-type or \texttt{x}-type argument. The assignment of a meaning to the \emph{function} is global.

\cs_generate_from_arg_count:NNnn \cs_generate_from_arg_count:NNnn (function) \{creator\}

\cs_generate_from_arg_count:NNnn \cs_generate_from_arg_count:NNnn (function) \{number\} \{code\}

Updated: 2012-01-14

Uses the \emph{creator} function (which should have signature \texttt{Npn}, for example \texttt{\cs_new:Npn}) to define a \emph{function} which takes \emph{number} arguments and has \emph{code} as replacement text. The \emph{number} of arguments is an integer expression, evaluated as detailed for \texttt{\int_eval:n}.

4.3.4 Copying control sequences

Control sequences (not just functions as defined above) can be set to have the same meaning using the functions described here. Making two control sequences equivalent means that the second control sequence is a \emph{copy} of the first (rather than a pointer to it). Thus the old and new control sequence are not tied together: changes to one are not reflected in the other.

In the following text “cs” is used as an abbreviation for “control sequence”.

19
Globally creates \langle control sequence \rangle and sets it to have the same meaning as \langle control sequence \rangle or \langle token \rangle. The second control sequence may subsequently be altered without affecting the copy.

Sets \langle control sequence \rangle to have the same meaning as \langle control sequence \rangle (or \langle token \rangle). The second control sequence may subsequently be altered without affecting the copy. The assignment of a meaning to the \langle control sequence \rangle is restricted to the current \TeX{} group level.

Globally sets \langle control sequence \rangle to have the same meaning as \langle control sequence \rangle (or \langle token \rangle). The second control sequence may subsequently be altered without affecting the copy. The assignment of a meaning to the \langle control sequence \rangle is not restricted to the current \TeX{} group level: the assignment is global.

### 4.3.5 Deleting control sequences

There are occasions where control sequences need to be deleted. This is handled in a very simple manner.

Sets \langle control sequence \rangle to be globally undefined.

This function expands to the meaning of the \langle control sequence \rangle control sequence. For a macro, this includes the \langle replacement text \rangle.

\TeX{}hackers note: This is the \TeX{} primitive \texttt{\meaning}. For tokens that are not control sequences, it is more logical to use \texttt{\token_to_meaning:N}. The c variant correctly reports undefined arguments.

Displays the definition of the \langle control sequence \rangle on the terminal.

\TeX{}hackers note: This is similar to the \TeX{} primitive \texttt{\show}, wrapped to a fixed number of characters per line.
\texttt{\cs_log:N} \langle control sequence \rangle
\texttt{\cs_log:c}

Writes the definition of the \langle control sequence \rangle in the log file. See also \texttt{\cs_show:N} which displays the result in the terminal.

\begin{verbatim}
\cs_log:N \cs_log:c
New: 2014-08-22
Updated: 2017-02-14
\end{verbatim}

4.3.7 Converting to and from control sequences

\texttt{\use:c \{\langle control sequence name \rangle\}}

Expands the \langle control sequence name \rangle until only characters remain, and then converts this into a control sequence. This process requires two expansions. As in other c-type arguments the \langle control sequence name \rangle must, when fully expanded, consist of character tokens, typically a mixture of category code 10 (space), 11 (letter) and 12 (other).

As an example of the \texttt{\use:c} function, both
\begin{verbatim}
\use:c \{ a \ b \ c \}
\end{verbatim}

and
\begin{verbatim}
\tl_new:N \l_my_tl
\tl_set:Nn \l_my_tl \{ a \ b \ c \}
\use:c \{ \tl_use:N \l_my_tl \}
\end{verbatim}

would be equivalent to
\begin{verbatim}
abc
\end{verbatim}

after two expansions of \texttt{\use:c}.

\texttt{\cs_if_exist_use:N \cs_if_exist_use:c \cs_if_exist_use:NTF \cs_if_exist_use:cTF}

Tests whether the \langle control sequence \rangle is currently defined according to the conditional \texttt{\cs_if_exist:NTF} (whether as a function or another control sequence type), and if it is inserts the \langle control sequence \rangle into the input stream followed by the \langle true code \rangle. Otherwise the \langle false code \rangle is used.

\texttt{\cs:w \cs:w \{\langle control sequence name \rangle\} \cs_end:}

\texttt{\cs_end:}

Converts the given \langle control sequence name \rangle into a single control sequence token. This process requires one expansion. The content for \langle control sequence name \rangle may be literal material or from other expandable functions. The \langle control sequence name \rangle must, when fully expanded, consist of character tokens which are not active: typically of category code 10 (space), 11 (letter) or 12 (other), or a mixture of these.

\TeXhackers note: These are the \TeX primitives \texttt{\csname} and \texttt{\endcsname}.

As an example of the \texttt{\cs:w} and \texttt{\cs_end:} functions, both
\begin{verbatim}
\cs:w a \ b \ c \cs_end:
\end{verbatim}

and
\begin{verbatim}
21
\end{verbatim}
\l_new:N \l_my_tl
\l_set:Nn \l_my_tl { a b c }
\cs:w \tl_use:N \l_my_tl \cs_end:

would be equivalent to

\abc

after one expansion of \cs:w.

\cs_to_str:N \cs_to_str:N \langle control sequence \rangle

Converts the given \langle control sequence \rangle into a series of characters with category code 12 (other), except spaces, of category code 10. The result does not include the current escape token, contrarily to \token_to_str:N. Full expansion of this function requires exactly 2 expansion steps, and so an e-type or x-type expansion, or two o-type expansions are required to convert the \langle control sequence \rangle to a sequence of characters in the input stream. In most cases, an f-expansion is correct as well, but this loses a space at the start of the result.

4.4 Analysing control sequences

\cs_split_function:N \cs_split_function:N \langle function \rangle

Splits the \langle function \rangle into the \langle name \rangle (i.e. the part before the colon) and the \langle signature \rangle (i.e. after the colon). This information is then placed in the input stream in three parts: the \langle name \rangle, the \langle signature \rangle and a logic token indicating if a colon was found (to differentiate variables from function names). The \langle name \rangle does not include the escape character, and both the \langle name \rangle and \langle signature \rangle are made up of tokens with category code 12 (other).

The next three functions decompose \TeX\ macros into their constituent parts: if the \langle token \rangle passed is not a macro then no decomposition can occur. In the latter case, all three functions leave \scan_stop: in the input stream.

\cs_prefix_spec:N \cs_prefix_spec:N \langle token \rangle

If the \langle token \rangle is a macro, this function leaves the applicable \TeX\ prefixes in input stream as a string of tokens of category code 12 (with spaces having category code 10). Thus for example

\cs_set:Npn \next:nn #1#2 \{ x #1-y #2 \}
\cs_prefix_spec:N \next:nn

leaves \textbackslash long in the input stream. If the \langle token \rangle is not a macro then \scan_stop: is left in the input stream.

\TeX\ hackers note: The prefix can be empty, \textbackslash long, \textbackslash protected or \textbackslash protected\textbackslash long with backslash replaced by the current escape character.
\cs_parameter_spec:N \token

If the \token is a macro, this function leaves the primitive \TeX parameter specification in input stream as a string of character tokens of category code 12 (with spaces having category code 10). Thus for example

\cs_set:Npn \next:nn \#1\#2 { x \#1 y \#2 }
\cs_parameter_spec:N \next:nn

leaves \#1\#2 in the input stream. If the \token is not a macro then \scan_stop: is left in the input stream.

\TeXhackers note: If the parameter specification contains the string ->, then the function produces incorrect results.

\cs_replacement_spec:N \cs_replacement_spec:c

If the \token is a macro, this function leaves the replacement text in input stream as a string of character tokens of category code 12 (with spaces having category code 10). Thus for example

\cs_set:Npn \next:nn \#1\#2 { x \#1-y \#2 }
\cs_replacement_spec:N \next:nn

leaves x\#1,\#2 in the input stream. If the \token is not a macro then \scan_stop: is left in the input stream.

\TeXhackers note: If the parameter specification contains the string ->, then the function produces incorrect results.

4.5 Using or removing tokens and arguments

Tokens in the input can be read and used or read and discarded. If one or more tokens are wrapped in braces then when absorbing them the outer set is removed. At the same time, the category code of each token is set when the token is read by a function (if it is read more than once, the category code is determined by the situation in force when first function absorbs the token).
\use:n \use:n \{ (group_1) \} \\
\use:nn \use:nn \{ (group_1) \} \{ (group_2) \} \\
\use:nnn \use:nnn \{ (group_1) \} \{ (group_2) \} \{ (group_3) \} \\
\use:nnnn \use:nnnn \{ (group_1) \} \{ (group_2) \} \{ (group_3) \} \{ (group_4) \} \\

As illustrated, these functions absorb between one and four arguments, as indicated by the argument specifier. The braces surrounding each argument are removed and the remaining tokens are left in the input stream. The category code of these tokens is also fixed by this process (if it has not already been by some other absorption). All of these functions require only a single expansion to operate, so that one expansion of

\use:nn \{ abc \} \{ \{ def \} \}

results in the input stream containing

abc \{ def \}

i.e. only the outer braces are removed.

\TeXhackers note: The \use:n function is equivalent to \LaTeX\ ε ’s \texttt{@firstofone}.  

24
These functions absorb a number \((n)\) arguments from the input stream. They then discard all arguments other than that indicated by the roman numeral, which is left in the input stream. For example, \(\texttt{use_i:nn}\) discards the second argument, and leaves the content of the first argument in the input stream. The category code of these tokens is also fixed (if it has not already been by some other absorption). A single expansion is needed for the functions to take effect.
This function absorbs three arguments and leaves the content of the first and second in the input stream. The category code of these tokens is also fixed (if it has not already been by some other absorption). A single expansion is needed for the function to take effect. An example:

\use_i_ii:nnn { abc } { { def } } { ghi }

results in the input stream containing

abc { def }

i.e. the outer braces are removed and the third group is removed.

This function absorbs two arguments and leaves the content of the second and first in the input stream. The category code of these tokens is also fixed (if it has not already been by some other absorption). A single expansion is needed for the function to take effect.

These functions absorb between one and nine groups from the input stream, leaving nothing on the resulting input stream. These functions work after a single expansion. One or more of the \texttt{n} arguments may be an unbraced single token (i.e. an \texttt{N} argument).

\textbf{TEXhackers note:} These are equivalent to \LaTeX{}'s \texttt{@gobble}, \texttt{@gobbletwo}, \texttt{etc.}

Fully expands the \texttt{token list} in an \texttt{e}-type manner, in which parameter character (usually \texttt{#}) need not be doubled, \textit{and} the function remains fully expandable.

\textbf{TEXhackers note:} \texttt{use:e} is a wrapper around the primitive \texttt{expanded}. It requires two expansions to complete its action.

### 4.5.1 Selecting tokens from delimited arguments

A different kind of function for selecting tokens from the token stream are those that use delimited arguments.

Absorb the \texttt{(balanced text)} from the input stream delimited by the marker given in the function name, leaving nothing in the input stream.
Absorb the \textit{balanced text} from the input stream delimited by the marker given in the function name, leaving \textit{inserted tokens} in the input stream for further processing.

### 4.6 Predicates and conditionals

\LaTeX{} has three concepts for conditional flow processing:

**Branching conditionals** Functions that carry out a test and then execute, depending on its result, either the code supplied as the \textit{true code} or the \textit{false code}. These arguments are denoted with \texttt{T} and \texttt{F}, respectively. An example would be

\begin{verbatim}
cs_if_free:cTF {abc} {\langle true code \rangle} {\langle false code \rangle}
\end{verbatim}

a function that turns the first argument into a control sequence (since it’s marked as \texttt{c}) then checks whether this control sequence is still free and then depending on the result carries out the code in the second argument (true case) or in the third argument (false case).

These type of functions are known as “conditionals”: whenever a \texttt{TF} function is defined it is usually accompanied by \texttt{T} and \texttt{F} functions as well. These are provided for convenience when the branch only needs to go a single way. Package writers are free to choose which types to define but the kernel definitions always provide all three versions.

Important to note is that these branching conditionals with \textit{true code} and/or \textit{false code} are always defined in a way that the code of the chosen alternative can operate on following tokens in the input stream.

These conditional functions may or may not be fully expandable, but if they are expandable they are accompanied by a “predicate” for the same test as described below.

**Predicates** “Predicates” are functions that return a special type of boolean value which can be tested by the boolean expression parser. All functions of this type are expandable and have names that end with \texttt{p} in the description part. For example,

\begin{verbatim}
cs_if_free_p:N
\end{verbatim}

would be a predicate function for the same type of test as the conditional described above. It would return “true” if its argument (a single token denoted by \texttt{N}) is still free for definition. It would be used in constructions like

\begin{verbatim}
\bool_if:nTF { \cs_if_free_p:N \l_tmpz_tl || \cs_if_free_p:N \g_tmpz_tl } {\langle true code \rangle} {\langle false code \rangle}
\end{verbatim}

For each predicate defined, a “branching conditional” also exists that behaves like a conditional described above.
Primitive conditionals

There is a third variety of conditional, which is the original concept used in plain \TeX and \LaTeX. Their use is discouraged in expl3 (although still used in low-level definitions) because they are more fragile and in many cases require more expansion control (hence more code) than the two types of conditionals described above.

4.6.1 Tests on control sequences

\begin{verbatim}
\cs_if_eq_p:NN \cs_if_eq_p:NN \cs_if_eq:NNTF \cs_if_eq:NNTF \cs_if_exist_p:N \cs_if_exist_p:c \cs_if_exist:NTF \cs_if_exist:cTF
\end{verbatim}

Compares the definition of two \textit{(control sequences)} and is logically \texttt{true} if they are the same, \textit{i.e.} if they have exactly the same definition when examined with \texttt{cs_show:N}.

\begin{verbatim}
\cs_if_exist_p:N \cs_if_exist_p:c \cs_if_exist:NTF \cs_if_exist:cIF \cs_if_free_p:N \cs_if_free_p:c \cs_if_free:NTF \cs_if_free:cIF
\end{verbatim}

Tests whether the \textit{(control sequence)} is currently defined (whether as a function or another control sequence type). Any definition of \textit{(control sequence)} other than \texttt{relax} evaluates as \texttt{true}.

4.6.2 Primitive conditionals

The \vTeX engine itself provides many different conditionals. Some expand whatever comes after them and others don't. Hence the names for these underlying functions often contains a \texttt{:w} part but higher level functions are often available. See for instance \texttt{int_compare_p:nNn} which is a wrapper for \texttt{if_int_compare:w}.

Certain conditionals deal with specific data types like boxes and fonts and are described there. The ones described below are either the universal conditionals or deal with control sequences. We prefix primitive conditionals with \texttt{if}, except for \texttt{if:w}.

\begin{verbatim}
\if_true: \if_true: \if_false: \else: \else: \reverse_if:N \reverse_if:N \reverse_if:N
\fi: \fi: \fi: \fi:
\end{verbatim}

\texttt{if_true:} always executes \texttt{(true code)}, while \texttt{if_false:} always executes \texttt{(false code)}. \texttt{reverse_if:N} reverses any two-way primitive conditional. \texttt{else:} and \texttt{fi:} delimit the branches of the conditional. The function \texttt{or:} is documented in \texttt{Bint} and used in case switches.

\textbf{\TeX hackers note:} \texttt{if_true:} and \texttt{if_false:} are equivalent to their corresponding \TeX primitives \texttt{iftrue} and \texttt{iffalse}; \texttt{else:} and \texttt{fi:} are the \TeX primitives \texttt{else} and \texttt{fi}; \texttt{reverse_if:N} is the \vTeX primitive \texttt{unless}.
\texttt{\textbackslash if\_meaning:w} \texttt{\textbackslash if\_meaning:w} \texttt{(arg\textsubscript{1}) (arg\textsubscript{2}) (true code)} \texttt{\textbackslash else: (false code) \textbackslash fi:}

\texttt{\textbackslash if\_meaning:w} executes \texttt{(true code)} when \texttt{(arg\textsubscript{1})} and \texttt{(arg\textsubscript{2})} are the same, otherwise it executes \texttt{(false code)}. \texttt{(arg\textsubscript{1})} and \texttt{(arg\textsubscript{2})} could be functions, variables, tokens; in all cases the 	extit{unexpanded} definitions are compared.

\textbf{\textit{\texttt{\textbackslash if\_meaning:w}} hack\textup{ers} note:} This is the \LaTeX{} primitive \texttt{\textbackslash ifx}.

\texttt{\textbackslash if:w} \texttt{\textbackslash if:w (token(s)) (true code)} \texttt{\textbackslash else: (false code) \textbackslash fi:}

\texttt{\textbackslash if\_charcode:w} \texttt{\textbackslash if\_charcode:w (token(s)) (true code) \textbackslash else: (false code) \textbackslash fi:}

\texttt{\textbackslash if\_catcode:w} \texttt{\textbackslash if\_catcode:w (token(s)) (true code) \textbackslash else: (false code) \textbackslash fi:}

\texttt{\textbackslash if\_charcode:w} is an alternative name for \texttt{\textbackslash if:w}. These conditionals expand \texttt{(token(s))} until two unexpandable tokens \texttt{(token\textsubscript{1})} and \texttt{(token\textsubscript{2})} are found; any further tokens up to the next unbalanced \texttt{\textbackslash else:} are the true branch, ending with \texttt{(true code)}. It is executed if the condition is fulfilled, otherwise \texttt{(false code)} is executed. You can omit \texttt{\textbackslash else:} when just in front of \texttt{\textbackslash fi:} and you can nest \texttt{\textbackslash if\ldots\textbackslash else\ldots\textbackslash fi:} constructs inside the true branch or the \texttt{(false code)}. With \texttt{\exp_not:N}, you can prevent the expansion of a token.

\texttt{\textbackslash if\_catcode:w} tests if \texttt{(token\textsubscript{1})} and \texttt{(token\textsubscript{2})} have the same category code whereas \texttt{\textbackslash if:w} and \texttt{\textbackslash if\_charcode:w} test if they have the same character code.

\textbf{\textit{\texttt{\textbackslash if:w}} hack\textup{ers} note:} \texttt{\textbackslash if:w} and \texttt{\textbackslash if\_charcode:w} are both the \LaTeX{} primitive \texttt{\textbackslash if}. \texttt{\textbackslash if\_catcode:w} is the \LaTeX{} primitive \texttt{\textbackslash ifcat}.

\texttt{\textbackslash if\_cs_exist:N} \texttt{\textbackslash if\_cs_exist:N (cs) (true code) \textbackslash else: (false code) \textbackslash fi:}

\texttt{\textbackslash if\_cs_exist:w} \texttt{\textbackslash if\_cs_exist:w (tokens) \cs_end: (true code) \textbackslash else: (false code) \textbackslash fi:}

Check if \texttt{(cs)} appears in the hash table or if the control sequence that can be formed from \texttt{(tokens)} appears in the hash table. The latter function does not turn the control sequence in question into \texttt{\scan_stop!:} This can be useful when dealing with control sequences which cannot be entered as a single token.

\textbf{\textit{\texttt{\textbackslash if\_cs_exist:N}} hack\textup{ers} note:} These are the \LaTeX{} primitives \texttt{\ifdefined} and \texttt{\ifcsname}.

\texttt{\textbackslash if\_mode_horizontal:} \texttt{\textbackslash if\_mode_horizontal: (true code) \textbackslash else: (false code) \textbackslash fi:}

\texttt{\textbackslash if\_mode_vertical:} \texttt{\textbackslash if\_mode_vertical: (true code) \textbackslash else: (false code) \textbackslash fi:}

\texttt{\textbackslash if\_mode_math:} \texttt{\textbackslash if\_mode_math: (true code) \textbackslash else: (false code) \textbackslash fi:}

\texttt{\textbackslash if\_mode_inner:} \texttt{\textbackslash if\_mode_inner: (true code) \textbackslash else: (false code) \textbackslash fi:}

\textbf{\textit{\texttt{\textbackslash if\_mode_horizontal:\textbackslash if\_mode_vertical:\textbackslash if\_mode_math:\textbackslash if\_mode_inner:}} hack\textup{ers} note:} These are the \LaTeX{} primitives \texttt{\ifhmode}, \texttt{\ifvmode}, \texttt{\ifmmode}, \texttt{\ifinner}, and \texttt{\ifinner}. 

29
4.7 Starting a paragraph

\mode_leave_vertical:

Ensures that \TeX{} is not in vertical (inter-paragraph) mode. In horizontal or math mode this command has no effect, in vertical mode it switches to horizontal mode, and inserts a box of width \parindent, followed by the \everypar token list.

\TeX{}hackers note: This results in the contents of the \everypar token register being inserted, after \mode_leave_vertical: is complete. Notice that in contrast to the \LaTeX{}2\epsilon \leavevmode approach, no box is used by the method implemented here.

4.8 Debugging support

\debug_on:n \debug_off:n { ⟨comma-separated list⟩ }

Turn on and off within a group various debugging code, some of which is also available as expl3 load-time options. The items that can be used in the ⟨list⟩ are

- check-declarations that checks all expl3 variables used were previously declared and that local/global variables (based on their name or on their first assignment) are only locally/globally assigned;
- check-expressions that checks integer, dimension, skip, and muskip expressions are not terminated prematurely;
- deprecation that makes deprecated commands produce errors;
- log-functions that logs function definitions and variable declarations;
- all that does all of the above.

Providing these as switches rather than options allows testing code even if it relies on other packages: load all other packages, call \debug_on:n, and load the code that one is interested in testing.

\debug_suspend: ... \debug_resume:

Suppress (locally) errors and logging from debug commands, except for the deprecation errors. These pairs of commands can be nested. This can be used around pieces of code that are known to fail checks, if such failures should be ignored. See for instance l3cctab and l3coffins.
Chapter 5

The \texttt{l3expan} module

Argument expansion

This module provides generic methods for expanding \TeX{} arguments in a systematic manner. The functions in this module all have prefix \texttt{exp}.

Not all possible variations are implemented for every base function. Instead only those that are used within the \LaTeX{}3 kernel or otherwise seem to be of general interest are implemented. Consult the module description to find out which functions are actually defined. The next section explains how to define missing variants.

5.1 Defining new variants

The definition of variant forms for base functions may be necessary when writing new functions or when applying a kernel function in a situation that we haven’t thought of before.

Internally preprocessing of arguments is done with functions of the form \texttt{\exp\_} ... . They all look alike, an example would be \texttt{\exp\_args:NNo}. This function has three arguments, the first and the second are a single tokens, while the third argument should be given in braces. Applying \texttt{\exp\_args:NNo} expands the content of third argument once before any expansion of the first and second arguments. If \texttt{\seq\_gpsh:No} was not defined it could be coded in the following way:

\begin{verbatim}
\exp\_args:NNo \seq\_gpsh:Nn \g_file_name_stack { \l_tmpa_tl }
\end{verbatim}

In other words, the first argument to \texttt{\exp\_args:NNo} is the base function and the other arguments are preprocessed and then passed to this base function. In the example the first argument to the base function should be a single token which is left unchanged while the second argument is expanded once. From this example we can also see how the variants are defined. They just expand into the appropriate \texttt{\exp\_} function followed by the desired base function, \textit{e.g.}.

\begin{verbatim}
\cs\_generate\_variant:Nn \seq\_gpsh:Nn { No }
\end{verbatim}

results in the definition of \texttt{\seq\_gpsh:No}
Providing variants in this way in style files is safe as the \cs_generate_variant:Nn function will only create new definitions if there is not already one available. Therefore adding such definition to later releases of the kernel will not make such style files obsolete.

The steps above may be automated by using the function \cs_generate_variant:Nn, described next.

\section{Methods for defining variants}

We recall the set of available argument specifiers.

- \texttt{N} is used for single-token arguments while \texttt{c} constructs a control sequence from its name and passes it to a parent function as an \texttt{N}-type argument.

- Many argument types extract or expand some tokens and provide it as an \texttt{n}-type argument, namely a braced multiple-token argument: \texttt{V} extracts the value of a variable, \texttt{v} extracts the value from the name of a variable, \texttt{n} uses the argument as it is, \texttt{o} expands once, \texttt{f} expands fully the front of the token list, \texttt{e} and \texttt{x} expand fully all tokens (differences are explained later).

- A few odd argument types remain: \texttt{T} and \texttt{F} for conditional processing, otherwise identical to \texttt{n}-type arguments, \texttt{p} for the parameter text in definitions, \texttt{w} for arguments with a specific syntax, and \texttt{D} to denote primitives that should not be used directly.
This function is used to define argument-specifier variants of the \emph{parent control sequence} for \TeX{} code-level macros. The \emph{parent control sequence} is first separated into the \emph{base name} and \emph{original argument specifier}. The comma-separated list of \emph{variant argument specifiers} is then used to define variants of the \emph{original argument specifier} if these are not already defined; entries which correspond to existing functions are silently ignored. For each \emph{variant} given, a function is created that expands its arguments as detailed and passes them to the \emph{parent control sequence}. So for example
\begin{verbatim}
cs_set:Npn \foo:Nn #1#2 { code here }
cs_generate_variant:Nn \foo:Nn { c }
\end{verbatim}
creates a new function \foo:cn which expands its first argument into a control sequence name and passes the result to \foo:Nn. Similarly
\begin{verbatim}
cs_generate_variant:Nn \foo:Nn \foo:Nn { NV , cV }
\end{verbatim}
generates the functions \foo:NV and \foo:cV in the same way. The \cs_generate_variant:Nn function should only be applied if the \emph{parent control sequence} is already defined. (This is only enforced if debugging support \texttt{check-declarations} is enabled.) If the \emph{parent control sequence} is protected or if the \emph{variant} involves any \texttt{x} argument, then the \emph{variant control sequence} is also protected. The \emph{variant} is created globally, as is any \texttt{exp_args:N(\emph{variant})} function needed to carry out the expansion. There is no need to re-apply \cs_generate_variant:Nn after changing the definition of the parent function: the variant will always use the current definition of the parent. Providing variants repeatedly is safe as \cs_generate_variant:Nn will only create new definitions if there is not already one available.

Only \texttt{n} and \texttt{N} arguments can be changed to other types. The only allowed changes are
\begin{itemize}
\item \texttt{c} variant of an \texttt{N} parent;
\item \texttt{o}, \texttt{V}, \texttt{v}, \texttt{f}, \texttt{e}, or \texttt{x} variant of an \texttt{n} parent;
\item \texttt{N}, \texttt{n}, \texttt{T}, \texttt{F}, or \texttt{p} argument unchanged.
\end{itemize}
This means the \emph{parent} of a \emph{variant} form is always unambiguous, even in cases where both an \texttt{n}-type parent and an \texttt{N}-type parent exist, such as for \texttt{\tl_count:n} and \texttt{\tl_count:N}.

When creating variants for conditional functions, \texttt{\prg_generate_conditional_variant:Nnn} provides a convenient way of handling the related function set.

For backward compatibility it is currently possible to make \texttt{n}, \texttt{o}, \texttt{V}, \texttt{v}, \texttt{f}, \texttt{e}, or \texttt{x}-type variants of an \texttt{N}-type argument or \texttt{N} or \texttt{c}-type variants of an \texttt{n}-type argument. Both are deprecated. The first because passing more than one token to an \texttt{N}-type argument will typically break the parent function’s code. The second because programmers who use that most often want to access the value of a variable given its name, hence should use a \texttt{V}-type or \texttt{v}-type variant instead of \texttt{c}-type. In those cases, using the lower-level \texttt{\exp_args:No} or \texttt{\exp_args:Nc} functions explicitly is preferred to defining confusing variants.
\exp_args_generate:n \exp_args_generate:n {(\textit{variant argument specifiers})}

Defines $\exp_args:N(\textit{variant})$ functions for each $\textit{variant}$ given in the comma list $\{(\textit{variant argument specifiers})\}$. Each $\textit{variant}$ should consist of the letters N, c, n, V, v, o, f, e, x, p and the resulting function is protected if the letter x appears in the $\textit{variant}$. This is only useful for cases where $\cs_generate_variant:Nn$ is not applicable.

### 5.3 Introducing the variants

The V type returns the value of a register, which can be one of tl, clist, int, skip, dim, muskip, or built-in TeX registers. The v type is the same except it first creates a control sequence out of its argument before returning the value.

In general, the programmer should not need to be concerned with expansion control. When simply using the content of a variable, functions with a V specifier should be used. For those referred to by (cs)name, the v specifier is available for the same purpose. Only when specific expansion steps are needed, such as when using delimited arguments, should the lower-level functions with o specifiers be employed.

The e type expands all tokens fully, starting from the first. More precisely the expansion is identical to that of TeX’s $\texttt{\message}$ (in particular # needs not be doubled). It relies on the primitive $\texttt{\expanded}$ hence is fast.

The x type expands all tokens fully, starting from the first. In contrast to e, all macro parameter characters # must be doubled, and omitting this leads to low-level errors. In addition this type of expansion is not expandable, namely functions that have x in their signature do not themselves expand when appearing inside e or x expansion.

The f type is so special that it deserves an example. It is typically used in contexts where only expandable commands are allowed. Then x-expansion cannot be used, and f-expansion provides an alternative that expands the front of the token list as much as can be done in such contexts. For instance, say that we want to evaluate the integer expression $3 + 4$ and pass the result 7 as an argument to an expandable function $\texttt{\example:n}$. For this, one should define a variant using $\cs_generate_variant:Nn \example:n { f }$, then do

\begin{verbatim}
\example:f \{ \int_eval:n \{ 3 + 4 \} \}
\end{verbatim}

Note that x-expansion would also expand $\texttt{\int_eval:n}$ fully to its result 7, but the variant $\example:x$ cannot be expandable. Note also that o-expansion would not expand $\texttt{\int_eval:n}$ fully to its result since that function requires several expansions. Besides the fact that x-expansion is protected rather than expandable, another difference between f-expansion and x-expansion is that f-expansion expands tokens from the beginning and stops as soon as a non-expandable token is encountered, while x-expansion continues expanding further tokens. Thus, for instance

\begin{verbatim}
\example:f \{ \int_eval:n \{ 1 + 2 \} , \int_eval:n \{ 3 + 4 \} \}
\end{verbatim}

results in the call

\begin{verbatim}
\example:n \{ 3 , \int_eval:n \{ 3 + 4 \} \}
\end{verbatim}

while using $\example:x$ or $\example:e$ instead results in

\begin{verbatim}
\example:n \{ 3 , 7 \}
\end{verbatim}
at the cost of being protected for x-type. If you use f type expansion in conditional processing then you should stick to using T\text{F} type functions only as the expansion does not finish any \textit{if}... \textit{fi}: itself!

It is important to note that both f- and o-type expansion are concerned with the expansion of tokens from left to right in their arguments. In particular, o-type expansion applies to the first token in the argument it receives: it is conceptually similar to

\begin{verbatim}
\exp_after:wN <base function> \exp_after:wN { <argument> }
\end{verbatim}

At the same time, f-type expansion stops at the first non-expandable token. This means for example that both

\begin{verbatim}
\tl_set:No \l_tmpa_tl \{ \{ \gtmpb_tl \} \}
\end{verbatim}

and

\begin{verbatim}
\tl_set:Nf \l_tmpa_tl \{ \{ \gtmpb_tl \} \}
\end{verbatim}

leave \texttt{\gtmpb_tl} unchanged: \texttt{\{} is the first token in the argument and is non-expandable.

It is usually best to keep the following in mind when using variant forms.

- Variants with x-type arguments (that are fully expanded before being passed to the n-type base function) are never expandable even when the base function is. Such variants cannot work correctly in arguments that are themselves subject to expansion. Consider using f or e expansion.

- In contrast, e expansion (full expansion, almost like x except for the treatment of #) does not prevent variants from being expandable (if the base function is).

- Finally f expansion only expands the front of the token list, stopping at the first non-expandable token. This may fail to fully expand the argument.

When speed is essential (for functions that do very little work and whose variants are used numerous times in a document) the following considerations apply because the speed of internal functions that expand the arguments of a base function depend on what needs doing with each argument and where this happens in the list of arguments:

- for fastest processing any c-type arguments should come first followed by all other modified arguments;

- unchanged N-type args that appear before modified ones have a small performance hit;

- unchanged n-type args that appear before modified ones have a relative larger performance hit.

\section{Manipulating the first argument}

These functions are described in detail: expansion of multiple tokens follows the same rules but is described in a shorter fashion.
This function absorbs two arguments (the \textit{function} name and the \textit{tokens}). The \textit{tokens} are expanded until only characters remain, and are then turned into a control sequence. The result is inserted into the input stream after reinsertion of the \textit{function}. Thus the \textit{function} may take more than one argument: all others are left unchanged.

The \texttt{:cc} variant constructs the \textit{function} name in the same manner as described for the \textit{tokens}.

This function absorbs two arguments (the \textit{function} name and the \textit{tokens}). The \textit{tokens} are expanded once, and the result is inserted in braces into the input stream after reinsertion of the \textit{function}. Thus the \textit{function} may take more than one argument: all others are left unchanged.

This function absorbs two arguments (the names of the \textit{function} and the \textit{variable}). The content of the \textit{variable} are recovered and placed inside braces into the input stream after reinsertion of the \textit{function}. Thus the \textit{function} may take more than one argument: all others are left unchanged.

This function absorbs two arguments (the \textit{function} name and the \textit{tokens}). The \textit{tokens} are expanded until only characters remain, and are then turned into a control sequence. This control sequence should be the name of a \textit{variable}. The content of the \textit{variable} are recovered and placed inside braces into the input stream after reinsertion of the \textit{function}. Thus the \textit{function} may take more than one argument: all others are left unchanged.

This function absorbs two arguments (the \textit{function} name and the \textit{tokens}) and exhaustively expands the \textit{tokens}. The result is inserted in braces into the input stream after reinsertion of the \textit{function}. Thus the \textit{function} may take more than one argument: all others are left unchanged.

This function absorbs two arguments (the \textit{function} name and the \textit{tokens}) and exhaustively expands the \textit{tokens}. The result is inserted in braces into the input stream after reinsertion of the \textit{function}. Thus the \textit{function} may take more than one argument: all others are left unchanged.
5.5 Manipulating two arguments

\begin{verbatim}
\exp_args:NNc ⟨token1⟩ ⟨token2⟩ \{\{tokens\}\}
\end{verbatim}

These optimized functions absorb three arguments and expand the second and third as detailed by their argument specifier. The first argument of the function is then the next item on the input stream, followed by the expansion of the second and third arguments.

\begin{verbatim}
\exp_args:Nnc ⟨token⟩ \{\{tokens1\}\} \{\{tokens2\}\}
\end{verbatim}

These functions absorb three arguments and expand the second and third as detailed by their argument specifier. The first argument of the function is then the next item on the input stream, followed by the expansion of the second and third arguments.

\begin{verbatim}
\exp_args:Nx ⟨token1⟩ ⟨token2⟩ \{\{tokens\}\}
\end{verbatim}

These functions absorb three arguments and expand the second and third as detailed by their argument specifier. The first argument of the function is then the next item on the input stream, followed by the expansion of the second and third arguments. These functions are not expandable due to their x-type argument.

5.6 Manipulating three arguments

\begin{verbatim}
\exp_args:NNNo ⟨token1⟩ ⟨token2⟩ ⟨token3⟩ \{\{tokens\}\}
\end{verbatim}

These optimized functions absorb four arguments and expand the second, third and fourth as detailed by their argument specifier. The first argument of the function is then the next item on the input stream, followed by the expansion of the second argument, etc.
These functions absorb four arguments and expand the second, third and fourth as detailed by their argument specifier. The first argument of the function is then the next item on the input stream, followed by the expansion of the second argument, etc.
5.7 Unbraced expansion

These functions absorb the number of arguments given by their specification, carry out the expansion indicated and leave the results in the input stream, with the last argument not surrounded by the usual braces. Of these, the \exp_last_unbraced:Nno, \exp_last_unbraced:Noo, \exp_last_unbraced:Nfo and \exp_last_unbraced:NnNo variants need slower processing.

\TeXhackers note: As an optimization, the last argument is unbraced by some of those functions before expansion. This can cause problems if the argument is empty: for instance, \exp_last_unbraced:Nf \foo_bar:w { } \q_stop leads to an infinite loop, as the quark is f-expanded.

\exp_last_unbraced:Nx \exp_last_unbraced:Nx (function) \{(tokens)\}

This function fully expands the \langle tokens \rangle and leaves the result in the input stream after reinsertion of the \langle function \rangle. This function is not expandable.

\exp_last_two_unbraced:Noo \exp_last_two_unbraced:Noo (token) \{(tokens)\} \{(tokens)\}

This function absorbs three arguments and expands the second and third once. The first argument of the function is then the next item on the input stream, followed by the expansion of the second and third arguments, which are not wrapped in braces. This function needs special (slower) processing.

\exp_after:wN \exp_after:wN (token1) (token2)

Carries out a single expansion of \langle token2 \rangle (which may consume arguments) prior to the expansion of \langle token1 \rangle. If \langle token2 \rangle has no expansion (for example, if it is a character) then it is left unchanged. It is important to notice that \langle token1 \rangle may be any single token, including group-opening and -closing tokens (\{ or \}) assuming normal \TeX\ category codes). Unless specifically required this should be avoided: expansion should be carried out using an appropriate argument specifier variant or the appropriate \exp_after:V variant function.

\TeXhackers note: This is the \TeX primitive \expandafter.
5.8 Preventing expansion

Despite the fact that the following functions are all about preventing expansion, they’re designed to be used in an expandable context and hence are all marked as being ‘expandable’ since they themselves disappear after the expansion has completed.

\exp_not:N \token

Prevents expansion of the \token in a context where it would otherwise be expanded, for example an e-type or x-type argument or the first token in an o-type or f-type argument.

\TeXhackers note: This is the \TeX primitive \noexpand. It only prevents expansion. At the beginning of an f-type argument, a space \token is removed even if it appears as \exp_not:N \c_space_token. In an e-expanding definition (\cs_new:Npe), a macro parameter introduces an argument even if it appears as \exp_not:N # 1. This differs from \exp_not:n.

\exp_not:c {\tokens}

Expands the \tokens until only characters remain, and then converts this into a control sequence. Further expansion of this control sequence is then inhibited using \exp_not:N.

\exp_not:n {\tokens}

Prevents expansion of the \tokens in an e-type or x-type argument. In all other cases the \tokens continue to be expanded, for example in the input stream or in other types of arguments such as c, f, v. The argument of \exp_not:n must be surrounded by braces.

\TeXhackers note: This is the \eTeX primitive \unexpanded. In an e-expanding definition (\cs_new:Npe), \exp_not:n {#1} is equivalent to ##1 rather than to #1, namely it inserts the two characters # and 1, and \exp_not:n {#} is equivalent to #, namely it inserts the character #.

\exp_not:o {\tokens}

Expands the \tokens once, then prevents any further expansion in e-type or x-type arguments using \exp_not:n.

\exp_not:V \variable

Recovering the content of the \variable, then prevents expansion of this material in e-type or x-type arguments using \exp_not:n.

\exp_not:v {\tokens}

Expands the \tokens until only characters remains, and then converts this into a control sequence which should be a \variable name. The content of the \variable is recovered, and further expansion in e-type or x-type arguments is prevented using \exp_not:n.
\[\exp_not:e \mathrel{\star} \exp_not:e\{\langle\text{tokens}\rangle}\]

Expands \textit{(tokens)} exhaustively, then protects the result of the expansion (including any tokens which were not expanded) from further expansion in \texttt{e}-type or \texttt{x}-type arguments using \texttt{\exp_not:n}. This is very rarely useful but is provided for consistency.

\[\exp_not:f \mathrel{\star} \exp_not:f\{\langle\text{tokens}\rangle}\]

Expands \textit{(tokens)} fully until the first unexpandable token is found (if it is a space it is removed). Expansion then stops, and the result of the expansion (including any tokens which were not expanded) is protected from further expansion in \texttt{e}-type or \texttt{x}-type arguments using \texttt{\exp_not:n}.

\[\exp_stop_f: \mathrel{\star} \foo_bar:f\{\langle\text{tokens}\rangle\} \exp_stop_f: \langle\text{more tokens}\rangle\}

This function terminates an \texttt{f}-type expansion. Thus if a function \texttt{\foo_bar:f} starts an \texttt{f}-type expansion and all of \textit{(tokens)} are expandable \texttt{\exp_stop_f:} terminates the expansion of tokens even if \textit{(more tokens)} are also expandable. The function itself is an implicit space token. Inside an \texttt{e}-type or \texttt{x}-type expansion, it retains its form, but when typeset it produces the underlying space (\texttt{␣}).

### 5.9 Controlled expansion

The \texttt{expl3} language makes all efforts to hide the complexity of \TeX{} expansion from the programmer by providing concepts that evaluate/expand arguments of functions prior to calling the “base” functions. Thus, instead of using many \texttt{\expandafter} calls and other trickery it is usually a matter of choosing the right variant of a function to achieve a desired result.

Of course, deep down \TeX{} is using expansion as always and there are cases where a programmer needs to control that expansion directly; typical situations are basic data manipulation tools. This section documents the functions for that level. These commands are used throughout the kernel code, but we hope that outside the kernel there will be little need to resort to them. Instead the argument manipulation methods document above should usually be sufficient.

While \texttt{\exp_after:wN} expands one token (out of order) it is sometimes necessary to expand several tokens in one go. The next set of commands provide this functionality. Be aware that it is absolutely required that the programmer has full control over the tokens to be expanded, i.e., it is not possible to use these functions to expand unknown input as part of \texttt{(expandable-tokens)} as that will break badly if unexpandable tokens are encountered in that place!
\texttt{\textbackslash exp:w \textbackslash exp_end:} \texttt{\textbackslash exp:w} \texttt{\langle expandable tokens \rangle} \texttt{\exp_end:} \texttt{\exp_end:} \\

Expands \texttt{\langle expandable-tokens \rangle} until reaching \texttt{\exp_end:} at which point expansion stops. The full expansion of \texttt{\langle expandable tokens \rangle} has to be empty. If any token in \texttt{\langle expandable tokens \rangle} or any token generated by expanding the tokens therein is not expandable the expansion will end prematurely and as a result \texttt{\exp_end:} will be misinterpreted later on.\footnote{Due to the implementation you might get the character in position 0 in the current font (typically ‘‘’’) in the output without any error message!}

In typical use cases the \texttt{\exp_end:} is hidden somewhere in the replacement text of \texttt{\langle expandable-tokens \rangle} rather than being on the same expansion level than \texttt{\exp:w}, e.g., you may see code such as

\texttt{\exp:w \textbackslash@@_case:NnTF \#1 \{ \} \{ \}}

where somewhere during the expansion of \texttt{\textbackslash@@_case:NnTF} the \texttt{\exp_end:} gets generated.

\textbf{\TeX}hackers note: The current implementation uses \texttt{\romannumeral} hence ignores space tokens and explicit signs $+$ and $-$ in the expansion of the \texttt{\langle expandable tokens \rangle}, but this should not be relied upon.

\texttt{\textbackslash exp:w \textbackslash exp_after:wN \{ \textbackslash exp:w \textbackslash exp_end_continue_f:w \textbackslash w \#2 \}}

where the \texttt{\exp_after:wN} triggers an \texttt{f}-expansion of the tokens in \texttt{\#2}. For technical reasons this has to happen using two tokens (if they would be hidden inside another command \texttt{\exp_after:wN} would only expand the command but not trigger any additional \texttt{f}-expansion).

You might wonder why there are two different approaches available, after all the effect of

\texttt{\exp:w \langle expandable-tokens \rangle \exp_end:}

\texttt{\exp:w \exp_end_continue_f:w \langle further-tokens \rangle}

\texttt{\exp_stop_f:}

\texttt{\exp_stop_f:}

The reason is simply that the first approach is slightly faster (one less token to parse and less expansion internally) so in places where such performance really matters and where we want to explicitly stop the expansion at a defined point the first form is preferable.

\texttt{\exp:w \textbackslash exp_end_continue_f:w \textbackslash w \langle further-tokens \rangle}

\texttt{\exp_stop_f:}
The difference to \texttt{exp_end_continue_f:wn} is that we first we pick up an argument which is then returned to the input stream. If \texttt{(further-tokens)} starts with space tokens then these space tokens are removed while searching for the argument. If it starts with a brace group then the braces are removed. Thus such spaces or braces will not terminate the \texttt{f-type} expansion.

5.10 Internal functions

\begin{verbatim}
\cs_new:Npn \exp_args:Ncnf { ::c ::o ::f ::: }
\cs_new:Npn \exp_last_unbraced:Nno { ::n ::o_unbraced ::: }
\end{verbatim}

Internal forms for the base expansion types. These names do \textit{not} conform to the general \LaTeX{} approach as this makes them more readily visible in the log and so forth. They should not be used outside this module.

\begin{verbatim}
\cs_new:Npn \exp_last_unbraced:Nno { ::n ::o_unbraced ::: }
\end{verbatim}

Internal forms for the expansion types which leave the terminal argument unbraced. These names do \textit{not} conform to the general \LaTeX{} approach as this makes them more readily visible in the log and so forth. They should not be used outside this module.

\footnote{In this particular case you may get a character into the output as well as an error message.}
Chapter 6

The l3sort module
Sorting functions

6.1 Controlling sorting

L3TeX comes with a facility to sort list variables (sequences, token lists, or comma-lists) according to some user-defined comparison. For instance,

\clist_set:Nn \l_foo_clist { 3 , 01 , -2 , 5 , +1 }
\clist_sort:Nn \l_foo_clist
{\int_compare:nNnTF { #1 } > { #2 }
{ \sort_return_swapped: }
{ \sort_return_same: }
}

results in \l_foo_clist holding the values \{ -2 , 01 , +1 , 3 , 5 \} sorted in non-decreasing order.

The code defining the comparison should call \sort_return_swapped: if the two items given as #1 and #2 are not in the correct order, and otherwise it should call \sort_return_same: to indicate that the order of this pair of items should not be changed.

For instance, a ⟨comparison code⟩ consisting only of \sort_return_same: with no test yields a trivial sort: the final order is identical to the original order. Conversely, using a ⟨comparison code⟩ consisting only of \sort_return_swapped: reverses the list (in a fairly inefficient way).

\textbf{TeXhackers note:} The current implementation is limited to sorting approximately 20000 items (40000 in LuaTeX), depending on what other packages are loaded.

Internally, the code from l3sort stores items in \toks registers allocated locally. Thus, the ⟨comparison code⟩ should not call \newtoks or other commands that allocate new \toks registers. On the other hand, altering the value of a previously allocated \toks register is not a problem.
\sort_return_same:  \seq_sort:Nn \seq \var
\sort_return_swapped:  \{ ... \sort_return_same: or \sort_return_swapped: ... \}  

Indicates whether to keep the order or swap the order of two items that are compared in the sorting code. Only one of the \sort_return... functions should be used by the code, according to the results of some tests on the items #1 and #2 to be compared.
Chapter 7

The \l3tl-analysis module

Analysing token lists

This module provides functions that are particularly useful in the \l3regex module for mapping through a token list one \texttt{token} at a time (including begin-group/end-group tokens). For \texttt{tl-analysis_map_inline:Nn} or \texttt{tl-analysis_map_inline:nn}, the token list is given as an argument; the analogous function \texttt{peek_analysis_map_inline:n} documented in \l3token finds tokens in the input stream instead. In both cases the user provides \texttt{(inline code)} that receives three arguments for each \texttt{token}:

- \texttt{tokens}, which both o-expand and e/x-expand to the \texttt{token}. The detailed form of \texttt{tokens} may change in later releases.
- \texttt{char code}, a decimal representation of the character code of the \texttt{token}, \(-1\) if it is a control sequence.
- \texttt{catcode}, a capital hexadecimal digit which denotes the category code of the \texttt{token} (0: control sequence, 1: begin-group, 2: end-group, 3: math shift, 4: alignment tab, 6: parameter, 7: superscript, 8: subscript, A: space, B: letter, C: other, D: active). This can be converted to an integer by writing \"\texttt{catcode}\".

In addition, there is a debugging function \texttt{tl-analysis_show:n}, very similar to the \texttt{ShowTokens} macro from the \texttt{ted} package.

\begin{verbatim}
\tl_analysis_show:N \tl_analysis_show:n \{token list\}
\tl_analysis_log:n \{token list\}
\tl_analysis_log:N \tl_analysis_log:N
\end{verbatim}

Displays to the terminal (or log) the detailed decomposition of the \texttt{token list} into tokens, showing the category code of each character token, the meaning of control sequences and active characters, and the value of registers.

\begin{verbatim}
\tl_analysis_map_inline:nn \tl_analysis_map_inline:nn \{token list\} \{inline function\}
\tl_analysis_map_inline:Nn
\end{verbatim}

Applies the \texttt{inline function} to each individual \texttt{token} in the \texttt{token list}. The \texttt{inline function} receives three arguments as explained above. As all other mappings the mapping is done at the current group level, \textit{i.e.} any local assignments made by the \texttt{inline function} remain in effect after the loop.

New: 2021-05-11
Updated: 2022-03-26

46
Chapter 8

The l3regex module

Regular expressions in \TeX

The l3regex module provides regular expression testing, extraction of submatches, splitting, and replacement, all acting on token lists. The syntax of regular expressions is mostly a subset of the PCRE syntax (and very close to POSIX), with some additions due to the fact that \TeX manipulates tokens rather than characters. For performance reasons, only a limited set of features are implemented. Notably, back-references are not supported.

Let us give a few examples. After

\begin{verbatim}
\tl_set:Nn \l_my_tl { That\text{-}cat. }
\regex_replace_once:nnN { at } { is } \l_my_tl
\end{verbatim}

the token list variable \l_my_tl holds the text “This cat.”, where the first occurrence of “at” was replaced by “is”. A more complicated example is a pattern to emphasize each word and add a comma after it:

\begin{verbatim}
\regex_replace_all:nnN { \w+ } { \c{emph}\cB\{ \0 \cE\} , } \l_my_tl
\end{verbatim}

The \w sequence represents any “word” character, and + indicates that the \w sequence should be repeated as many times as possible (at least once), hence matching a word in the input token list. In the replacement text, \0 denotes the full match (here, a word). The command \emph is inserted using \c{emph}, and its argument \0 is put between braces \cB\{ and \cE\}.

If a regular expression is to be used several times, it can be compiled once, and stored in a regex variable using \regex_set:Nn. For example,

\begin{verbatim}
\regex_new:N \l_foo_regex
\regex_set:Nn \l_foo_regex { \c{begin} \cB. (\c[^BE].*) \cE. }
\end{verbatim}

stores in \l_foo_regex a regular expression which matches the starting marker for an environment: \begin, followed by a begin-group token (\cB.), then any number of tokens which are neither begin-group nor end-group character tokens (\c[^BE].*), ending with an end-group token (\cE.). As explained in the next section, the parentheses “capture” the result of \c[^BE].*, giving us access to the name of the environment when doing replacements.
8.1 Syntax of regular expressions

8.1.1 Regular expression examples

We start with a few examples, and encourage the reader to apply \regex_show:n to these regular expressions.

- **Cat** matches the word “Cat” capitalized in this way, but also matches the beginning of the word “Cattle”: use \bCat\b to match a complete word only.

- **[abc]** matches one letter among “a”, “b”, “c”; the pattern (a|b|c) matches the same three possible letters (but see the discussion of submatches below).

- **[A-Za-z]** matches any number (due to the quantifier *) of Latin letters (not accented).

- \c{[A-Za-z]**} matches a control sequence made of Latin letters.

- \_.*?\_ matches an underscore, any number of characters other than underscore, and another underscore; it is equivalent to \.\*?\_ where . matches arbitrary characters and the lazy quantifier *? means to match as few characters as possible, thus avoiding matching underscores.

- \[\+\-]?\d+ matches an explicit integer with at most one sign.

- \[\+\-\]\*\d+\* matches an explicit integer with any number of + and − signs, with spaces allowed except within the mantissa, and surrounded by spaces.

- \[\+\-\]\*\(\d+\)\d+\* matches an explicit integer or decimal number; using [.,] instead of \. would allow the comma as a decimal marker.

- \[\+\-\]\*\(\d+\|\d*\.|\d+\)\* matches an explicit dimension with any unit that \TeX{} knows, where (?i) means to treat lowercase and uppercase letters identically.

- \[\+\-\]\*\(\(?i\)nan\|inf\|\(\d+\|\d*\.|\d+\)\(\(\(?i\)pt\|in\|\(cem\|m\|ex\|bs\|dn\)\|\(pcn\)\)\)\*\ matches an explicit floating point number or the special values nan and inf (with signs and spaces allowed).

- \[\+\-\]\*\(\d+\|\cC\.)\* matches an explicit integer or control sequence (without checking whether it is an integer variable).

- \G.*?\K at the beginning of a regular expression matches and discards (due to \K) everything between the end of the previous match (\G) and what is matched by the rest of the regular expression; this is useful in \regex_replace_all:nnN when the goal is to extract matches or submatches in a finer way than with \regex_extract_all:nnN.

While it is impossible for a regular expression to match only integer expressions, \(\(+\-\)(\d+\)\*\(+\+\-\)\)(\(+\-\)(\*\d+\)\)\* matches among other things all valid integer expressions (made only with explicit integers). One should follow it with further testing.
8.1.2 Characters in regular expressions

Most characters match exactly themselves, with an arbitrary category code. Some characters are special and must be escaped with a backslash (e.g., \* matches a star character). Some escape sequences of the form backslash-letter also have a special meaning (for instance \d matches any digit). As a rule,

- every alphanumeric character (A–Z, a–z, 0–9) matches exactly itself, and should not be escaped, because \A, \B, ... have special meanings;
- non-alphanumeric printable ascii characters can (and should) always be escaped: many of them have special meanings (e.g., use \, \), \?, \\;,
- spaces should always be escaped (even in character classes);
- any other character may be escaped or not, without any effect: both versions match exactly that character.

Note that these rules play nicely with the fact that many non-alphanumeric characters are difficult to input into \TeX\ under normal category codes. For instance, \abc\% matches the characters \abc\% (with arbitrary category codes), but does not match the control sequence \abc followed by a percent character. Matching control sequences can be done using the \c{⟨regex⟩} syntax (see below).

Any special character which appears at a place where its special behaviour cannot apply matches itself instead (for instance, a quantifier appearing at the beginning of a string), after raising a warning.

Characters.
\x{hh...} Character with hex code hh...
\xhh Character with hex code hh.
\a Alarm (hex 07).
\e Escape (hex 1B).
\f Form-feed (hex 0C).
\n New line (hex 0A).
\r Carriage return (hex 0D).
\t Horizontal tab (hex 09).

8.1.3 Characters classes

Character properties.
.
\d Any decimal digit.
\h Any horizontal space character, equivalent to [\ \^^I\]: space and tab.
\s Any space character, equivalent to [\ \^^I\\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^-\^
\v Any vertical space character, equivalent to [\^M\^L\^K\^J]. Note that \^K is a vertical space, but not a space, for compatibility with Perl.

\w Any word character, *i.e.*, alphanumerics and underscore, equivalent to the explicit class [A-Za-z0-9\_].

\D Any token not matched by \d.

\H Any token not matched by \h.

\N Any token other than the \n character (hex 0A).

\S Any token not matched by \s.

\V Any token not matched by \v.

\W Any token not matched by \w.

Of those, \D, \H, \N, \S, \V, and \W match arbitrary control sequences. Character classes match exactly one token in the subject.

[... ] Positive character class. Matches any of the specified tokens.

[^... ] Negative character class. Matches any token other than the specified characters.

[x-y] Within a character class, this denotes a range (can be used with escaped characters).

[:\(name):] Within a character class (one more set of brackets), this denotes the POSIX character class \(\text{name}\), which can be alnum, alpha, ascii, blank, cntrl, digit, graph, lower, print, punct, space, upper, word, or xdigit.

[:^\(name):] Negative POSIX character class.

For instance, [a-oq-z\cC.] matches any lowercase latin letter except p, as well as control sequences (see below for a description of \c).

In character classes, only [, ^, ], \ and spaces are special, and should be escaped. Other non-alphanumeric characters can still be escaped without harm. Any escape sequence which matches a single character (\d, \D, etc.) is supported in character classes. If the first character is ^, then the meaning of the character class is inverted; ^ appearing anywhere else in the range is not special. If the first character (possibly following a leading ^) is ] then it does not need to be escaped since ending the range there would make it empty. Ranges of characters can be expressed using -, for instance, [\D 0-5] and [^6-9] are equivalent.

8.1.4 Structure: alternatives, groups, repetitions

Quantifiers (repetition).

? 0 or 1, greedy.

?? 0 or 1, lazy.

* 0 or more, greedy.

*? 0 or more, lazy.

+ 1 or more, greedy.
+? 1 or more, lazy.
{n} Exactly \(n\).
{n,} \(n\) or more, greedy.
{n,}? \(n\) or more, lazy.
{n, m} At least \(n\), no more than \(m\), greedy.
{n, m}? At least \(n\), no more than \(m\), lazy.

For greedy quantifiers the regex code will first investigate matches that involve as many repetitions as possible, while for lazy quantifiers it investigates matches with as few repetitions as possible first.

Alternation and capturing groups.

A|B|C Either one of A, B, or C, investigating A first.

(... ) Capturing group.

(? : . . ) Non-capturing group.

(?! . . ) Non-capturing group which resets the group number for capturing groups in each alternative. The following group is numbered with the first unused group number.

Capturing groups are a means of extracting information about the match. Parenthesized groups are labelled in the order of their opening parenthesis, starting at 1. The contents of those groups corresponding to the “best” match (leftmost longest) can be extracted and stored in a sequence of token lists using for instance \regex_extract_once.nnNTF.

The \K escape sequence resets the beginning of the match to the current position in the token list. This only affects what is reported as the full match. For instance,

\regex_extract_all.nnN { a \K . } { a123aaxyz } \l_foo_seq
results in \l_foo_seq containing the items \{1\} and \{a\}: the true matches are \{a1\} and \{aa\}, but they are trimmed by the use of \K. The \K command does not affect capturing groups: for instance,

\regex_extract_once.nnN { (. \K c)+ \d } { acbc3 } \l_foo_seq
results in \l_foo_seq containing the items \{c3\} and \{bc\}: the true match is \{acbc3\}, with first submatch \{bc\}, but \K resets the beginning of the match to the last position where it appears.

8.1.5 Matching exact tokens
The \c escape sequence allows to test the category code of tokens, and match control sequences. Each character category is represented by a single uppercase letter:

- C for control sequences;
- B for begin-group tokens;
- E for end-group tokens;
• M for math shift;
• T for alignment tab tokens;
• P for macro parameter tokens;
• U for superscript tokens (up);
• D for subscript tokens (down);
• S for spaces;
• L for letters;
• 0 for others; and
• A for active characters.

The \c escape sequence is used as follows.

\c\{⟨regex⟩\} A control sequence whose csname matches the ⟨regex⟩, anchored at the beginning and end, so that \c\{begin\} matches exactly \begin, and nothing else.

\cX Applies to the next object, which can be a character, escape character sequence such as \x{0A}, character class, or group, and forces this object to only match tokens with category X (any of CBEMTPUDSLOA). For instance, \cL[A-Z\d] matches uppercase letters and digits of category code letter, \cC matches any control sequence, and \c0(abc) matches abc where each character has category other.\footnote{This last example also captures “abc” as a regex group; to avoid this use a non-capturing group \c0(?:abc).}

\c[XYZ] Applies to the next object, and forces it to only match tokens with category X, Y, or Z (each being any of CBEMTPUDSLOA). For instance, \c[LSO](...) matches two tokens of category letter, space, or other.

\c[^XYZ] Applies to the next object and prevents it from matching any token with category X, Y, or Z (each being any of CBEMTPUDSLOA). For instance, \c[^O]\d matches digits which have any category different from other.

The category code tests can be used inside classes; for instance, [\cO\d \c[LO][A-F]] matches what \TeX considers as hexadecimal digits, namely digits with category other, or uppercase letters from \texttt{A} to \texttt{F} with category either letter or other. Within a group affected by a category code test, the outer test can be overridden by a nested test: for instance, \cL(ab\c0\*cd) matches ab*cd where all characters are of category letter, except \* which has category other.

The \u escape sequence allows to insert the contents of a token list directly into a regular expression or a replacement, avoiding the need to escape special characters. Namely, \u\{(var name)\} matches the exact contents (both character codes and category codes) of the variable \texttt{(var name)}, which are obtained by applying \exp_not:v \{(var name)\} at the time the regular expression is compiled. Within a \c\{\ldots\} control sequence matching, the \u escape sequence only expands its argument once, in effect performing \tl_to_str:v. Quantifiers are supported.

The \ur escape sequence allows to insert the contents of a regex variable into a larger regular expression. For instance, A\ur{l_tmpa_regex}D matches the tokens A and
D separated by something that matches the regular expression \l_tmpa_regex. This behaves as if a non-capturing group were surrounding \l_tmpa_regex, and any group contained in \l_tmpa_regex is converted to a non-capturing group. Quantifiers are supported.

For instance, if \l_tmpa_regex has value B|C, then A\ur{l_tmpa_regex}D is equivalent to A(?:B|C)D (matching ABD or ACD) and not to AB|CD (matching AB or CD). To get the latter effect, it is simplest to use \TeX{}’s expansion machinery directly: if \l_{-}mymodule_BC_tl contains B|C then the following two lines show the same result:

\begin{verbatim}
\regex_show:n { A \u{l_mymodule_BC_tl} D }
\regex_show:n { A B | C D }
\end{verbatim}

8.1.6 Miscellaneous

Anchors and simple assertions.

\b Word boundary: either the previous token is matched by \w and the next by \W, or the opposite. For this purpose, the ends of the token list are considered as \W.

\B Not a word boundary: between two \w tokens or two \W tokens (including the boundary).

\^ or \A Start of the subject token list.

\$ or \z End of the subject token list.

\G Start of the current match. This is only different from ^ in the case of multiple matches: for instance \regex_count:nnN { \G a } { aaba } \l_tmpa_int yields 2, but replacing \G by ^ would result in \l_tmpa_int holding the value 1.

The option (?i) makes the match case insensitive (treating A–Z and a–z as equivalent, with no support yet for Unicode case changing). This applies until the end of the group in which it appears, and can be reverted using (?-i). For instance, in (?i)(a(?-i)b|c)d, the letters a and d are affected by the i option. Characters within ranges and classes are affected individually: (?i)\[?\~B\] is equivalent to \[\~ABab\] (and differs from the much larger class \[?\~b\]), and (?i)\[^aeiou\] matches any character which is not a vowel. The i option has no effect on \c{…}, on \u{…}, on character properties, or on character classes, for instance it has no effect at all in (?i)\u{l_foo_tl}\d\d\[[:lower:]].

8.2 Syntax of the replacement text

Most of the features described in regular expressions do not make sense within the replacement text. Backslash introduces various special constructions, described further below:

- \0 is the whole match;
- \1 is the submatch that was matched by the first (capturing) group (…); similarly for \2, …, \9 and \g{number};
- \u inserts a space (spaces are ignored when not escaped);
• \a, \e, \f, \n, \r, \t, \xhh, \x{hhh} correspond to single characters as in regular expressions;
• \c{⟨es name⟩} inserts a control sequence;
• \c{⟨category⟩⟨character⟩} (see below);
• \u{⟨tl var name⟩} inserts the contents of the ⟨tl var⟩ (see below).

Characters other than backslash and space are simply inserted in the result (but since the replacement text is first converted to a string, one should also escape characters that are special for TeX, for instance use \#). Non-alphanumeric characters can always be safely escaped with a backslash.

For instance,
\tl_set:Nn \l_my_tl { Hello,-world! }
\regex_replace_all:nnN { (\[er\]?l|o) . } { (\0--\1) } \l_my_tl
results in \l_my_tl holding H(ell--el)(o,--o) w(or--o)(ld--l)!

The submatches are numbered according to the order in which the opening parenthesis of capturing groups appear in the regular expression to match. The n-th submatch is empty if there are fewer than n capturing groups or for capturing groups that appear in alternatives that were not used for the match. In case a capturing group matches several times during a match (due to quantifiers) only the last match is used in the replacement text. Submatches always keep the same category codes as in the original token list.

By default, the category code of characters inserted by the replacement are determined by the prevailing category code regime at the time where the replacement is made, with two exceptions:
• space characters (with character code 32) inserted with \␣ or \x20 or \x{20} have category code 10 regardless of the prevailing category code regime;
• if the category code would be 0 (escape), 5 (newline), 9 (ignore), 14 (comment) or 15 (invalid), it is replaced by 12 (other) instead.

The escape sequence \c allows to insert characters with arbitrary category codes, as well as control sequences.
\cX(...) Produces the characters “...” with category X, which must be one of CBEMTPUDSLOA as in regular expressions. Parentheses are optional for a single character (which can be an escape sequence). When nested, the innermost category code applies, for instance \cL(Hello\cS\ world)! gives this text with standard category codes.

\c{⟨text⟩} Produces the control sequence with csname ⟨text⟩. The ⟨text⟩ may contain references to the submatches \0, \1, and so on, as in the example for \u below.

The escape sequence \u{⟨var name⟩} allows to insert the contents of the variable with name ⟨var name⟩ directly into the replacement, giving an easier control of category codes. When nested in \c{...} and \u{...} constructions, the \u and \c escape sequences perform \tl_to_str:v, namely extract the value of the control sequence and turn it into a string. Matches can also be used within the arguments of \c and \u. For instance,
\tl_set:Nn \l_my_one_tl { first }
\tl_set:Nn \l_my_two_tl { \emph{second} }
\tl_set:Nn \l_my_tl { one , two , one , one }
\regex_replace_all:nnN { [^,]+ } { \u{l_my_\0_tl} } \l_my_tl
results in \l_my_tl holding first, emph{second}, first, first.

Regex replacement is also a convenient way to produce token lists with arbitrary category codes. For instance
\begin{verbatim}
\tl_clear:N \l_tmpa_tl
\regex_replace_all:nnN { } { \cU\% \cA\~ } \l_tmpa_tl
\end{verbatim}
results in \l_tmpa_tl containing the percent character with category code 7 (superscript) and an active tilde character.

### 8.3 Pre-compiling regular expressions

If a regular expression is to be used several times, it is better to compile it once rather than doing it each time the regular expression is used. The compiled regular expression is stored in a variable. All of the l3regex module’s functions can be given their regular expression argument either as an explicit string or as a compiled regular expression.

\begin{verbatim}
\regex_new:N \l_my_regex
\regex_set:Nn \l_my_regex { my\ (simple\ )? reg(ex|ular\ expression) }
\end{verbatim}

\begin{verbatim}
\regex_new:N \l_my_regex
\regex_set:Nn \l_my_regex { my\ (simple\ )? reg(ex|ular\ expression) }
\end{verbatim}

\begin{verbatim}
\regex_const:Nn \l_my_regex \regex_set:Nn \l_my_regex { my\ (simple\ )? reg(ex|ular\ expression) }
\end{verbatim}

\begin{verbatim}
\regex_show:n \regex_log:n \regex_log:N
\end{verbatim}

Displays in the terminal or writes in the log file (respectively) how l3regex interprets the \regex. For instance, \regex_show:n {\A X|Y} shows

\begin{verbatim}
+-branch
  anchor at start (\A)
  char code 88 (X)
+-branch
  char code 89 (Y)
\end{verbatim}

indicating that the anchor \A only applies to the first branch: the second branch is not anchored to the beginning of the match.
8.4 Matching

All regular expression functions are available in both :n and :N variants. The former require a “standard” regular expression, while the latter require a compiled expression as generated by \regex_set:Nn.

\regex_match:nnTF \{ ⟨regex⟩ \} \{ ⟨token list⟩ \} \{ ⟨true code⟩ \} \{ ⟨false code⟩ \}

Tests whether the (regular expression) matches any part of the (token list). For instance,

\regex_match:nnTF \{ b [cde]* \} \{ abecdx \} \{ TRUE \} \{ FALSE \}
\regex_match:nnTF \{ [b-dq-w] \} \{ example \} \{ TRUE \} \{ FALSE \}

leaves TRUE then FALSE in the input stream.

\regex_count:nnN \{ ⟨regex⟩ \} \{ ⟨token list⟩ \} \{ ⟨int var⟩ \}

Sets ⟨int var⟩ within the current \TeX{} group level equal to the number of times (regular expression) appears in (token list). The search starts by finding the left-most longest match, respecting greedy and lazy (non-greedy) operators. Then the search starts again from the character following the last character of the previous match, until reaching the end of the token list. Infinite loops are prevented in the case where the regular expression can match an empty token list: then we count one match between each pair of characters. For instance,

\int_new:N \l_foo_int
\regex_count:nnN \{ (b+|c) \} \{ abbababcbbb \} \l_foo_int

results in \l_foo_int taking the value 5.

\regex_match_case:nnTF \{ ⟨regex⟩ \} \{ ⟨code case⟩ \} \{ ⟨true code⟩ \} \{ ⟨false code⟩ \}

Determines which of the (regular expressions) matches at the earliest point in the (token list), and leaves the corresponding (code e) followed by the (true code) in the input stream. If several (regex) match starting at the same point, then the first one in the list is selected and the others are discarded. If none of the (regex) match, the (false code) is left in the input stream. Each (regex) can either be given as a regex variable or as an explicit regular expression.

In detail, for each starting position in the (token list), each of the (regex) is searched in turn. If one of them matches then the corresponding (code e) is used and everything else is discarded, while if none of the (regex) match at a given position then the next starting position is attempted. If none of the (regex) match anywhere in the (token list) then nothing is left in the input stream. Note that this differs from nested \regex_match:nnTF statements since all (regex) are attempted at each position rather than attempting to match (regex) at every position before moving on to (regex)2.

56
8.5 Submatch extraction

\regex_extract_once:nnN \regex_extract_once:nVN \regex_extract_once:nnN \regex_extract_once:nVN \regex_extract_once:NnN \regex_extract_once:NVN
\regex_extract_all:nnN \regex_extract_all:nVN \regex_extract_all:nnN ... \regex_extract_all:nVN \regex_extract_all:NnN \regex_extract_all:NVN

\regex_extract_once:nnN \regex_extract_once:nVN \regex_extract_once:nnN \regex_extract_once:nVN \regex_extract_once:NnN \regex_extract_once:NVN
\regex_extract_all:nnN \regex_extract_all:nVN \regex_extract_all:nnN ... \regex_extract_all:nVN \regex_extract_all:NnN \regex_extract_all:NVN

Finds the first match of the (regular expression) in the (token list). If it exists, the match is stored as the first item of the (seq var), and further items are the contents of capturing groups, in the order of their opening parenthesis. The (seq var) is assigned locally. If there is no match, the (seq var) is cleared. The testing versions insert the (true code) into the input stream if a match was found, and the (false code) otherwise.

For instance, assume that you type

\regex_extract_once:nnNTF \regex_extract_once:nVN \regex_extract_once:nnNTF \regex_extract_once:NnN \regex_extract_once:NVN
\regex_extract_all:nnNTF \regex_extract_all:nVN \regex_extract_all:nnNTF \regex_extract_all:nVN \regex_extract_all:NnN \regex_extract_all:NVN

Then the regular expression (anchored at the start with \A and at the end with \Z) must match the whole token list. The first capturing group, (La)?, matches La, and the second capturing group, (!*), matches !!!. Thus, \l_foo_seq contains as a result the items {LaTeX!!!}, {La}, and {!!!}, and the true branch is left in the input stream. Note that the n-th item of \l_foo_seq, as obtained using \seq_item:Nn, correspond to the submatch numbered (n − 1) in functions such as \regex_replace_once:nnN.

\regex_extract_all:nnNTF \regex_extract_all:nVN \regex_extract_all:nnNTF \regex_extract_all:nVN \regex_extract_all:NnN \regex_extract_all:NVN
\regex_extract_all:nnNTF \regex_extract_all:nVN \regex_extract_all:nnNTF \regex_extract_all:nVN \regex_extract_all:NnN \regex_extract_all:NVN

Finds all matches of the (regular expression) in the (token list), and stores all the submatch information in a single sequence (concatenating the results of multiple \regex_extract_once:nnN calls). The (seq var) is assigned locally. If there is no match, the (seq var) is cleared. The testing versions insert the (true code) into the input stream if a match was found, and the (false code) otherwise. For instance, assume that you type

\regex_extract_all:nnNTF \regex_extract_all:nVN \regex_extract_all:nnNTF \regex_extract_all:nVN \regex_extract_all:NnN \regex_extract_all:NVN
\regex_extract_all:nnNTF \regex_extract_all:nVN \regex_extract_all:nnNTF \regex_extract_all:nVN \regex_extract_all:NnN \regex_extract_all:NVN

Then the regular expression matches twice, the resulting sequence contains the two items {Hello} and {world}, and the true branch is left in the input stream.
Splits the ⟨token list⟩ into a sequence of parts, delimited by matches of the ⟨regular expression⟩. If the ⟨regular expression⟩ has capturing groups, then the token lists that they match are stored as items of the sequence as well. The assignment to ⟨seq var⟩ is local. If no match is found the resulting ⟨seq var⟩ has the ⟨token list⟩ as its sole item. If the ⟨regular expression⟩ matches the empty token list, then the ⟨token list⟩ is split into single tokens. The testing versions insert the ⟨true code⟩ into the input stream if a match was found, and the ⟨false code⟩ otherwise. For example, after\[
\seq_new:N \l_path_seq
\regex_split:nnNTF { / } { the/path/for/this/file.tex } \l_path_seq
\{ true \} \{ false \}
\]the sequence \l_path_seq contains the items {the}, {path}, {for}, {this}, and {file.tex}, and the true branch is left in the input stream.

## 8.6 Replacement

Searches for the ⟨regular expression⟩ in the contents of the ⟨tl var⟩ and replaces the first match with the ⟨replacement⟩. In the ⟨replacement⟩, \0 represents the full match, \1 represent the contents of the first capturing group, \2 of the second, etc. The result is assigned locally to ⟨tl var⟩.

Replaces all occurrences of the ⟨regular expression⟩ in the contents of the ⟨tl var⟩ by the ⟨replacement⟩, where \0 represents the full match, \1 represent the contents of the first capturing group, \2 of the second, etc. Every match is treated independently, and matches cannot overlap. The result is assigned locally to ⟨tl var⟩.
Replaces the earliest match of the regular expression \( ?(?\texttt{regex}_1)\ldots(?\texttt{regex}_n) \) in the \( \langle \text{token list variable} \rangle \) by the \( \langle \text{replacement} \rangle \) corresponding to which \( \texttt{regex}_i \) matched, then leaves the \( \langle \text{true code} \rangle \) in the input stream. If none of the \( \texttt{regex}_i \) match, then the \( \langle \text{tl var} \rangle \) is not modified, and the \( \langle \text{false code} \rangle \) is left in the input stream. Each \( \texttt{regex}_i \) can either be given as a \texttt{regex} variable or as an explicit regular expression.

In detail, for each starting position in the \( \langle \text{token list} \rangle \), each of the \( \langle \texttt{regex} \rangle \) is searched in turn. If one of them matches then it is replaced by the corresponding \( \langle \text{replacement} \rangle \) as described for \texttt{regex_replace_once:nN}. This is equivalent to checking with \texttt{regex_match_case:nn} which \( \texttt{regex}_i \) matches, then performing the replacement with \texttt{regex_replace_once:nN}.

Replaces all occurrences of all \( \langle \texttt{regex} \rangle \) in the \( \langle \text{token list} \rangle \) by the corresponding \( \langle \text{replacement} \rangle \). Every match is treated independently, and matches cannot overlap. The result is assigned locally to \( \langle \text{tl var} \rangle \), and the \( \langle \text{true code} \rangle \) or \( \langle \text{false code} \rangle \) is left in the input stream depending on whether any replacement was made or not.

In detail, for each starting position in the \( \langle \text{token list} \rangle \), each of the \( \langle \texttt{regex} \rangle \) is searched in turn. If one of them matches then it is replaced by the corresponding \( \langle \text{replacement} \rangle \), and the search resumes at the position that follows this match (and replacement). For instance

\begin{verbatim}
\tl_set:Nn \l_tmpa_tl { Hello,-world! }
\regex_replace_case_all:nN
  { \[A-Za-z\]+ } { ''\0'' }
  { \b } { --- }
  { . } { \[\0\] }
\l_tmpa_tl
\end{verbatim}

results in \texttt{\l_tmpa_tl} having the contents ‘‘Hello’’---[,][,]‘‘world’’---[!]'. Note in particular that the word-boundary assertion \texttt{\b} did not match at the start of words because the case \texttt{[A-Za-z\]+} matched at these positions. To change this, one could simply swap the order of the two cases in the argument of \texttt{regex_replace_case_all:nN}.
8.7 Scratch regular expressions

\l_tmpa_regex
\l_tmpb_regex

Scratch regex for local assignment. These are never used by the kernel code, and so are safe for use with any \LaTeX3-defined function. However, they may be overwritten by other non-kernel code and so should only be used for short-term storage.

\l_tmpa_regex
\l_tmpb_regex

Scratch regex for global assignment. These are never used by the kernel code, and so are safe for use with any \LaTeX3-defined function. However, they may be overwritten by other non-kernel code and so should only be used for short-term storage.

8.8 Bugs, misfeatures, future work, and other possibilities

The following need to be done now.

- Rewrite the documentation in a more ordered way, perhaps add a BNF?
  Additional error-checking to come.
- Clean up the use of messages.
- Cleaner error reporting in the replacement phase.
- Add tracing information.
- Detect attempts to use back-references and other non-implemented syntax.
- Test for the maximum register \c_max_register_int.
- Find out whether the fact that \W and friends match the end-marker leads to bugs. Possibly update \_\_regex_item_reverse:n.
- The empty cs should be matched by \c{}, not by \c{csname.?endcsname\s?}.
  Code improvements to come.
- Shift arrays so that the useful information starts at position 1.
- Only build \c{}\ldots{} once.
- Use arrays for the left and right state stacks when compiling a regex.
- Should \_\_regex_action_free_group:n only be used for greedy {n,} quantifier? (I think not.)
- Quantifiers for \u and assertions.
- When matching, keep track of an explicit stack of \curr_state and \curr_submatches.
- If possible, when a state is reused by the same thread, kill other subthreads.
• Use an array rather than \g__regex_balance_tl to build the function \__regex_replacement_balance_one_match:n.
• Reduce the number of epsilon-transitions in alternatives.
• Optimize simple strings: use less states (abcade should give two states, for abc and ade). [Does that really make sense?]
• Optimize groups with no alternative.
• Optimize states with a single \__regex_action_free:n.
• Optimize the use of \__regex_action_success: by inserting it in state 2 directly instead of having an extra transition.
• Optimize the use of \int_step.... functions.
• Groups don’t capture within regexes for csnames; optimize and document.
• Better “show” for anchors, properties, and catcode tests.
• Does \K really need a new state for itself?
• When compiling, use a boolean in_cs and less magic numbers.

The following features are likely to be implemented at some point in the future.

• General look-ahead/behind assertions.
• Regex matching on external files.
• Conditional subpatterns with look ahead/behind: “if what follows is […], then […].”
• (.*...) and (?..) sequences to set some options.
• UTF-8 mode for pdfTeX.
• Newline conventions are not done. In particular, we should have an option for . not to match newlines. Also, \A should differ from ^, and \Z, \z and $ should differ.
• Unicode properties: \p{..} and \P{..}; \X which should match any “extended” Unicode sequence. This requires to manipulate a lot of data, probably using tree-boxes.

The following features of PCRE or Perl may or may not be implemented.

• Callout with (?C...) or other syntax: some internal code changes make that possible, and it can be useful for instance in the replacement code to stop a regex replacement when some marker has been found; this raises the question of a potential \regex_break: and then of playing well with \tl_map_break: called from within the code in a regex. It also raises the question of nested calls to the regex machinery, which is a problem since \fontdimen are global.
• Conditional subpatterns (other than with a look-ahead or look-behind condition): this is non-regular, isn’t it?
• Named subpatterns: \TeX{} programmers have lived so far without any need for named macro parameters.

The following features of \texttt{pcre} or Perl will definitely not be implemented.

• Back-references: non-regular feature, this requires backtracking, which is prohibitively slow.

• Recursion: this is a non-regular feature.

• Atomic grouping, possessive quantifiers: those tools, mostly meant to fix catastrophic backtracking, are unnecessary in a non-backtracking algorithm, and difficult to implement.

• Subroutine calls: this syntactic sugar is difficult to include in a non-backtracking algorithm, in particular because the corresponding group should be treated as atomic.

• Backtracking control verbs: intrinsically tied to backtracking.

• \texttt{\textbackslash ddd}, matching the character with octal code \texttt{ddd}: we already have \texttt{\textbackslash x{...}} and the syntax is confusingly close to what we could have used for backreferences (\texttt{\textbackslash1, \textbackslash2, ...}), making it harder to produce useful error message.

• \texttt{\textbackslash cx}, similar to \TeX{}’s own \texttt{\textasciitilde x}.

• Comments: \TeX{} already has its own system for comments.

• \texttt{\textbackslash Q...\textbackslash E} escaping: this would require to read the argument verbatim, which is not in the scope of this module.

• \texttt{\textbackslash c single byte in UTF-8 mode: Xe\TeX{} and Lua\TeX{} serve us characters directly, and splitting those into bytes is tricky, encoding dependent, and most likely not useful anyways.}
Chapter 9

The l3prg module

Control structures

Conditional processing in \input{tex3} is defined as something that performs a series of tests, possibly involving assignments and calling other functions that do not read further ahead in the input stream. After processing the input, a state is returned. The states returned are $\langle \text{true} \rangle$ and $\langle \text{false} \rangle$.

\input{tex3} has two forms of conditional flow processing based on these states. The first form is predicate functions that turn the returned state into a boolean $\langle \text{true} \rangle$ or $\langle \text{false} \rangle$. For example, the function $\text{cs_if_free_p:N}$ checks whether the control sequence given as its argument is free and then returns the boolean $\langle \text{true} \rangle$ or $\langle \text{false} \rangle$ values to be used in testing with $\text{if_predicate:w}$ or in functions to be described below. The second form is the kind of functions choosing a particular argument from the input stream based on the result of the testing as in $\text{cs_if_free:NTF}$ which also takes one argument (the \text{N}) and then executes either $\text{true}$ or $\text{false}$ depending on the result.

\TeX hackers note: The arguments are executed after exiting the underlying $\text{if...fi}$ structure.

9.1 Defining a set of conditional functions

\prg_new_conditional:Npnn $\langle \text{name} \rangle$:\langle arg spec \rangle \langle parameters \rangle \{(conditions)\} \{(code)\}
\prg_set_conditional:Npnn $\langle \text{name} \rangle$:\langle arg spec \rangle \{(conditions)\} \{(code)\}
\prg_gset_conditional:Npnn $\langle \text{name} \rangle$:\langle arg spec \rangle \{(conditions)\} \{(code)\}

These functions create a family of conditionals using the same $\langle \text{code} \rangle$ to perform the test created. Those conditionals are expandable if $\langle \text{code} \rangle$ is. The \text{new} versions check for existing definitions and perform assignments globally (cf. $\text{cs_new:Npn}$) whereas the \text{set} versions do no check and perform assignments locally (cf. $\text{cs_set:Npn}$). The conditionals created are dependent on the comma-separated list of $\langle \text{conditions} \rangle$, which should be one or more of \text{p}, \text{T}, \text{F} and \text{TF}.
These functions create a family of protected conditionals using the same \{code\} to perform the test created. The \{code\} does not need to be expandable. The new version check for existing definitions and perform assignments globally (cf. \cs_new:Npn) whereas the set version do not (cf. \cs_set:Npn). The conditionals created are depended on the comma-separated list of \{conditions\}, which should be one or more of $T$, $F$ and $TF$ (not $p$).

The conditionals are defined by \prg_new_protected_conditional:Npnn and friends as:

- \langle\text{name}\rangle_p:⟨\text{arg spec}⟩ — a predicate function which will supply either a logical true or logical false. This function is intended for use in cases where one or more logical tests are combined to lead to a final outcome. This function cannot be defined for protected conditionals.

- \langle\text{name}\rangle:⟨\text{arg spec}⟩T — a function with one more argument than the original \langle\text{arg spec}\rangle demands. The \langle true branch\rangle code in this additional argument will be left on the input stream only if the test is true.

- \langle\text{name}\rangle:⟨\text{arg spec}⟩F — a function with one more argument than the original \langle\text{arg spec}\rangle demands. The \langle false branch\rangle code in this additional argument will be left on the input stream only if the test is false.

- \langle\text{name}\rangle:⟨\text{arg spec}⟩TF — a function with two more argument than the original \langle\text{arg spec}\rangle demands. The \langle true branch\rangle code in the first additional argument will be left on the input stream if the test is true, while the \langle false branch\rangle code in the second argument will be left on the input stream if the test is false.

The \langle code\rangle of the test may use \langle parameters\rangle as specified by the second argument to \prg_set_protected_conditional:Nnn: this should match the \langle argument specification\rangle but this is not enforced. The Nnn versions infer the number of arguments from the argument specification given (cf. \cs_new:Nn, etc.). Within the \langle code\rangle, the functions \prg_return_true: and \prg_return_false: are used to indicate the logical outcomes of the test.

An example can easily clarify matters here:

\begin{verbatim}
\prg_set_protected_conditional:Nnn \foo_if_bar:Nn #1#2 { p , T , TF } {
  \if_meaning:w \l_tmpa_tl #1  \prg_return_true:
  \else:
    \if_meaning:w \l_tmpa_tl #2  \prg_return_true:
      \else:
        \prg_return_false:
      \fi:
  \fi:
\end{verbatim}
This defines the function \texttt{\textbackslash foo\_if\_bar\_p:NN}, \texttt{\textbackslash foo\_if\_bar:NTF} and \texttt{\textbackslash foo\_if\_bar:NT} but not \texttt{\textbackslash foo\_if\_bar:NNF} (because \texttt{F} is missing from the \texttt{(conditions)} list). The return statements take care of resolving the remaining \texttt{\textbackslash else:} and \texttt{\textbackslash fi:} before returning the state. There must be a return statement for each branch; failing to do so will result in erroneous output if that branch is executed.

The special case where the code of a conditional ends with \texttt{\textbackslash prg\_return\_true:} \texttt{\textbackslash else:} \texttt{\textbackslash prg\_return\_false:} \texttt{\textbackslash fi:} is optimized.

These functions copy a family of conditionals. The \texttt{\textbackslash new} version checks for existing definitions (cf. \texttt{\textbackslash cs\_new\_eq:NN}) whereas the set version does not (cf. \texttt{\textbackslash cs\_set\_eq:NN}). The conditionals copied are dependent on the comma-separated list of \texttt{(conditions)}, which should be one or more of \texttt{p}, \texttt{T}, \texttt{F} and \texttt{TF}.

These “return” functions define the logical state of a conditional statement. They appear within the code for a conditional function generated by \texttt{\textbackslash prg\_set\_conditional:Npnn}, etc., to indicate when a true or false branch should be taken. While they may appear multiple times each within the code of such conditionals, the execution of the conditional must result in the expansion of one of these two functions \textit{exactly once}.

The return functions trigger what is internally an f-expansion process to complete the evaluation of the conditional. Therefore, after \texttt{\textbackslash prg\_return\_true:} or \texttt{\textbackslash prg\_return\_false:} there must be no non-expandable material in the input stream for the remainder of the expansion of the conditional code. This includes other instances of either of these functions.

Defines argument-specifier variants of conditionals. This is equivalent to running \texttt{\textbackslash cs\_generate\_variant:NN \textbackslash conditional \{\textit{variant argument specifiers}\} \{\textit{condition specifiers}\}} on each \texttt{(conditional)} described by the \texttt{(condition specifiers)}. These base-form \texttt{(conditionals)} are obtained from the \texttt{(name)} and \texttt{\textbackslash arg spec} as described for \texttt{\textbackslash prg\_new\_conditional:Npnn}, and they should be defined.

\section{The boolean data type}

This section describes a boolean data type which is closely connected to conditional processing as sometimes you want to execute some code depending on the value of a switch (e.g., draft/final) and other times you perhaps want to use it as a predicate function in an \texttt{\textbackslash if\_predicate:w} test. The problem of the primitive \texttt{\textbackslash if\_false:} and
\if_true: tokens is that it is not always safe to pass them around as they may interfere with scanning for termination of primitive conditional processing. Therefore, we employ two canonical booleans: \texttt{\c_true_bool} or \texttt{\c_false_bool}. Besides preventing problems as described above, it also allows us to implement a simple boolean parser supporting the logical operations And, Or, Not, etc. which can then be used on both the boolean type and predicate functions.

All conditional \texttt{\bool_} functions except assignments are expandable and expect the input to also be fully expandable (which generally means being constructed from predicate functions and booleans, possibly nested).

\TeXhackers note: The bool data type is not implemented using the \texttt{\iffalse}/\texttt{\iftrue} primitives, in contrast to \texttt{\newif}, etc., in plain \TeX, \LaTeX\ and so on. Programmers should not base use of \texttt{bool} switches on any particular expectation of the implementation.

\begin{Verbatim}
\bool_new:N \bool_new:c
\bool_const:Nn \bool_const:cn
\bool_set_false:N \bool_set_false:c \bool_gset_false:N \bool_gset_false:c
\bool_set_true:N \bool_set_true:c \bool_gset_true:N \bool_gset_true:c
\bool_set_eq:NN \bool_set_eq:(cN|Nc|cc) \bool_gset_eq:NN \bool_gset_eq:(cN|Nc|cc)
\bool_set:Nn \bool_set:cn \bool_gset:Nn \bool_gset:cn
\bool_set_inverse:N \bool_set_inverse:c \bool_gset_inverse:N \bool_gset_inverse:c
\end{Verbatim}
Tests the current truth of \texttt{boolean}, and continues expansion based on this result.

\bool_to_str:N \bool_to_str:c \bool_to_str:n \texttt{Updated: 2021-11-01} \texttt{Updated: 2023-11-14}

Expands to the string \texttt{true} or \texttt{false} depending on the logical truth of the \texttt{boolean} or \texttt{boolean expression}.

\bool_show:N \bool_show:c \bool_show:n \texttt{New: 2012-02-09} \texttt{Updated: 2021-04-29}

Displays the logical truth of the \texttt{boolean expression} on the terminal.

\bool_log:N \bool_log:c \bool_log:n \texttt{New: 2014-08-22} \texttt{Updated: 2021-04-29}

Writes the logical truth of the \texttt{boolean} in the log file.

\bool_if_exist_p:N \bool_if_exist_p:c \bool_if_exist:N \bool_if_exist:c \texttt{New: 2012-03-03} \texttt{New: 2012-03-03}

Tests whether the \texttt{boolean} is currently defined. This does not check that the \texttt{boolean} really is a boolean variable.

\texttt{Updated: 2017-07-15}

\texttt{9.2.1 Constant and scratch booleans}

\texttt{\c_true_bool \c_false_bool}

Constants that represent \texttt{true} and \texttt{false}, respectively. Used to implement predicates.

\texttt{\l_tmpa_bool \l_tmpb_bool}

A scratch boolean for local assignment. It is never used by the kernel code, and so is safe for use with any \LaTeX3-defined function. However, it may be overwritten by other non-kernel code and so should only be used for short-term storage.
A scratch boolean for global assignment. It is never used by the kernel code, and so is safe for use with any \LaTeX-defined function. However, it may be overwritten by other non-kernel code and so should only be used for short-term storage.

### 9.3 Boolean expressions

As we have a boolean datatype and predicate functions returning boolean \langle true or false\rangle values, it seems only fitting that we also provide a parser for \langle boolean expressions\rangle.

A boolean expression is an expression which given input in the form of predicate functions and boolean variables, return boolean \langle true or false\rangle. It supports the logical operations And, Or and Not as the well-known infix operators \&\& and || and prefix ! with their usual precedences (namely, \&\& binds more tightly than ||). In addition to this, parentheses can be used to isolate sub-expressions. For example,

\begin{verbatim}
\int_compare_p:n { 1 = 1 } \&\&
 ( \int_compare_p:n { 2 = 3 } ||
 \int_compare_p:n { 4 <= 4 } ||
 \str_if_eq_p:nn { abc } { def } ) \&\&
 ! \int_compare_p:n { 2 = 4 }
\end{verbatim}

is a valid boolean expression.

Contrarily to some other programming languages, the operators \&\& and || evaluate both operands in all cases, even when the first operand is enough to determine the result. This “eager” evaluation should be contrasted with the “lazy” evaluation of \texttt{bool_lazy-...} functions.

\textbf{\TeX hackers note:} The eager evaluation of boolean expressions is unfortunately necessary in \TeX. Indeed, a lazy parser can get confused if \&\& or || or parentheses appear as (unbraced) arguments of some predicates. For instance, the innocuous-looking expression below would break (in a lazy parser) if \#1 were a closing parenthesis and \l_tmpa_bool were true.

\begin{verbatim}
( \l_tmpa_bool || \token_if_eq_meaning_p:NN X \#1 )
\end{verbatim}

Minimal (lazy) evaluation can be obtained using the conditionals \texttt{bool_lazy-all:nTF}, \texttt{bool_lazy_and:nnTF}, \texttt{bool_lazy_any:nTF}, or \texttt{bool_lazy_or:nnTF}, which only evaluate their boolean expression arguments when they are needed to determine the resulting truth value. For example, when evaluating the boolean expression

\begin{verbatim}
\bool_lazy_and_p:nn
 { \bool_lazy_any_p:n
   { \int_compare_p:n { 2 = 3 } }
   { \int_compare_p:n { 4 <= 4 } }
   { \int_compare_p:n { 1 = \error } } % skipped
   }
 { ! \int_compare_p:n { 2 = 4 } }
\end{verbatim}

68
the line marked with skipped is not expanded because the result of `bool_lazy_any_p:n` is known once the second boolean expression is found to be logically true. On the other hand, the last line is expanded because its logical value is needed to determine the result of `bool_lazy_and_p:nn`.

\[ \text{Updated: 2017-07-15} \]
\[ \text{New: 2015-11-15} \]

Tests the current truth of `\langle \text{boolean expression} \rangle`, and continues expansion based on this result. The `\langle \text{boolean expression} \rangle` should consist of a series of predicates or boolean variables with the logical relationship between these defined using `&&` ("And"), `||` ("Or"), `!` ("Not") and parentheses. The logical Not applies to the next predicate or group.

\[ \text{Updated: 2017-07-15} \]
\[ \text{New: 2015-11-15} \]

Implements the “And” operation on the `\langle \text{boolean expressions} \rangle`, hence is true if all of them are true and false if any of them is false. Contrarily to the infix operator `&&`, only the `\langle \text{boolean expressions} \rangle` which are needed to determine the result of `\text{bool_lazy_all:nTF}` are evaluated. See also `\text{bool_lazy_all:nTF}` when there are only two `\langle \text{boolean expressions} \rangle`.

\[ \text{Updated: 2017-07-15} \]
\[ \text{New: 2015-11-15} \]

Implements the “And” operation between two boolean expressions, hence is true if both are true. Contrarily to the infix operator `&&`, the `\langle \text{boolean expressions} \rangle` is only evaluated if it is needed to determine the result of `\text{bool_lazy_and:nnTF}`. See also `\text{bool_lazy_all:nTF}` when there are more than two `\langle \text{boolean expressions} \rangle`.

\[ \text{Updated: 2017-07-15} \]
\[ \text{New: 2015-11-15} \]

Implements the “Or” operation on the `\langle \text{boolean expressions} \rangle`, hence is true if any of them is true and false if all of them are false. Contrarily to the infix operator `||`, only the `\langle \text{boolean expressions} \rangle` which are needed to determine the result of `\text{bool_lazy_any:nTF}` are evaluated. See also `\text{bool_lazy_all:nTF}` when there are more than two `\langle \text{boolean expressions} \rangle`.

\[ \text{Updated: 2017-07-15} \]
\[ \text{New: 2015-11-15} \]

Implements the “Or” operation between two boolean expressions, hence is true if either one is true. Contrarily to the infix operator `||`, the `\langle \text{boolean expressions} \rangle` is only evaluated if it is needed to determine the result of `\text{bool_lazy_or:nnTF}`. See also `\text{bool_lazy_any:nTF}` when there are more than two `\langle \text{boolean expressions} \rangle`.

\[ \text{Updated: 2017-07-15} \]

Function version of `!(\langle \text{boolean expression} \rangle)` within a boolean expression.
\bool_xor_p:nn \{boolexpr1\} \{boolexpr2\}
\bool_xor:nnTF \{boolexpr1\} \{boolexpr2\} \{true code\} \{false code\}

Implements an “exclusive or” operation between two boolean expressions. There is no infix operation for this logical operation.

9.4 Logical loops

Loops using either boolean expressions or stored boolean values.

\bool_do_until:Nn \{boolean\} \{code\}
\bool_do_until:cn \{boolean\} \{code\}

Places the \{code\} in the input stream for \TeX{} to process, and then checks the logical value of the \{boolean\}. If it is \textit{false} then the \{code\} is inserted into the input stream again and the process loops until the \{boolean\} is \textit{true}.

\bool_until_do:Nn \{boolean\} \{code\}
\bool_until_do:cn \{boolean\} \{code\}

This function first checks the logical value of the \{boolean\}. If it is \textit{false} the \{code\} is placed in the input stream and expanded. After the completion of the \{code\} the truth of the \{boolean\} is re-evaluated. The process then loops until the \{boolean\} is \textit{true}.

\bool_while_do:Nn \{boolean\} \{code\}
\bool_while_do:cn \{boolean\} \{code\}

This function first checks the logical value of the \{boolean\}. If it is \textit{true} the \{code\} is placed in the input stream and expanded. After the completion of the \{code\} the truth of the \{boolean\} is re-evaluated. The process then loops until the \{boolean\} is \textit{false}.

\bool_do_until:nn \{boolean expression\} \{code\}
\bool_do_until:cn \{boolean expression\} \{code\}

Places the \{code\} in the input stream for \TeX{} to process, and then checks the logical value of the \{boolean expression\} as described for \bool_if:nTF. If it is \textit{false} then the \{code\} is inserted into the input stream again and the process loops until the \{boolean expression\} evaluates to \textit{true}.

\bool_do_while:nn \{boolean expression\} \{code\}
\bool_do_while:cn \{boolean expression\} \{code\}

Places the \{code\} in the input stream for \TeX{} to process, and then checks the logical value of the \{boolean expression\} as described for \bool_if:nTF. If it is \textit{true} then the \{code\} is inserted into the input stream again and the process loops until the \{boolean expression\} evaluates to \textit{false}.

\bool_until_do:nn \{boolean expression\} \{code\}
\bool_until_do:cn \{boolean expression\} \{code\}

This function first checks the logical value of the \{boolean expression\} (as described for \bool_if:nTF). If it is \textit{false} the \{code\} is placed in the input stream and expanded. After the completion of the \{code\} the truth of the \{boolean expression\} is re-evaluated. The process then loops until the \{boolean expression\} is \textit{true}.
\bool_while_do:nn \bool_while_do:nn \{
⟨boolean expression⟩\} \{⟨code⟩\}

This function first checks the logical value of the ⟨boolean expression⟩ (as described for \bool_if:nTF). If it is true the ⟨code⟩ is placed in the input stream and expanded. After the completion of the ⟨code⟩ the truth of the ⟨boolean expression⟩ is re-evaluated. The process then loops until the ⟨boolean expression⟩ is false.

\bool_case:n \bool_case:n \{\}
\{
\}\{\}
{
\}\{\}
{
\}\{\}
{
\}\{\}

Evaluates in turn each of the ⟨boolean expression cases⟩ until the first one that evaluates to true. The ⟨code⟩ associated to this first case is left in the input stream, followed by the ⟨true code⟩, and other cases are discarded. If none of the cases match then only the ⟨false code⟩ is inserted. The function \bool_case:n, which does nothing if there is no match, is also available. For example

\bool_case:nF
{\}
{\dim_compare_p:n \\_\_\_mypkg_wd_dim <= 10pt } \{ Fits \}
{\int_compare_p:n \\_\_\_mypkg_total_int >= 10 } \{ Many \}
{\\_\_\_mypkg_special_bool \ Special \}
{ No idea! }

leaves “Fits” or “Many” or “Special” or “No idea!” in the input stream, in a way similar to some other language’s “if ... elseif ... elseif ... else ...”.

9.5 Producing multiple copies

\prg_replicate:nn \prg_replicate:nn \{\}
\{\}

Evaluates the ⟨integer expression⟩ (which should be zero or positive) and creates the resulting number of copies of the ⟨tokens⟩. The function is both expandable and safe for nesting. It yields its result after two expansion steps.

9.6 Detecting \TeX’s mode

\mode_if_horizontal_p: \mode_if_horizontal_p:
\mode_if_horizontal:TF \mode_if_horizontal:TF \{\}
\{\}

Detects if \TeX{} is currently in horizontal mode.
\mode_if_inner_p: \mode_if_inner:TF \{\text{(true code)}\} \{\text{(false code)}\}

Detects if \TeX{} is currently in inner mode.

\mode_if_math_p: \mode_if_math:TF \{\text{(true code)}\} \{\text{(false code)}\}

Detects if \TeX{} is currently in maths mode.

\mode_if_vertical_p: \mode_if_vertical:TF \{\text{(true code)}\} \{\text{(false code)}\}

Detects if \TeX{} is currently in vertical mode.

### 9.7 Primitive conditionals

\if_predicate:w \{\text{predicate}\} \{\text{true code}\} \text{else:} \{\text{false code}\} \text{\textbackslash f i:}

This function takes a predicate function and branches according to the result. (In practice this function would also accept a single boolean variable in place of the \textit{predicate} but to make the coding clearer this should be done through \texttt{\if_bool:N}.)

\if_bool:N \{\text{boolean}\} \{\text{true code}\} \text{else:} \{\text{false code}\} \text{\textbackslash f i:}

This function takes a boolean variable and branches according to the result.

### 9.8 Nestable recursions and mappings

There are a number of places where recursion or mapping constructs are used in expl3. At a low-level, these typically require insertion of tokens at the end of the content to allow “clean up”. To support such mappings in a nestable form, the following functions are provided.

\prg_break_point:Nn \{\text{type}\}_map_break: \{\text{code}\}

Used to mark the end of a recursion or mapping: the functions \texttt{\{type\}_map_break:} and \texttt{\{type\}_map_break:n} use this to break out of the loop (see \texttt{\prg_map_break:Nn} for how to set these up). After the loop ends, the \textit{code} is inserted into the input stream. This occurs even if the break functions are not applied: \texttt{\prg_break_point:Nn} is functionally-equivalent in these cases to \texttt{\use_ii:n}.
Breaks a recursion in mapping contexts, inserting in the input stream the \textit{(user code)} after the \textit{(ending code)} for the loop. The function breaks loops, inserting their \textit{(ending code)}, until reaching a loop with the same \textit{(type)} as its first argument. This \texttt{\textbackslash (type)_map-break:} argument must be defined; it is simply used as a recognizable marker for the \textit{(type)}.

For types with mappings defined in the kernel, \texttt{\textbackslash (type)_map-break:} and \texttt{\textbackslash (type)_map-break:n} are defined as \texttt{\textbackslash prg_map_break:Nn (type)_map_break: \{\}} and the same with \{\} omitted.

9.8.1 Simple mappings

In addition to the more complex mappings above, non-nestable mappings are used in a number of locations and support is provided for these.

\begin{verbatim}
\prg_break:n {\langle code\rangle} ... \prg_break_point:
\prg_break:n \{\langle code\rangle\}
\prg_break_point:
\prg_break:n \{\}
\prg_break_point:
\prg_break:n \{\}
\end{verbatim}

Breaks a recursion which has no \textit{(ending code)} and which is not a user-breakable mapping (see for instance implementation of \texttt{\textbackslash int_step_function:nnnN}), and inserts the \textit{(code)} in the input stream.

9.9 Internal programming functions

\begin{verbatim}
\group_align_safe_begin: ...
\group_align_safe_end: ...
\end{verbatim}

These functions are used to enclose material in a \TeX\ alignment environment within a specially-constructed group. This group is designed in such a way that it does not add brace groups to the output but does act as a group for the \texttt{&} token inside \texttt{\halign}. This is necessary to allow grabbing of tokens for testing purposes, as \TeX\ uses group level to determine the effect of alignment tokens. Without the special grouping, the use of a function such as \texttt{\textbackslash peek_after:Nw} would result in a forbidden comparison of the internal \texttt{\endtemplate} token, yielding a fatal error. Each \texttt{\group_align_safe_begin:} must be matched by a \texttt{\group_align_safe_end:}, although this does not have to occur within the same function.
Chapter 10

The \texttt{l3sys} module
System/runtime functions

10.1 The name of the job

\begin{verbatim}
c_sys_jobname_str
\end{verbatim}

Constant that gets the “job name” assigned when \TeX{} starts.

\textbf{\TeX{}hackers note:} This is the \TeX{} primitive \texttt{\jobname{}}. For technical reasons, the string here is not of the same internal form as other, but may be manipulated using normal string functions.

10.2 Date and time

\begin{verbatim}
c_sys_minute_int
c_sys_hour_int
c_sys_day_int
c_sys_month_int
c_sys_year_int
\end{verbatim}

The date and time at which the current job was started: these are all reported as integers.

\textbf{\TeX{}hackers note:} Whilst the underlying \TeX{} primitives \texttt{\time{}}, \texttt{\day{}}, \texttt{\month{}}, and \texttt{\year{}} can be altered by the user, this interface to the time and date is intended to be the “real” values.

\begin{verbatim}
c_sys_timestamp_str
\end{verbatim}

The timestamp for the current job: the format is as described for \texttt{\file_timestamp:n}.
10.3 Engine

\sys_if_engine_latex_p: ⋆  \sys_if_engine_pdfTeX_p: ⋆ \sys_if_engine_pdftex: TF ⋆ \sys_if_engine_pTeX_p: ⋆ \sys_if_engine_ptex: TF ⋆ \sys_if_engine_uptex_p: ⋆ \sys_if_engine_uptex: TF ⋆ \sys_if_engine_xetex_p: ⋆ \sys_if_engine_xetex: TF ⋆

New: 2015-09-07

\c_sys_engine_str

The current engine given as a lower case string: one of \texttt{luatex}, \texttt{pdftex}, \texttt{pTeX}, \texttt{upTeX} or \texttt{xetex}.

\c_sys_engine_exec_str

The name of the standard executable for the current \TeX{} engine given as a lower case string: one of \texttt{luatex}, \texttt{luahbtex}, \texttt{pdftex}, \texttt{pTeX}, \texttt{upTeX} or \texttt{xetex}.

\c_sys_engine_format_str

The name of the preloaded format for the current \TeX{} run given as a lower case string: one of \texttt{lualatex} (or \texttt{dvilualatex}), \texttt{pdflatex} (or \texttt{latex}), \texttt{platex}, \texttt{uplatex} or \texttt{xelatex} for \LaTeX{}, similar names for plain \TeX{} (except pdf\TeX{} in DVI mode yields \texttt{etex}), and \texttt{cont-en} for Con\TeX{} (i.e. the \texttt{fmtname}).

\c_sys_engine_version_str

The version string of the current engine, in the same form as given in the banner issued when running a job. For pdf\TeX{} and Lua\TeX{} this is of the form

\( \langle \text{major} \rangle.\langle \text{minor} \rangle.\langle \text{revision} \rangle \)

For X\LaTeX{}, the form is

\( \langle \text{major} \rangle.\langle \text{minor} \rangle \)

For \TeX{} and up\TeX{}, only releases since \TeX{} Live 2018 make the data available, and the form is more complex, as it comprises the \TeX{} version, the up\TeX{} version and the e-p\TeX{} version.

\( p(\langle \text{major} \rangle.\langle \text{minor} \rangle.\langle \text{revision} \rangle)-u(\langle \text{major} \rangle.\langle \text{minor} \rangle)-epTeX \)

where the \texttt{u} part is only present for up\TeX{}.

\sys_timer: ⋆ \sys_timer:

Expands to the current value of the engine’s timer clock, a non-negative integer. This function is only defined for engines with timer support. This command measures not just CPU time but real time (including time waiting for user input). The unit are scaled seconds \(2^{-16} \) seconds.
Tests whether current engine has timer support.

### 10.4 Output format

Conditionals which give the current output mode the TeX run is operating in. This is always one of two outcomes, DVI mode or PDF mode. The two sets of conditionals are thus complementary and are both provided to allow the programmer to emphasise the most appropriate case.

\c_sys_output_str

The current output mode given as a lower case string: one of dvi or pdf.

### 10.5 Platform

Conditionals which allow platform-specific code to be used. The names follow the Lua os.type() function, i.e. all Unix-like systems are unix (including Linux and MacOS).

\c_sys_platform_str

The current platform given as a lower case string: one of unix, windows or unknown.

### 10.6 Random numbers

Expands to the current value of the engine's random seed, a non-negative integer. In engines without random number support this expands to 0.
\gset_rand_seed:n \{(int expr)\}

Globally sets the seed for the engine's pseudo-random number generator to the \textit{integer expression}. This random seed affects all \ldots_rand functions (such as \pseudorand or \clist_rand_item:n) as well as other packages relying on the engine's random number generator. In engines without random number support this produces an error.

\textbf{\textit{\TeX} hackers note}: While a 32-bit (signed) integer can be given as a seed, only the absolute value is used and any number beyond \(2^{28}\) is divided by an appropriate power of 2. We recommend using an integer in \([0, 2^{28} - 1]\).

### 10.7 Access to the shell

\texttt{\sys_get_shell:nN} \texttt{\sys_get_shell:nN \{(shell command)\} \{(setup)\} \{tl var\}}

\texttt{\sys_get_shell:nNTF \{(shell command)\} \{(setup)\} \{tl var\} \{\textit{true code}\} \{\textit{false code}\}}

Defines \(\{\texttt{tl var}\}\) to the text returned by the \textit{shell command}. The \textit{shell command} is converted to a string using \texttt{\tl_to_str:n}. Category codes may need to be set appropriately via the \textit{setup} argument, which is run just before running the \textit{shell command} (in a group). If shell escape is disabled, the \(\{\texttt{tl var}\}\) will be set to \texttt{\q_no_value} in the non-branching version. Note that quote characters (\"\) cannot be used inside the \textit{shell command}. The \texttt{\sys_get_shell:nNTF} conditional inserts the \textit{true code} if the shell is available and no quote is detected, and the \textit{false code} otherwise.

\textit{Note}: It is not possible to tell from \TeX{} if a command is allowed in restricted shell escape. If restricted escape is enabled, the \texttt{true} branch is taken: if the command is forbidden at this stage, a low-level \TeX{} error will arise.

\texttt{\c_sys_shell_escape_int}

This variable exposes the internal triple of the shell escape status. The possible values are:

- 0 Shell escape is disabled
- 1 Unrestricted shell escape is enabled
- 2 Restricted shell escape is enabled

\texttt{\sys_if_shell_p: \sys_if_shell:TF}

\texttt{\sys_if_shell:TF \{(true code)\} \{(false code)\}}

Performs a check for whether shell escape is enabled. This returns true if either of restricted or unrestricted shell escape is enabled.

\texttt{\sys_if_shell_unrestricted_p: \sys_if_shell_unrestricted:TF}

\texttt{\sys_if_shell_unrestricted:TF \{(true code)\} \{(false code)\}}

Performs a check for whether \textit{unrestricted} shell escape is enabled.
Perform a check for whether restricted shell escape is enabled. This returns false if unrestricted shell escape is enabled. Unrestricted shell escape is not considered a superset of restricted shell escape in this case. To find whether any shell escape is enabled use \sys_if_shell:TF.

\sys_shell_now:n \sys_shell_now:e
\sys_shell_shipout:n \sys_shell_shipout:e

10.8 System queries

Some queries can be made about the file system, etc., without needing to use unrestricted shell escape. This is carried out using the script l3sys-query, which is documented separately. The wrappers here use this script, if available, to obtain system information that is not directly available within the \TeX run. Note that if restricted shell escape is disabled, no results can be obtained.

\sys_get_query:nN \sys_get_query:nnN \sys_get_query:nnnN
\sys_get_query:nN \sys_get_query:nnN \sys_get_query:nnnN

\sys_get_query:nN \sys_get_query:nN \sys_get_query:nN \sys_get_query:nN
\sys_get_query:nN \sys_get_query:nN \sys_get_query:nN \sys_get_query:nN

Sets the \(\texttt{tl var}\) to the information returned by the l3sys-query \(\langle\texttt{cmd}\rangle\), potentially supplying the \(\langle\texttt{options}\rangle\) and \(\langle\texttt{spec}\rangle\) to the query call. The valid \(\langle\texttt{cmd}\rangle\) names are at present

- `pwd` Returns the present working directory
- `ls` Returns a directory listing, using the \(\langle\texttt{spec}\rangle\) to select files and applying the \(\langle\texttt{options}\rangle\) if given

The \(\langle\texttt{spec}\rangle\) is likely to contain the wildcards \`*\` or \`?\`, and will automatically be passed to the script without shell expansion. In a glob is needed within the \(\langle\texttt{options}\rangle\), this will need to be protected from shell expansion using \`\` tokens.

The \(\langle\texttt{spec}\rangle\) and \(\langle\texttt{options}\rangle\), if given, are expanded fully before passing to the underlying script.

Spaces in the output are stored as active tokens, allowing them to be replaced by for example a visible space easily. Other non-letter characters in the ASCII range are set to category code 12. The category codes for characters out of the ASCII range are left unchanged: typically this will mean that with an 8-bit engine, accented values can be typeset directly whilst in Unicode engines, standard category code setup will apply.

If more than one line of text is returned by the \(\langle\texttt{cmd}\rangle\), these will be separated by character 13 (\`\^\M\`) tokens of category code 12. In most cases, \sys_split_query:nnN should be preferred when multi-line output is expected.
\texttt{\sys_split_query:n} \quad \texttt{\sys_split_query:nn} \quad \texttt{\sys_split_query:nnn}

Works as described for \texttt{\sys_split_query:nnn}, but sets the \texttt{\seq} to contain one entry for each line returned by \texttt{l3sys-query}. This function should therefore be preferred where multi-line return is expected, e.g. for the \texttt{ls} command.

### 10.9 Loading configuration data

\texttt{\sys_load_backend:n} \quad \texttt{\sys_load_backend:n \{\texttt{\backend}\}}

Loads the additional configuration file needed for backend support. If the \texttt{\backend} is empty, the standard backend for the engine in use will be loaded. This command may only be used once.

\texttt{\sys_ensure_backend:} \quad \texttt{\sys_ensure_backend:}

Ensures that a backend has been loaded by calling \texttt{\sys_load_backend:n} if required.

\texttt{\c_sys_backend_str} \quad Set to the name of the backend in use by \texttt{\sys_load_backend:n} when issued. Possible values are

- \texttt{pdftex}
- \texttt{luatex}
- \texttt{xetex}
- \texttt{dvips}
- \texttt{dvipdfmx}
- \texttt{dvisvgm}

\texttt{\sys_load_debug:} \quad \texttt{\sys_load_debug:}

Load the additional configuration file for debugging support.

### 10.9.1 Final settings

\texttt{\sys_finalise:} \quad \texttt{\sys_finalise:}

Finalises all system-dependent functionality: required before loading a backend.
Chapter 11

The l3msg module

Messages

Messages need to be passed to the user by modules, either when errors occur or to indicate how the code is proceeding. The l3msg module provides a consistent method for doing this (as opposed to writing directly to the terminal or log).

The system used by l3msg to create messages divides the process into two distinct parts. Named messages are created in the first part of the process; at this stage, no decision is made about the type of output that the message will produce. The second part of the process is actually producing a message. At this stage a choice of message class has to be made, for example error, warning or info.

By separating out the creation and use of messages, several benefits are available. First, the messages can be altered later without needing details of where they are used in the code. This makes it possible to alter the language used, the detail level and so on. Secondly, the output which results from a given message can be altered. This can be done on a message class, module or message name basis. In this way, message behaviour can be altered and messages can be entirely suppressed.

11.1 Creating new messages

All messages have to be created before they can be used. The text of messages is automatically wrapped to the length available in the console. As a result, formatting is only needed where it helps to show meaning. In particular, \ may be used to force a new line and \ forces an explicit space. Additionally, \#, \%, \~ and \- can be used to produce the corresponding character.

Messages may be subdivided by one level using the / character. This is used within the message filtering system to allow for example the \LaTeX\ kernel messages to belong to the module \LaTeX\ while still being filterable at a more granular level. Thus for example

\msg_new:nnnn { mymodule } { submodule / message } ...

will allow to filter out specifically messages from the submodule.

Some authors may find the need to include spaces as - characters tedious. This can be avoided by locally resetting the category code of -.
\char_set_catcode_space:n { ‘ \ }
\msg_new:nnn { foo } { bar }
{Some message text using ’#1’ and usual message shorthands \{ \ \ \}.}
\char_set_catcode_ignore:n { ‘ \ }

although in general this may be confusing; simply writing the messages using ~ characters is the method favored by the team.

\msg_new:nnnn {⟨module⟩} {⟨message⟩} {⟨text⟩} {⟨more text⟩}
Creates a ⟨message⟩ for a given ⟨module⟩. The message is defined to first give ⟨text⟩ and then ⟨more text⟩ if the user requests it. If no ⟨more text⟩ is available then a standard text is given instead. Within ⟨text⟩ and ⟨more text⟩ four parameters (#1 to #4) can be used: these will be supplied at the time the message is used. An error is raised if the ⟨message⟩ already exists.

\msg_set:nnnn {⟨module⟩} {⟨message⟩} {⟨text⟩} {⟨more text⟩}
Sets up the text for a ⟨message⟩ for a given ⟨module⟩. The message is defined to first give ⟨text⟩ and then ⟨more text⟩ if the user requests it. If no ⟨more text⟩ is available then a standard text is given instead. Within ⟨text⟩ and ⟨more text⟩ four parameters (#1 to #4) can be used: these will be supplied at the time the message is used.

\msg_if_exist_p:nn {⟨module⟩} {⟨message⟩}
\msg_if_exist:nnTF {⟨module⟩} {⟨message⟩} {⟨true code⟩} {⟨false code⟩}
Tests whether the ⟨message⟩ for the ⟨module⟩ is currently defined.

11.2 Customizable information for message modules

\msg_module_name:n * \msg_module_name:n {⟨module⟩}
Expands to the public name of the ⟨module⟩ as defined by \g_msg_module_name_prop (or otherwise leaves the ⟨module⟩ unchanged).

\msg_module_type:n * \msg_module_type:n {⟨module⟩}
Expands to the description which applies to the ⟨module⟩, for example a Package or Class. The information here is defined in \g_msg_module_type_prop, and will default to Package if an entry is not present.

\g_msg_module_name_prop
Provides a mapping between the module name used for messages, and that for documentation.

\g_msg_module_type_prop
Provides a mapping between the module name used for messages, and that type of module. For example, for \LATEX3 core messages, an empty entry is set here meaning that they are not described using the standard Package text.

81
11.3 Contextual information for messages

\msg_line_context: ★\msg_line_context:
Prints the current line number when a message is given, and thus suitable for giving context to messages. The number itself is proceeded by the text on line.

\msg_line_number: ★\msg_line_number:
Prints the current line number when a message is given.

\msg_fatal_text:n {⟨module⟩}
Produce the standard text

Fatal Package ⟨module⟩ Error

This function can be redefined to alter the language in which the message is given, using #1 as the name of the ⟨module⟩ to be included. Any redefinition must produce output containing the ⟨module⟩ name, and will affect all messages using the expl3 mechanism.

\msg_critical_text:n {⟨module⟩}
Produce the standard text

Critical Package ⟨module⟩ Error

This function can be redefined to alter the language in which the message is given, using #1 as the name of the ⟨module⟩ to be included. Any redefinition must produce output containing the ⟨module⟩ name, and will affect all messages using the expl3 mechanism.

\msg_error_text:n {⟨module⟩}
Produce the standard text

Package ⟨module⟩ Error

This function can be redefined to alter the language in which the message is given, using #1 as the name of the ⟨module⟩ to be included. Any redefinition must produce output containing the ⟨module⟩ name, and will affect all messages using the expl3 mechanism.

\msg_warning_text:n {⟨module⟩}
Produce the standard text

Package ⟨module⟩ Warning

This function can be redefined to alter the language in which the message is given, using #1 as the name of the ⟨module⟩ to be included. The ⟨type⟩ of ⟨module⟩ may be adjusted: Package is the standard outcome: see \msg_module_type:n. Any redefinition must produce output containing the ⟨module⟩ name, and will affect all messages using the expl3 mechanism.
\msg_info_text:n \msg_info_text:n \langle\text{module}\rangle

Produces the standard text:

\textbf{Package} \langle\text{module}\rangle\text{ Info}

This function can be redefined to alter the language in which the message is given, using \texttt{#1} as the name of the \langle\text{module}\rangle to be included. The \texttt{\langle\text{type}\rangle} of \langle\text{module}\rangle may be adjusted: \texttt{Package} is the standard outcome: see \msg_module_type:n. Any redefinition must produce output containing the \langle\text{module}\rangle name, and will affect all messages using the expl3 mechanism.

\msg_see_documentation_text:n \msg_see_documentation_text:n \langle\text{module}\rangle

\texttt{Updated: 2018-09-30}

Produces the standard text

\textbf{See the} \langle\text{module}\rangle\text{ documentation for further information.}

This function can be redefined to alter the language in which the message is given, using \texttt{#1} as the name of the \langle\text{module}\rangle to be included. The name of the \langle\text{module}\rangle is produced using \msg_module_name:n.

11.4 Issuing messages

Messages behave differently depending on the message class. In all cases, the message may be issued supplying 0 to 4 arguments. If the number of arguments supplied here does not match the number in the definition of the message, extra arguments are ignored, or empty arguments added (of course the sense of the message may be impaired). The four arguments are converted to strings before being added to the message text: the e-type variants should be used to expand material. Note that this expansion takes place with the standard definitions in effect, which means that shorthands such as \texttt{\~} or \texttt{\\} are \textit{not} available; instead one should use \texttt{iow_char:N \~} and \texttt{iow_newline:}, respectively. The following message classes exist:

- \texttt{fatal}, ending the \TeX\ run;
- \texttt{critical}, ending the file being input;
- \texttt{error}, interrupting the \TeX\ run without ending it;
- \texttt{warning}, written to terminal and log file, for important messages that may require corrections by the user;
- \texttt{note} (less common than \texttt{info}) for important information messages written to the terminal and log file;
- \texttt{info} for normal information messages written to the log file only;
- \texttt{term} and \texttt{log} for un-decorated messages written to the terminal and log file, or to the log file only;
- \texttt{none} for suppressed messages.
\msg_fatal:nnnnnn
\msg_fatal:nneeee
\msg_fatal:(nneee|nnnee)
\msg_fatal:nnnn
\msg_fatal:(nnVV|nnVn|nnnV|nnee|nnne)
\msg_fatal:nnn
\msg_fatal:(nnV|nne)
\msg_fatal:nn

Updated: 2012-08-11

Issues \module error \message, passing \arg one to \arg four to the text-creating functions. After issuing a fatal error the \TeX run halts. No PDF file will be produced in this case (DVI mode runs may produce a truncated DVI file).

\msg_critical:nnnnnn
\msg_critical:nneeee
\msg_critical:(nneee|nnnee)
\msg_critical:nnnn
\msg_critical:(nnVV|nnVn|nnnV|nnee|nnne)
\msg_critical:nnn
\msg_critical:(nnV|nne)
\msg_critical:nn

Updated: 2012-08-11

Issues \module error \message, passing \arg one to \arg four to the text-creating functions. After issuing a critical error, \TeX stops reading the current input file. This may halt the \TeX run (if the current file is the main file) or may abort reading a sub-file.

\TeXhackers note: The \TeX \endinput primitive is used to exit the file. In particular, the rest of the current line remains in the input stream.

\msg_error:nnnnnn
\msg_error:nneeee
\msg_error:(nneee|nnnee)
\msg_error:nnnn
\msg_error:(nnVV|nnVn|nnnV|nnee|nnne)
\msg_error:nnn
\msg_error:(nnV|nne)
\msg_error:nn

Updated: 2012-08-11

Issues \module error \message, passing \arg one to \arg four to the text-creating functions. The error interrupts processing and issues the text at the terminal. After user input, the run continues.
Issues \textit{\texttt{module}} warning \textit{\texttt{message}}, passing \textit{\texttt{arg one}} to \textit{\texttt{arg four}} to the text-creating functions. The warning text is added to the log file and the terminal, but the \TeX run is not interrupted.

Issues \textit{\texttt{module}} information \textit{\texttt{message}}, passing \textit{\texttt{arg one}} to \textit{\texttt{arg four}} to the text-creating functions. For the more common \texttt{\msg_info}, the information text is added to the log file only, while \texttt{\msg_note} adds the info text to both the log file and the terminal. The \TeX run is not interrupted.
Issues *(module)* information *(message)*, passing *(arg one)* to *(arg four)* to the text-creating functions. The output is briefer than \texttt{\textbackslash msg\_info:nnnnnn}, omitting for instance the module name. It is added to the log file by \texttt{\textbackslash msg\_log:nnnnnn} while \texttt{\textbackslash msg\_term:nnnnnn} also prints it on the terminal.

\begin{verbatim}
\msg_term:nnnnnn
\msg_term:nneeee
\msg_term:nnnnn
\msg_term:(nnee|nnnee)
\msg_term:nnnn
\msg_term:(nnV\|nnV|nnnV|nnee|nnne)
\msg_term:nnn
\msg_term:(nnV\|nne)
\msg_term:nneee
\msg_term:nnnn
\msg_term:(nnee|nnnee)
\msg_term:nnnn
\msg_term:(nnV\|nnV|nnnV|nnee|nnne)
\msg_term:nnn
\msg_term:(nnV\|nne)
\msg_term:nn
\end{verbatim}

\textbf{Updated: 2012-08-11}

Does nothing: used as a message class to prevent any output at all (see the discussion of message redirection).

\begin{verbatim}
\msg_none:nnnnnn
\msg_none:nneeee
\msg_none:nnnnn
\msg_none:(nnee|nnnee)
\msg_none:nnnnn
\msg_none:(nnV\|nnV|nnnV|nnee|nnne)
\msg_none:nnn
\msg_none:(nnV\|nne)
\msg_none:nn
\end{verbatim}

\textbf{Updated: 2012-08-11}
11.4.1 Messages for showing material

\msg_show:nnnnnn \msg_show:nneeee \msg_show:nnee \msg_show:nne \msg_show:nn \msg_show:nnew

Issues \module information \message, passing \arg one to \arg four to the text-creating functions. The information text is shown on the terminal and the \TeX run is interrupted in a manner similar to \tl_show:n. This is used in conjunction with \msg_show_item:n and similar functions to print complex variable contents completely. If the formatted text does not contain >- at the start of a line, an additional line >-, will be put at the end. In addition, a final period is added if not present.

\seq_map_function:NN \l_tmpa_seq \msg_show_item:n
\prop_map_function:NN \l_tmpa_seq \msg_show_item:nn

Used in the text of messages for \msg_show:nnnnnn to show or log a list of items or key–value pairs. The output of \msg_show_item:n produces a newline, the prefix >, two spaces, then the braced string representation of its argument. The two-argument versions separates the key and value using \\>\rightarrow\\>, and the unbraced versions don’t print the surrounding braces.

These functions are suitable for usage with iterator functions like \seq_map_function:NN, \prop_map_function:NN, etc. For example, with a sequence \l_tmma_seq containing a, \{b\} and \c,

\seq_map_function:NN \l_tmma_seq \msg_show_item:n

would expand to three lines:

>\l_tmma{a}
>\l_tmma{\{b\}}
>\l_tmma{\c}

11.4.2 Expandable error messages

In very rare cases it may be necessary to produce errors in an expansion-only context. The functions in this section should only be used if there is no alternative approach using \msg_error:nnnnnn or other non-expandable commands from the previous section. Despite having a similar interface as non-expandable messages, expandable errors must be handled internally very differently from normal error messages, as none of the tools
to print to the terminal or the log file are expandable. As a result, short-hands such as \{ or \ do not work, and messages must be very short (with default settings, they are truncated after approximately 50 characters). It is advisable to ensure that the message is understandable even when truncated, by putting the most important information up front. Another particularity of expandable messages is that they cannot be redirected or turned off by the user.

\msg\_expandable\_error:nnnnnn \{module\} \{message\} \{arg one\} \{arg two\} \{arg three\} \{arg four\}
\msg\_expandable\_error:nnffff  \{module\} \{message\} \{arg one\} \{arg two\}  \{arg three\} \{arg four\}
\msg\_expandable\_error:nnnnnn
\msg\_expandable\_error:nnffff
\msg\_expandable\_error:nnnn
\msg\_expandable\_error:nnn
\msg\_expandable\_error:nnf
\msg\_expandable\_error:nn

Issues an “Undefined error” message from \TeX itself using the undefined control sequence \?? then prints “! \{module\}: "\{error message\}, which should be short. With default settings, anything beyond approximately 60 characters long (or bytes in some engines) is cropped. A leading space might be removed as well.

### 11.5 Redirecting messages

Each message has a “name”, which can be used to alter the behaviour of the message when it is given. Thus we might have

\msg\_new:nnnn \{ module \} \{ my-message \} { Some-text } { Some-more-text }

to define a message, with

\msg\_error:nn \{ module \} \{ my-message \}

when it is used. With no filtering, this raises an error. However, we could alter the behaviour with

\msg\_redirect\_class:nn \{ error \} \{ warning \}

to turn all errors into warnings, or with

\msg\_redirect\_module:nnn \{ module \} \{ error \} \{ warning \}

to alter only messages from that module, or even

\msg\_redirect\_name:nnn \{ module \} \{ my-message \} \{ warning \}

to target just one message. Redirection applies first to individual messages, then to messages from one module and finally to messages of one class. Thus it is possible to select out an individual message for special treatment even if the entire class is already redirected.

Multiple redirections are possible. Redirections can be cancelled by providing an empty argument for the target class. Redirection to a missing class raises an error.
immediately. Infinite loops are prevented by eliminating the redirection starting from the target of the redirection that caused the loop to appear. Namely, if redirections are requested as \( A \rightarrow B, B \rightarrow C \) and \( C \rightarrow A \) in this order, then the \( A \rightarrow B \) redirection is cancelled.

\[
\texttt{msg\_redirect\_class:nn} \\{\langle \text{class\ one} \rangle \} \{\langle \text{class\ two} \rangle \}
\]

Changes the behaviour of messages of \( \langle \text{class\ one} \rangle \) so that they are processed using the code for those of \( \langle \text{class\ two} \rangle \). Each \( \langle \text{class} \rangle \) can be one of fatal, critical, error, warning, note, info, term, log, none.

\[
\texttt{msg\_redirect\_module:nnn} \\{\langle \text{module} \rangle \} \{\langle \text{class\ one} \rangle \} \{\langle \text{class\ two} \rangle \}
\]

Redirects message of \( \langle \text{class\ one} \rangle \) for \( \langle \text{module} \rangle \) to act as though they were from \( \langle \text{class\ two} \rangle \). Messages of \( \langle \text{class\ one} \rangle \) from sources other than \( \langle \text{module} \rangle \) are not affected by this redirection. This function can be used to make some messages “silent” by default. For example, all of the warning messages of \( \langle \text{module} \rangle \) could be turned off with:

\[
\texttt{msg\_redirect\_module:nnn} \\{\langle \text{module} \rangle \} \{\langle \text{warning} \rangle \} \{\langle \text{none} \rangle \}
\]

\[
\texttt{msg\_redirect\_name:nnn} \\{\langle \text{module} \rangle \} \{\langle \text{message} \rangle \} \{\langle \text{class} \rangle \}
\]

Redirects a specific \( \langle \text{message} \rangle \) from a specific \( \langle \text{module} \rangle \) to act as a member of \( \langle \text{class} \rangle \) of messages. No further redirection is performed. This function can be used to make a selected message “silent” without changing global parameters:

\[
\texttt{msg\_redirect\_name:nnn} \\{\langle \text{module} \rangle \} \{\langle \text{annoying\-message} \rangle \} \{\langle \text{none} \rangle \}
\]
Chapter 12

The l3file module
File and I/O operations

This module provides functions for working with external files. Some of these functions apply to an entire file, and have prefix \file_, while others are used to work with files on a line by line basis and have prefix \ior_ (reading) or \iow_ (writing).

It is important to remember that when reading external files \TeX attempts to locate them using both the operating system path and entries in the \TeX file database (most \TeX systems use such a database). Thus the “current path” for \TeX is somewhat broader than that for other programs.

For functions which expect a ⟨file name⟩ argument, this argument may contain both literal items and expandable content, which should on full expansion be the desired file name. Active characters (as declared in \l_char_active_seq) are not expanded, allowing the direct use of these in file names. Quote tokens (") are not permitted in file names as they are reserved for internal use by some \TeX primitives.

Spaces are trimmed at the beginning and end of the file name: this reflects the fact that some file systems do not allow or interact unpredictably with spaces in these positions. When no extension is given, this will trim spaces from the start of the name only.

12.1 Input–output stream management

As \TeX engines have a limited number of input and output streams, direct use of the streams by the programmer is not supported in \LaTeX3. Instead, an internal pool of streams is maintained, and these are allocated and deallocated as needed by other modules. As a result, the programmer should close streams when they are no longer needed, to release them for other processes.

Note that I/O operations are global: streams should all be declared with global names and treated accordingly.
Globally reserves the name of the (stream), either for reading or for writing as appropriate. The (stream) is not opened until the appropriate \..._open:Nn function is used. Attempting to use a (stream) which has not been opened is an error, and the (stream) will behave as the corresponding \c_term_....

\ior_open:N
\ior_open:cn
Opens (file name) for reading using (stream) as the control sequence for file access. If the (stream) was already open it is closed before the new operation begins. The (stream) is available for access immediately and will remain allocated to (file name) until an \ior_close:N instruction is given or the \TeX run ends. If the file is not found, an error is raised.

\ior_open:Nn
\ior_open:cn
Opens (file name) for writing using (stream) as the control sequence for file access. If the (stream) was already open it is closed before the new operation begins. The (stream) is available for access immediately and will remain allocated to (file name) until a \ior_close:N instruction is given or the \TeX run ends. Opening a file for writing clears any existing content in the file (i.e. writing is not additive).

\ior_shell_open:Nn
\ior_shell_open:cn
Opens the pseudo-file created by the output of the (shell command) for reading using (stream) as the control sequence for access. If the (stream) was already open it is closed before the new operation begins. The (stream) is available for access immediately and will remain allocated to (shell command) until a \ior_close:N instruction is given or the \TeX run ends. If piped system calls are disabled an error is raised.

For details of handling of the (shell command), see \sys_get_shell:nnNTF.

\iow_shell_open:Nn
\iow_shell_open:cn
Opens the pseudo-file created by the output of the (shell command) for writing using (stream) as the control sequence for access. If the (stream) was already open it is closed before the new operation begins. The (stream) is available for access immediately and will remain allocated to (shell command) until an \iow_close:N instruction is given or the \TeX run ends. If piped system calls are disabled an error is raised.

For details of handling of the (shell command), see \sys_get_shell:nnNTF.
Closes the \texttt{stream}. Streams should always be closed when they are finished with as this ensures that they remain available to other programmers.

Display (to the terminal or log file) the file name associated to the (read or write) \texttt{stream}.

Display (to the terminal or log file) a list of the file names associated with each open (read or write) stream. This is intended for tracking down problems.

### 12.1.1 Reading from files

Reading from files and reading from the terminal are separate processes in expl3. The functions \texttt{\ior_get:NN} and \texttt{\ior_str_get:NN}, and their branching equivalents, are designed to work with \texttt{files}. 
Function that reads one or more lines (until an equal number of left and right braces are found) from the file input \langle stream \rangle and stores the result locally in the \langle token list \rangle variable. The material read from the \langle stream \rangle is tokenized by \TeX{} according to the category codes and \texttt{\endlinechar} in force when the function is used. Assuming normal settings, any lines which do not end in a comment character % have the line ending converted to a space, so for example input
\begin{verbatim}
a b c
\end{verbatim}
results in a token list \texttt{a b c}. Any blank line is converted to the token \texttt{\par}. Therefore, blank lines can be skipped by using a test such as
\begin{verbatim}
\ior_get:NN \l_my_stream \l_tmpa_tl
\tl_set:Nn \l_tmpb_tl { \par }
\tl_if_eq:NNF \l_tmpa_tl \l_tmpb_tl
...
\end{verbatim}
Also notice that if multiple lines are read to match braces then the resulting token list can contain \texttt{\par} tokens. In the non-branching version, where the \langle stream \rangle is not open the \langle tl var \rangle is set to \texttt{\q_no_value}.

\textbf{\TeX{}hackers note:} This protected macro is a wrapper around the \TeX{} primitive \texttt{\read}. Regardless of settings, \TeX{} replaces trailing space and tab characters (character codes 32 and 9) in each line by an end-of-line character (character code \texttt{\endlinechar}, omitted if \texttt{\endlinechar} is negative or too large) before turning characters into tokens according to current category codes. With default settings, spaces appearing at the beginning of lines are also ignored.
by the \textit{function} or \textit{code} discussed below remain in effect after the loop.

\ior_map_inline:Nn \ior_map_inline:Nn \ior_map_inline:Nn \ior_map_inline:Nn (stream) \{(inline function)\}
\hspace{1em} New: 2012-02-11
Applies the \textit{inline function} to each set of \textit{lines} obtained by calling \ior_get:NN until reaching the end of the file. \TeX{} ignores any trailing new-line marker from the file it reads. The \textit{inline function} should consist of code which receives the \textit{line} as #1.

\ior_str_map_inline:Nn \ior_str_map_inline:Nn \ior_str_map_inline:Nn \ior_str_map_inline:Nn (stream) \{(inline function)\}
\hspace{1em} New: 2012-02-11
Applies the \textit{inline function} to every \textit{line} in the \textit{stream}. The material is read from the \textit{stream} as a series of tokens with category code 12 (other), with the exception of space characters which are given category code 10 (space). The \textit{inline function} should consist of code which receives the \textit{line} as #1. Note that \TeX{} removes trailing space and tab characters (character codes 32 and 9) from every line upon input. \TeX{} also ignores any trailing new-line marker from the file it reads.

\ior_map_variable:NNn \ior_map_variable:NNn \ior_map_variable:NNn \ior_map_variable:NNn (stream) \{tl var\} \{(code)\}
\hspace{1em} New: 2019-01-13
For each set of \textit{lines} obtained by calling \ior_get:NN until reaching the end of the file, stores the \textit{lines} in the \textit{tl var} then applies the \textit{code}. The \textit{code} will usually make use of the \textit{variable}, but this is not enforced. The assignments to the \textit{variable} are local. Its value after the loop is the last set of \textit{lines}, or its original value if the \textit{stream} is empty. \TeX{} ignores any trailing new-line marker from the file it reads. This function is typically faster than \ior_map_inline:Nn.

\ior_str_map_variable:NNn \ior_str_map_variable:NNn \ior_str_map_variable:NNn \ior_str_map_variable:NNn (stream) \{variable\} \{(code)\}
\hspace{1em} New: 2019-01-13
For each \textit{line} in the \textit{stream}, stores the \textit{line} in the \textit{variable} then applies the \textit{code}. The material is read from the \textit{stream} as a series of tokens with category code 12 (other), with the exception of space characters which are given category code 10 (space). The \textit{code} will usually make use of the \textit{variable}, but this is not enforced. The assignments to the \textit{variable} are local. Its value after the loop is the last \textit{line}, or its original value if the \textit{stream} is empty. Note that \TeX{} removes trailing space and tab characters (character codes 32 and 9) from every line upon input. \TeX{} also ignores any trailing new-line marker from the file it reads. This function is typically faster than \ior_str_map_inline:Nn.
\ior_map_break: Used to terminate a \ior_map_... function before all lines from the \langle stream \rangle have been processed. This normally takes place within a conditional statement, for example

\ior_map_inline:Nn \l_my_ior
{ \str_if_eq:nnTF { #1 } { bingo } \ior_map_break: }
{ % Do something useful }
}

Use outside of a \ior_map_... scenario leads to low level \TeX errors.

\textbf{\TeX hackers note:} When the mapping is broken, additional tokens may be inserted before further items are taken from the input stream. This depends on the design of the mapping function.

\ior_map_break:n Used to terminate a \ior_map_... function before all lines in the \langle stream \rangle have been processed, inserting the \langle code \rangle after the mapping has ended. This normally takes place within a conditional statement, for example

\ior_map_inline:Nn \l_my_ior
{ \str_if_eq:nnTF { #1 } { bingo } \ior_map_break:n \langle code \rangle }
{ % Do something useful }
}

Use outside of a \ior_map_... scenario leads to low level \TeX errors.

\textbf{\TeX hackers note:} When the mapping is broken, additional tokens may be inserted before the \langle code \rangle is inserted into the input stream. This depends on the design of the mapping function.

\ior_if_eof_p:N \ior_if_eof:NTF \langle stream \rangle
\langle true code \rangle \langle false code \rangle

Tests if the end of a file \langle stream \rangle has been reached during a reading operation. The test also returns a \texttt{true} value if the \langle stream \rangle is not open.
12.1.2 Reading from the terminal

\ior_get_term:nN \ior_get_str_term:nN
\ior_get_term:nN \{prompt\} \{token list variable\}

Function that reads one or more lines (until an equal number of left and right braces are found) from the terminal and stores the result locally in the \textit{token list} variable. Tokenization occurs as described for \ior_get:NN or \ior_str_get:NN, respectively. When the \textit{prompt} is empty, \TeX{} will wait for input without any other indication: typically the programmer will have provided a suitable text using e.g. \ior_term:n.

Where the \textit{prompt} is given, it will appear in the terminal followed by an =, e.g.

\begin{verbatim}
prompt=
\end{verbatim}

12.1.3 Writing to files

\iow_now:Nn \iow_now:NN \{stream\} \{tokens\}
\iow_now:Nn \iow_now:NN \{stream\} \{tokens\}

This function writes \textit{tokens} to the specified \textit{stream} immediately (i.e. the write operation is called on expansion of \iow_now:Nn).

\iow_log:n \iow_log:e \iow_log:n \{tokens\}

This function writes the given \textit{tokens} to the log (transcript) file immediately: it is a dedicated version of \iow_now:Nn.

\iow_term:n \iow_term:e \iow_term:n \{tokens\}

This function writes the given \textit{tokens} to the terminal file immediately: it is a dedicated version of \iow_now:Nn.

\iow_shipout:Nn \iow_shipout:NN \{stream\} \{tokens\}
\iow_shipout:Nn \iow_shipout:NN \{stream\} \{tokens\}

This function writes \textit{tokens} to the specified \textit{stream} when the current page is finalised (i.e. at shipout). The \texttt{e}-type variants expand the \textit{tokens} at the point where the function is used but not when the resulting tokens are written to the \textit{stream} (cf. \iow_shipout_e:Nn).

\textbf{\TeX{}hackers note:} When using expl3 with a format other than \LaTeX{}, new line characters inserted using \iow_newline: or using the line-wrapping code \iow_wrap:nnnN are not recognized in the argument of \iow_shipout:Nn. This may lead to the insertion of additional unwanted line-breaks.
\iow_shipout_e:Nn \iow_shipout_e: (Ne|cn|ce)

This function writes \langle tokens \rangle to the specified \langle stream \rangle when the current page is finalised (\textit{i.e.} at shipout). The \langle tokens \rangle are expanded at the time of writing in addition to any expansion when the function is used. This makes these functions suitable for including material finalised during the page building process (such as the page number integer).

\textbf{\textsc{Texhackers note:}} This is a wrapper around the \TeX{} primitive \texttt{\write}. When using expl3 with a format other than \LaTeX{}, new line characters inserted using \texttt{\iow_newline:} or using the line-wrapping code \texttt{\iow-wrap:nnnN} are not recognized in the argument of \texttt{\iow_shipout:Nn}. This may lead to the insertion of additional unwanted line-breaks.

\iow_char:N \langle char \rangle

Inserts \langle char \rangle into the output stream. Useful when trying to write difficult characters such as \%, \{, \}, \textit{etc.} in messages, for example:

\iow_now:Ne \g_my_iow \{ \iow_char:N \{ text \iow_char:N \} \}

The function has no effect if writing is taking place without expansion (\textit{e.g.} in the second argument of \iow_now:Nn).

\iow_newline:

Function to add a new line within the \langle tokens \rangle written to a file. The function has no effect if writing is taking place without expansion (\textit{e.g.} in the second argument of \iow_now:Nn).

\textbf{\textsc{Texhackers note:}} When using expl3 with a format other than \LaTeX{}, the character inserted by \texttt{\iow_newline:} is not recognized by \TeX{}, which may lead to the insertion of additional unwanted line-breaks. This issue only affects \texttt{\iow_shipout:Nn, \iow_shipout_e:Nn} and direct uses of primitive operations.
12.1.4 Wrapping lines in output

\iow_wrap:nnn \iow_wrap:nnN \iow_wrap:nenN

This function wraps the ⟨text⟩ to a fixed number of characters per line. At the start of each line which is wrapped, the ⟨run-on text⟩ is inserted. The line character count targeted is the value of \l_iow_line_count_int minus the number of characters in the ⟨run-on text⟩ for all lines except the first, for which the target number of characters is simply \l_iow_line_count_int since there is no run-on text. The ⟨text⟩ and ⟨run-on text⟩ are exhaustively expanded by the function, with the following substitutions:

- \ or \iow_newline: may be used to force a new line,
- \␣ may be used to represent a forced space (for example after a control sequence),
- \#, \%, \{, \} may be used to represent the corresponding character,
- \iow_wrap_allow_break: may be used to allow a line-break without inserting a space,
- \iow_indent:n may be used to indent a part of the ⟨text⟩ (not the ⟨run-on text⟩).

Additional functions may be added to the wrapping by using the ⟨set up⟩, which is executed before the wrapping takes place: this may include overriding the substitutions listed.

Any expandable material in the ⟨text⟩ which is not to be expanded on wrapping should be converted to a string using \token_to_str:N, \tl_to_str:n, \tl_to_str:N, etc.

The result of the wrapping operation is passed as a braced argument to the ⟨function⟩, which is typically a wrapper around a write operation. The output of \iow_wrap:nnn (i.e. the argument passed to the ⟨function⟩) consists of characters of category “other” (category code 12), with the exception of spaces which have category “space” (category code 10). This means that the output does not expand further when written to a file.

\iow_wrap_allow_break:

\iow_wrap_allow_break:

In the first argument of \iow_wrap:nnn (for instance in messages), inserts a break-point that allows a line break. If no break occurs, this function adds nothing to the output.

\iow_indent:n

\iow_indent:n ⟨(text)⟩

In the first argument of \iow_wrap:nnn (for instance in messages), indents ⟨text⟩ by four spaces. This function does not cause a line break, and only affects lines which start within the scope of the ⟨text⟩. In case the indented ⟨text⟩ should appear on separate lines from the surrounding text, use \ to force line breaks.
The maximum number of characters in a line to be written by the \texttt{\iow\_wrap:nnnN} function. This value depends on the \TeX system in use: the standard value is 78, which is typically correct for unmodified \TeX\ Live and MiK\TeX systems.

12.1.5 Constant input–output streams, and variables

\texttt{\tmpa\_ior} \texttt{\tmpb\_ior} Scratch input stream for global use. These are never used by the kernel code, and so are safe for use with any \TeX3-defined function. However, they may be overwritten by other non-kernel code and so should only be used for short-term storage.

\texttt{\log\_ior} \texttt{\term\_ior} Scratch output stream for global use. These are never used by the kernel code, and so are safe for use with any \TeX3-defined function. However, they may be overwritten by other non-kernel code and so should only be used for short-term storage.

12.1.6 Primitive conditionals

\texttt{\if\_eof:w} \langle \textit{stream} \rangle \langle \textit{true code} \rangle \texttt{\else:} \langle \textit{false code} \rangle \texttt{\fi:} Tests if the \langle \textit{stream} \rangle returns “end of file”, which is true for non-existent files. The \texttt{\else:} branch is optional.

\TeXhackers note: This is the \TeX primitive \texttt{\ifeof}.

12.2 File operations

12.2.1 Basic file operations

\texttt{\file\_curr\_dir\_str} \texttt{\file\_curr\_name\_str} \texttt{\file\_curr\_ext\_str} Contain the directory, name and extension of the current file. The directory is empty if the file was loaded without an explicit path (\textit{i.e.} if it is in the \TeX search path), and does not end in / other than the case that it is exactly equal to the root directory. The \langle \textit{name} \rangle and \langle \textit{ext} \rangle parts together make up the file name, thus the \langle \textit{name} \rangle part may be thought of as the “job name” for the current file.

Note that \TeX does not provide information on the \langle \textit{dir} \rangle and \langle \textit{ext} \rangle part for the main (top level) file and that this file always has empty \langle \textit{dir} \rangle and \langle \textit{ext} \rangle components. Also, the \langle \textit{name} \rangle here will be equal to \texttt{\sys\_jobname\_str}, which may be different from the real file name (if set using \texttt{\--jobname}, for example).
Each entry is the path to a directory which should be searched when seeking a file. Each path can be relative or absolute, and need not include the trailing slash. Spaces need not be quoted.

\l_file_search_path_seq

Rev: 2017-06-18
Updated: 2023-06-15

Each entry is the path to a directory which should be searched when seeking a file. Each path can be relative or absolute, and need not include the trailing slash. Spaces need not be quoted.

**T\hackerX hackers note:** When working as a package in L\TeX\, expl3 will automatically append the current \input@path to the set of values from \l_file_search_path_seq.

\l_file_search_path_seq

Updated: 2023-09-18

100

12.2.2 Information about files and file contents

Functions in this section return information about files as expl3 str data, except that the non-expandable functions set their return token list to \q_no_value if the file requested is not found. As such, comparison of file names, hashes, sizes, etc., should use \str_-if_eq:nnTF rather than \tl_if_eq:nnTF and so on.

\file_hex_dump:n
\file_hex_dump:V
\file_hex_dump:nnn
\file_hex_dump:Vnn

Searches for \langle file name \rangle using the current \TeX\ search path and the additional paths controlled by \l_file_search_path_seq. It then expands to leave the hexadecimal dump of the file content in the input stream. The file is read as bytes, which means that in contrast to most \TeX\ behaviour there will be a difference in result depending on the line endings used in text files. The same file will produce the same result between different engines: the algorithm used is the same in all cases. When the file is not found, the result of expansion is empty. The \{\langle start index \rangle\} and \{\langle end index \rangle\} values work as described for \str_range:nnn.

\file_get_hex_dump:nN
\file_get_hex_dump:VN
\file_get_hex_dump:nN
\file_get_hex_dump:VN
\file_get_hex_dump:nnnN
\file_get_hex_dump:VnnN
\file_get_hex_dump:nnnN
\file_get_hex_dump:VnnN

Sets the \langle tl var \rangle to the result of applying \file_hex_dump:n/\file_hex_dump:nnn to the \langle file \rangle. If the file is not found, the \langle tl var \rangle will be set to \q_no_value.

\file_if_exist_p:n
\file_if_exist_p:V
\file_if_exist:nTF
\file_if_exist:VTF

Tests if \langle file name \rangle is found in the path as detailed for \file_if_exist:nTF.
\file_mdfive_hash:n \file_mdfive_hash:V

Searches for \file{n} using the current \TeX\ search path and the additional paths
controlled by \l_file_search_path_seq. It then expands to leave the MD5 sum
generated from the contents of the file in the input stream. The file is read as bytes, which
means that in contrast to most \TeX\ behaviour there will be a difference in result de-
pending on the line endings used in text files. The same file will produce the same result
between different engines: the algorithm used is the same in all cases. When the file is
not found, the result of expansion is empty.

\file_get_mdfive_hash:nN \file_get_mdfive_hash:VN

\file_get_mdfive_hash:nN \file_get_mdfive_hash:V

\file_get_mdfive_hash:nN

\file_get_mdfive_hash:VN

\file_get_mdfive_hash:nN

\file_get_mdfive_hash:VN

\file_size:n \file_size:V

Searches for \file{n} using the current \TeX\ search path and the additional paths
controlled by \l_file_search_path_seq. It then expands to leave the size of the file in
bytes in the input stream. When the file is not found, the result of expansion is empty.

\file_get_size:nN \file_get_size:VN

\file_get_size:nN \file_get_size:V

\file_get_size:nN

\file_get_size:VN

\file_get_size:nN

\file_get_size:VN

\file_timestamp:n \file_timestamp:V

Searches for \file{n} using the current \TeX\ search path and the additional
paths controlled by \l_file_search_path_seq. It then expands to leave the modification
timestamp of the file in the input stream. The timestamp is of the form
D:\year\month\day\hour\minute\second\offset, where the latter may be Z
(UTC) or \plusminus\hours'\minutes'. When the file is not found, the result
of expansion is empty. This is not available in older versions of \TeX\.

\file_get_timestamp:nN \file_get_timestamp:VN

\file_get_timestamp:nN \file_get_timestamp:V

\file_get_timestamp:nN

\file_get_timestamp:VN

\file_get_timestamp:nN

\file_get_timestamp:VN

\file_get_timestamp:nN

\file_get_timestamp:VN

\file_get_timestamp:nN

\file_get_timestamp:VN
\file_compare_timestamp_p:nNn \file_compare_timestamp_p: (nNV|VNn|VNV) \file_compare_timestamp:nNnTF \file_compare_timestamp: (nNV|VNn|VNV) TF

Compares the file stamps on the two \files as indicated by the \comparator, and inserts either the \true code or \false case as required. A file which is not found is treated as older than any file which is found. This allows for example the construct

\file_compare_timestamp:nNnT { source-file } > { derived-file }
{ % Code to regenerate derived file }

to work when the derived file is entirely absent. The timestamp of two absent files is regarded as different. This is not available in older versions of X\TeX.

\file_get_full_name:nN \file_get_full_name:VN \file_get_full_name:nNTF \file_get_full_name:VNTF

Searches for \file name in the path as detailed for \file_if_exist:nTF, and if found leaves the fully-qualified name of the file, \textit{i.e.} the path and file name. This includes an extension .\texttt{tex} when the given \file name has no extension but the file found has that extension. In the non-branching version, the \tl var will be set to \texttt{\no_value} in the case that the file does not exist.

\file_full_name:n \file_full_name:V

Searches for \file name in the path as detailed for \file_if_exist:nTF, and if found leaves the fully-qualified name of the file, \textit{i.e.} the path and file name, in the input stream. This includes an extension .\texttt{tex} when the given \file name has no extension but the file found has that extension. If the file is not found on the path, the expansion is empty.

\file_parse_full_name:nNNN \file_parse_full_name:VNNN

 Parses the \full name and splits it into three parts, each of which is returned by setting the appropriate local string variable:

- The \dir: everything up to the last / (path separator) in the \file path. As with system PATH variables and related functions, the \dir does not include the trailing / unless it points to the root directory. If there is no path (only a file name), \dir is empty.

- The \name: everything after the last / up to the last ., where both of those characters are optional. The \name may contain multiple . characters. It is empty if \full name consists only of a directory name.

- The \ext: everything after the last . (including the dot). The \ext is empty if there is no . after the last /.

Before parsing, the \full name is expanded until only non-expandable tokens remain, except that active characters are also not expanded. Quotes (") are invalid in file names and are discarded from the input.
Parses the \texttt{\file_parse_full_name:n} as described for \texttt{\file_parse_full_name:nNNN}, and leaves \texttt{\dir}, \texttt{\name}, and \texttt{\ext} in the input stream, each inside a pair of braces.

\texttt{\file_parse_full_name:apply:nN} \quad \texttt{\file_parse_full_name:apply:VN}

Parses the \texttt{\full_name} as described for \texttt{\file_parse_full_name:nNNN}, and passes \texttt{\dir}, \texttt{\name}, and \texttt{\ext} as arguments to \texttt{\function}, as an \texttt{n}-type argument each, in this order.

### 12.2.3 Accessing file contents

\texttt{\file_get:nnN} \quad \texttt{\file_get:VnN} \quad \texttt{\file_get:nnNTF} \quad \texttt{\file_get:VnNTF}

Defines \texttt{\tl} to the contents of \texttt{\filename}. Category codes may need to be set appropriately via the \texttt{\setup} argument. The non-branching version sets the \texttt{\tl} to \texttt{\q_\text{_no_value}} if the file is not found. The branching version runs the \texttt{\true_code} after the assignment to \texttt{\tl} if the file is found, and \texttt{\false_code} otherwise. The file content will be tokenized using the current category code régime.

\texttt{\file_input:n} \quad \texttt{\file_input:V}

Searches for \texttt{\filename} in the path as detailed for \texttt{\file_if_exist:nTF}, and if found reads in the file as additional \LaTeX{} source. All files read are recorded for information and the file name stack is updated by this function. An error is raised if the file is not found.

\texttt{\file_input_raw:n} \quad \texttt{\file_input_raw:V}

Searches for \texttt{\filename} in the path as detailed for \texttt{\file_if_exist:nTF}, and if found reads in the file as additional \LaTeX{} source. No data concerning the file is tracked. If the file is not found, no action is taken.

\TeX{}hackers note: This function is intended only for contexts where files must be read purely by expansion, for example at the start of a table cell in an \texttt{\halign}. 

\texttt{\file_if_exist_input:n} \quad \texttt{\file_if_exist_input:V} \quad \texttt{\file_if_exist_input:nF} \quad \texttt{\file_if_exist_input:VF}

Searches for \texttt{\filename} using the current \TeX{} search path and the additional paths included in \texttt{\_file_search_path_seq}. If found then reads in the file as additional \TeX{} source as described for \texttt{\file_input:n}, otherwise inserts the \texttt{\false_code}. Note that these functions do not raise an error if the file is not found, in contrast to \texttt{\file_input:n}. 

103
\file_input_stop: \file_input_stop:  

\verb|\file_input:nn| or similar before the end of the file is reached. Where the file reading is being terminated due to an error, \msg_-
critical:nn(nn) should be preferred.  

\TeXhackers note: This function must be used on a line on its own: \TeX reads files line-by-line and so any additional tokens in the "current" line will still be read.  
This is also true if the function is hidden inside another function (which will be the normal case), i.e., all tokens on the same line in the source file are still processed. Putting it on a line by itself in the definition doesn’t help as it is the line where it is used that counts!

\file_show_list: \file_show_list:  
\file_log_list: \file_log_list:  

These functions list all files loaded by \LaTeXe commands that populate \filelist or by \file_input:n. While \file_show_list: displays the list in the terminal, \file_log_list: outputs it to the log file only.
Chapter 13

The \texttt{luatex} module

\texttt{Lua\TeX}-specific functions

The \texttt{Lua\TeX} engine provides access to the Lua programming language, and with it access to the “internals” of \TeX. In order to use this within the framework provided here, a family of functions is available. When used with pdf\TeX, \pdf\TeX, up\TeX or \Xe\TeX these raise an error: use \texttt{\system_if_engine_luatex:T} to avoid this. Details on using Lua with the \texttt{Lua\TeX} engine are given in the \texttt{Lua\TeX} manual.

13.1 Breaking out to Lua

\begin{verbatim}
lua_now:n {⟨token list⟩}
lua_now:e {⟨token list⟩}
\end{verbatim}

The \texttt{⟨token list⟩} is first tokenized by \TeX, which includes converting line ends to spaces in the usual \TeX manner and which respects currently-applicable \TeX category codes. The resulting \texttt{⟨Lua input⟩} is passed to the Lua interpreter for processing. Each \texttt{\lua_now:n} block is treated by Lua as a separate chunk. The Lua interpreter executes the \texttt{⟨Lua input⟩} immediately, and in an expandable manner.

\texttt{\TeXXhacker note:} \texttt{\lua_now:e} is a macro wrapper around \texttt{\directlua}: when \texttt{Lua\TeX} is in use two expansions are required to yield the result of the Lua code.

\begin{verbatim}
lua_shipout:n {⟨token list⟩}
lua_shipout_e:n {⟨token list⟩}
\end{verbatim}

The \texttt{⟨token list⟩} is first tokenized by \TeX, which includes converting line ends to spaces in the usual \TeX manner and which respects currently-applicable \TeX category codes. The resulting \texttt{⟨Lua input⟩} is passed to the Lua interpreter when the current page is finalised (\textit{i.e.} at shipout). Each \texttt{\lua_shipout:n} block is treated by Lua as a separate chunk. The Lua interpreter will execute the \texttt{⟨Lua input⟩} during the page-building routine: no \TeX expansion of the \texttt{⟨Lua input⟩} will occur at this stage.

In the case of the \texttt{\lua_shipout_e:n} version the input is fully expanded by \TeX in an \texttt{e}-type manner during the shipout operation.

\texttt{\TeXXhacker note:} At a \TeX level, the \texttt{⟨Lua input⟩} is stored as a “whatsit”.

\footnote{New: 2018-06-18}
\lua_escape:n  \lua_escape:e

Converts the \texttt{token list} such that it can safely be passed to Lua: embedded backslashes, double and single quotes, and newlines and carriage returns are escaped. This is done by prepending an extra token consisting of a backslash with category code 12, and for the line endings, converting them to \texttt{\textbackslash n} and \texttt{\textbackslash r}, respectively.

\textbf{\textsc{TpXhackers note}}: \lua_escape:e is a macro wrapper around \luaescapestring: when \LaTeX{} is in use two expansions are required to yield the result of the Lua code.

\lua_load_module:n

Loads a Lua module into the Lua interpreter. \\lua_now:n passes its \texttt{\{token list\}} argument to the Lua interpreter as a single line, with characters interpreted under the current catcode regime. These two facts mean that \lua_now:n rarely behaves as expected for larger pieces of code. Therefore, package authors should \textbf{not} write significant amounts of Lua code in the arguments to \lua_now:n. Instead, it is strongly recommended that they write the majority of their Lua code in a separate file, and then load it using \lua_load_module:n.

\textbf{\textsc{TpXhackers note}}: This is a wrapper around the Lua call \texttt{require \langle module\rangle}.

### 13.2 Lua interfaces

As well as interfaces for \LaTeX{}, there are a small number of Lua functions provided here.

\texttt{\texttt{ltx.utils}}

Most public interfaces provided by the module are stored within the \texttt{ltx.utils} table.

\begin{verbatim}
\texttt{ltx.utils.filedump} \langle dump \rangle = \texttt{ltx.utils.filedump(\langle file\rangle,\langle offset\rangle,\langle length\rangle)}
\end{verbatim}

Returns the uppercase hexadecimal representation of the content of the \texttt{\langle file\rangle} read as bytes. If the \texttt{\langle length\rangle} is given, only this part of the file is returned; similarly, one may specify the \texttt{\langle offset\rangle} from the start of the file. If the \texttt{\langle length\rangle} is not given, the entire file is read starting at the \texttt{\langle offset\rangle}.

\begin{verbatim}
\texttt{ltx.utils.filemd5sum} \langle hash \rangle = \texttt{ltx.utils.filemd5sum(\langle file\rangle)}
\end{verbatim}

Returns the MD5 sum of the file contents read as bytes; note that the result will depend on the nature of the line endings used in the file, in contrast to normal \LaTeX{} behaviour. If the \texttt{\langle file\rangle} is not found, nothing is returned with \texttt{no error raised}.

\begin{verbatim}
\texttt{ltx.utils.filemoddate} \langle date \rangle = \texttt{ltx.utils.filemoddate(\langle file\rangle)}
\end{verbatim}

Returns the date/time of last modification of the \texttt{\langle file\rangle} in the format

\begin{verbatim}
D:\langle year\rangle\langle month\rangle\langle day\rangle\langle hour\rangle\langle minute\rangle\langle second\rangle\langle offset\rangle
\end{verbatim}

where the latter may be \texttt{Z} (UTC) or \texttt{\langle plus-minus\rangle\langle hours\rangle\langle minutes\rangle}. If the \texttt{\langle file\rangle} is not found, nothing is returned with \texttt{no error raised}.
\texttt{ltx.utils.filesize} \texttt{size} = \texttt{ltx.utils.filesize(\texttt{file})}

Returns the size of the \texttt{file} in bytes. If the \texttt{file} is not found, nothing is returned with \textit{no error raised}. 
Chapter 14

The \texttt{l3}legacy module

Interfaces to legacy concepts

There are a small number of \TeX or \LaTeX\ \texttt{2e} concepts which are not used in \exp\texttt{l3} code but which need to be manipulated when working as a \LaTeX\ \texttt{2e} package. To allow these to be integrated cleanly into \exp\texttt{l3} code, a set of legacy interfaces are provided here.

\begin{verbatim}
\legacy_if_p:n \ legacy_if_p:n \{name\}
\legacy_if:nTF \ legacy_if:nTF \{name\} \{(true code)\} \{(false code)\}
\end{verbatim}

Tests if the \LaTeX\ \texttt{2e}/plain \TeX\ conditional (generated by \texttt{\newif}) is \texttt{true} or \texttt{false} and branches accordingly. The \{name\} of the conditional should \emph{omit} the leading \texttt{if}.

\begin{verbatim}
\legacy_if_set_true:n \ legacy_if_set_true:n \{name\}
\legacy_if_set_false:n \ legacy_if_set_false:n \{name\}
\legacy_if_gset_true:n \ legacy_if_gset_true:n \{name\}
\legacy_if_gset_false:n \ legacy_if_gset_false:n \{name\}
\end{verbatim}

Sets the \LaTeX\ \texttt{2e}/plain \TeX\ conditional \texttt{\if\{name\}} (generated by \texttt{\newif}) to be \texttt{true} or \texttt{false}.

\begin{verbatim}
\legacy_if_set:nn \ legacy_if_set:nn \{name\} \{boolean expr\}
\legacy_if_gset:nn \ legacy_if_gset:nn \{name\} \{boolean expr\}
\end{verbatim}

Sets the \LaTeX\ \texttt{2e}/plain \TeX\ conditional \texttt{\if\{name\}} (generated by \texttt{\newif}) to the result of evaluating the \texttt{boolean expression}. 

\texttt{Nov: 2021-05-10}
Part IV
Data types
Chapter 15

The \l3tl module

Token lists

\TeX{} works with tokens, and \LaTeX{} therefore provides a number of functions to deal with lists of tokens. Token lists may be present directly in the argument to a function:

\begin{verbatim}
\foo:n \{ a collection of \tokens \}
\end{verbatim}

or may be stored in a so-called “token list variable”, which have the suffix \texttt{tl}: a token list variable can also be used as the argument to a function, for example

\begin{verbatim}
\foo:N \l_{\text{some\_tl}}
\end{verbatim}

In both cases, functions are available to test and manipulate the lists of tokens, and these have the module prefix \texttt{tl}. In many cases, functions which can be applied to token list variables are paired with similar functions for application to explicit lists of tokens: the two “views” of a token list are therefore collected together here.

A token list (explicit, or stored in a variable) can be seen either as a list of “items”, or a list of “tokens”. An item is whatever \texttt{\use:n} would grab as its argument: a single non-space token or a brace group, with optional leading explicit space characters (each item is thus itself a token list). A token is either a normal \texttt{N} argument, or \texttt{␣}, \texttt{\{}, or \texttt{\}} (assuming normal \TeX{} category codes). Thus for example

\begin{verbatim}
\{ Hello \} - world
\end{verbatim}

contains six items (Hello, w, o, r, l and d), but thirteen tokens ({\texttt{\}}, \texttt{\#}, \texttt{e}, \texttt{l}, \texttt{1}, \texttt{o}, \texttt{\}, \texttt{␣}, \texttt{\w}, \texttt{\o}, \texttt{\r}, \texttt{\l} and \texttt{\d}). Functions which act on items are often faster than their analogue acting directly on tokens.

15.1 Creating and initialising token list variables

\begin{verbatim}
\tl_new:N \tl_new:N \tl_new:C
\end{verbatim}

\texttt{\tl_new:N} \tl_new:N \(\langle \text{tl var} \rangle\)

\begin{verbatim}
\tl_new:C
\end{verbatim}

Creates a new \(\langle \text{tl var} \rangle\) or raises an error if the name is already taken. The declaration is global. The \(\langle \text{tl var} \rangle\) is initially empty.
\tl const:Nn  \tl const:(Ne|cn|ce)
Creates a new constant \langle tl var \rangle or raises an error if the name is already taken. The
value of the \langle tl var \rangle is set globally to the \langle tokens \rangle.

\tl clear:N  \tl clear:c  \tl gclear:N
\tl clear:c
Clears all entries from the \langle tl var \rangle.

\tl clear_new:N  \tl clear_new:c  \tl gclear_new:N  \tl gclear_new:c
\tl clear_new:N
Ensures that the \langle tl var \rangle exists globally by applying \tl new:N if necessary, then ap-
plies \tl (g)clear:N to leave the \langle tl var \rangle empty.

\tl set_eq:NN  \tl set_eq:(cN|Nc|cc)
\tl gset_eq:NN  \tl gset_eq:(cN|Nc|cc)
\tl set_eq:NN
Sets the content of \langle tl var \rangle equal to that of \langle tl var \rangle.

\tl concat:NNN  \tl concat:ccc  \tl gconcat:NNN  \tl gconcat:ccc
\tl concat:NNN
Concatenates the content of \langle tl var \rangle and \langle tl var \rangle together and saves the result in
\langle tl var \rangle. The \langle tl var \rangle is placed at the left side of the new token list.

\tl if_exist_p:N  \tl if_exist:NTF  \tl if_exist:p:c  \tl if_exist:TF
\tl if_exist:p:c  \tl if_exist:TF
\tl if_exist_p:N
Tests whether the \langle tl var \rangle is currently defined. This does not check that the \langle tl var \rangle
really is a token list variable.

15.2 Adding data to token list variables

\tl set:Nn  \tl set:(NV|NV|Ne|Ne|Nf|Nf|cn|cn|cV|cV|co|co|ce|ce|cf)
\tl gset:Nn  \tl gset:(NV|NV|Ne|Ne|Nf|Nf|cn|cn|cV|cV|co|co|ce|ce|cf)
\tl set:Nn
Sets \langle tl var \rangle to contain \langle tokens \rangle, removing any previous content from the variable.

\tl put_left:Nn  \tl put_left:(NV|NV|Ne|Ne|Nf|Nf|cn|cn|cV|cV|co|co|ce|ce|cf)
\tl gput_left:Nn  \tl gput_left:(NV|NV|Ne|Ne|Nf|Nf|cn|cn|cV|cV|co|co|ce|ce|cf)
\tl put_left:Nn
Appends \langle tokens \rangle to the left side of the current content of \langle tl var \rangle.
Appends \( \text{tokens} \) to the right side of the current content of \( \langle \text{tl var} \rangle \).

\[ \text{\texttt{\textbackslash tl\_put\_right:Nn}} \]
\[ \langle \text{tl var} \rangle \{\{\text{tokens}\}\} \]

15.3 Token list conditionals

\[ \text{\texttt{\textbackslash tl\_if\_blank\_p:n}} \]
\[ \langle \text{token list} \rangle \{\{\text{true code}\}\} \{\{\text{false code}\}\} \]

Tests if the \( \langle \text{token list} \rangle \) consists only of blank spaces (i.e. contains no item). The test is true if \( \langle \text{token list} \rangle \) is zero or more explicit space characters (explicit tokens with character code 32 and category code 10), and is false otherwise.

\[ \text{\texttt{\textbackslash tl\_if\_empty\_p:N}} \]
\[ \langle \text{tl var} \rangle \{\{\text{true code}\}\} \{\{\text{false code}\}\} \]

Tests if the \( \langle \text{tl var} \rangle \) is entirely empty (i.e. contains no tokens at all).

\[ \text{\texttt{\textbackslash tl\_if\_eq\_p:NN}} \]
\[ \langle \text{tl var} \rangle \{\{\text{token list}\}\} \{\{\text{true code}\}\} \{\{\text{false code}\}\} \]

Compares the content of \( \langle \text{tl var} \rangle \) and \( \langle \text{token list} \rangle \) and is logically true if the two contain the same list of tokens (i.e. identical in both the list of characters they contain and the category codes of those characters). Thus for example

\[
\begin{align*}
\text{\texttt{\textbackslash tl\_set:Nn}} & \text{ \_l\_tmpa\_tl} \ {\{\text{abc}\}} \\
\text{\texttt{\textbackslash tl\_set:Ne}} & \text{ \_l\_tmpb\_tl} \ {\{\text{abc}\}} \\
\text{\texttt{\textbackslash tl\_if\_eq:NN}} & \text{ \l\_tmpa\_tl} \ \text{\_l\_tmpb\_tl} \ {\{\text{true}\}} \ {\{\text{false}\}}
\end{align*}
\]

yields false. See also \( \text{\texttt{\textbackslash str\_if\_eq:nnTF}} \) for a comparison that ignores category codes.

\[ \text{\texttt{\textbackslash tl\_if\_eq:NN}} \]
\[ \text{\texttt{\textbackslash tl\_if\_eq:NN}} \{\{\text{token list}\}_1\} \{\{\text{true code}\}\} \{\{\text{false code}\}\} \]

Tests if the \( \langle \text{token list}_1 \rangle \) and the \( \langle \text{token list}_2 \rangle \) contain the same list of tokens, both in respect of character codes and category codes. This conditional is not expandable: see \( \text{\texttt{\textbackslash tl\_if\_eq:NNTF}} \) for an expandable version when both token lists are stored in variables, or \( \text{\texttt{\textbackslash str\_if\_eq:nnTF}} \) if category codes are not important.
\tl_if_in:Nn \tl_if_in:nn Tests if \langle token list_1 \rangle and \langle token list_2 \rangle contain the same list of tokens, both in respect of character codes and category codes. This conditional is not expandable: see \tl_if_eq:nnTF for an expandable version when token lists are stored in variables, or \str_if_eq:nnTF if category codes are not important.

\tl_if_in:Nn \tl_if_in:nn Tests if the \langle token list \rangle is found in the content of the \langle tl var \rangle. The \langle token list \rangle cannot contain the tokens \{,\} or \# (more precisely, explicit character tokens with category code 1 (begin-group) or 2 (end-group), and tokens with category code 6). The search does not enter brace (category code 1/2) groups.

\tl_if_in:nn \tl_if_in:nn Tests if \langle token list_2 \rangle is found inside \langle token list_1 \rangle. The \langle token list_2 \rangle cannot contain the tokens \{,\} or \# (more precisely, explicit character tokens with category code 1 (begin-group) or 2 (end-group), and tokens with category code 6). The search does not enter brace (category code 1/2) groups.

\tl_if_novalue:p:n \tl_if_novalue:nTF Tests if the \langle token list \rangle and the special \c_novalue_tl marker contain the same list of tokens, both in respect of character codes and category codes. This means that \exp_args:No \tl_if_novalue:nTF {\c_novalue_tl} is logically true but \tl_if_novalue:nTF {\c_novalue_tl} is logically false. This function is intended to allow construction of flexible document interface structures in which missing optional arguments are detected.

\tl_if_single_p:N \tl_if_single_p:c \tl_if_single_p:N \tl_if_single_p:c Tests if the content of the \langle tl var \rangle consists of a single \langle item \rangle, i.e. is a single normal token (neither an explicit space character nor a begin-group character) or a single brace group, surrounded by optional spaces on both sides. In other words, such a token list has token count 1 according to \tl_count:N.

\tl_if_single_p:n \tl_if_single_p:n Tests if the \langle token list \rangle has exactly one \langle item \rangle, i.e. is a single normal token (neither an explicit space character nor a begin-group character) or a single brace group, surrounded by optional spaces on both sides. In other words, such a token list has token count 1 according to \tl_count:n.

\tl_if_single_token_p:n \tl_if_single_token_p:n \tl_if_single_token_p:n \tl_if_single_token_p:n Tests if the token list consists of exactly one token, i.e. is either a single space character or a single normal token. Token groups \langle \ldots \rangle are not single tokens.
15.3.1 Testing the first token

Tests if the first \langle token \rangle in the \langle token list \rangle has the same category code as the \langle test token \rangle. In the case where the \langle token list \rangle is empty, the test is always \textit{false}.

Tests if the first \langle token \rangle in the \langle token list \rangle has the same character code as the \langle test token \rangle. In the case where the \langle token list \rangle is empty, the test is always \textit{false}.

Tests if the first \langle token \rangle in the \langle token list \rangle has the same meaning as the \langle test token \rangle. In the case where the \langle token list \rangle is empty, the test is always \textit{false}.

Tests if the first \langle token \rangle in the \langle token list \rangle is an explicit begin-group character (with category code 1 and any character code), in other words, if the \langle token list \rangle starts with a brace group. In particular, the test is \textit{false} if the \langle token list \rangle starts with an implicit token such as \texttt{\c_group_begin_token}, or if it is empty. This function is useful to implement actions on token lists on a token by token basis.

Tests if the first \langle token \rangle in the \langle token list \rangle is a normal N-type argument. In other words, it is neither an explicit space character (explicit token with character code 32 and category code 10) nor an explicit begin-group character (with category code 1 and any character code). An empty argument yields \textit{false}, as it does not have a normal first token. This function is useful to implement actions on token lists on a token by token basis.
Tests if the first \emph{token} in the \emph{token list} is an explicit space character (explicit token with character code 32 and category code 10). In particular, the test is \texttt{false} if the \emph{token list} starts with an implicit token such as \texttt{\textbackslash c\space_token}, or if it is empty. This function is useful to implement actions on token lists on a token by token basis.

\section{Working with token lists as a whole}

\subsection{Using token lists}

\texttt{\tl_to_str:n} \hl{\texttt{\tl_to_str:n \{\token list\}}} \hl{\texttt{\tl_to_str:n \{\token list\}}} \hl{\texttt{\tl_to_str:n \{\token list\}}} \hl{\texttt{\tl_to_str:n \{\token list\}}}

Converts the \emph{token list} to a \emph{string}, leaving the resulting character tokens in the input stream. A \emph{string} is a series of tokens with category code 12 (other) with the exception of spaces, which retain category code 10 (space). The base function requires only a single expansion. Its argument \textit{must} be braced.

\textbf{\TeXhackers note:} This is the \texttt{\textbackslash detokenize} primitive. Converting a \emph{token list} to a \emph{string} yields a concatenation of the string representations of every token in the \emph{token list}. The string representation of a control sequence is

- an escape character, whose character code is given by the internal parameter \texttt{\escapechar}, absent if the \texttt{\escapechar} is negative or greater than the largest character code;
- the control sequence name, as defined by \texttt{\cs_to_str:N};
- a space, unless the control sequence name is a single character whose category at the time of expansion of \texttt{\tl_to_str:n} is not “letter”.

The string representation of an explicit character token is that character, doubled in the case of (explicit) macro parameter characters (normally \#). In particular, the string representation of a token list may depend on the category codes in effect when it is evaluated, and the value of the \texttt{\escapechar}: for instance \texttt{\tl_to_str:n \{\texttt{a}\}} normally produces the three character “backslash”, “lower-case a”, “space”, but it may also produce a single “lower-case a” if the escape character is negative and \texttt{a} is currently not a letter.

\texttt{\tl_use:N} \hl{\texttt{\tl_use:N \{\tl var\}}} \hl{\texttt{\tl_use:N \{\tl var\}}} \hl{\texttt{\tl_use:N \{\tl var\}}} \hl{\texttt{\tl_use:N \{\tl var\}}}

Recovers the content of a \emph{tl var} and places it directly in the input stream. An error is raised if the variable does not exist or if it is invalid. Note that it is possible to use a \emph{tl var} directly without an accessor function.
15.4.2 Counting and reversing token lists

\tl_count:n \{\langle token list\rangle\}  
Counts the number of \langle\textit{items}\rangle in the \langle\textit{token list}\rangle and leaves this information in the input stream. Unbraced tokens count as one element as do each token group \{(\ldots)\}. This process ignores any unprotected spaces within the \langle\textit{token list}\rangle. See also \tl_count:N. This function requires three expansions, giving an \langle\textit{integer denotation}\rangle.

\tl_count:N \{tl var\}  
Counts the number of \langle\textit{items}\rangle in the \langle\textit{tl var}\rangle and leaves this information in the input stream. Unbraced tokens count as one element as do each token group \{(\ldots)\}. This process ignores any unprotected spaces within the \langle\textit{tl var}\rangle. See also \tl_count:n. This function requires three expansions, giving an \langle\textit{integer denotation}\rangle.

\tl_count_tokens:n \{\langle token list\rangle\}  
Counts the number of \TeX tokens in the \langle\textit{token list}\rangle and leaves this information in the input stream. Every token, including spaces and braces, contributes one to the total; thus for instance, the token count of \texttt{a-\{bc\}} is 6. This function requires three expansions, giving an \langle\textit{integer denotation}\rangle.

\tl_reverse:n \{\langle token list\rangle\}  
Reverses the order of the \langle\textit{items}\rangle in the \langle\textit{token list}\rangle, so that \langle\textit{item}_1\rangle\langle\textit{item}_2\rangle\langle\textit{item}_3\rangle\ldots\langle\textit{item}_n\rangle becomes \langle\textit{item}_n\rangle\ldots\langle\textit{item}_3\rangle\langle\textit{item}_2\rangle\langle\textit{item}_1\rangle. This process preserves unprotected space within the \langle\textit{token list}\rangle. Tokens are not reversed within braced token groups, which keep their outer set of braces. In situations where performance is important, consider \tl_reverse_items:n. See also \tl_reverse:N.

\textbf{\TeXhackers note:} The result is returned within \unexpanded, which means that the token list does not expand further when appearing in an \texttt{e}-type or \texttt{x}-type argument expansion.

\tl_reverse:N \{\langle token list\rangle\}  
Sets the \langle\textit{tl var}\rangle to contain the result of reversing the order of its \langle\textit{items}\rangle, so that \langle\textit{item}_1\rangle\langle\textit{item}_2\rangle\langle\textit{item}_3\rangle\ldots\langle\textit{item}_n\rangle becomes \langle\textit{item}_n\rangle\ldots\langle\textit{item}_3\rangle\langle\textit{item}_2\rangle\langle\textit{item}_1\rangle. This process preserves unprotected spaces within the \langle\textit{tl var}\rangle. Braced token groups are copied without reversing the order of tokens, but keep the outer set of braces. This is equivalent to a combination of an assignment and \tl_reverse:V. See also \tl_reverse:V for improved performance.

\tl_reverse_items:n \{\langle token list\rangle\}  
Reverses the order of the \langle\textit{items}\rangle in the \langle\textit{token list}\rangle, so that \langle\textit{item}_1\rangle\langle\textit{item}_2\rangle\langle\textit{item}_3\rangle\ldots\langle\textit{item}_n\rangle becomes \{(\langle\textit{item}_n\rangle)\ldots\{(\langle\textit{item}_3\rangle)\langle\textit{item}_2\rangle\langle\textit{item}_1\rangle). This process removes any unprotected space within the \langle\textit{token list}\rangle. Braced token groups are copied without reversing the order of tokens, and keep the outer set of braces. Items which are initially not braced are copied with braces in the result. In cases where preserving spaces is important, consider the slower function \tl_reverse:n.

\textbf{\TeXhackers note:} The result is returned within \unexpanded, which means that the token list does not expand further when appearing in an \texttt{e}-type or \texttt{x}-type argument expansion.
\tl_trim_spaces:n \⋆ \tl_trim_spaces:n (\{\text{token list}\})

Removes any leading and trailing explicit space characters (explicit tokens with character code 32 and category code 10) from the (\text{token list}) and leaves the result in the input stream.

\text{TeXhacker note:} The result is returned within \texttt{\unexpanded}, which means that the token list does not expand further when appearing in an \texttt{e}-type or \texttt{x}-type argument expansion.

\tl_trim_spaces_apply:nN \⋆ \tl_trim_spaces_apply:nN (\{\text{token list}\} \{\text{function}\})

Removes any leading and trailing explicit space characters (explicit tokens with character code 32 and category code 10) from the (\text{token list}) and passes the result to the (\text{function}) as an \texttt{n}-type argument.

\tl_trim_spaces:N \tl_trim_spaces:c \tl_gtrim_spaces:N \tl_gtrim_spaces:c

Sets the (\text{tl var}) to contain the result of removing any leading and trailing explicit space characters (explicit tokens with character code 32 and category code 10) from its contents.

### 15.4.3 Viewing token lists

\tl_show:N \tl_show:c

Displays the content of the (\text{tl var}) on the terminal.

\text{TeXhacker note:} This is similar to the \TeX primitive \texttt{\show}, wrapped to a fixed number of characters per line.

\tl_show:n \tl_show:e

Displays the (\text{token list}) on the terminal.

\text{TeXhacker note:} This is similar to the \texttt{\showtokens} primitive, wrapped to a fixed number of characters per line.

\tl_log:N \tl_log:c

Writes the content of the (\text{tl var}) in the log file. See also \tl_show:N which displays the result in the terminal.

\tl_log:n \tl_log:e\texttt{(e|x)}

Writes the (\text{token list}) in the log file. See also \tl_show:n which displays the result in the terminal.
15.5 Manipulating items in token lists

15.5.1 Mapping over token lists

All mappings are done at the current group level, i.e. any local assignments made by the \langle function \rangle or \langle code \rangle discussed below remain in effect after the loop.

\tl_map_function:NN \tl_map_function:NN (tl var) \langle function \rangle

Applies \langle function \rangle to every \langle item \rangle in the \langle tl var \rangle. The \langle function \rangle receives one argument for each iteration. This may be a number of tokens if the \langle item \rangle was stored within braces. Hence the \langle function \rangle should anticipate receiving n-type arguments. See also \tl_map_function:nN.

\tl_map_function:nN \tl_map_function:cN \tl_map_function:NN

Updated: 2012-06-29

\tl_map_function:nN \{\langle token list \rangle\} \langle function \rangle

Applies \langle function \rangle to every \langle item \rangle in the \langle token list \rangle. The \langle function \rangle receives one argument for each iteration. This may be a number of tokens if the \langle item \rangle was stored within braces. Hence the \langle function \rangle should anticipate receiving n-type arguments. See also \tl_map_function:nN.

\tl_map_inline:Nn \tl_map_inline:cn

Updated: 2012-06-29

\tl_map_inline:Nn \tl_map_inline:nn \tl_map_tokens:Nn

\tl_map_tokens:nn \tl_map_tokens:cn

\tl_map_tokens:nn \tl_map_tokens:NN

\tl_map_tokens:NN \tl_map_tokens:cN

\tl_map_variable:NNn \tl_map_variable:NN \tl_map_variable:cn

Updated: 2012-06-29

\tl_map_variable:NNn \tl_map_variable:NN (tl var) \langle variable \rangle \langle code \rangle

Stores each \langle item \rangle of the \langle tl var \rangle in turn in the \langle token list \rangle \langle variable \rangle and applies the \langle code \rangle. The \langle code \rangle will usually make use of the \langle variable \rangle, but this is not enforced. The assignments to the \langle variable \rangle are local. Its value after the loop is the last \langle item \rangle in the \langle tl var \rangle, or its original value if the \langle tl var \rangle is blank. See also \tl_map_inline:Nn.

\tl_map_tokens:NN \tl_my_tl \{ \prg_replicate:nn \{ 2 \} \}

expands to twice each \langle item \rangle in the \langle tl var \rangle: for each \langle item \rangle in \tl_my_tl the function \prg_replicate:nn receives 2 and \langle item \rangle as its two arguments. The function \tl_map_inline:Nn is typically faster but is not expandable.
\texttt{\textbackslash tl\_map\_variable:nNn}  \texttt{\textbackslash tl\_map\_variable:nNn \{\texttt{\textbackslash token\ list}\}\ \{\texttt{\textbackslash variable}\} \{\texttt{\textbackslash code}\}}

Stores each \texttt{\textit{item}} of the \texttt{\textit{token\ list}} in turn in the \texttt{\textit{token\ list}} \texttt{\textit{variable}} and applies the \texttt{\textit{code}}. The \texttt{\textit{code}} will usually make use of the \texttt{\textit{variable}}, but this is not enforced. The assignments to the \texttt{\textit{variable}} are local. Its value after the loop is the last \texttt{\textit{item}} in the \texttt{\textit{tl\ var}}, or its original value if the \texttt{\textit{tl\ var}} is blank. See also \texttt{\textbackslash tl\_map\_inline:nn}.

\texttt{\textbackslash tl\_map\_break: } \texttt{\textbackslash tl\_map\_break:}

Used to terminate a \texttt{\textit{tl\_map\_}}\ldots function before all entries in the \texttt{\textit{token\ list}} have been processed. This normally takes place within a conditional statement, for example

\texttt{\textbackslash tl\_map\_inline:Nn \l\_my\_tl}
\{
\texttt{\str\_if\_eq:nnT \{ #1 \} \{ bingo \} \{ \texttt{\textbackslash tl\_map\_break: } \}}
\%
\texttt{Do something useful}
\}

See also \texttt{\textbackslash tl\_map\_break:n}. Use outside of a \texttt{\textit{tl\_map\_}}\ldots scenario leads to low level \TeX{} errors.

\texttt{T\TeX{}hackers note:} When the mapping is broken, additional tokens may be inserted before further items are taken from the input stream. This depends on the design of the mapping function.

\texttt{\textbackslash tl\_map\_break:n}  \texttt{\textbackslash tl\_map\_break:n \{\texttt{\textbackslash code}\}}

Used to terminate a \texttt{\textit{tl\_map\_}}\ldots function before all entries in the \texttt{\textit{token\ list}} have been processed, inserting the \texttt{\textit{code}} after the mapping has ended. This normally takes place within a conditional statement, for example

\texttt{\textbackslash tl\_map\_inline:Nn \l\_my\_tl}
\{
\texttt{\str\_if\_eq:nnT \{ #1 \} \{ bingo \}}
\{ \texttt{\textbackslash tl\_map\_break:n \{ <\texttt{code}> \} } \}
\%
\texttt{Do something useful}
\}

Use outside of a \texttt{\textit{tl\_map\_}}\ldots scenario leads to low level \TeX{} errors.

\texttt{T\TeX{}hackers note:} When the mapping is broken, additional tokens may be inserted before the \texttt{\textit{code}} is inserted into the input stream. This depends on the design of the mapping function.

15.5.2 Head and tail of token lists

Functions which deal with either only the very first item (balanced text or single normal token) in a token list, or the remaining tokens.
Leaves in the input stream the first \textit{item} in the \texttt{(token list)}, discarding the rest of the \texttt{(token list)}. All leading explicit space characters (explicit tokens with character code 32 and category code 10) are discarded; for example

\begin{verbatim}
\tl_head:n { abc }
\end{verbatim}

and

\begin{verbatim}
\tl_head:n { - abc }
\end{verbatim}

both leave \texttt{a} in the input stream. If the “head” is a brace group, rather than a single token, the braces are removed, and so

\begin{verbatim}
\tl_head:n { - { - ab } c }
\end{verbatim}

yields \texttt{ab}. A blank \texttt{(token list)} (see \texttt{\tl_if_blank:nTF}) results in \texttt{\tl_head:n} leaving nothing in the input stream.

\textbf{TEXhackers note:} The result is returned within \texttt{\exp_not:n}, which means that the token list does not expand further when appearing in an \texttt{e}-type or \texttt{x}-type argument expansion.

\begin{verbatim}
\tl_head:w \{ \token list \} \{ \q_stop
\end{verbatim}

Leaves in the input stream the first \textit{item} in the \texttt{(token list)}, discarding the rest of the \texttt{(token list)}. All leading explicit space characters (explicit tokens with character code 32 and category code 10) are discarded. A blank \texttt{(token list)} (which consists only of space characters) results in a low-level \TeX error, which may be avoided by the inclusion of an empty group in the input (as shown), without the need for an explicit test. Alternatively, \texttt{\tl_if_blank:nF} may be used to avoid using the function with a “blank” argument. This function requires only a single expansion, and thus is suitable for use within an \texttt{o}-type expansion. In general, \texttt{\tl_head:n} should be preferred if the number of expansions is not critical.

\begin{verbatim}
\tl_tail:N \{ \token list \} \\
\tl_tail:n \{ \token list \} \\
\tl_tail:w \{ \token list \} \{ \q_stop
\end{verbatim}

Discards all leading explicit space characters (explicit tokens with character code 32 and category code 10) and the first \textit{item} in the \texttt{(token list)}, and leaves the remaining tokens in the input stream. Thus for example

\begin{verbatim}
\tl_tail:n { a - {bc} d }
\end{verbatim}

and

\begin{verbatim}
\tl_tail:n { - a - {bc} d }
\end{verbatim}

both leave \texttt{(bc)d} in the input stream. A blank \texttt{(token list)} (see \texttt{\tl_if_blank:nTF}) results in \texttt{\tl_tail:n} leaving nothing in the input stream.

\textbf{TEXhackers note:} The result is returned within \texttt{\exp_not:n}, which means that the token list does not expand further when appearing in an \texttt{e}-type or \texttt{x}-type argument expansion.

If you wish to handle token lists where the first token may be a space, and this
needs to be treated as the head/tail, this can be accomplished using \tl_if_head_is_space:nTF, for example

\exp_last_unbraced:NNo
\cs_new:Npn \_mypkg_gobble_space:w \c_space_tl { }
\cs_new:Npn \mypkg_tl_head_keep_space:n #1
{ \tl_if_head_is_space:nTF {#1}
{ ~ }
{ \tl_head:n {#1} }
}
\cs_new:Npn \mypkg_tl_tail_keep_space:n #1
{ \tl_if_head_is_space:nTF {#1}
{ \exp_not:o { \__mypkg_gobble_space:w #1 } }
{ \tl_tail:n {#1} }
}

15.5.3 Items and ranges in token lists

\tl_item:nn ⟨token list⟩ ⟨integer expression⟩
Indexing items in the ⟨token list⟩ from 1 on the left, this function evaluates the ⟨integer expression⟩ and leaves the appropriate item from the ⟨token list⟩ in the input stream. If the ⟨integer expression⟩ is negative, indexing occurs from the right of the token list, starting at −1 for the right-most item. If the index is out of bounds, then the function expands to nothing.

\tl_item:Nn ⋆
\tl_item:cn ⋆
New: 2014-07-17

\tl_rand_item:N ⟨tl var⟩
\tl_rand_item:c ⟨token list⟩
\tl_rand_item:n ⟨token list⟩
Selects a pseudo-random item of the ⟨token list⟩. If the ⟨token list⟩ is blank, the result is empty. This is not available in older versions of XeLaTeX.

\TeXhackers note: The result is returned within the \unexpanded primitive (\exp_not:n), which means that the ⟨item⟩ does not expand further when appearing in an e-type or x-type argument expansion.

\tl_rand_item:N ⋆
\tl_rand_item:c ⋆
\tl_rand_item:n ⋆
New: 2016-12-06
Leaves in the input stream the items from the \textit{(start index)} to the \textit{(end index)} inclusive. Spaces and braces are preserved between the items returned (but never at either end of the list). Here \textit{(start index)} and \textit{(end index)} should be \textit{(integer expressions)}. For describing in detail the functions’ behavior, let \textit{m} and \textit{n} be the start and end index respectively. If either is 0, the result is empty. A positive index means ‘start counting from the left end’, and a negative index means ‘from the right end’. Let \textit{l} be the count of the token list.

The \textit{actual start point} is determined as \textit{M} = \textit{m} if \textit{m} > 0 and as \textit{M} = \textit{l} + \textit{m} + 1 if \textit{m} < 0. Similarly the \textit{actual end point} is \textit{N} = \textit{n} if \textit{n} > 0 and \textit{N} = \textit{l} + \textit{n} + 1 if \textit{n} < 0. If \textit{M} > \textit{N}, the result is empty. Otherwise it consists of all items from position \textit{M} to position \textit{N} inclusive; for the purpose of this rule, we can imagine that the token list extends at infinity on either side, with void items at positions \textit{s} for \textit{s} ≤ 0 or \textit{s} > \textit{l}.

Spaces in between items in the actual range are preserved. Spaces at either end of the token list will be removed anyway (think to the token list being passed to \texttt{\tl_trim_spaces:n} to begin with.

Thus, with \textit{l} = 7 as in the examples below, all of the following are equivalent and result in the whole token list

\begin{verbatim}
\tl_range:nnn { abcd-e{}}fg } { 1 } { 7 }
\tl_range:nnn { abcd-e{}}fg } { 1 } { 12 }
\tl_range:nnn { abcd-e{}}fg } { -7 } { 7 }
\tl_range:nnn { abcd-e{}}fg } { -12 } { 7 }
\end{verbatim}

Here are some more interesting examples. The calls

\begin{verbatim}
\iow_term:e { \tl_range:nnn { abcd-e{}}fg } { 2 } { 5 } }
\iow_term:e { \tl_range:nnn { abcd-e{}}fg } { 2 } { -3 } }
\iow_term:e { \tl_range:nnn { abcd-e{}}fg } { -6 } { 5 } }
\iow_term:e { \tl_range:nnn { abcd-e{}}fg } { -6 } { -3 } }
\end{verbatim}

are all equivalent and will print \texttt{bcd\{e\}} on the terminal; similarly

\begin{verbatim}
\iow_term:e { \tl_range:nnn { abcd-e{}}fg } { 2 } { 5 } }
\iow_term:e { \tl_range:nnn { abcd-e{}}fg } { 2 } { -3 } }
\iow_term:e { \tl_range:nnn { abcd-e{}}fg } { -6 } { 5 } }
\iow_term:e { \tl_range:nnn { abcd-e{}}fg } { -6 } { -3 } }
\end{verbatim}

are all equivalent and will print \texttt{bcd \{e\}} on the terminal (note the space in the middle).

To the contrary,

\begin{verbatim}
\tl_range:nnn { abcd-e{}}f } { 2 } { 4 }
\end{verbatim}

will discard the space after ‘d’.

If we want to get the items from, say, the third to the last in a token list \texttt{<tl>}, the call is \texttt{\tl_range:nnn { <tl> } { 3 } { -1 }}. Similarly, for discarding the last item, we can do \texttt{\tl_range:nnn { <tl> } { 1 } { -2 }).

\texttt{TeXhackers note}: The result is returned within the \texttt{\unexpanded} primitive \texttt{(\exp_not:n)}, which means that the \textit{(item)} does not expand further when appearing in an \texttt{e}-type or \texttt{x}-type argument expansion.
15.5.4 Sorting token lists

\tl_sort:Nn \tl_sort:cn
\tl_gsort:Nn \tl_gsort:cn

\tl_sort:Nn \tl_sort:cn \langle \tl var \rangle \{\langle \text{comparison code} \rangle\}

Sorts the items in the \langle \tl var \rangle according to the \langle \text{comparison code} \rangle, and assigns the result to \langle \tl var \rangle. The details of sorting comparison are described in Section 6.1.

\tl_sort:nN \tl_sort:nN *

\tl_sort:nN \tl_sort:nN \{\langle \text{token list} \rangle\} \{\langle \text{conditional} \rangle\}

Sorts the items in the \langle \text{token list} \rangle, using the \langle \text{conditional} \rangle to compare items, and leaves the result in the input stream. The \langle \text{conditional} \rangle should have signature :nnTF, and return true if the two items being compared should be left in the same order, and false if the items should be swapped. The details of sorting comparison are described in Section 6.1.

\TeX{}hackers note: The result is returned within \expnot:n, which means that the token list does not expand further when appearing in an e-type or x-type argument expansion.

15.6 Manipulating tokens in token lists

15.6.1 Replacing tokens

Within token lists, replacement takes place at the top level: there is no recursion into brace groups (more precisely, within a group defined by a category code 1/2 pair).

\tl_replace_once:Nnn \tl_replace_once:NNn \tl_replace_once:nNn \tl_greplace_once:Nnn \tl_greplace_once:NNn \tl_greplace_once:nNn

\tl_replace_once:Nnn \tl_greplace_once:Nnn \tl_replace_once:NNn \tl_greplace_once:NNn \tl_replace_once:nNn \tl_greplace_once:nNn \tl_greplace_once:nnn \tl_greplace_once:nnn

\tl_replace_once:Nnn \tl_greplace_once:Nnn \tl_replace_once:NNn \tl_greplace_once:NNn \tl_replace_once:nNn \tl_greplace_once:nNn \tl_greplace_once:nnn \tl_greplace_once:nnn

\tl_replace_once:Nnn \tl_greplace_once:Nnn \tl_replace_once:NNn \tl_greplace_once:NNn \tl_replace_once:nNn \tl_greplace_once:nNn \tl_greplace_once:nnn \tl_greplace_once:nnn

\tl_replace_once:Nnn \tl_greplace_once:Nnn \tl_replace_once:NNn \tl_greplace_once:NNn \tl_replace_once:nNn \tl_greplace_once:nNn \tl_greplace_once:nnn \tl_greplace_once:nnn

Updated: 2011-08-11

\tl_replace_once:Nnn \tl_greplace_once:Nnn \tl_replace_once:NNn \tl_greplace_once:NNn \tl_replace_once:nNn \tl_greplace_once:nNn \tl_greplace_once:nnn \tl_greplace_once:nnn

\tl_replace_once:Nnn \tl_greplace_once:Nnn \tl_replace_once:NNn \tl_greplace_once:NNn \tl_replace_once:nNn \tl_greplace_once:nNn \tl_greplace_once:nnn \tl_greplace_once:nnn

\tl_replace_once:Nnn \tl_greplace_once:Nnn \tl_replace_once:NNn \tl_greplace_once:NNn \tl_replace_once:nNn \tl_greplace_once:nNn \tl_greplace_once:nnn \tl_greplace_once:nnn

\tl_replace_once:Nnn \tl_greplace_once:Nnn \tl_replace_once:NNn \tl_greplace_once:NNn \tl_replace_once:nNn \tl_greplace_once:nNn \tl_greplace_once:nnn \tl_greplace_once:nnn

\tl_replace_once:Nnn \tl_greplace_once:Nnn \tl_replace_once:NNn \tl_greplace_once:NNn \tl_replace_once:nNn \tl_greplace_once:nNn \tl_greplace_once:nnn \tl_greplace_once:nnn

Replaces the first (leftmost) occurrence of \langle \text{old tokens} \rangle in the \langle \tl var \rangle with \langle \text{new tokens} \rangle. \langle \text{Old tokens} \rangle cannot contain \{, \} or # (more precisely, explicit character tokens with category code 1 (begin-group) or 2 (end-group), and tokens with category code 6).
\tl_replace_all:Nnn \tl_replace_all:(N\text{n}|N\text{nV}|N\text{en}|N\text{ne}|c\text{V}|c\text{nV}|c\text{en}|c\text{ne}|c\text{e}e)
\tl_greplace_all:Nnn \tl_greplace_all:(N\text{n}|N\text{nV}|N\text{en}|N\text{ne}|N\text{ee}|N\text{ene}|c\text{n}|c\text{Vn}|c\text{en}|c\text{ne}|c\text{ee})

Replaces all occurrences of \langle \text{old tokens} \rangle in the \langle \text{tl var} \rangle with \langle \text{new tokens} \rangle. \langle \text{Old tokens} \rangle cannot contain \{, \} or \# (more precisely, explicit character tokens with category code 1 (begin-group) or 2 (end-group), and tokens with category code 6). As this function operates from left to right, the pattern \langle \text{old tokens} \rangle may remain after the replacement (see \tl_remove_all:Nn for an example).

\tl_remove_once:Nn \tl_remove_once:(N\text{V}|N\text{e}|c\text{V}|c\text{e})
\tl_gremove_once:Nn \tl_gremove_once:(N\text{V}|N\text{e}|c\text{V}|c\text{e})

Removes the first (leftmost) occurrence of \langle \text{tokens} \rangle from the \langle \text{tl var} \rangle. The \langle \text{tokens} \rangle cannot contain \{, \} or \# (more precisely, explicit character tokens with category code 1 (begin-group) or 2 (end-group), and tokens with category code 6).

\tl_remove_all:Nn \tl_remove_all:(N\text{V}|N\text{e}|c\text{V}|c\text{e})
\tl_gremove_all:Nn \tl_gremove_all:(N\text{V}|N\text{e}|c\text{V}|c\text{e})

Removes all occurrences of \langle \text{tokens} \rangle from the \langle \text{tl var} \rangle. The \langle \text{tokens} \rangle cannot contain \{, \} or \# (more precisely, explicit character tokens with category code 1 (begin-group) or 2 (end-group), and tokens with category code 6). As this function operates from left to right, the pattern \langle \text{tokens} \rangle may remain after the removal, for instance,

\tl_set:Nn \l_tmpa_tl {abbcdd} \tl_remove_all:Nn \l_tmpa_tl {bc}

results in \l_tmpa_tl containing abcd.

15.6.2 Reassigning category codes

These functions allow the rescanning of tokens: re-apply \TeX{}'s tokenization process to apply category codes different from those in force when the tokens were absorbed. Whilst this functionality is supported, it is often preferable to find alternative approaches to achieving outcomes rather than rescanning tokens (for example construction of token lists token-by-token with intervening category code changes or using \char_generate:nn).
Sets \langle \text{tl var} \rangle to contain \langle \text{tokens} \rangle, applying the category code régime specified in the \langle \text{setup} \rangle before carrying out the assignment. (Category codes applied to tokens not explicitly covered by the \langle \text{setup} \rangle are those in force at the point of use of \texttt{\tl_set_rescan:Nnn}.) This allows the \langle \text{tl var} \rangle to contain material with category codes other than those that apply when \langle \text{tokens} \rangle are absorbed. The \langle \text{setup} \rangle is run within a group and may contain any valid input, although only changes in category codes, such as uses of \texttt{\cctab_select:N}, are relevant. See also \texttt{\tl_rescan:nn}.

\textbf{\TeXhackers note:} The \langle \text{tokens} \rangle are first turned into a string (using \texttt{\tl_to_str:n}). If the string contains one or more characters with character code \texttt{\newlinechar} (set equal to \texttt{\endlinechar} unless that is equal to 32, before the user \langle \text{setup} \rangle), then it is split into lines at these characters, then read as if reading multiple lines from a file, ignoring spaces (catcode 10) at the beginning and spaces and tabs (character code 32 or 9) at the end of every line. Otherwise, spaces (and tabs) are retained at both ends of the single-line string, as if it appeared in the middle of a line read from a file.

\textbf{\TeXhackers note:} The \langle \text{tokens} \rangle are first turned into a string (using \texttt{\tl_to_str:n}). If the string contains one or more characters with character code \texttt{\newlinechar} (set equal to \texttt{\endlinechar} unless that is equal to 32, before the user \langle \text{setup} \rangle), then it is split into lines at these characters, then read as if reading multiple lines from a file, ignoring spaces (catcode 10) at the beginning and spaces and tabs (character code 32 or 9) at the end of every line. Otherwise, spaces (and tabs) are retained at both ends of the single-line string, as if it appeared in the middle of a line read from a file.

Contrarily to the \texttt{\scantokens \LaTeX} primitive, \texttt{\tl_rescan:nn} tokenizes the whole string in the same category code regime rather than one token at a time, so that directives such as \texttt{\verb} that rely on changing category codes will not function properly.

### 15.7 Constant token lists

\texttt{\c_empty_tl} Constant that is always empty.
\texttt{\textbackslash c\_novalue\_tl} A marker for the absence of an argument. This constant \texttt{tl} can safely be typeset (cf. \texttt{\textbackslash q\_nil}), with the result being \texttt{-NoValue-}. It is important to note that \texttt{\textbackslash c\_novalue\_tl} is constructed such that it will not match the simple text input \texttt{-NoValue-}, \textit{i.e.} that
\begin{verbatim}
\tl\_if\_eq\_:NnTF \texttt{\textbackslash c\_novalue\_tl} { \texttt{-NoValue-} }
\end{verbatim}
is logically \textit{false}. The \texttt{\textbackslash c\_novalue\_tl} marker is intended for use in creating document-level interfaces, where it serves as an indicator that an (optional) argument was omitted. In particular, it is distinct from a simple empty \texttt{tl}.

\texttt{\textbackslash c\_space\_tl} An explicit space character contained in a token list (compare this with \texttt{\textbackslash c\_space\_token}). For use where an explicit space is required.

### 15.8 Scratch token lists

\texttt{\textbackslash l\_tmpa\_tl} Scratch token lists for local assignment. These are never used by the kernel code, and so are safe for use with any \texttt{\LaTeX3}-defined function. However, they may be overwritten by other non-kernel code and so should only be used for short-term storage.

\texttt{\textbackslash l\_tmpb\_tl}

\texttt{\textbackslash g\_tmpa\_tl} Scratch token lists for global assignment. These are never used by the kernel code, and so are safe for use with any \texttt{\LaTeX3}-defined function. However, they may be overwritten by other non-kernel code and so should only be used for short-term storage.

\texttt{\textbackslash g\_tmpb\_tl}
Chapter 16

The l3tl-build module

Piecewise tl constructions

16.1 Constructing \langle tl var \rangle by accumulation

When creating a \langle tl var \rangle by accumulation of many tokens, the performance available using a combination of \tl_set:Nn and \tl_put_right:Nn or similar begins to become an issue. To address this, a set of functions are available to “build” a \langle tl var \rangle. The performance of this approach is much more efficient than the standard \tl_put_right:Nn, but the constructed token list cannot be accessed during construction other than by methods provided in this section.

Whilst the exact performance difference is dependent on the size of each added block of tokens and the total number of blocks, in general, the \tl_build_(g)put... functions will out-perform the basic \tl_(g)put... equivalent if more than 100 non-empty addition operations occur. See https://github.com/latex3/latex3/issues/1393#issuecomment-1880164756 for a more detailed analysis.

\tl_build_begin:N \tl_build_gbegin:N

Clears the \langle tl var \rangle and sets it up to support other \tl_build_... functions. Until \tl_build_end:N \langle tl var \rangle or \tl_build_gend:N \langle tl var \rangle is called, applying any function from l3tl other than \tl_build_... will lead to incorrect results. The begin and gbegin functions must be used for local and global \langle tl var \rangle respectively.

\tl_build_put_left:Nn \tl_build_gput_left:Nn \tl_build_put_left:Ne \tl_build_gput_left:Ne

\tl_build_put_right:Nn \tl_build_gput_right:Nn \tl_build_put_right:Ne \tl_build_gput_right:Ne

\tl_build_put_left:N \langle tl var \rangle \{(tokens)\}
\tl_build_put_right:N \langle tl var \rangle \{(tokens)\}

Adds \langle tokens \rangle to the left or right side of the current contents of \langle tl var \rangle. The \langle tl var \rangle must have been set up with \tl_build_begin:N or \tl_build_gbegin:N. The put and gput functions must be used for local and global \langle tl var \rangle respectively. The right functions are about twice faster than the left functions.
\tl_build_end:N \tl_build_gend:N

\tl_build_end:N \tl_var

 Gets the contents of \tl_var and stores that into the \tl_var using \tl_set:Nn or \tl_gset:Nn. The \tl_var must have been set up with \tl_build_begin:N or \tl_build_gbegin:N. The \texttt{end} and \texttt{gend} functions must be used for local and global \tl_var respectively. These functions completely remove the setup code that enabled \tl_var to be used for other \tl_build_... functions. After the action of \texttt{end/gend}, the \tl_var may be manipulated using standard \tl functions.

\tl_build_get_intermediate:NN \tl_build_get_intermediate:NN \tl_var_1 \tl_var_2

\texttt{New: 2018-04-01}

Stores the contents of the \tl_var_1 in the \tl_var_2. The \tl_var_1 must have been set up with \tl_build_begin:N or \tl_build_gbegin:N. The \tl_var_2 is a “normal” token list variable, assigned locally using \tl_set:Nn.
Chapter 17

The \texttt{l3str} module

Strings

\TeX{} associates each character with a category code: as such, there is no concept of a “string” as commonly understood in many other programming languages. However, there are places where we wish to manipulate token lists while in some sense “ignoring” category codes: this is done by treating token lists as strings in a \TeX{} sense.

A \TeX{} string (and thus an \texttt{expl3} string) is a series of characters which have category code 12 (“other”) with the exception of space characters which have category code 10 (“space”). Thus at a technical level, a \TeX{} string is a token list with the appropriate category codes. In this documentation, these are simply referred to as strings.

String variables are simply specialised token lists, but by convention should be named with the suffix \ldots\texttt{str}. Such variables should contain characters with category code 12 (other), except spaces, which have category code 10 (blank space). All the functions in this module which accept a token list argument first convert it to a string using \texttt{\tl_to_str:n} for internal processing, and do not treat a token list or the corresponding string representation differently.

As a string is a subset of the more general token list, it is sometimes unclear when one should be used over the other. Use a string variable for data that isn’t primarily intended for typesetting and for which a level of protection from unwanted expansion is suitable. This data type simplifies comparison of variables since there are no concerns about expansion of their contents.

The functions \texttt{\cs_to_str:N}, \texttt{\tl_to_str:n}, \texttt{\tl_to_str:N} and \texttt{\token_to_str:N} (and variants) generate strings from the appropriate input: these are documented in \texttt{l3basics}, \texttt{l3tl} and \texttt{l3token}, respectively.

Most expandable functions in this module come in three flavours:

- \texttt{\str\ldots:N}, which expect a token list or string variable as their argument;
- \texttt{\str\ldots:n}, taking any token list (or string) as an argument;
- \texttt{\str\ldots\_ignore_spaces:n}, which ignores any space encountered during the operation: these functions are typically faster than those which take care of escaping spaces appropriately.
17.1 Creating and initialising string variables

\texttt{\str_new:N \str_new:c}\new:2015-09-18

Creates a new \langle str var \rangle or raises an error if the name is already taken. The declaration is global. The \langle str var \rangle is initially empty.

\texttt{\str_new:N \str_new:c \str_const:Nn \str_const:Nn (\texttt{NV}|\texttt{Ne}|\texttt{cn}|\texttt{cV}|\texttt{ce})}\new:2015-09-18

\texttt{\str_const:Nn \str_const:Nn ( token list )}\new:2015-09-18

Creates a new constant \langle str var \rangle or raises an error if the name is already taken. The value of the \langle str var \rangle is set globally to the \langle token list \rangle, converted to a string.

\texttt{\str_clear:N \str_clear:c \str_gclear:N \str_gclear:c}\new:2015-09-18

Clears the content of the \langle str var \rangle.

\texttt{\str_clear:N \str_clear:c \str_gclear:N \str_gclear:c}\new:2015-09-18

\texttt{\str_gclear:N \str_gclear:c}\new:2015-09-18

Ensures that the \langle str var \rangle exists globally by applying \str_new:N if necessary, then applies \str_(g)clear:N to leave the \langle str var \rangle empty.

\texttt{\str_set_eq:NN \str_set_eq:cc \str_gset_eq:NN \str_gset_eq:cc}\new:2015-09-18

Sets the content of \langle str var_1 \rangle equal to that of \langle str var_2 \rangle.

\texttt{\str_set_eq:NN \str_set_eq:cc \str_set_eq:cc}\new:2015-09-18

\texttt{\str_set_eq:NN \str_set_eq:cc \str_set_eq:cc}\new:2015-09-18

\texttt{\str_concat:NNN \str_concat:ccc \str_gconcat:NNN \str_gconcat:ccc}\new:2017-10-08

Concatenates the content of \langle str var_2 \rangle and \langle str var_3 \rangle together and saves the result in \langle str var_1 \rangle. The \langle str var_2 \rangle is placed at the left side of the new string variable. The \langle str var_2 \rangle and \langle str var_3 \rangle must indeed be strings, as this function does not convert their contents to a string.

\texttt{\str_if_exist_p:N \str_if_exist_p:c \str_if_exist:N \str_if_exist:N TF \str_if_exist:NTF \str_if_exist:NTF cTF cTF}\new:2015-09-18

Tests whether the \langle str var \rangle is currently defined. This does not check that the \langle str var \rangle really is a string.
17.2 Adding data to string variables

\texttt{\textbackslash str\_set:Nn} \langle \texttt{str\_var} \rangle \{\langle \texttt{token\_list} \rangle\}

Converts the \langle \texttt{token\_list} \rangle to a \langle \texttt{string} \rangle, and stores the result in \langle \texttt{str\_var} \rangle.

\texttt{\textbackslash str\_set:Nn} \texttt{\textbackslash str\_set:} (\texttt{NV|Ne|cn|cV|ce})
\texttt{\textbackslash str\_gset:Nn} \texttt{\textbackslash str\_gset:} (\texttt{NV|Ne|cn|cV|ce})

\texttt{\textbackslash str\_put\_left:Nn} \langle \texttt{str\_var} \rangle \{\langle \texttt{token\_list} \rangle\}
\texttt{\textbackslash str\_put\_left:Nn} \texttt{\textbackslash str\_put\_left:} (\texttt{NV|Ne|cn|cV|ce})
\texttt{\textbackslash str\_gput\_left:Nn} \texttt{\textbackslash str\_gput\_left:} (\texttt{NV|Ne|cn|cV|ce})

Converts the \langle \texttt{token\_list} \rangle to a \langle \texttt{string} \rangle, and prepends the result to \langle \texttt{str\_var} \rangle. The current contents of the \langle \texttt{str\_var} \rangle are not automatically converted to a string.

\texttt{\textbackslash str\_put\_right:Nn} \langle \texttt{str\_var} \rangle \{\langle \texttt{token\_list} \rangle\}
\texttt{\textbackslash str\_put\_right:Nn} \texttt{\textbackslash str\_put\_right:} (\texttt{NV|Ne|cn|cV|ce})
\texttt{\textbackslash str\_gput\_right:Nn} \texttt{\textbackslash str\_gput\_right:} (\texttt{NV|Ne|cn|cV|ce})

Converts the \langle \texttt{token\_list} \rangle to a \langle \texttt{string} \rangle, and appends the result to \langle \texttt{str\_var} \rangle. The current contents of the \langle \texttt{str\_var} \rangle are not automatically converted to a string.

17.3 String conditionals

\texttt{\textbackslash str\_if\_empty\_p:N} \langle \texttt{str\_var} \rangle
\texttt{\textbackslash str\_if\_empty\_p:p} \langle \texttt{str\_var} \rangle
\texttt{\textbackslash str\_if\_empty\_p:NTF} \langle \texttt{str\_var} \rangle \{\langle \texttt{true\_code} \rangle\} \{\langle \texttt{false\_code} \rangle\}

Tests if the \langle \texttt{string\_variable} \rangle is entirely empty (\textit{i.e.} contains no characters at all).

\texttt{\textbackslash str\_if\_eq\_p:NN} \langle \texttt{str\_var\_1} \rangle \langle \texttt{str\_var\_2} \rangle
\texttt{\textbackslash str\_if\_eq:NNTF} \langle \texttt{str\_var\_1} \rangle \langle \texttt{str\_var\_2} \rangle \{\langle \texttt{true\_code} \rangle\} \{\langle \texttt{false\_code} \rangle\}

\texttt{\textbackslash str\_if\_eq\_p:NN} \langle \texttt{str\_var\_1} \rangle \langle \texttt{str\_var\_2} \rangle
\texttt{\textbackslash str\_if\_eq:NN} \langle \texttt{str\_var\_1} \rangle \langle \texttt{str\_var\_2} \rangle

Compares the content of two \langle \texttt{str\_variables} \rangle and is logically \texttt{true} if the two contain the same characters in the same order. See \texttt{\textbackslash tl\_if\_eq:NNTF} to compare tokens (including their category codes) rather than characters.
\str_if_eq_p:nn \str_if_eq_p:(Vn|on|no|nV|VV|vn|nv|ee) \str_if_eq_p:nnTF
\str_if_eq:nnTF
\str_if_eq_p:nn ⋆ \str_if_eq_p:(Vn|on|no|nV|VV|vn|nv|ee) ⋆

Compares the two \textit{token lists} on a character by character basis (namely after converting them to strings), and is \textit{true} if the two \textit{strings} contain the same characters in the same order. Thus for example

\texttt{\str_if_eq_p:no \{ abc \} \{ \tl_to_str:n \{ abc \} \}}

is logically \textit{true}. See \texttt{\tl_if_eq:nnTF} to compare tokens (including their category codes) rather than characters.

\str_if_in:NnTF \str_if_in:cnTF \str_if_in:nnTF

Converts the \textit{token list} to a \textit{string} and tests if that \textit{string} is found in the content of the \textit{str var}.

\str_case:nn \str_case:nnTF \str_case:nn TF \str_case:nnTF ⋆ \str_case:nnTF ⋆

Compares the \textit{test string} in turn with each of the \textit{string case} s (all token lists are converted to strings). If the two are equal (as described for \texttt{\str_if_eq:nnTF}) then the associated \textit{code} is left in the input stream and other cases are discarded. If any of the cases are matched, the \textit{true code} is also inserted into the input stream (after the code for the appropriate case), while if none match then the \textit{false code} is inserted. The function \texttt{\str_case:nn}, which does nothing if there is no match, is also available.

This set of functions performs no expansion on each \textit{string case} argument, so any variable in there will be compared as a string. If expansion is needed in the \textit{string case} s, then \texttt{\str_case_e:nn(TF)} should be used instead.
\str_case_en * \str_case_enTF \{\langle test string\rangle\}
\str_case_en * \{ 
  \{\langle string case\rangle\} \{\langle code case\rangle\}
  \{\langle string case\rangle\} \{\langle code case\rangle\}
  ... 
  \{\langle string case\rangle\} \{\langle code case\rangle\} 
\}\{\langle true code\rangle\} 
\{\langle false code\rangle\}

Compares the full expansion of the \langle test string \rangle in turn with the full expansion of the \langle string case \rangles (all token lists are converted to strings). If the two full expansions are equal (as described for \str_if_eq:eeTF) then the associated \langle code \rangle is left in the input stream and other cases are discarded. If any of the cases are matched, the \langle true code \rangle is also inserted into the input stream (after the code for the appropriate case), while if none match then the \langle false code \rangle is inserted. In \str_case_enTF, the \langle test string \rangle is expanded in each comparison, and must always yield the same result: for example, random numbers must not be used within this string.

\str_compare_p:nNn * \str_compare_p:eNe * \str_compare:nNnTF * \str_compare:eNeTF *
\{\langle token list\rangle\} \langle relation \rangle \{\langle token list\rangle\} \{\langle true code\rangle\} \{\langle false code\rangle\}

Compares the two \langle token lists \rangle on a character by character basis (namely after converting them to strings) in a lexicographic order according to the character codes of the characters. The \langle relation \rangle can be <, =, or > and the test is true under the following conditions:

- for <, if the first string is earlier than the second in lexicographic order;
- for =, if the two strings have exactly the same characters;
- for >, if the first string is later than the second in lexicographic order.

Thus for example the following is logically true:
\str_compare_p:nNn \{ ab \} < \{ abc \}

\TeXhackers note: This is a wrapper around the \TeX primitive \langle pdf \rangle strcmp. It is meant for programming and not for sorting textual contents, as it simply considers character codes and not more elaborate considerations of grapheme clusters, locale, etc.

17.4 Mapping over strings

All mappings are done at the current group level, \textit{i.e.} any local assignments made by the \langle function \rangle or \langle code \rangle discussed below remain in effect after the loop.

\str_map_function:nN * \str_map_function:nN {\langle token list\rangle} \langle function \rangle 
\str_map_function:nN * \str_map_function:NN {\langle str var \rangle} \langle function \rangle 
\str_map_function:cN * 
\{\langle token list\rangle\} to a \langle string \rangle then applies \langle function \rangle to every \langle character \rangle in the \langle string \rangle including spaces.
\str_map_inline:nn \str_map_inline:Nn \str_map_inline:cn
Converts the \langle token list \rangle to a \langle string \rangle then applies the \langle inline function \rangle to every \langle character \rangle in the \langle str var \rangle including spaces. The \langle inline function \rangle should consist of code which receives the \langle character \rangle as \#1.

\str_map_tokens:nn \str_map_tokens:Nn \str_map_tokens:cn
Converts the \langle token list \rangle to a \langle string \rangle then applies \langle code \rangle to every \langle character \rangle in the \langle string \rangle including spaces. The \langle code \rangle receives each character as a trailing brace group. This is equivalent to \str_map_function:nN if the \langle code \rangle consists of a single function.

\str_map_variable:nNn \str_map_variable:NNn \str_map_variable:cNn
Converts the \langle token list \rangle to a \langle string \rangle then stores each \langle character \rangle in the \langle string \rangle (including spaces) in turn in the \langle string \rangle \langle variable \rangle and applies the \langle code \rangle. The \langle code \rangle will usually make use of the \langle variable \rangle, but this is not enforced. The assignments to the \langle variable \rangle are local. Its value after the loop is the last \langle character \rangle in the \langle string \rangle, or its original value if the \langle string \rangle is empty. See also \str_map_inline:nn.

\str_map_break: \str_map_break:
Used to terminate a \str_map_... function before all characters in the \langle string \rangle have been processed. This normally takes place within a conditional statement, for example

\str_map_inline:Nn \_my_str
{\str_if_eq:nnT { #1 } { bingo } { \str_map_break: }
% Do something useful
}

See also \str_map_break:n. Use outside of a \str_map_... scenario leads to low level \TeX errors.

\TeXhackers note: When the mapping is broken, additional tokens may be inserted before continuing with the code that follows the loop. This depends on the design of the mapping function.
\str_map_break:n \str_map_break:n \langle \text{code} \rangle

Used to terminate a \str_map\ldots function before all characters in the \langle\text{string}\rangle have been processed, inserting the \langle\text{code}\rangle after the mapping has ended. This normally takes place within a conditional statement, for example

\str_map_inline:Nn \l_my_str
{\str_if_eq:nnt { #1 } { bingo }{ % Do something useful
{ \str_map_break:n \langle \text{code} \rangle
}
}

Use outside of a \str_map\ldots scenario leads to low level \TeX errors.

\TeXhackersnote: When the mapping is broken, additional tokens may be inserted before the \langle\text{code}\rangle is inserted into the input stream. This depends on the design of the mapping function.

17.5 Working with the content of strings

\str_use:N \str_use:N \langle \text{str var} \rangle

Recover the content of a \langle\text{str var}\rangle and places it directly in the input stream. An error is raised if the variable does not exist or if it is invalid. Note that it is possible to use a \langle\text{str}\rangle directly without an accessor function.

\str_count:n \str_count:n \langle \text{token list} \rangle
\str_count:n \str_count:n \langle \text{token list} \rangle
\str_count:n \str_count:n \langle \text{token list} \rangle
\str_count:n \str_count:n \langle \text{token list} \rangle

Leaves in the input stream the number of characters in the string representation of \langle\text{token list}\rangle, as an integer denotation. The functions differ in their treatment of spaces. In the case of \str_count:n and \str_count:n, all characters including spaces are counted. The \str_count:n function leaves the number of non-space characters in the input stream.

\str_count_spaces:n \str_count_spaces:n \langle \text{token list} \rangle
\str_count_spaces:n \str_count_spaces:n \langle \text{token list} \rangle
\str_count_spaces:n \str_count_spaces:n \langle \text{token list} \rangle

Leaves in the input stream the number of space characters in the string representation of \langle\text{token list}\rangle, as an integer denotation. Of course, this function has no \_ignore_spaces variant.
\str_head:N * \str_head:n {\langle token list\rangle}
\str_head:c *
\str_head:n *
\str_head_ignore_spaces:n *

New: 2015-09-18

Converts the \langle token list\rangle into a \langle string\rangle. The first character in the \langle string\rangle is then left in the input stream, with category code “other”. The functions differ if the first character is a space: \str_head:N and \str_head:n return a space token with category code 10 (blank space), while the \str_head_ignore_spaces:n function ignores this space character and leaves the first non-space character in the input stream. If the \langle string\rangle is empty (or only contains spaces in the case of the \_ignore_spaces function), then nothing is left on the input stream.

\str_tail:N * \str_tail:n {\langle token list\rangle}
\str_tail:c *
\str_tail:n *
\str_tail_ignore_spaces:n *

New: 2015-09-18

Converts the \langle token list\rangle to a \langle string\rangle, removes the first character, and leaves the remaining characters (if any) in the input stream, with category codes 12 and 10 (for spaces). The functions differ in the case where the first character is a space: \str_tail:N and \str_tail:n only trim that space, while \str_tail_ignore_spaces:n removes the first non-space character and any space before it. If the \langle token list\rangle is empty (or blank in the case of the \_ignore_spaces variant), then nothing is left on the input stream.

\str_item:Nn * \str_item:nn {\langle token list\rangle} {\langle integer expression\rangle}
\str_item:nn *
\str_item_ignore_spaces:nn *

New: 2015-09-18

Converts the \langle token list\rangle to a \langle string\rangle, and leaves in the input stream the character in position \langle integer expression\rangle of the \langle string\rangle, starting at 1 for the first (left-most) character. In the case of \str_item:Nn and \str_item:nn, all characters including spaces are taken into account. The \str_item Ignore Spaces:nn function skips spaces when counting characters. If the \langle integer expression\rangle is negative, characters are counted from the end of the \langle string\rangle. Hence, −1 is the right-most character, etc.
\texttt{\str{range}:nnn} \star \texttt{\str{range}:nnn} \{\langle\text{token list}\rangle\} \{\langle\text{start index}\rangle\} \{\langle\text{end index}\rangle\} \\
\texttt{\str{range}:cnn} \star \\
\texttt{\str{range}:nnn} \star \\
\texttt{\str{range}_ignore\_spaces:nnn} \star \\

New: 2015-09-18

Converts the \langle\text{token list}\rangle to a \langle\text{string}\rangle, and leaves in the input stream the characters from the \langle\text{start index}\rangle to the \langle\text{end index}\rangle inclusive. Spaces are preserved and counted as items (contrast this with \texttt{\str{tl\_range}:nnn} where spaces are not counted as items and are possibly discarded from the output).

Here \langle\text{start index}\rangle and \langle\text{end index}\rangle should be integer denotations. For describing in detail the functions’ behavior, let \(m\) and \(n\) be the start and end index respectively. If either is 0, the result is empty. A positive index means ‘start counting from the left end’, a negative index means ‘start counting from the right end’. Let \(l\) be the count of the token list.

The actual start point is determined as \(M = m\) if \(m > 0\) and as \(M = l + m + 1\) if \(m < 0\). Similarly the actual end point is \(N = n\) if \(n > 0\) and \(N = l + n + 1\) if \(n < 0\). If \(M > N\), the result is empty. Otherwise it consists of all items from position \(M\) to position \(N\) inclusive; for the purpose of this rule, we can imagine that the token list extends at infinity on either side, with void items at positions \(s\) for \(s \leq 0\) or \(s > l\). For instance,

\begin{verbatim}
\iow{term}{e}{ \str{range}:nnn { abcdef } { 2 } { 5 } }
\iow{term}{e}{ \str{range}:nnn { abcdef } { -4 } { -1 } }
\iow{term}{e}{ \str{range}:nnn { abcdef } { -2 } { -1 } }
\iow{term}{e}{ \str{range}:nnn { abcdef } { 0 } { -1 } }
\end{verbatim}

prints bcde, cdef, ef, and an empty line to the terminal. The \langle\text{start index}\rangle must always be smaller than or equal to the \langle\text{end index}\rangle: if this is not the case then no output is generated. Thus

\begin{verbatim}
\iow{term}{e}{ \str{range}:nnn { abcdef } { 5 } { 2 } }
\iow{term}{e}{ \str{range}:nnn { abcdef } { -1 } { -4 } }
\end{verbatim}

both yield empty strings.

The behavior of \texttt{\str{range}_ignore\_spaces:nnn} is similar, but spaces are removed before starting the job. The input

\begin{verbatim}
\iow{term}{e}{ \str{range}_ignore\_spaces:nnn { abcdef } { 2 } { 5 } }
\iow{term}{e}{ \str{range}_ignore\_spaces:nnn { abcdef } { 2 } { -3 } }
\iow{term}{e}{ \str{range}_ignore\_spaces:nnn { abcdef } { -6 } { 5 } }
\iow{term}{e}{ \str{range}_ignore\_spaces:nnn { abcdef } { -6 } { -3 } }
\iow{term}{e}{ \str{range}_ignore\_spaces:nnn { abc-efg } { 2 } { 5 } }
\iow{term}{e}{ \str{range}_ignore\_spaces:nnn { abc-efg } { 2 } { -3 } }
\iow{term}{e}{ \str{range}_ignore\_spaces:nnn { abc-efg } { -6 } { 5 } }
\iow{term}{e}{ \str{range}_ignore\_spaces:nnn { abc-efg } { -6 } { -3 } }
\end{verbatim}

137
\iow_term:e \{ \str_range_ignore_spaces:nnn \{ abcd-efg \} \{ 2 \} \{ 5 \} \}
\iow_term:e \{ \str_range_ignore_spaces:nnn \{ abcd-efg \} \{ 2 \} \{ -3 \} \}
\iow_term:e \{ \str_range_ignore_spaces:nnn \{ abcd-efg \} \{ -6 \} \{ 5 \} \}
\iow_term:e \{ \str_range_ignore_spaces:nnn \{ abcd-efg \} \{ -6 \} \{ -3 \} \}

will print four instances of bcde, four instances of bc e and eight instances of bcde.

17.6 Modifying string variables

\str_replace_once:Nnn \str_replace_once:cnn \str_greplace_once:Nnn \str_greplace_once:cnn

\str_replace_once:Nnn (str var) \{(old)} \{(new)}
Converts the (old) and (new) token lists to strings, then replaces the first (leftmost) occurrence of (old string) in the (str var) with (new string).

\str_replace_all:Nnn \str_replace_all:cnn \str_greplace_all:Nnn \str_greplace_all:cnn

\str_replace_all:Nnn (str var) \{(old)} \{(new)}
Converts the (old) and (new) token lists to strings, then replaces all occurrences of (old string) in the (str var) with (new string). As this function operates from left to right, the pattern (old string) may remain after the replacement (see \str_remove_all:Nn for an example).

\str_remove_once:Nn \str_remove_once:cn \str_gremove_once:Nn \str_gremove_once:cn

\str_remove_once:Nn (str var) \{(token list)}
Converts the (token list) to a (string) then removes the first (leftmost) occurrence of (string) from the (str var).

\str_remove_all:Nn \str_remove_all:cn \str_gremove_all:Nn \str_gremove_all:cn

\str_remove_all:Nn (str var) \{(token list)}
Converts the (token list) to a (string) then removes all occurrences of (string) from the (str var). As this function operates from left to right, the pattern (string) may remain after the removal, for instance,

\str_set:Nn \l_tmpa_str {abbccd} \str_remove_all:Nn \l_tmpa_str \{bc\}

results in \l_tmpa_str containing abcd.


\subsection{String manipulation}

\begin{itemize}
  \item \texttt{\str_lowercase:n \{\textbf{tokens}\}}
  \item \texttt{\str_uppercase:n \{\textbf{tokens}\}}
\end{itemize}

Converts the input \texttt{\{\textbf{tokens}\}} to their string representation, as described for \texttt{\tl_to_str:n}, and then to the lower or upper case representation using a one-to-one mapping as described by the Unicode Consortium file \texttt{UnicodeData.txt}.

These functions are intended for case changing programmatic data in places where upper/lower case distinctions are meaningful. One example would be automatically generating a function name from user input where some case changing is needed. In this situation the input is programmatic, not textual, case does have meaning and a language-independent one-to-one mapping is appropriate. For example

\begin{verbatim}
\cs_new_protected:Npn \myfunc:nn #1#2
{\cs_set_protected:cpn
    \{user
        \str_uppercase:f \{ \tl_head:n \{#1\} \}
        \str_lowercase:f \{ \tl_tail:n \{#1\} \}
    \}
    \{ #2 \}
}
\end{verbatim}

would be used to generate a function with an auto-generated name consisting of the upper case equivalent of the supplied name followed by the lower case equivalent of the rest of the input.

These functions should not be used for

- Caseless comparisons: use \texttt{\str_casefold:n} for this situation (case folding is distinct from lower casing).

- Case changing text for typesetting: see the \texttt{\text_lowercase:n(n)}, \texttt{\text_uppercase:n(n)} and \texttt{\text_titlecase_(all|once):n(n)} functions which correctly deal with context-dependence and other factors appropriate to text case changing.
\texttt{\str_casefold:n} \{\texttt{tokens}\} \new{2022-10-16}

Converts the input \texttt{\langle tokens\rangle} to their string representation, as described for \texttt{\tl_to_str:n}, and then folds the case of the resulting \texttt{\langle string\rangle} to remove case information. The result of this process is left in the input stream.

String folding is a process used for material such as identifiers rather than for “text”. The folding provided by \texttt{\str_casefold:n} follows the mappings provided by the Unicode Consortium, who state:

Case folding is primarily used for caseless comparison of text, such as identifiers in a computer program, rather than actual text transformation. Case folding in Unicode is based on the lowercase mapping, but includes additional changes to the source text to help make it language-insensitive and consistent. As a result, case-folded text should be used solely for internal processing and generally should not be stored or displayed to the end user.

The folding approach implemented by \texttt{\str_casefold:n} follows the “full” scheme defined by the Unicode Consortium (\textit{e.g.} SSfolds to SS). As case-folding is a language-insensitive process, there is no special treatment of Turkic input (\textit{i.e.} I always folds to i and not to ı).

\texttt{\str_mdfive_hash:n} \{\texttt{tl}\} \new{2023-05-19}

Expands to the MD5 sum generated from the \texttt{\langle tl\rangle}, which is converted to a \texttt{\langle string\rangle} as described for \texttt{\tl_to_str:n}.

### 17.8 Viewing strings

\texttt{\str_show:N} \{\texttt{str var}\} \new{2015-09-18} \updated{2021-04-29}

Displays the content of the \texttt{\langle str var\rangle} on the terminal.

\texttt{\str_log:N} \{\texttt{str var}\} \new{2019-02-15} \updated{2021-04-29}

Writes the content of the \texttt{\langle str var\rangle} in the log file.
17.9 Constant strings

\c_\text{ampersand\_str} \quad \text{Constant strings, containing a single character token, with category code 12.}
\c_\text{@\_str} \c_\text{backslash\_str} \c_\text{left\_brace\_str} \c_\text{right\_brace\_str} \c_\text{circumflex\_str} \c_\text{colon\_str} \c_\text{dollar\_str} \c_\text{hash\_str} \c_\text{percent\_str} \c_\text{tilde\_str} \c_\text{underscore\_str} \c_\text{zero\_str}

New: 2015-09-19
Updated: 2020-12-22

\c_\text{empty\_str} \quad \text{Constant that is always empty.}

New: 2023-12-07

17.10 Scratch strings

\l_\text{tmpa\_str} \quad \text{Scratch strings for local assignment. These are never used by the kernel code, and so}
\l_\text{tmpb\_str} \quad \text{are safe for use with any \LaTeX\defined function. However, they may be overwritten by}
\l_\text{tmpa\_str} \quad \text{other non-kernel code and so should only be used for short-term storage.}
\l_\text{tmpb\_str}

\g_\text{tmpa\_str} \quad \text{Scratch strings for global assignment. These are never used by the kernel code, and so}
\g_\text{tmpb\_str} \quad \text{are safe for use with any \LaTeX\defined function. However, they may be overwritten by}
\g_\text{tmpa\_str} \quad \text{other non-kernel code and so should only be used for short-term storage.}
\g_\text{tmpb\_str}
Chapter 18

The \texttt{l3str-convert} module
String encoding conversions

18.1 Encoding and escaping schemes

Traditionally, string encodings only specify how strings of characters should be stored as bytes. However, the resulting lists of bytes are often to be used in contexts where only a restricted subset of bytes are permitted (e.g., PDF string objects, URLs). Hence, storing a string of characters is done in two steps.

- The code points ("character codes") are expressed as bytes following a given "encoding". This can be \texttt{utf-16}, \texttt{iso 8859-1}, \textit{etc}. See Table 1 for a list of supported encodings.\footnote{Encodings and escapings will be added as they are requested.}

- Bytes are translated to \TeX{} tokens through a given "escaping". Those are defined for the most part by the \texttt{pdf} file format. See Table 2 for a list of escaping methods supported.\footnote{Encodings and escapings will be added as they are requested.}
Table 1: Supported encodings. Non-alphanumeric characters are ignored, and capital letters are lower-cased before searching for the encoding in this list.

<table>
<thead>
<tr>
<th>Encoding</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>utf8</td>
<td>UTF-8</td>
</tr>
<tr>
<td>utf16</td>
<td>UTF-16, with byte-order mark</td>
</tr>
<tr>
<td>utf16be</td>
<td>UTF-16, big-endian</td>
</tr>
<tr>
<td>utf16le</td>
<td>UTF-16, little-endian</td>
</tr>
<tr>
<td>utf32</td>
<td>UTF-32, with byte-order mark</td>
</tr>
<tr>
<td>utf32be</td>
<td>UTF-32, big-endian</td>
</tr>
<tr>
<td>utf32le</td>
<td>UTF-32, little-endian</td>
</tr>
<tr>
<td>iso88591, latin1</td>
<td>ISO 8859-1</td>
</tr>
<tr>
<td>iso88592, latin2</td>
<td>ISO 8859-2</td>
</tr>
<tr>
<td>iso88593, latin3</td>
<td>ISO 8859-3</td>
</tr>
<tr>
<td>iso88594, latin4</td>
<td>ISO 8859-4</td>
</tr>
<tr>
<td>iso88595</td>
<td>ISO 8859-5</td>
</tr>
<tr>
<td>iso88596</td>
<td>ISO 8859-6</td>
</tr>
<tr>
<td>iso88597</td>
<td>ISO 8859-7</td>
</tr>
<tr>
<td>iso88598</td>
<td>ISO 8859-8</td>
</tr>
<tr>
<td>iso88599, latin5</td>
<td>ISO 8859-9</td>
</tr>
<tr>
<td>iso885910, latin6</td>
<td>ISO 8859-10</td>
</tr>
<tr>
<td>iso885911</td>
<td>ISO 8859-11</td>
</tr>
<tr>
<td>iso885913, latin7</td>
<td>ISO 8859-13</td>
</tr>
<tr>
<td>iso885914, latin8</td>
<td>ISO 8859-14</td>
</tr>
<tr>
<td>iso885915, latin9</td>
<td>ISO 8859-15</td>
</tr>
<tr>
<td>iso885916, latin10</td>
<td>ISO 8859-16</td>
</tr>
<tr>
<td>clist</td>
<td>Comma-list of integers</td>
</tr>
<tr>
<td>(empty)</td>
<td>Native (Unicode) string</td>
</tr>
<tr>
<td>default</td>
<td>Like utf8 with 8-bit engines, and like native with unicode-engines</td>
</tr>
</tbody>
</table>

Table 2: Supported escapings. Non-alphanumeric characters are ignored, and capital letters are lower-cased before searching for the escaping in this list.

<table>
<thead>
<tr>
<th>Escaping</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>bytes, or empty</td>
<td>Arbitrary bytes</td>
</tr>
<tr>
<td>hex, hexadecimal name</td>
<td>Byte = two hexadecimal digits</td>
</tr>
<tr>
<td>string</td>
<td>See \pdfescaapename</td>
</tr>
<tr>
<td>url</td>
<td>Encoding used in URLs</td>
</tr>
</tbody>
</table>
18.2 Conversion functions

\str_set_convert:Nnnn \str_set_convert:Nnnn (str var) \{\langle string\rangle\} \{\langle name 1\rangle\} \{\langle name 2\rangle\}

This function converts the \langle string\rangle from the encoding given by \langle name 1\rangle to the encoding given by \langle name 2\rangle, and stores the result in the \langle str var\rangle. Each \langle name\rangle can have the form \langle encoding\rangle or \langle encoding\rangle/\langle escaping\rangle, where the possible values of \langle encoding\rangle and \langle escaping\rangle are given in Tables 1 and 2, respectively. The default escaping is to input and output bytes directly. The special case of an empty \langle name\rangle indicates the use of “native” strings, 8-bit for pdfTeX, and Unicode strings for the other two engines.

For example,

\str_set_convert:Nnnn \l_foo_str { Hello! } { } { utf16/hex }

results in the variable \l_foo_str holding the string FEFF00480065006C006C006F0021. This is obtained by converting each character in the (native) string Hello! to the UTF-16 encoding, and expressing each byte as a pair of hexadecimal digits. Note the presence of a (big-endian) byte order mark “FEFF, which can be avoided by specifying the encoding utf16be/hex.

An error is raised if the \langle string\rangle is not valid according to the \langle escaping 1\rangle and \langle encoding 1\rangle, or if it cannot be reencoded in the \langle encoding 2\rangle and \langle escaping 2\rangle (for instance, if a character does not exist in the \langle encoding 2\rangle). Erroneous input is replaced by the Unicode replacement character “FFFD, which can be avoided by specifying the encoding utf16be/hex.

18.3 Conversion by expansion (for PDF contexts)

A small number of expandable functions are provided for use in PDF string/name contexts. These assume UTF-8 and no escaping in the input.

\str_convert_pdfname:n \str_convert_pdfname:n \{\langle string\rangle\}

As \str_set_convert:Nnnn, converts the \langle string\rangle on a byte-by-byte basis with non-ASCII codepoints escaped using hashes.

18.4 Possibilities, and things to do

Encoding/escaping-related tasks.
In XeTeX/LuaTeX, would it be better to use the approach to build a string from a given list of character codes? Namely, within a group, assign 0-9a-f and all characters we want to category “other”, then assign ~ the category superscript, and use \scantokens.

- Change \str_set_convert:Nnnn to expand its last two arguments.

- Describe the internal format in the code comments. Refuse code points in [",",":","] in the internal representation?

- Add documentation about each encoding and escaping method, and add examples.

- The hex unescaping should raise an error for odd-token count strings.

- Decide what bytes should be escaped in the url escaping. Perhaps the characters !=)/( are safe, and all other characters should be escaped?

- Automate generation of 8-bit mapping files.

- Change the framework for 8-bit encodings: for decoding from 8-bit to Unicode, use 256 integer registers; for encoding, use a tree-box.

- More encodings (see Heiko’s stringenc). CESU?

- More escapings: ascii85, shell escapes, lua escapes, etc.?
Chapter 19

The \texttt{l3quark} module

Quarks and scan marks

Two special types of constants in \LaTeX{} are “quarks” and “scan marks”. By convention all constants of type quark start out with \texttt{\q}, and scan marks start with \texttt{\s}.

19.1 Quarks

Quarks are control sequences (and in fact, token lists) that expand to themselves and should therefore \textit{never} be executed directly in the code. This would result in an endless loop!

They are meant to be used as delimiter in weird functions, the most common use case being the ‘stop token’ (i.e. \texttt{\q\_stop}). For example, when writing a macro to parse a user-defined date

\begin{verbatim}
\date_parse:n {19/June/1981}
\end{verbatim}

one might write a command such as

\begin{verbatim}
\cs_new:Npn \date_parse:n #1 { \date_parse_aux:w #1 \q_stop }
\cs_new:Npn \date_parse_aux:w #1 / #2 / #3 \q_stop
  { <do something with the date> }
\end{verbatim}

Quarks are sometimes also used as error return values for functions that receive erroneous input. For example, in the function \texttt{\prop\_get:NnN} to retrieve a value stored in some key of a property list, if the key does not exist then the return value is the quark \texttt{\q\_no\_value}. As mentioned above, such quarks are extremely fragile and it is imperative when using such functions that code is carefully written to check for pathological cases to avoid leakage of a quark into an uncontrolled environment.

Quarks also permit the following ingenious trick when parsing tokens: when you pick up a token in a temporary variable and you want to know whether you have picked up a particular quark, all you have to do is compare the temporary variable to the quark using \texttt{\tl\_if\_eq:NNTF}. A set of special quark testing functions is set up below. All the quark testing functions are expandable although the ones testing only single tokens are much faster.
\section*{19.2 Defining quarks}

\begin{verbatim}
\quark_new:N \quark_new:N \quark
\end{verbatim}

Creates a new \texttt{quark} which expands only to \texttt{quark}. The \texttt{quark} is defined globally, and an error message is raised if the name was already taken.

\begin{verbatim}
\quark_stop
\end{verbatim}

Used as a marker for delimited arguments, such as
\begin{verbatim}
\cs_set:Npn \tmp:w #1#2 \q_stop \{#1}
\end{verbatim}

\begin{verbatim}
\q_stop
\end{verbatim}

Used as a marker for delimited arguments when \texttt{\q_stop} is already in use.

\begin{verbatim}
\q_mark
\end{verbatim}

Quark to mark a null value in structured variables or functions. Used as an end delimiter when this may itself need to be tested (in contrast to \texttt{\q_stop}, which is only ever used as a delimiter).

\begin{verbatim}
\q_nil
\end{verbatim}

A canonical value for a missing value, when one is requested from a data structure. This is therefore used as a “return” value by functions such as \texttt{prop_get:NnN} if there is no data to return.

\begin{verbatim}
\q_no_value
\end{verbatim}

\section*{19.3 Quark tests}

The method used to define quarks means that the single token (N) tests are faster than the multi-token (n) tests. The latter should therefore only be used when the argument can definitely take more than a single token.

\begin{verbatim}
\quark_if_nil_p:N \quark_if_nil_p:N \quark
\end{verbatim}\

Tests if the \texttt{token} is equal to \texttt{\q_nil}.

\begin{verbatim}
\quark_if_nil_p:n \quark_if_nil_p:n \quark
\end{verbatim}\

Tests if the \texttt{token list} contains only \texttt{\q_nil} (distinct from \texttt{token list} being empty or containing \texttt{\q_nil} plus one or more other tokens).

\begin{verbatim}
\quark_if_no_value_p:N \quark_if_no_value_p:N \quark
\end{verbatim}\

Tests if the \texttt{token} is equal to \texttt{\q_no_value}.
19.4 Recursion

This module provides a uniform interface to intercepting and terminating loops as when one is doing tail recursion. The building blocks follow below and an example is shown in Section 19.4.1.

\q_recursion_tail

This quark is appended to the data structure in question and appears as a real element there. This means it gets any list separators around it.

\q_recursion_stop

This quark is added after the data structure. Its purpose is to make it possible to terminate the recursion at any point easily.

\quark_if_recursion_tail_stop:N \q_recursion_tail \q_recursion_stop:N \langle token \rangle

Tests if \langle token \rangle contains only the marker \q_recursion_tail, and if so uses \use\-none_delimit_by_q_recursion_stop:w to terminate the recursion that this belongs to. The recursion input must include the marker tokens \q_recursion_tail and \q_recursion_stop as the last two items.

\quark_if_recursion_tail_stop:n \q_recursion_tail \q_recursion_stop:o \langle token list \rangle

Tests if the \langle token list \rangle contains only \q_recursion_tail, and if so uses \use\-i_delimit_by_q_recursion_stop:w to terminate the recursion that this belongs to. The recursion input must include the marker tokens \q_recursion_tail and \q_recursion_stop as the last two items. The \langle insertion \rangle code is then added to the input stream after the recursion has ended.

\quark_if_recursion_tail_stop_do:Nn \q_recursion_tail \q_recursion_stop:do:Nn \langle token \rangle \{ \langle insertion \rangle \}

Tests if \langle token \rangle contains only \q_recursion_tail, and if so uses \use\-i_delimit_by_q_recursion_stop:w to terminate the recursion that this belongs to. The recursion input must include the marker tokens \q_recursion_tail and \q_recursion_stop as the last two items. The \langle insertion \rangle code is then added to the input stream after the recursion has ended.
Tests if \( \text{token list} \) contains only \q_recursion_tail\, and if so terminates the recursion using \( \langle \text{(type)} \rangle \_\text{map} \_\text{break} \). The recursion end should be marked by \prg_break_point\.

19.4.1 An example of recursion with quarks

Quarks are mainly used internally in the expl3 code to define recursion functions such as \tl_map_inline:nn and so on. Here is a small example to demonstrate how to use quarks in this fashion. We shall define a command called \my_map_dbl:nn which takes a token list and applies an operation to every pair of tokens. For example, \my_map_dbl:nn \{abcd\} \{|--#1--#2--\} would produce “[-a-b-] [-c-d-]”. Using quarks to define such functions simplifies their logic and ensures robustness in many cases.

Here’s the definition of \my_map_dbl:nn. First of all, define the function that does the processing based on the inline function argument \#2. Then initiate the recursion using an internal function. The token list \#1 is terminated using \q_recursion_tail\, with delimiters according to the type of recursion (here a pair of \q_recursion_tail\), concluding with \q_recursion_stop\.

\begin{verbatim}
\cs_new:Npn \my_map_dbl:nn #1#2 \{
    \cs_set:Npn \__my_map_dbl_fn:nn ##1 ##2 {#2}
    \__my_map_dbl:nn #1 \q_recursion_tail \q_recursion_tail
    \q_recursion_stop
}\end{verbatim}

The definition of the internal recursion function follows. First check if either of the input tokens are the termination quarks. Then, if not, apply the inline function to the two arguments.

\begin{verbatim}
\cs_new:Nn \__my_map_dbl:nn \{
    \quark_if_recursion_tail_stop:n {#1}
    \quark_if_recursion_tail_stop:n {#2}
    \__my_map_dbl_fn:nn {#1} {#2}
\}\end{verbatim}

Finally, recurse:

\begin{verbatim}
\__my_map_dbl:nn
\end{verbatim}

Note that contrarily to \LaTeX3 built-in mapping functions, this mapping function cannot be nested, since the second map would overwrite the definition of \__my_map_dbl_fn:nn.
19.5 Scan marks

Scan marks are control sequences set equal to \texttt{\textbackslash scan\_stop}; hence never expand in an expansion context and are (largely) invisible if they are encountered in a typesetting context.

Like quarks, they can be used as delimiters in weird functions and are often safer to use for this purpose. Since they are harmless when executed by \TeX{} in non-expandable contexts, they can be used to mark the end of a set of instructions. This allows to skip to that point if the end of the instructions should not be performed (see \texttt{l3regex}).

\begin{itemize}
  \item \texttt{\textbackslash scan\_new:N} \hspace{1cm}\texttt{\textbackslash scan\_new:N \langle scan mark \rangle}
  \hspace{1cm}New: 2018-04-01
  \hspace{1cm}Creates a new \texttt{\langle scan mark \rangle} which is set equal to \texttt{\textbackslash scan\_stop:}. The \texttt{\langle scan mark \rangle} is defined globally, and an error message is raised if the name was already taken by another scan mark.
  \item \texttt{\textbackslash s\_stop}
  \hspace{1cm}New: 2018-04-01
  \hspace{1cm}Used at the end of a set of instructions, as a marker that can be jumped to using \texttt{\textbackslash use\_none\_delimit\_by\_s\_stop:w}.
  \item \texttt{\textbackslash use\_none\_delimit\_by\_s\_stop:w \ast \textbackslash use\_none\_delimit\_by\_s\_stop:w \langle tokens \rangle \textbackslash s\_stop}
  \hspace{1cm}New: 2018-04-01
  \hspace{1cm}Removes the \texttt{\langle tokens \rangle} and \texttt{\textbackslash s\_stop} from the input stream. This leads to a low-level \TeX{} error if \texttt{\textbackslash s\_stop} is absent.
\end{itemize}
Chapter 20

The l3seq module
Sequences and stacks

l3seqX3 implements a “sequence” data type, which contain an ordered list of entries which may contain any balanced text. It is possible to map functions to sequences such that the function is applied to every item in the sequence.

Sequences are also used to implement stack functions in l3seqX3. This is achieved using a number of dedicated stack functions.

20.1 Creating and initialising sequences

\seq_new:N \seq_new:c
Creates a new \seq or raises an error if the name is already taken. The declaration is global. The \seq initially contains no items.

\seq_clear:N \seq_clear:c \seq_gclear:N \seq_gclear:c
Clears all items from the \seq.

\seq_clear_new:N \seq_clear_new:c \seq_gclear_new:N \seq_gclear_new:c
Ensures that the \seq exists globally by applying \seq_new:N if necessary, then applies \seq(g)clear:N to leave the \seq empty.

\seq_set_eq:NN \seq_set_eq:cnNcc \seq_gset_eq:NN \seq_gset_eq:cnNcc
Sets the content of \seq equal to that of \seq.
\seq_set_from_clist:NN \seq_set_from_clist:NN \seq_set_from_clist:NN \seq_set_from_clist:NN \seq_set_from_clist:NN \seq_set_from_clist:NN
New: 2014-07-17
Converting the data in the \{
\textit{comma list}\} into a \{
\textit{seq var}\}: the original \{
\textit{comma list}\} is unchanged.

\seq_set_from_clist:NN \seq_set_from_clist:NN \seq_set_from_clist:NN \seq_set_from_clist:NN \seq_set_from_clist:NN \seq_set_from_clist:NN
New: 2017-11-28
Creates a new constant \{
\textit{seq var}\} or raises an error if the name is already taken. The \{
\textit{seq var}\} is set globally to contain the items in the \{
\textit{comma list}\}.

\seq_set_split:Nnn \seq_set_split:Nnn \seq_set_split:Nnn \seq_set_split:Nnn \seq_set_split:Nnn \seq_set_split:Nnn
New: 2011-08-16
Updated: 2012-07-02
Splits the \{
\textit{token list}\} into \{
\textit{items}\} separated by \{
\textit{delimiter}\}, and assigns the result to the \{
\textit{seq var}\}. Spaces on both sides of each \{
\textit{item}\} are ignored, then one set of outer braces is removed (if any); this space trimming behaviour is identical to that of \l3clist functions. Empty \{
\textit{items}\} are preserved by \seq_set_split:Nnn, and can be removed afterwards using \seq_remove_all:Nn \{
\textit{seq var}\} \{\}. The \{
\textit{delimiter}\} may not contain \{, \} or \# (assuming \TeX’s normal category code régime). If the \{
\textit{delimiter}\} is empty, the \{
\textit{token list}\} is split into \{
\textit{items}\} as a \{
\textit{token list}\}. See also \seq_set_split:--keep_spaces:Nnn, which omits space stripping.

\seq_set_split_keep_spaces:Nnn \seq_set_split_keep_spaces:Nnn \seq_set_split_keep_spaces:Nnn \seq_set_split_keep_spaces:Nnn \seq_set_split_keep_spaces:Nnn
New: 2021-03-24
Splits the \{
\textit{token list}\} into \{
\textit{items}\} separated by \{
\textit{delimiter}\}, and assigns the result to the \{
\textit{seq var}\}. One set of outer braces is removed (if any) but any surrounding spaces are retained: any braces \textit{inside} one or more spaces are therefore kept. Empty \{
\textit{items}\} are preserved by \seq_set_split_keep_spaces:Nnn, and can be removed afterwards using \seq_remove_all:Nn \{
\textit{seq var}\} \{\}. The \{
\textit{delimiter}\} may not contain \{, \} or \# (assuming \TeX’s normal category code régime). If the \{
\textit{delimiter}\} is empty, the \{
\textit{token list}\} is split into \{
\textit{items}\} as a \{
\textit{token list}\}. See also \seq_set_split:Nnn, which removes spaces around the delimiters.
\seq_set_filter:NNn \seq_gset_filter:NNn
New: 2012-06-15
Evaluates the \inline boolexpr\ for every \item stored within the \seq var\2. The \inline boolexpr\ receives the \item as #1. The sequence of all \items for which the \inline boolexpr\ evaluated to true is assigned to \seq var\1.

\TeXhackers note: Contrarily to other mapping functions, \seq_map_break: cannot be used in this function, and would lead to low-level \TeX errors.

\seq_concat:NNN \seq_concat:ccc
\seq_gconcat:NNN \seq_gconcat:ccc
\seq_if_exist_p:N \seq_if_exist:NTF \seq_if_exist:cTF
\seq_if_exist_p:c \seq_if_exist:N TF \seq_if_exist:c TF
Updated: 2012-03-03
Tests whether the \seq is currently defined. This does not check that the \seq really is a sequence variable.

20.2 Appending data to sequences

\seq_put_left:Nn \seq_put_left:N \seq_put_left:(NV|Nv|Ne|No|cn|cV|cv|ce|co) \seq_gput_left:Nn \seq_gput_left:cN \seq_gput_left:N \seq_gput_left:(NV|Nv|Ne|No|cn|cV|cv|ce|co)
Appends the \item to the left of the \seq var.

\seq_put_right:Nn \seq_put_right:N \seq_put_right:(NV|Nv|Ne|No|cn|cV|cv|ce|co) \seq_gput_right:Nn \seq_gput_right:N \seq_gput_right:(NV|Nv|Ne|No|cn|cV|cv|ce|co)
Appends the \item to the right of the \seq var.

20.3 Recovering items from sequences

Items can be recovered from either the left or the right of sequences. For implementation reasons, the actions at the left of the sequence are faster than those acting on the right. These functions all assign the recovered material locally, i.e. setting the \token list variable \used with \tl_set:Nn and never \tl_gset:Nn.

\seq_get_left:NN \seq_get_left:N \seq_get_left:ch
Updated: 2012-05-14
Stores the left-most item from a \seq in the \token list variable without removing it from the \seq var. The \token list variable is assigned locally. If \seq var is empty the \token list variable is set to the special marker \q_no_ value.
\seq_get_right:NN \seq_get_right:CN
Stores the right-most item from a \seq var in the \token list variable without removing it from the \seq var. The \token list variable is assigned locally. If \seq var is empty the \token list variable is set to the special marker \q_no_value.

\seq_pop_left:NN \seq_pop_left:CN
Pops the left-most item from a \seq var into the \token list variable, i.e. removes the item from the sequence and stores it in the \token list variable. Both of the variables are assigned locally. If \seq var is empty the \token list variable is set to the special marker \q_no_value.

\seq_gpop_left:NN \seq_gpop_left:CN
Pops the left-most item from a \seq var into the \token list variable, i.e. removes the item from the sequence and stores it in the \token list variable. The \seq var is modified globally, while the assignment of the \token list variable is local. If \seq var is empty the \token list variable is set to the special marker \q_no_value.

\seq_pop_right:NN \seq_pop_right:CN
Pops the right-most item from a \seq var into the \token list variable, i.e. removes the item from the sequence and stores it in the \token list variable. Both of the variables are assigned locally. If \seq var is empty the \token list variable is set to the special marker \q_no_value.

\seq_gpop_right:NN \seq_gpop_right:CN
Pops the right-most item from a \seq var into the \token list variable, i.e. removes the item from the sequence and stores it in the \token list variable. The \seq var is modified globally, while the assignment of the \token list variable is local. If \seq var is empty the \token list variable is set to the special marker \q_no_value.

\seq_item:Nn \seq_item:N \seq_item:n \seq_item:cN \seq_item:cn \seq_item:cV \seq_item:ce
Indexing items in the \seq var from 1 at the top (left), this function evaluates the \integer expression and leaves the appropriate item from the sequence in the input stream. If the \integer expression is negative, indexing occurs from the bottom (right) of the sequence. If the \integer expression is larger than the number of items in the \seq var (as calculated by \seq_count:N) then the function expands to nothing.

\TeXhackers note: The result is returned within the \unexpanded primitive \exp_not:n, which means that the \item does not expand further when appearing in an e-type or x-type argument expansion.
\seq_rand_item:N \seq_rand_item:c

Selects a pseudo-random item of the \seq. If the \seq is empty the result is empty. This is not available in older versions of Xe\TeX.

\textbf{\LaTeX{}hackers note:} The result is returned within the \unexpanded primitive (\exp_not:n), which means that the \item does not expand further when appearing in an e-type or x-type argument expansion.

## 20.4 Recovering values from sequences with branching

The functions in this section combine tests for non-empty sequences with recovery of an item from the sequence. They offer increased readability and performance over separate testing and recovery phases.

\seq_get_left:NNTF \seq_get_left:NN \seq_get_left:cN

If the \seq is empty, leaves the \false in the input stream. The value of the \token list variable is not defined in this case and should not be relied upon. If the \seq is non-empty, stores the left-most item from the \seq in the \token list variable without removing it from the \seq, then leaves the \true in the input stream. The \token list variable is assigned locally.

\seq_get_right:NNTF \seq_get_right:NN \seq_get_right:cN

If the \seq is empty, leaves the \false in the input stream. The value of the \token list variable is not defined in this case and should not be relied upon. If the \seq is non-empty, stores the right-most item from the \seq in the \token list variable without removing it from the \seq, then leaves the \true in the input stream. The \token list variable is assigned locally.

\seq_pop_left:NNTF \seq_pop_left:NN \seq_pop_left:cN

If the \seq is empty, leaves the \false in the input stream. The value of the \token list variable is not defined in this case and should not be relied upon. If the \seq is non-empty, pops the left-most item from the \seq in the \token list variable, i.e. removes the item from the \seq, then leaves the \true in the input stream. Both the \seq and the \token list variable are assigned locally.

\seq_gpop_left:NNTF \seq_gpop_left:NN \seq_gpop_left:cN

If the \seq is empty, leaves the \false in the input stream. The value of the \token list variable is not defined in this case and should not be relied upon. If the \seq is non-empty, pops the left-most item from the \seq in the \token list variable, i.e. removes the item from the \seq, then leaves the \true in the input stream. The \seq is modified globally, while the \token list variable is assigned locally.
If the \texttt{seq var} is empty, leaves the \texttt{false code} in the input stream. The value of the \texttt{token list variable} is not defined in this case and should not be relied upon. If the \texttt{seq var} is non-empty, pops the right-most item from the \texttt{seq var} in the \texttt{token list variable}, i.e. removes the item from the \texttt{seq var}, then leaves the \texttt{true code} in the input stream. Both the \texttt{seq var} and the \texttt{token list variable} are assigned locally.

If the \texttt{seq var} is empty, leaves the \texttt{false code} in the input stream. The value of the \texttt{token list variable} is not defined in this case and should not be relied upon. If the \texttt{seq var} is non-empty, pops the right-most item from the \texttt{seq var} in the \texttt{token list variable}, i.e. removes the item from the \texttt{seq var}, then leaves the \texttt{true code} in the input stream. The \texttt{seq var} is modified globally, while the \texttt{token list variable} is assigned locally.

\section{Modifying sequences}

While sequences are normally used as ordered lists, it may be necessary to modify the content. The functions here may be used to update sequences, while retaining the order of the unaffected entries.

Removes duplicate items from the \texttt{seq var}, leaving the left most copy of each item in the \texttt{seq var}. The \texttt{item} comparison takes place on a token basis, as for \texttt{\tl_if_eq:nnTF}.

\texttt{\TeX} hackers note: This function iterates through every item in the \texttt{seq var} and does a comparison with the \texttt{items} already checked. It is therefore relatively slow with large sequences.

Removes every occurrence of \texttt{item} from the \texttt{seq var}. The \texttt{item} comparison takes place on a token basis, as for \texttt{\tl_if_eq:nnTF}.

Removes the item of \texttt{seq var} at the position given by evaluating the \texttt{int expr} and replaces it by \texttt{item}. Items are indexed from 1 on the left/top of the \texttt{seq var}, or from \texttt{-1} on the right/bottom. If the \texttt{int expr} is zero or is larger (in absolute value) than the number of items in the sequence, the \texttt{seq var} is not modified. In these cases, \texttt{\seq_set_item:Nnn} raises an error while \texttt{\seq_set_item:NnnTF} runs the \texttt{false code}. In cases where the assignment was successful, \texttt{true code} is run afterwards.
\seq_reverse:N \seq_reverse:c
\seqgrepse:N \seqgrepse:c
Reverses the order of the items stored in the \seq var.

\seq_sort:N \seq_sort:cn \seqgsort:N \seqgsort:cn
Sorts the items in the \seq var according to the \comparison code, and assigns the result to \seq var. The details of sorting comparison are described in Section 6.1.

\seq_shuffle:N \seq_shuffle:c \seqgshuffle:N \seqgshuffle:cn
Sets the \seq var to the result of placing the items of the \seq var in a random order. Each item is (roughly) as likely to end up in any given position.

\seq_if_empty_p:N \seq_if_empty_p:c \seq_if_empty:N \seq_if_empty:c
\seq_if_in:NnTF \seq_if_in:nTF \seq_if_in:nTF\{item\} \{true code\} \{false code\}
Tests if the \seq var is empty (containing no items).

\seq_map_function:NN \seq_map_function:cn
Applies \function to every \item stored in the \seq var. The \function will receive one argument for each iteration. The \items are returned from left to right.
To pass further arguments to the \function, see \seq_map_tokens:Nn. The function \seq_map_inline:NN is faster than \seq_map_function:NN for sequences with more than about 10 items.

\TeXhackers note: For sequences with more than 13 items or so, only a small proportion of all possible permutations can be reached, because the random seed \sys_rand_seed: only has 28-bits. The use of \toks internally means that sequences with more than 32767 or 65535 items (depending on the engine) cannot be shuffled.

20.6 Sequence conditionals

20.7 Mapping over sequences

All mappings are done at the current group level, \textit{i.e.} any local assignments made by the \function or \code discussed below remain in effect after the loop.
\seq_map_inline:Nn \seq_map_inline:Nn (seq var) \{ (inline function) \}

Applies (inline function) to every (item) stored within the (seq var). The (inline function) should consist of code which will receive the (item) as #1. The (items) are returned from left to right.

\seq_map_tokens:Nn \seq_map_tokens:Nn (seq var) \{ (code) \}

Analogue of \seq_map_function:NN which maps several tokens instead of a single function. The (code) receives each item in the (seq var) as a trailing brace group. For instance,

\seq_map_tokens:Nn \l_my_seq \{ \prg_replicate:nn { 2 } \}

expands to twice each item in the (seq var): for each item in \l_my_seq the function \prg_replicate:nn receives 2 and (item) as its two arguments. The function \seq_map_inline:Nn is typically faster but it is not expandable.

\seq_map_variable:NNn \seq_map_variable:Nn (seq var) (variable) \{ (code) \}

Stores each (item) of the (seq var) in turn in the (token list) (variable) and applies the (code). The (code) will usually make use of the (variable), but this is not enforced. The assignments to the (variable) are local. Its value after the loop is the last (item) in the (seq var), or its original value if the (seq var) is empty. The (items) are returned from left to right.

\seq_map_indexed_function:NN \seq_map_indexed_function:NN (seq var) (function)

Applies (function) to every entry in the (seq var). The (function) should have signature :nn. It receives two arguments for each iteration: the (index) (namely 1 for the first entry, then 2 and so on) and the (item).

\seq_map_indexed_inline:Nn \seq_map_indexed_inline:Nn (seq var) \{ (inline function) \}

Applies (inline function) to every entry in the (seq var). The (inline function) should consist of code which receives the (index) (namely 1 for the first entry, then 2 and so on) as #1 and the (item) as #2.

\seq_map_pairwise_function:NNN \seq_map_pairwise_function:NNN (seq_1) (seq_2) (function)

Applies (function) to every pair of items (seq_1-item)--(seq_2-item) from the two sequences, returning items from both sequences from left to right. The (function) receives two n-type arguments for each iteration. The mapping terminates when the end of either sequence is reached (i.e. whichever sequence has fewer items determines how many iterations occur).
\seq_map_break: \seq_map_break:

Used to terminate a \seq_map\ldots function before all entries in the \textit{seq var} have been processed. This normally takes place within a conditional statement, for example

\seq_map_inline:Nn \l_my_seq
{\str_if_eq:nnTF { #1 } { bingo }  
{ \seq_map_break: }  
{\% Do something useful  
}  
}

Use outside of a \seq_map\ldots scenario leads to low level \TeX errors.

\TeXhackers note: When the mapping is broken, additional tokens may be inserted before further items are taken from the input stream. This depends on the design of the mapping function.

\seq_map_break:n \seq_map_break:n \langle code\rangle

Used to terminate a \seq_map\ldots function before all entries in the \textit{seq var} have been processed, inserting the \textit{code} after the mapping has ended. This normally takes place within a conditional statement, for example

\seq_map_inline:Nn \l_my_seq
{\str_if_eq:nnTF { #1 } { bingo }  
{ \seq_map_break:n \langle code\rangle }  
{\% Do something useful  
}  
}

Use outside of a \seq_map\ldots scenario leads to low level \TeX errors.

\TeXhackers note: When the mapping is broken, additional tokens may be inserted before the \textit{code} is inserted into the input stream. This depends on the design of the mapping function.

\seq_set_map:NNn \seq_gset_map:NNn

Applies \textit{inline function} to every \textit{item} stored within the \textit{seq var}. The \textit{inline function} should consist of code which will receive the \textit{item} as #1. The sequence resulting applying \textit{inline function} to each \textit{item} is assigned to \textit{seq var}.

\TeXhackers note: Contrarily to other mapping functions, \seq_map_break: cannot be used in this function, and would lead to low-level \TeX errors.
\seq_map_e:NNn \seq_gset_map_e:NNn

Applies \textit{inline function} to every \textit{item} stored within the \textit{seq var}. The \textit{inline function} should consist of code which will receive the \textit{item} as #1. The sequence resulting from e-expanding \textit{inline function} applied to each \textit{item} is assigned to \textit{seq var}. As such, the code in \textit{inline function} should be expandable.

\textbf{\TeX hackers note:} Contrarily to other mapping functions, \texttt{\seq_map_break}: cannot be used in this function, and would lead to low-level \TeX{} errors.

\seq_count:N * \seq_count:N \texttt{\textit{seq var}}

Leaves the number of items in the \textit{seq var} in the input stream as an \textit{integer denotation}. The total number of items in a \textit{seq var} includes those which are empty and duplicates, i.e. every item in a \textit{seq var} is unique.

\textbf{20.8 Using the content of sequences directly}

\seq_use:Nnnn \seq_use:Nnnn \texttt{\textit{seq var}} \{\texttt{\textit{separator between two}}\}
\seq_use:ccnnn \{\texttt{\textit{separator between more than two}}\} \{\texttt{\textit{separator between final two}}\}

Places the contents of the \textit{seq var} in the input stream, with the appropriate \texttt{\textit{separator}} between the items. Namely, if the sequence has more than two items, the \texttt{\textit{separator between more than two}} is placed between each pair of items except the last, for which the \texttt{\textit{separator between final two}} is used. If the sequence has exactly two items, then they are placed in the input stream separated by the \texttt{\textit{separator between two}}. If the sequence has a single item, it is placed in the input stream, and an empty sequence produces no output. An error is raised if the variable does not exist or if it is invalid.

For example,

\begin{verbatim}
\seq_set_split:Nnn \l_tmpa_seq { | } { a | b | c | {de} | f }
\seq_use:Nnnn \l_tmpa_seq { ~and~ } { ,~ } { ,~and~ }
\end{verbatim}

inserts “a, b, c, de, and f” in the input stream. The first separator argument is not used in this case because the sequence has more than 2 items.

\textbf{\TeX hackers note:} The result is returned within the \texttt{\unexpanded} primitive \texttt{\exp_not:n}, which means that the \textit{items} do not expand further when appearing in an \texttt{e}-type or \texttt{x}-type argument expansion.
\useseq{\var}{\seq}{(\textit{separator})}

Places the contents of the \seq{} in the input stream, with the \textit{separator} between the items. If the sequence has a single item, it is placed in the input stream with no \textit{separator}, and an empty sequence produces no output. An error is raised if the variable does not exist or if it is invalid.

For example,

\begin{verbatim}
\seq_set_split:Nnn \l_tmpa_seq { | } { a | b | c | \{de\} | f }
\seq_use:Nn \l_tmpa_seq { \textit{and} }
\end{verbatim}

inserts “a and b and c and de and f” in the input stream.

\textbf{TeXhackers note:} The result is returned within the \texttt{\unexpanded} primitive (\texttt{\exp_not:n}), which means that the \textit{items} do not expand further when appearing in an \texttt{e}-type or \texttt{x}-type argument expansion.

### 20.9 Sequences as stacks

Sequences can be used as stacks, where data is pushed to and popped from the top of the sequence. (The left of a sequence is the top, for performance reasons.) The stack functions for sequences are not intended to be mixed with the general ordered data functions detailed in the previous section: a sequence should either be used as an ordered data type or as a stack, but not in both ways.

\begin{verbatim}
\useseq{\var}{\seq}{\tokenlist}
\end{verbatim}

Reads the top item from a \seq{} into the \tokenlist{} without removing it from the \seq{}. The \tokenlist{} is assigned locally. If \seq{} is empty the \tokenlist{} is set to the special marker \qno_value.

\begin{verbatim}
\popseq{\var}{\seq}{\tokenlist}
\end{verbatim}

Pops the top item from a \seq{} into the \tokenlist{}. Both of the variables are assigned locally. If \seq{} is empty the \tokenlist{} is set to the special marker \qno_value.

\begin{verbatim}
\gpopseq{\var}{\seq}{\tokenlist}
\end{verbatim}

Pops the top item from a \seq{} into the \tokenlist{}. The \seq{} is modified globally, while the \tokenlist{} is assigned locally. If \seq{} is empty the \tokenlist{} is set to the special marker \qno_value.

\begin{verbatim}
\getseq:NNF \seq{} \tokenlist{} \{\textit{true code}\} \{\textit{false code}\}
\end{verbatim}

If the \seq{} is empty, leaves the \textit{false code} in the input stream. The value of the \textit{token list variable} is not defined in this case and should not be relied upon. If the \seq{} is non-empty, stores the top item from a \seq{} in the \tokenlist{} without removing it from the \seq{}. The \tokenlist{} is assigned locally.
If the \( \langle \text{seq var} \rangle \) is empty, leaves the \( \langle \text{false code} \rangle \) in the input stream. The value of the \( \langle \text{token list variable} \rangle \) is not defined in this case and should not be relied upon.

If the \( \langle \text{seq var} \rangle \) is non-empty, pops the top item from the \( \langle \text{seq var} \rangle \) in the \( \langle \text{token list variable} \rangle \), \textit{i.e.} removes the item from the \( \langle \text{seq var} \rangle \). Both the \( \langle \text{seq var} \rangle \) and the \( \langle \text{token list variable} \rangle \) are assigned locally.

\[
\text{seq_pop:NN} \quad \text{seq_pop:cN}
\]

\[
\text{New: 2012-05-14} \quad \text{Updated: 2012-05-19}
\]

\[
\text{seq_gpop:NN} \quad \text{seq_gpop:cN}
\]

\[
\text{New: 2012-05-14} \quad \text{Updated: 2012-05-19}
\]

\[
\text{seq_push:Nn} \quad \text{seq_gpush:Nn}
\]

\[
\text{New: 2012-05-14} \quad \text{Updated: 2012-05-19}
\]

\[
\text{Add the \{item\} to the top of the \langle seq var \rangle.}
\]

\[\text{20.10 Sequences as sets}\]

Sequences can also be used as sets, such that all of their items are distinct. Usage of sequences as sets is not currently widespread, hence no specific set function is provided. Instead, it is explained here how common set operations can be performed by combining several functions described in earlier sections. When using sequences to implement sets, one should be careful not to rely on the order of items in the sequence representing the set.

Sets should not contain several occurrences of a given item. To make sure that a \( \langle \text{seq var} \rangle \) only has distinct items, use \texttt{\seq_remove_duplicates:N \langle \text{seq var} \rangle}. This function is relatively slow, and to avoid performance issues one should only use it when necessary.

Some operations on a set \( \langle \text{seq var} \rangle \) are straightforward. For instance, \texttt{\seq_count:N \langle \text{seq var} \rangle} expands to the number of items, while \texttt{\seq_if_in:NnTF \langle \text{seq var} \rangle \{\langle \text{item} \rangle\}} tests if the \{\text{item}\} is in the set.

Adding an \{\text{item}\} to a set \( \langle \text{seq var} \rangle \) can be done by appending it to the \langle \text{seq var} \rangle:

\[
\text{\seq_if_in:NnF \langle seq var \rangle \{\langle item \rangle\} \{ \seq_put_right:Nn (seq var) \{\langle item \rangle\} \}}
\]

Removing an \{\text{item}\} from a set \( \langle \text{seq var} \rangle \) can be done using \texttt{\seq_remove_all:Nn}.

\[
\text{\seq_remove_all:Nn \langle seq var \rangle \{\langle item \rangle\}}
\]

The intersection of two sets \( \langle \text{seq var}_1 \rangle \) and \( \langle \text{seq var}_2 \rangle \) can be stored into \( \langle \text{seq var}_3 \rangle \) by collecting items of \( \langle \text{seq var}_1 \rangle \) which are in \( \langle \text{seq var}_2 \rangle \).
The code as written here only works if \texttt{seq var$_3$} is different from the other two sequence variables. To cover all cases, items should first be collected in a sequence \texttt{l\_\_\_(pkg)\_internal_seq}, then \texttt{seq var$_3$} should be set equal to this internal sequence. The same remark applies to other set functions.

The union of two sets \texttt{seq var$_1$} and \texttt{seq var$_2$} can be stored into \texttt{seq var$_3$} through

\begin{verbatim}
\seq_concat:NNN \seq_var$_3$ \seq_var$_1$ \seq_var$_2$
\end{verbatim}
or by adding items to (a copy of) \texttt{seq var$_3$} one by one

\begin{verbatim}
\seq_set_eq:NN \seq_var$_3$ \seq_var$_1$
\seq_map_inline:Nn \seq_var$_2$
{ \seq_if_in:NnF \seq_var$_3$ {#1}
  { \seq_put_right:Nn \seq_var$_3$ {#1} }
}
\end{verbatim}

The second approach is faster than the first when the \texttt{seq var$_2$} is short compared to \texttt{seq var$_1$}.

The difference of two sets \texttt{seq var$_1$} and \texttt{seq var$_2$} can be stored into \texttt{seq var$_3$} by removing items of the \texttt{seq var$_2$} from (a copy of) the \texttt{seq var$_1$} one by one.

\begin{verbatim}
\seq_set_eq:NN \seq_var$_3$ \seq_var$_1$
\seq_map_inline:Nn \seq_var$_2$
{ \seq_remove_all:Nn \seq_var$_3$ {#1} }
\end{verbatim}

The symmetric difference of two sets \texttt{seq var$_1$} and \texttt{seq var$_2$} can be stored into \texttt{seq var$_3$} by computing the difference between \texttt{seq var$_1$} and \texttt{seq var$_2$} and storing the result as \texttt{l\_\_\_(pkg)\_internal_seq}, then the difference between \texttt{seq var$_2$} and \texttt{seq var$_1$}, and finally concatenating the two differences to get the symmetric differences.

\begin{verbatim}
\seq_set_eq:NN \l\_\_\_(pkg)\_internal_seq \seq_var$_1$
\seq_map_inline:Nn \seq_var$_2$
{ \seq_remove_all:Nn \l\_\_\_(pkg)\_internal_seq {#1} }
\seq_set_eq:NN \seq_var$_3$ \seq_var$_2$
\seq_map_inline:Nn \seq_var$_1$
{ \seq_remove_all:Nn \seq_var$_3$ {#1} }
\seq_concat:NNN \seq_var$_1$ \seq_var$_3$ \l\_\_\_(pkg)\_internal_seq
\end{verbatim}

20.11 Constant and scratch sequences

\texttt{c\_empty_seq} Constant that is always empty.

\texttt{Rev: 2012-07-02}
Scratch sequences for local assignment. These are never used by the kernel code, and so are safe for use with any \LaTeX3-defined function. However, they may be overwritten by other non-kernel code and so should only be used for short-term storage.

\l_tmpa_seq \l_tmpb_seq

Scratch sequences for global assignment. These are never used by the kernel code, and so are safe for use with any \LaTeX3-defined function. However, they may be overwritten by other non-kernel code and so should only be used for short-term storage.

\g_tmpa_seq \g_tmpb_seq

\section{20.12 Viewing sequences}

\seq_show:N \seq_show:N \langle \textit{seq var} \rangle

Displays the entries in the \langle \textit{seq var} \rangle in the terminal.

\seq_log:N \seq_log:N \langle \textit{seq var} \rangle

Writes the entries in the \langle \textit{seq var} \rangle in the log file.
Chapter 21

The \texttt{l3int} module

Integers

Calculation and comparison of integer values can be carried out using literal numbers, \texttt{int} registers, constants and integers stored in token list variables. The standard operators $+$, $-$, $/$ and $\ast$ and parentheses can be used within such expressions to carry arithmetic operations. This module carries out these functions on \texttt{integer expressions} ("\langle \texttt{int expr} \rangle").

21.1 Integer expressions

Throughout this module, (almost) all \texttt{n}-type argument allow for an \texttt{\langle intexpr \rangle} argument with the following syntax. The \texttt{\langle integer expression \rangle} should consist, after expansion, of $+$, $-$, $\ast$, $/$, $(,)$ and of course integer operands. The result is calculated by applying standard mathematical rules with the following peculiarities:

- $/$ denotes division rounded to the closest integer with ties rounded away from zero;
- there is an error and the overall expression evaluates to zero whenever the absolute value of any intermediate result exceeds $2^{31} - 1$, except in the case of scaling operations $a \ast b / c$, for which $a \ast b$ may be arbitrarily large (but the operands $a$, $b$, $c$ are still constrained to an absolute value at most $2^{31} - 1$);
- parentheses may not appear after unary $+$ or $-$, namely placing $+(+$ or $-($ at the start of an expression or after $+, -, \ast, /$ or $(+$ leads to an error.

Each integer operand can be either an integer variable (with no need for \texttt{\int_use:N}) or an integer denotation. For example both

\begin{verbatim}
\int_show:n \{ 5 + 4 * 3 - (3 + 4 * 5) \}
\end{verbatim}


and

\begin{verbatim}
\tl_new:N \l_my_tl
\tl_set:Nn \l_my_tl \{ 5 \}
\int_new:N \l_my_int
\int_set:Nn \l_my_int \{ 4 \}
\int_show:n \{ \l_my_tl + \l_my_int * 3 - (3 + 4 * 5) \}
\end{verbatim}
show the same result \(-6\) because \verb+\l_my_tl+ expands to the integer denotation \(5\) while the integer variable \verb+\l_my_int+ takes the value \(4\). As the \(<\textbf{integer expression}\>) is fully expanded from left to right during evaluation, fully expandable and restricted-expandable functions can both be used, and \verb+\exp_not:n+ and its variants have no effect while \verb+\exp_not:N+ may incorrectly interrupt the expression.

\textbf{TeXhackers note:} Exactly two expansions are needed to evaluate \verb+\int_eval:n+. The result is \emph{not} an \(<\textbf{internal integer}\>\), and therefore should be terminated by a space if used in \verb+\int_value:w+ or in a \TeX-\textit{style} integer assignment.

As all \TeX integers, integer operands can also be: \verb+\value+(\langle L\TeX\ 2ε\ counter\rangle); dimension or skip variables, converted to integers in \verb+sp+; the character code of some character given as \langle\texttt{char}\rangle or \texttt{'\langle char\rangle}; octal numbers given as \texttt{'} followed by digits from \(0\) to \(7\); or hexadecimal numbers given as \texttt{'\texttt{a}\texttt{h}\texttt{b}\texttt{f}\texttt{A}\texttt{H}\texttt{C}\texttt{D}
\int_eval:n * \int_eval:n \{\langle \text{int expr} \rangle \}

Evaluates the \langle \text{int expr} \rangle and leaves the result in the input stream as an integer denotation: for positive results an explicit sequence of decimal digits not starting with 0, for negative results \(-\) followed by such a sequence, and 0 for zero. The \langle \text{int expr} \rangle should consist, after expansion, of \(+\), \(-\), \(*\), \(/\), \((\) and of course integer operands. The result is calculated by applying standard mathematical rules with the following peculiarities:

- / denotes division rounded to the closest integer with ties rounded away from zero;
- there is an error and the overall expression evaluates to zero whenever the absolute value of any intermediate result exceeds \(2^{31} - 1\), except in the case of scaling operations \(a*b/c\), for which \(a*b\) may be arbitrarily large;
- parentheses may not appear after unary \(+\) or \(-\), namely placing \(+\{\) or \(-\{\) at the start of an expression or after \(+\), \(-\), \(*\), \(/\) or \((\) leads to an error.

Each integer operand can be either an integer variable (with no need for \(\\int_{\text{use}}:\text{N}\)) or an integer denotation. For example both

\begin{verbatim}
\int_eval:n \{ 5 + 4 * 3 - ( 3 + 4 * 5 ) \}
\end{verbatim}

and

\begin{verbatim}
\tl_new:N \l_my_tl
\tl_set:Nn \l_my_tl \{ 5 \}
\int_new:N \l_my_int
\int_set:Nn \l_my_int \{ 4 \}
\int_eval:n \{ \l_my_tl + \l_my_int * 3 - ( 3 + 4 * 5 ) \}
\end{verbatim}

evaluate to \(-6\) because \(\l_my_tl\) expands to the integer denotation 5. As the \langle \text{int expr} \rangle is fully expanded from left to right during evaluation, fully expandable and restricted-expandable functions can both be used, and \exp_not:n and its variants have no effect while \exp_not:N may incorrectly interrupt the expression.

\TeX hackers note: Exactly two expansions are needed to evaluate \(\int_{\text{eval}}:\text{n}\). The result is not an \langle \text{internal integer} \rangle, and therefore requires suitable termination if used in a \TeX-style integer assignment.

As all \TeX integers, integer operands can also be dimension or skip variables, converted to integers in \text{sp}, or octal numbers given as \(\text{'}\) followed by digits other than 8 and 9, or hexadecimal numbers given as \(\text{"}\) followed by digits or upper case letters from \text{A} to \text{F}, or the character code of some character or one-character control sequence, given as \(\text{'}\langle \text{char} \rangle\).

\int_eval:w * \int_eval:w \{ \text{int expr} \}

New: 2018-03-30

Evaluates the \langle \text{int expr} \rangle as described for \(\int_{\text{eval}}:\text{n}\). The end of the expression is the first token encountered that cannot form part of such an expression. If that token is \scan_stop: it is removed, otherwise not. Spaces do not terminate the expression. However, spaces terminate explicit integers, and this may terminate the expression: for instance, \(\int_{\text{eval}}:w \ 1_\text{c} + \tl_\text{c} 9\) (with explicit space tokens inserted using \(~\) in a code setting) expands to 29 since the digit 9 is not part of the expression. Expansion details, etc., are as given for \(\int_{\text{eval}}:\text{n}\).
\texttt{\textbackslash int\_sign:n} \texttt{\{\textit{int expr}\}} \texttt{\textbackslash int\_sign:n} \texttt{\{\textit{int expr}\}}

Evaluates the \textit{\textit{int expr}} then leaves 1 or 0 or $-1$ in the input stream according to the sign of the result.

\texttt{\textbackslash int\_abs:n} \texttt{\{\textit{int expr}\}} \texttt{\textbackslash int\_abs:n} \texttt{\{\textit{int expr}\}}

Evaluates the \textit{\textit{int expr}} as described for \texttt{\textbackslash int\_eval:n} and leaves the absolute value of the result in the input stream as an \textit{integer denotation} after two expansions.

\texttt{\textbackslash int\_div\_round:nn} \texttt{\{\textit{int expr}\}_1} \texttt{\{\textit{int expr}\}_2} \texttt{\textbackslash int\_div\_round:nn} \texttt{\{\textit{int expr}\}_1} \texttt{\{\textit{int expr}\}_2}

Evaluates the two \textit{\textit{int expr}}s as described earlier, then divides the first value by the second, and rounds the result to the closest integer. Ties are rounded away from zero. Note that this is identical to using / directly in an \textit{int expr}. The result is left in the input stream as an \textit{integer denotation} after two expansions.

\texttt{\textbackslash int\_div\_truncate:nn} \texttt{\{\textit{int expr}\}_1} \texttt{\{\textit{int expr}\}_2} \texttt{\textbackslash int\_div\_truncate:nn} \texttt{\{\textit{int expr}\}_1} \texttt{\{\textit{int expr}\}_2}

Evaluates the two \textit{\textit{int expr}}s as described earlier, then divides the first value by the second, and rounds the result towards zero. Note that division using / rounds to the closest integer instead. The result is left in the input stream as an \textit{integer denotation} after two expansions.

\texttt{\textbackslash int\_max:nn} \texttt{\{\textit{int expr}\}_1} \texttt{\{\textit{int expr}\}_2} \texttt{\textbackslash int\_max:nn} \texttt{\{\textit{int expr}\}_1} \texttt{\{\textit{int expr}\}_2}

\texttt{\textbackslash int\_min:nn} \texttt{\{\textit{int expr}\}_1} \texttt{\{\textit{int expr}\}_2} \texttt{\textbackslash int\_min:nn} \texttt{\{\textit{int expr}\}_1} \texttt{\{\textit{int expr}\}_2}

Evaluates the \textit{\textit{int expr}}s as described for \texttt{\textbackslash int\_eval:n} and leaves either the larger or smaller value in the input stream as an \textit{integer denotation} after two expansions.

\texttt{\textbackslash int\_mod:nn} \texttt{\{\textit{int expr}\}_1} \texttt{\{\textit{int expr}\}_2} \texttt{\textbackslash int\_mod:nn} \texttt{\{\textit{int expr}\}_1} \texttt{\{\textit{int expr}\}_2}

Evaluates the two \textit{\textit{int expr}}s as described earlier, then calculates the integer remainder of dividing the first expression by the second. This is obtained by subtracting \texttt{\textbackslash int\_div\_truncate:nn} \texttt{\{\textit{int expr}\}_1} \texttt{\{\textit{int expr}\}_2} \times \texttt{\{\textit{int expr}\}_2} from \texttt{\{\textit{int expr}\}_1}. Thus, the result has the same sign as \texttt{\{\textit{int expr}\}_1} and its absolute value is strictly less than that of \texttt{\{\textit{int expr}\}_2}. The result is left in the input stream as an \textit{integer denotation} after two expansions.

### 21.2 Creating and initialising integers

\texttt{\textbackslash int\_new:N} \texttt{\textbackslash int\_new:N} \texttt{\{\textit{integer}\}} \texttt{\textbackslash int\_new:N} \texttt{\{\textit{integer}\}}

\texttt{\textbackslash int\_new:c} \texttt{\textbackslash int\_new:c}

Creates a new \textit{\textit{integer}} or raises an error if the name is already taken. The declaration is global. The \textit{\textit{integer}} is initially equal to 0.

\texttt{\textbackslash int\_const:Nn} \texttt{\textbackslash int\_const:Nn} \texttt{\{\textit{integer}\}} \texttt{\textbackslash int\_const:Nn} \texttt{\{\textit{integer}\}} \texttt{\textbackslash int\_const:Nn} \texttt{\{\textit{integer}\}} \texttt{\textbackslash int\_const:Nn} \texttt{\{\textit{integer}\}}

\texttt{\textbackslash int\_const:cn} \texttt{\textbackslash int\_const:cn}

Updated: 2011-10-22

Creates a new constant \textit{\textit{integer}} or raises an error if the name is already taken. The value of the \textit{\textit{integer}} is set globally to the \textit{\textit{int expr}}.
\mint{integer} to 0.

\mint{integer} \mint{c}

\mint{integer} \mint{c}

Ensures that the \mint{integer} exists globally by applying \mint{new} if necessary, then applies \mint{(g)zero} to leave the \mint{integer} set to zero.

\mint{integer} \mint{c}

\mint{integer} \mint{c}

Updated: 2011-12-13

\mint{integer} \mint{integer1} \mint{integer2}

Sets the content of \mint{integer1} equal to that of \mint{integer2}.

\mint{integer} \mint{c}

\mint{integer} \mint{c}

\mint{integer} \mint{c}

\mint{integer} \mint{c}

\mint{integer} \mint{c}

\mint{integer} \mint{c}

\mint{integer} \mint{c}

\mint{integer} \mint{c}

\mint{integer} \mint{c}

\mint{integer} \mint{c}

\mint{integer} \mint{c}

\mint{integer} \mint{c}

\mint{integer} \mint{c}

\mint{integer} \mint{c}

\mint{integer} \mint{c}

\mint{integer} \mint{c}

\mint{integer} \mint{c}

\mint{integer} \mint{c}

21.3 Setting and incrementing integers

\mint{integer} \mint{integer} \{\mint{expr}\}

 Adds the result of the \mint{expr} to the current content of the \mint{integer}.

\mint{integer} \mint{c}

\mint{integer} \mint{c}

\mint{integer} \mint{c}

\mint{integer} \mint{c}

\mint{integer} \mint{c}

\mint{integer} \mint{c}

\mint{integer} \mint{c}

\mint{integer} \mint{c}

Updated: 2011-10-22

\mint{integer} \mint{integer} \{\mint{expr}\}

 Sets \mint{integer} to the value of \mint{expr}, which must evaluate to an integer (as described for \mint{eval} ).
\int_sub:Nn \int_sub:cn \int_gsub:Nn \int_gsub:cn
\text{Updated: 2011-10-22}

\section*{21.4 Using integers}

\int_use:N \int_use:c
\text{Updated: 2011-10-22}

\section*{21.5 Integer expression conditionals}

\int_compare_p:nNn \int_compare:nNnTF
\text{Updated: 2011-10-22}
This function evaluates the \( \text{int expr} \)s as described for \texttt{\int_eval:n} and compares consecutive result using the corresponding \( \text{relation} \), namely it compares \( \text{int expr}_1 \) and \( \text{int expr}_2 \) using the \( \text{relation}_1 \), then \( \text{int expr}_2 \) and \( \text{int expr}_3 \) using the \( \text{relation}_2 \), until finally comparing \( \text{int expr}_N \) and \( \text{int expr}_{N+1} \) using the \( \text{relation}_N \). The test yields \texttt{true} if all comparisons are \texttt{true}. Each \( \text{int expr} \) is evaluated only once, and the evaluation is lazy, in the sense that if one comparison is \texttt{false}, then no other \( \text{integer expression} \) is evaluated and no other comparison is performed. The \( \text{relations} \) can be any of the following:

\[
\begin{align*}
\text{Equal} & \quad = \quad \text{or} \quad == \\
\text{Greater than or equal to} & \quad >= \\
\text{Greater than} & \quad > \\
\text{Less than or equal to} & \quad <= \\
\text{Less than} & \quad < \\
\text{Not equal} & \quad !=
\end{align*}
\]

This function is more flexible than \texttt{\int_compare:nNnTF} but around 5 times slower.
\begin{verbatim}
\int_case:nn \int_case:nnTF \{test int expr\} {
\{int expr case1\} \{code case1\}
\{int expr case2\} \{code case2\}
\ldots
\{int expr case_n\} \{code case_n\}
\}
\{true code\}
\{false code\}
\end{verbatim}

This function evaluates the \texttt{test int expr} and compares this in turn to each of the \texttt{int expr cases}. If the two are equal then the associated \texttt{code} is left in the input stream and other cases are discarded. If any of the cases are matched, the \texttt{true code} is also inserted into the input stream (after the code for the appropriate case), while if none match then the \texttt{false code} is inserted. The function \texttt{int_case:nn}, which does nothing if there is no match, is also available. For example

\begin{verbatim}
\int_case:nnF
\{ 2 * 5 \}
{ 5 } \{ Small \}
{ 4 + 6 } \{ Medium \}
{ -2 * 10 } \{ Negative \}
{ No idea! }
\end{verbatim}

leaves “Medium” in the input stream.

\begin{verbatim}
\int_if_odd_p:n \int_if_odd:nTF \{int expr\} {
\{true code\} \{false code\}
\end{verbatim}

This function first evaluates the \texttt{int expr} as described for \texttt{int_eval:n}. It then evaluates if this is odd or even, as appropriate.

\begin{verbatim}
\int_if_zero_p:n \int_if_zero:nTF \{int expr\} {
\{true code\} \{false code\}
\end{verbatim}

This function first evaluates the \texttt{int expr} as described for \texttt{int_eval:n}. It then evaluates if this is zero or not.

\section{21.6 Integer expression loops}

\begin{verbatim}
\int_do_until:nNnn \{int expr\} \{relation\} \{int expr2\} \{code\}
\end{verbatim}

Places the \texttt{code} in the input stream for \LaTeX{} to process, and then evaluates the relationship between the two \texttt{int expr}s as described for \texttt{int_compare:nNnTF}. If the test is \texttt{false} then the \texttt{code} is inserted into the input stream again and a loop occurs until the \texttt{relation} is true.
\int_do_while:nNnn \star \int_do_while:nNnn \{(int\ expr_1)\} \{relation\} \{(int\ expr_2)\} \{(code)\}

Places the \langle code \rangle in the input stream for \TeX to process, and then evaluates the relationship between the two \langle int expr \rangle s as described for \int_compare:nNnTF. If the test is \texttt{true} then the \langle code \rangle is inserted into the input stream again and a loop occurs until the \langle relation \rangle is \texttt{false}.

\int_until_do:nNnn \star \int_until_do:nNnn \{(int\ expr_1)\} \{relation\} \{(int\ expr_2)\} \{(code)\}

Evaluates the relationship between the two \langle int expr \rangle s as described for \int_compare:nNnTF, and then places the \langle code \rangle in the input stream if the \langle relation \rangle is \texttt{false}. After the \langle code \rangle has been processed by \TeX the test is repeated, and a loop occurs until the test is \texttt{false}.

\int_while_do:nNnn \star \int_while_do:nNnn \{(int\ expr_1)\} \{relation\} \{(int\ expr_2)\} \{(code)\}

Evaluates the relationship between the two \langle int expr \rangle s as described for \int_compare:nNnTF, and then places the \langle code \rangle in the input stream if the \langle relation \rangle is \texttt{true}. After the \langle code \rangle has been processed by \TeX the test is repeated, and a loop occurs until the test is \texttt{false}.

\int_do_until:nn \star \int_do_until:nn \{(integer\ relation)\} \{(code)\}

\texttt{Updated: 2013-01-13}

Places the \langle code \rangle in the input stream for \TeX to process, and then evaluates the \langle integer relation \rangle as described for \int_compare:nTF. If the test is \texttt{false} then the \langle code \rangle is inserted into the input stream again and a loop occurs until the \langle relation \rangle is \texttt{true}.

\int_do_until:nn \star \int_do_until:nn \{(integer\ relation)\} \{(code)\}

\texttt{Updated: 2013-01-13}

Places the \langle code \rangle in the input stream for \TeX to process, and then evaluates the \langle integer relation \rangle as described for \int_compare:nTF. If the test is \texttt{true} then the \langle code \rangle is inserted into the input stream again and a loop occurs until the \langle relation \rangle is \texttt{false}.

\int_until_do:nn \star \int_until_do:nn \{(integer\ relation)\} \{(code)\}

\texttt{Updated: 2013-01-13}

Evaluates the \langle integer relation \rangle as described for \int_compare:nTF, and then places the \langle code \rangle in the input stream if the \langle relation \rangle is \texttt{false}. After the \langle code \rangle has been processed by \TeX the test is repeated, and a loop occurs until the test is \texttt{true}.

\int_while_do:nn \star \int_while_do:nn \{(integer\ relation)\} \{(code)\}

\texttt{Updated: 2013-01-13}

Evaluates the \langle integer relation \rangle as described for \int_compare:nTF, and then places the \langle code \rangle in the input stream if the \langle relation \rangle is \texttt{true}. After the \langle code \rangle has been processed by \TeX the test is repeated, and a loop occurs until the test is \texttt{false}.


21.7 Integer step functions

\int_step_function:n \{\text{final value}\} \{\text{function}\}
\int_step_function:nn \{\text{initial value}\} \{\text{final value}\} \{\text{function}\}
\int_step_function:nNn \{\text{initial value}\} \{\text{step}\} \{\text{final value}\} \{\text{function}\}

This function first evaluates the \langle initial value \rangle, \langle step \rangle and \langle final value \rangle, all of which should be integer expressions. The \langle function \rangle is then placed in front of each \langle value \rangle from the \langle initial value \rangle to the \langle final value \rangle in turn (using \langle step \rangle between each \langle value \rangle). The \langle step \rangle must be non-zero. If the \langle step \rangle is positive, the loop stops when the \langle value \rangle becomes larger than the \langle final value \rangle. If the \langle step \rangle is negative, the loop stops when the \langle value \rangle becomes smaller than the \langle final value \rangle. The \langle function \rangle should absorb one numerical argument. For example

\begin{verbatim}
\cs_set:Npn \my_func:n #1 { \[I saw #1\] \quad }
\int_step_function:nNn \{1\} \{1\} \{5\} \my_func:n
\end{verbatim}

would print

\begin{verbatim}
[I saw 1] [I saw 2] [I saw 3] [I saw 4] [I saw 5]
\end{verbatim}

The functions \int_step_function:nN and \int_step_function:nnN both use a fixed \langle step \rangle of 1, and in the case of \int_step_function:nN the \langle initial value \rangle is also fixed as 1. These functions are provided as simple short-cuts for code clarity.

\int_step_inline:nn \{\text{final value}\} \{\text{code}\}
\int_step_inline:nn \{\text{initial value}\} \{\text{final value}\} \{\text{code}\}
\int_step_inline:nnn \{\text{initial value}\} \{\text{step}\} \{\text{final value}\} \{\text{code}\}

This function first evaluates the \langle initial value \rangle, \langle step \rangle and \langle final value \rangle, all of which should be integer expressions. Then for each \langle value \rangle from the \langle initial value \rangle to the \langle final value \rangle in turn (using \langle step \rangle between each \langle value \rangle), the \langle code \rangle is inserted into the input stream with \#1 replaced by the current \langle value \rangle. Thus the \langle code \rangle should define a function of one argument (\#1).

The functions \int_step_inline:nn and \int_step_inline:nnn both use a fixed \langle step \rangle of 1, and in the case of \int_step_inline:nn the \langle initial value \rangle is also fixed as 1. These functions are provided as simple short-cuts for code clarity.

\int_step_variable:nN \{\text{final value}\} \{\text{tl var}\} \{\text{code}\}
\int_step_variable:nn \{\text{initial value}\} \{\text{final value}\} \{\text{tl var}\} \{\text{code}\}
\int_step_variable:nnNn \{\text{initial value}\} \{\text{step}\} \{\text{final value}\} \{\text{tl var}\} \{\text{code}\}

This function first evaluates the \langle initial value \rangle, \langle step \rangle and \langle final value \rangle, all of which should be integer expressions. Then for each \langle value \rangle from the \langle initial value \rangle to the \langle final value \rangle in turn (using \langle step \rangle between each \langle value \rangle), the \langle code \rangle is inserted into the input stream, with the \langle tl var \rangle defined as the current \langle value \rangle. Thus the \langle code \rangle should make use of the \langle tl var \rangle.

The functions \int_step_variable:nN and \int_step_variable:nnNn both use a fixed \langle step \rangle of 1, and in the case of \int_step_variable:nN the \langle initial value \rangle is also fixed as 1. These functions are provided as simple short-cuts for code clarity.
21.8 Formatting integers

Integers can be placed into the output stream with formatting. These conversions apply to any integer expressions.

\texttt{\int_to_arabic:n \{int expr\}}

Places the value of the \texttt{\{int expr\}} in the input stream as digits, with category code 12 (other).

\texttt{\int_to_alph:n \{int expr\}}

Evaluates the \texttt{\{int expr\}} and converts the result into a series of letters, which are then left in the input stream. The conversion rule uses the 26 letters of the English alphabet, in order, adding letters when necessary to increase the total possible range of representable numbers. Thus

\texttt{\int_to_alph:n \{ 1 \}}

places a in the input stream,

\texttt{\int_to_alph:n \{ 26 \}}

is represented as z and

\texttt{\int_to_alph:n \{ 27 \}}

is converted to aa. For conversions using other alphabets, use \texttt{\int_to_symbols:nnn} to define an alphabet-specific function. The basic \texttt{\int_to_alph:n} and \texttt{\int_to_Alph:n} functions should not be modified. The resulting tokens are digits with category code 12 (other) and letters with category code 11 (letter).

\texttt{\int_to_symbols:nnn \{\{\text{int expr}\} \{\text{total symbols}\}\} \{\text{value to symbol mapping}\}}

This is the low-level function for conversion of an \texttt{\{int expr\}} into a symbolic form (often letters). The \texttt{\{total symbols\}} available should be given as an integer expression. Values are actually converted to symbols according to the \texttt{\{value to symbol mapping\}}. This should be given as \texttt{\{total symbols\}} pairs of entries, a number and the appropriate symbol. Thus the \texttt{\int_to_alph:n} function is defined as

\texttt{\cs_new:Npn \int_to_alph:n \#1}

\begin{verbatim}
  \int_to_symbols:nnn \{\#1\} \{ 26 \}
  \{
    \{ 1 \} \{ a \}
    \{ 2 \} \{ b \}
    \ldots
    \{ 26 \} \{ z \}
  \}
\end{verbatim}
\int_to_bin:n \{\textit{int \ expr}\} \quad \text{New: 2014-02-11}
Calculates the value of the $\langle \textit{int \ expr} \rangle$ and places the binary representation of the result in the input stream.

\int_to_hex:n \quad \int_to_Hex:n \quad \text{New: 2014-02-11}
Calculates the value of the $\langle \textit{int \ expr} \rangle$ and places the hexadecimal (base 16) representation of the result in the input stream. Letters are used for digits beyond 9: lower case letters for $\int_to_hex:n$ and upper case ones for $\int_to_Hex:n$. The resulting tokens are digits with category code 12 (other) and letters with category code 11 (letter).

\int_to_oct:n \quad \text{New: 2014-02-11}
Calculates the value of the $\langle \textit{int \ expr} \rangle$ and places the octal (base 8) representation of the result in the input stream. The resulting tokens are digits with category code 12 (other) and letters with category code 11 (letter).

\int_to_base:nn \quad \int_to_Base:nn \quad \text{Updated: 2014-02-11}
Calculates the value of the $\langle \textit{int \ expr} \rangle$ and converts it into the appropriate representation in the $\langle \textit{base} \rangle$. The later may be given as an integer expression. For bases greater than 10 the higher “digits” are represented by letters from the English alphabet: lower case letters for $\int_to_base:n$ and upper case ones for $\int_to_Base:n$. The maximum $\langle \textit{base} \rangle$ value is 36. The resulting tokens are digits with category code 12 (other) and letters with category code 11 (letter).

\textsf{TeXhackers note:} This is a generic version of $\int_to_bin:n$, etc.

\int_to_roman:n \quad \int_to_Roman:n \quad \text{Updated: 2011-10-22}
Places the value of the $\langle \textit{int \ expr} \rangle$ in the input stream as Roman numerals, either lower case ($\int_to_roman:n$) or upper case ($\int_to_Roman:n$). If the value is negative or zero, the output is empty. The Roman numerals are letters with category code 11 (letter). The letters used are mdclxvi, repeated as needed: the notation with bars (such as $\overline{v}$ for 5000) is not used. For instance $\int_to_roman:n \{8249\}$ expands to mmmmmmccclxix.

21.9 Converting from other formats to integers

\int_from_alph:n \quad \int_from_alph:n \{\langle \textit{letters} \rangle\} \quad \text{Updated: 2014-08-25}
Converts the $\langle \textit{letters} \rangle$ into the integer (base 10) representation and leaves this in the input stream. The $\langle \textit{letters} \rangle$ are first converted to a string, with no expansion. Lower and upper case letters from the English alphabet may be used, with “a” equal to 1 through to “z” equal to 26. The function also accepts a leading sign, made of + and -. This is the inverse function of $\int_to_alph:n$ and $\int_to_Alph:n$.

\int_from_bin:n \quad \int_from_bin:n \{\langle \textit{binary number} \rangle\} \quad \text{New: 2014-02-11, Updated: 2014-08-25}
Converts the $\langle \textit{binary number} \rangle$ into the integer (base 10) representation and leaves this in the input stream. The $\langle \textit{binary number} \rangle$ is first converted to a string, with no expansion. The function accepts a leading sign, made of + and -, followed by binary digits. This is the inverse function of $\int_to_bin:n$. 

176
\int_from_hex:n \{\langle hexadecimal \ number \rangle\}

Converts the \langle hexadecimal \ number \rangle into the integer (base 10) representation and leaves this in the input stream. Digits greater than 9 may be represented in the \langle hexadecimal \ number \rangle by upper or lower case letters. The \langle hexadecimal \ number \rangle is first converted to a string, with no expansion. The function also accepts a leading sign, made of + and -. This is the inverse function of \int_to_hex:n and \int_to_Hex:n.

\int_from_oct:n \{\langle octal \ number \rangle\}

Converts the \langle octal \ number \rangle into the integer (base 10) representation and leaves this in the input stream. The \langle octal \ number \rangle is first converted to a string, with no expansion. The function accepts a leading sign, made of + and -, followed by octal digits. This is the inverse function of \int_to_oct:n.

\int_from_roman:n \{\langle roman \ numeral \rangle\}

Converts the \langle roman \ numeral \rangle into the integer (base 10) representation and leaves this in the input stream. The \langle roman \ numeral \rangle may be in upper or lower case; if the numeral contains characters besides mdclxvi or MDCLXVI then the resulting value is -1. This is the inverse function of \int_to_roman:n and \int_to_Roman:n.

\int_from_base:nn \{\langle number \rangle\} \{\langle base \rangle\}

Converts the \langle number \rangle expressed in \langle base \rangle into the appropriate value in base 10. The \langle number \rangle is first converted to a string, with no expansion. The \langle number \rangle should consist of digits and letters (either lower or upper case), plus optionally a leading sign. The maximum \langle base \rangle value is 36. This is the inverse function of \int_to_base:nn and \int_to_Base:nn.

### 21.10 Random integers

\int_rand:nn \{\langle int \ expr_1 \rangle\} \{\langle int \ expr_2 \rangle\}

Evaluates the two \langle int \ expr \rangles and produces a pseudo-random number between the two (with bounds included). This is not available in older versions of \XeTeX.

\int_rand:n \{\langle int \ expr \rangle\}

Evaluates the \langle int \ expr \rangle then produces a pseudo-random number between 1 and the \langle int \ expr \rangle (included). This is not available in older versions of \XeTeX.

### 21.11 Viewing integers

\int_show:N \langle integer \rangle

Displays the value of the \langle integer \rangle on the terminal.
\int_show:n \{int \ expr\} \int_log:N \{integer\}

Displays the result of evaluating the \textit{int expr} on the terminal.

\int_log:N \int_log:c \int_log:n \{int \ expr\}

Writes the value of the \textit{integer} in the log file.

\c_zero_int \c_one_int \c_max_int \c_max_register_int \c_max_char_int

Integer values used with primitive tests and assignments: their self-terminating nature makes these more convenient and faster than literal numbers.

The maximum value that can be stored as an integer.

Maximum number of registers.

Maximum character code completely supported by the engine.

\l_tmpa_int \l_tmpb_int \g_tmpa_int \g_tmpb_int

Scratch integer for local assignment. These are never used by the kernel code, and so are safe for use with any \LaTeX3-defined function. However, they may be overwritten by other non-kernel code and so should only be used for short-term storage.

Scratch integer for global assignment. These are never used by the kernel code, and so are safe for use with any \LaTeX3-defined function. However, they may be overwritten by other non-kernel code and so should only be used for short-term storage.
21.14 Direct number expansion

\int_value:w \star \int_value:w \langle \text{integer} \rangle \int_value:w \langle \text{integer denotation} \rangle \ (\text{optional space})

Expands the following tokens until an \langle \text{integer} \rangle is formed, and leaves a normalized form (no leading sign except for negative numbers, no leading digit 0 except for zero) in the input stream as category code 12 (other) characters. The \langle \text{integer} \rangle can consist of any number of signs (with intervening spaces) followed by

- an integer variable (in fact, any \TeX register except \texttt{\toks}) or
- explicit digits (or by \langle \text{octal digits} \rangle or \langle \text{hexadecimal digits} \rangle or \langle \text{character} \rangle).

In this last case expansion stops once a non-digit is found; if that is a space it is removed as in \texttt{f}\texttt{-expansion}, and so \texttt{\exp_stop_f}: may be employed as an end marker. Note that protected functions \texttt{are} expanded by this process.

This function requires exactly one expansion to produce a value, and so is suitable for use in cases where a number is required “directly”. In general, \texttt{\int_eval:n} is the preferred approach to generating numbers.

\TeXhackers\texttt{note}: This is the \TeX primitive \texttt{\number}.

21.15 Primitive conditionals

\texttt{\if_int_compare:w \star \if_int_compare:w \langle \text{integer}_1 \rangle \ (\text{relation}) \langle \text{integer}_2 \rangle \langle \text{true code} \rangle \else: \langle \text{false code} \rangle \fi:}

Compare two integers using \langle \text{relation} \rangle, which must be one of \texttt{=} \text{,} \texttt{<} or \texttt{>} with category code 12. The \texttt{\else:} branch is optional.

\TeXhackers\texttt{note}: This is the \TeX primitive \texttt{\ifnum}.

\texttt{\if_case:w \star \if_case:w \langle \text{integer} \rangle \ (\text{case}_0) \or: \langle \text{case}_1 \rangle \or: \ldots \else: \langle \text{default} \rangle \fi:}

Selects a case to execute based on the value of the \langle \text{integer} \rangle. The first case (\langle \text{case}_0 \rangle) is executed if \langle \text{integer} \rangle is 0, the second (\langle \text{case}_1 \rangle) if the \langle \text{integer} \rangle is 1, etc. The \langle \text{integer} \rangle may be a literal, a constant or an integer expression (\textit{e.g.} using \texttt{\int_eval:n}).

\TeXhackers\texttt{note}: These are the \TeX primitives \texttt{\ifcase} and \texttt{\or}.
\if_int_odd:w \if_int_odd:w (tokens) (optional space)
\else:
\fi:
Expands (tokens) until a non-numeric token or a space is found, and tests whether the resulting (integer) is odd. If so, (true code) is executed. The \else: branch is optional.

\TeX{} hackers note: This is the \TeX{} primitive \texttt{\textbackslash ifodd}. 
Chapter 22

The l3flag module

Expandable flags

Flags are the only data-type that can be modified in expansion-only contexts. This module is meant mostly for kernel use: in almost all cases, booleans or integers should be preferred to flags because they are very significantly faster.

A flag can hold any (small) non-negative value, which we call its (height). In expansion-only contexts, a flag can only be “raised”: this increases the (height) by 1. The (height) can also be queried expandably. However, decreasing it, or setting it to zero requires non-expandable assignments.

Flag variables are always local.

A typical use case of flags would be to keep track of whether an exceptional condition has occurred during expandable processing, and produce a meaningful (non-expandable) message after the end of the expandable processing. This is exemplified by l3str-convert, which for performance reasons performs conversions of individual characters expandably and for readability reasons produces a single error message describing incorrect inputs that were encountered.

Flags should not be used without carefully considering the fact that raising a flag takes a time and memory proportional to its height and that the memory cannot be reclaimed even if the flag is cleared. Flags should not be used unless it is unavoidable.

In earlier versions, flags were referenced by an n-type (flag name) such as fp_-overflow, used as part of \use:c constructions. All of the commands described below have n-type analogues that can still appear in old code, but the N-type commands are to be preferred moving forward. The n-type (flag name) is simply mapped to \l_(flag name)_flag, which makes it easier for packages using public flags (such as l3fp) to retain backwards compatibility.

22.1 Setting up flags

| \flag_new:N | \flag_new:N (flag var) |
| \flag_new:c | Creates a new (flag var), or raises an error if the name is already taken. The declaration is global, but flags are always local variables. The (flag var) initially has zero height.
\flag_clear:N  \flag_clear:c
Sets the height of the \texttt{\flag var} to zero. The assignment is local.

\flag_clear_new:N  \flag_clear_new:c
Ensures that the \texttt{\flag var} exists globally by applying \flag_new:N if necessary, then applies \flag_clear:N, setting the height to zero locally.

\flag_show:N  \flag_show:c
Displays the height of the \texttt{\flag var} in the terminal.

\flag_log:N  \flag_log:c
Writes the height of the \texttt{\flag var} in the log file.

\flag_if_exist_p:N  \flag_if_exist_p:c
This function returns \texttt{true} if the \texttt{\flag var} is currently defined, and \texttt{false} otherwise. This does not check that the \texttt{\flag var} really is a flag variable.

\flag_if_raised_p:N  \flag_if_raised_p:c
This function returns \texttt{true} if the \texttt{\flag var} has non-zero height, and \texttt{false} if the \texttt{\flag var} has zero height.

\flag_height:N  \flag_height:c
Expands to the height of the \texttt{\flag var} as an integer denotation.

\flag_raise:N  \flag_raise:c
The height of \texttt{\flag var} is increased by 1 locally.

\flag_ensure_raised:N  \flag_ensure_raised:c
Ensures the \texttt{\flag var} is raised by making its height at least 1, locally.

22.2 Expandable flag commands
\l_tmpa_flag Scratch flag for local assignment. These are never used by the kernel code, and so are safe for use with any \TeX3-defined function. However, they may be overwritten by other non-kernel code and so should only be used for short-term storage.
\l_tmpb_flag

Rev: 2024-01-12
Chapter 23

The \l3clist module
Comma separated lists

Comma lists (in short, \clist) contain ordered data where items can be added to the left or right end of the list. This data type allows basic list manipulations such as adding/removing items, applying a function to every item, removing duplicate items, extracting a given item, using the comma list with specified separators, and so on. Sequences (defined in \l3seq) are safer, faster, and provide more features, so they should often be preferred to comma lists. Comma lists are mostly useful when interfacing with \LaTeX\ or other code that expects or provides items separated by commas.

Several items can be added at once. To ease input of comma lists from data provided by a user outside an \texttt{\ExplSyntaxOn ... \ExplSyntaxOff} block, spaces are removed from both sides of each comma-delimited argument upon input. Blank arguments are ignored, to allow for trailing commas or repeated commas (which may otherwise arise when concatenating comma lists “by hand”). In addition, a set of braces is removed if the result of space-trimming is braced: this allows the storage of any item in a comma list. For instance,

\begin{verbatim}
\clist_new:N \l_my_clist
\clist_put_left:Nn \l_my_clist { -a- , -\{b\} , c\-d }\n\clist_put_right:Nn \l_my_clist { -e- , , \{f\} , }
\end{verbatim}

results in \texttt{\l_my_clist} containing \texttt{a,b,c\-d,e\-}, \texttt{\{f\}} namely the five items \texttt{a}, \texttt{b}, \texttt{c\-d}, \texttt{e\-} and \texttt{f}. Comma lists normally do not contain empty or blank items so the following gives an empty comma list:

\begin{verbatim}
\clist_clear_new:N \l_my_clist
\clist_set:Nn \l_my_clist { , - , , }
\clist_if_empty:NTF \l_my_clist { true } { false }
\end{verbatim}

and it leaves \texttt{true} in the input stream. To include an “unsafe” item (empty, or one that contains a comma, or starts or ends with a space, or is a single brace group), surround it with braces.

Any \texttt{n}-type token list is a valid comma list input for \l3clist functions, which will split the token list at every comma and process the items as described above. On the other hand, \texttt{N}-type functions expect comma list variables, which are particular token list variables in which this processing of items (and removal of blank items) has already
occurred. Because comma list variables are token list variables, expanding them once yields their items separated by commas, and \texttt{\texttt{l3tl}} functions such as \texttt{\texttt{tl\_show:N}} can be applied to them. (These functions often have \texttt{l3clist} analogues, which should be preferred.)

Almost all operations on comma lists are noticeably slower than those on sequences so converting the data to sequences using \texttt{\texttt{seq\_set\_from\_clist:Nn}} (see \texttt{l3seq}) may be advisable if speed is important. The exception is that \texttt{\texttt{clist\_if\_in:NnTF}} and \texttt{\texttt{clist\_remove\_duplicates:N}} may be faster than their sequence analogues for large lists. However, these functions work slowly for “unsafe” items that must be braced, and may produce errors when their argument contains {, } or \# (assuming the usual \TeX\ category codes apply). The sequence data type should thus certainly be preferred to comma lists to store such items.

\section{Creating and initialising comma lists}

\begin{verbatim}
\clist_new:N \clist_new:N \clist_new:c \clist_new:nn
\end{verbatim}

Creates a new \texttt{\texttt{clist var}} or raises an error if the name is already taken. The declaration is global. The \texttt{\texttt{clist var}} initially contains no items.

\begin{verbatim}
\clist_const:Nn \clist_const:(Ne|cn|ce)
\end{verbatim}

Creates a new constant \texttt{\texttt{clist var}} or raises an error if the name is already taken. The value of the \texttt{\texttt{clist var}} is set globally to the \texttt{\texttt{comma list}}.

\begin{verbatim}
\clist_clear:N \clist_clear:c \clist_gclear:N \clist_gclear:c
\end{verbatim}

Clears all items from the \texttt{\texttt{clist var}}.

\begin{verbatim}
\clist_clear_new:N \clist_clear_new:c \clist_gclear_new:N \clist_gclear_new:c
\end{verbatim}

Ensures that the \texttt{\texttt{clist var}} exists globally by applying \texttt{\texttt{clist\_new:N}} if necessary, then applies \texttt{\texttt{clist\_gclear:N}} to leave the list empty.

\begin{verbatim}
\clist_set_eq:NN \clist_set_eq:nn \clist_gset_eq:NN \clist_gset_eq:nn
\end{verbatim}

Sets the content of \texttt{\texttt{clist var}} equal to that of \texttt{\texttt{clist var}}. To set a token list variable equal to a comma list variable, use \texttt{\texttt{tl\_set\_eq:NN}}. Conversely, setting a comma list variable to a token list is unadvisable unless one checks space-trimming and related issues.

\begin{verbatim}
\clist_set_from_seq:NN \clist_set_from_seq:nn \clist_gset_from_seq:NN \clist_gset_from_seq:nn
\end{verbatim}

Converts the data in the \texttt{\texttt{seq var}} into a \texttt{\texttt{clist var}}: the original \texttt{\texttt{seq var}} is unchanged. Items which contain either spaces or commas are surrounded by braces.
23.2 Adding data to comma lists

\clist_set:Nn \clist_set:N \clist_set:Nn \clist_set:Nn
\clist_set:cccc
\clist_gset:Nn \clist_gset:Nn \clist_gset:Nn \clist_gset:ccc
\clist_if_exist_p:N \clist_if_exist:NTF \clist_if_exist_p:NTF \clist_if_exist:NTF
\clist_if_exist:p:N \clist_if_exist:TF \clist_if_exist:TF \clist_if_exist:TF
\clist_gset:Nn \clist_gset:Nn \clist_gset:Nn \clist_gset:ccc
\clist_gput_left:Nn \clist_gput_left:Nn \clist_gput_left:Nn \clist_gput_left:TF
\clist_gput_right:Nn \clist_gput_right:Nn \clist_gput_right:Nn \clist_gput_right:TF

Added: 2011-09-06
Sets \clist var to contain the \items, removing any previous content from the variable. Blank items are omitted, spaces are removed from both sides of each item, then a set of braces is removed if the resulting space-trimmed item is braced. To store some \tokens as a single \item even if the \tokens contain commas or spaces, add a set of braces: \clist_set:Nn \clist var \{ \tokens \}.

Updated: 2011-09-06
Apends the \items to the left of the \clist var. Blank items are omitted, spaces are removed from both sides of each item, then a set of braces is removed if the resulting space-trimmed item is braced. To append some \tokens as a single \item even if the \tokens contain commas or spaces, add a set of braces: \clist_put_left:Nn \clist var \{ \tokens \}.

Updated: 2011-09-06
Apends the \items to the right of the \clist var. Blank items are omitted, spaces are removed from both sides of each item, then a set of braces is removed if the resulting space-trimmed item is braced. To append some \tokens as a single \item even if the \tokens contain commas or spaces, add a set of braces: \clist_put_right:Nn \clist var \{ \tokens \}.
23.3 Modifying comma lists

While comma lists are normally used as ordered lists, it may be necessary to modify the content. The functions here may be used to update comma lists, while retaining the order of the unaffected entries.

\clist_remove_duplicates:N \clist_remove_duplicates:N \clist_var
\clist_remove_duplicates:c
\clist_gremove_duplicates:N
\clist_gremove_duplicates:c

Removes duplicate items from the \clist_var, leaving the left most copy of each item in the \clist_var. The \item comparison takes place on a token basis, as for \tl_if_eq:nnTF.

\texttt{\TeXhacker{n}ote}: This function iterates through every item in the \clist_var and does a comparison with the \item already checked. It is therefore relatively slow with large comma lists. Furthermore, it may fail if any of the items in the \clist_var contains \texttt{,}, or \texttt{#} (assuming the usual \TeX{} category codes apply).

\clist_remove_all:Nn \clist_remove_all:Nn \clist_var \{\item\}
\clist_remove_all:cn \clist_remove_all:cn \clist_var
\clist_gremove_all:Nn \clist_gremove_all:Nn \clist_var
\clist_gremove_all:cn \clist_gremove_all:cn \clist_var

Updated: 2011-09-06

Removes every occurrence of \item from the \clist_var. The \item comparison takes place on a token basis, as for \tl_if_eq:nnTF.

\texttt{\TeXhacker{n}ote}: The function may fail if the \item contains \texttt{,}, or \texttt{#} (assuming the usual \TeX{} category codes apply).

\clist_reverse:N \clist_reverse:N \clist_var
\clist_reverse:c
\clist_greverse:N
\clist_greverse:c

Nov: 2014-07-18

Reverses the order of items stored in the \clist_var.

\clist_reverse:n \clist_reverse:n \{\comma list\}

Nov: 2014-07-18

Leaves the items in the \comma list in the input stream in reverse order. Contrarily to other what is done for other \texttt{n-type} \comma list arguments, braces and spaces are preserved by this process.

\texttt{\TeXhacker{n}ote}: The result is returned within \texttt{\unexpanded}, which means that the comma list does not expand further when appearing in an \texttt{e-type} or \texttt{x-type} argument expansion.
23.4 Comma list conditionals

\clist_sort:Nn \clist_sort:cn \clist_gsort:Nn \clist_gsort:cn

Sorts the items in the ⟨clist var ⟩ according to the ⟨comparison code⟩, and assigns the result to ⟨clist var ⟩. The details of sorting comparison are described in Section 6.1.

\clist_sort:Nn \clist_sort:cn \clist_gsort:Nn \clist_gsort:cn

New: 2017-02-06

Tests if the ⟨clist var ⟩ is empty (containing no items).

\clist_if_empty_p:N \clist_if_empty_p:c \clist_if_empty:NTF \clist_if_empty:c
\clist_if_empty_p:n \clist_if_empty:nTF \clist_if_empty:n

Tests if the ⟨clist var ⟩ is empty (containing no items). The rules for space trimming are as for other n-type comma-list functions, hence the comma list {-,{},} (without outer braces) is empty, while {-,{},} (without outer braces) contains one element, which happens to be empty: the comma-list is not empty.

\clist_if_in:NnTF \clist_if_in:N \clist_if_in:NTF \clist_if_in:N
\clist_if_in:nTF \clist_if_in:n \clist_if_in:n

Tests if the ⟨item ⟩ is present in the ⟨clist var ⟩. In the case of an n-type ⟨comma list⟩, the usual rules of space trimming and brace stripping apply. Hence,

\clist_if_in:nTF \clist_if_in:NmTF { a , {b} , {b} , c } { b } {true} {false}

yields true.

\TeX hackers note: The function may fail if the ⟨item ⟩ contains ⟨,⟩, or ⟨#⟩ (assuming the usual \TeX category codes apply).

23.5 Mapping over comma lists

The functions described in this section apply a specified function to each item of a comma list. All mappings are done at the current group level, i.e. any local assignments made by the ⟨function⟩ or ⟨code⟩ discussed below remain in effect after the loop.

When the comma list is given explicitly, as an n-type argument, spaces are trimmed around each item. If the result of trimming spaces is empty, the item is ignored. Otherwise, if the item is surrounded by braces, one set is removed, and the result is passed to the mapped function. Thus, if the comma list that is being mapped is {a, ⟨{b}⟩, ⟨{b}⟩, ⟨{c}⟩}, then the arguments passed to the mapped function are ‘a’, ‘⟨{b}⟩’, an empty argument, and ‘c’.
When the comma list is given as an \texttt{N}-type argument, spaces have already been trimmed on input, and items are simply stripped of one set of braces if any. This case is more efficient than using \texttt{n}-type comma lists.

\begin{verbatim}
\clist_map_function:NN \clist_map_function:cn \clist_map_function:nn \clist_map_function:nN
\end{verbatim}

Applies \texttt{function} to every \texttt{item} stored in the \texttt{clist var}. The \texttt{items} are returned from left to right. The function \texttt{\clist_map_inline:Nn} is in general more efficient than \texttt{\clist_map_function:NN}.

\begin{verbatim}
\clist_map_inline:Nn \clist_map_inline:cn \clist_map_inline:nn
\end{verbatim}

Applies \texttt{inline function} to every \texttt{item} stored within the \texttt{clist var}. The \texttt{inline function} should consist of code which receives the \texttt{item} as \texttt{#1}. The \texttt{items} are returned from left to right.

\begin{verbatim}
\clist_map_variable:NNn \clist_map_variable:cNn \clist_map_variable:nNn
\end{verbatim}

Stores each \texttt{item} of the \texttt{clist var} in turn in the (token list) \texttt{variable} and applies the \texttt{code}. The \texttt{code} will usually make use of the \texttt{variable}, but this is not enforced. The assignments to the \texttt{variable} are local. Its value after the loop is the last \texttt{item} in the \texttt{clist var}, or its original value if there were no \texttt{item}. The \texttt{items} are returned from left to right.

\begin{verbatim}
\clist_map_tokens:Nn \clist_map_tokens:cn \clist_map_tokens:nn
\end{verbatim}

Calls \texttt{code} \texttt{\{item\}} for every \texttt{item} stored in the \texttt{clist var}. The \texttt{code} receives each \texttt{item} as a trailing brace group. If the \texttt{code} consists of a single function this is equivalent to \texttt{\clist_map_function:nN}.

\begin{verbatim}
\clist_map_break:
\end{verbatim}

Used to terminate a \texttt{\clist_map_...} function before all entries in the \texttt{comma list} have been processed. This normally takes place within a conditional statement, for example

\begin{verbatim}
\clist_map_inline:Nn \l_my_clist
{ \str_if_eq:nnTF { #1 } { bingo } { \clist_map_break: }
  % Do something useful
}
\end{verbatim}

Use outside of a \texttt{\clist_map_...} scenario leads to low level \TeX errors.

\textbf{\TeX hackers note:} When the mapping is broken, additional tokens may be inserted before further items are taken from the input stream. This depends on the design of the mapping function.
\texttt{\clist_map_break:n \{\textit{code}\}}

Used to terminate a \texttt{\clist_map...} function before all entries in the \texttt{\langle comma list\rangle} have been processed, inserting the \texttt{\langle code\rangle} after the mapping has ended. This normally takes place within a conditional statement, for example

\begin{verbatim}
\clist_map_inline:Nn \l_my_clist
  \{\str_if_eq:nnTF { #1 } { bingo } \{ \clist_map_break:n \{ <code> \} \%
    \text{Do something useful} \\}
\}
\end{verbatim}

Use outside of a \texttt{\clist_map...} scenario leads to low level \TeX{} errors.

\textbf{\TeX{}hackers note:} When the mapping is broken, additional tokens may be inserted before the \texttt{\langle code\rangle} is inserted into the input stream. This depends on the design of the mapping function.

\begin{verbatim}
\clist_count:N \langle clist var\rangle
\clist_count:c \langle clist var\rangle
\clist_count:n \langle clist var\rangle
\clist_count:e \langle clist var\rangle
\end{verbatim}

Leaves the number of items in the \texttt{\langle clist var\rangle} in the input stream as an \texttt{\langle integer\rangle} \texttt{\langle denotation\rangle}. The total number of items in a \texttt{\langle clist var\rangle} includes those which are duplicates, i.e. every item in a \texttt{\langle clist var\rangle} is counted.
23.6 Using the content of comma lists directly

\clist_use:Nnnn \clist_use:Nnnn \clist var \{\langle \text{separator between two} \rangle\}
\clist_use:cn \clist_use:cn \{\langle \text{separator between more than two} \rangle\} \{\langle \text{separator between final two} \rangle\}

Places the contents of the \clist var in the input stream, with the appropriate \langle \text{separator} \rangle between the items. Namely, if the comma list has more than two items, the \langle \text{separator between more than two} \rangle is placed between each pair of items except the last, for which the \langle \text{separator between final two} \rangle is used. If the comma list has exactly two items, then they are placed in the input stream separated by the \langle \text{separator between two} \rangle. If the comma list has a single item, it is placed in the input stream, and a comma list with no items produces no output. An error is raised if the variable does not exist or if it is invalid.

For example,

\clist_set:Nn \l_tmpa_clist { a , b , , c , \{de\} , f }
\clist_use:Nnnn \l_tmpa_clist { ~and~ } { ,~ } { ,~and~ }

inserts “a, b, c, de, and f” in the input stream. The first separator argument is not used in this case because the comma list has more than 2 items.

\TeXhackers note: The result is returned within the \unexpanded primitive (\exp_not:n), which means that the \langle \text{items} \rangle do not expand further when appearing in an e-type or x-type argument expansion.

\clist_use:Nn \clist_use:NN \clist var \{\langle \text{separator} \rangle\}
\clist_use:cn \clist_use:cn \{\langle \text{separator} \rangle\}

Places the contents of the \clist var in the input stream, with the \langle \text{separator} \rangle between the items. If the comma list has a single item, it is placed in the input stream, and a comma list with no items produces no output. An error is raised if the variable does not exist or if it is invalid.

For example,

\clist_set:Nn \l_tmpa_clist { a , b , , c , \{de\} , f }
\clist_use:Nn \l_tmpa_clist { -and- } { , - } { , -and- }

inserts “a and b and c and de and f” in the input stream.

\TeXhackers note: The result is returned within the \unexpanded primitive (\exp_not:n), which means that the \langle \text{items} \rangle do not expand further when appearing in an e-type or x-type argument expansion.
Places the contents of the ⟨comma list⟩ in the input stream, with the appropriate ⟨separator⟩ between the items. As for \clist_set:Nn, blank items are omitted, spaces are removed from both sides of each item, then a set of braces is removed if the resulting space-trimmed item is braced. The ⟨separators⟩ are then inserted in the same way as for \clist_use:Nnnn and \clist_use:Nn, respectively.

\emph{\TeXhackers note:} The result is returned within the \unexpanded primitive (\exp_not:n), which means that the ⟨items⟩ do not expand further when appearing in an e-type or x-type argument expansion.

### 23.7 Comma lists as stacks

Comma lists can be used as stacks, where data is pushed to and popped from the top of the comma list. (The left of a comma list is the top, for performance reasons.) The stack functions for comma lists are not intended to be mixed with the general ordered data functions detailed in the previous section: a comma list should either be used as an ordered data type or as a stack, but not in both ways.

\begin{quote}
\clist_get:NN \clist_get:cn \clist_get:NNTF
\end{quote}

Stores the left-most item from a ⟨clist var⟩ in the ⟨token list variable⟩ without removing it from the ⟨clist var⟩. The ⟨token list variable⟩ is assigned locally. In the non-branching version, if the ⟨clist var⟩ is empty the ⟨token list variable⟩ is set to the marker value \q_no_value.

\begin{quote}
\clist_pop:NN \clist_pop:cn
\end{quote}

Pops the left-most item from a ⟨clist var⟩ into the ⟨token list variable⟩, i.e. removes the item from the comma list and stores it in the ⟨token list variable⟩. Both of the variables are assigned locally.

\begin{quote}
\clist_gpop:NN \clist_gpop:cn
\end{quote}

Pops the left-most item from a ⟨clist var⟩ into the ⟨token list variable⟩, i.e. removes the item from the comma list and stores it in the ⟨token list variable⟩. The ⟨clist var⟩ is modified globally, while the assignment of the ⟨token list variable⟩ is local.

\begin{quote}
\clist_pop:NNTF \clist_pop:cnTF
\end{quote}

If the ⟨clist var⟩ is empty, leaves the ⟨false code⟩ in the input stream. The value of the ⟨token list variable⟩ is not defined in this case and should not be relied upon. If the ⟨clist var⟩ is non-empty, pops the top item from the ⟨clist var⟩ in the ⟨token list variable⟩, i.e. removes the item from the ⟨clist var⟩. Both the ⟨clist var⟩ and the ⟨token list variable⟩ are assigned locally.
If the ⟨clist var⟩ is empty, leaves the ⟨false code⟩ in the input stream. The value of the ⟨token list variable⟩ is not defined in this case and should not be relied upon. If the ⟨clist var⟩ is non-empty, pops the top item from the ⟨clist var⟩ in the ⟨token list variable⟩, i.e. removes the item from the ⟨clist var⟩. The ⟨clist var⟩ is modified globally, while the ⟨token list variable⟩ is assigned locally.

\clist_push:Nn ⟨clist var⟩ {⟨items⟩}
\clist_push:Nn ⟨clist var⟩ {⟨int expr⟩}
\clist_item:Nn ⟨clist var⟩ {⟨int expr⟩}
\clist_rand_item:N ⟨clist var⟩ ⟨comma list⟩
\clist_show:N ⟨clist var⟩

23.8 Using a single item

Indexing items in the ⟨clist var⟩ from 1 at the top (left), this function evaluates the ⟨int expr⟩ and leaves the appropriate item from the comma list in the input stream. If the ⟨int expr⟩ is negative, indexing occurs from the bottom (right) of the comma list. When the ⟨int expr⟩ is larger than the number of items in the ⟨clist var⟩ (as calculated by \clist_count:N) then the function expands to nothing.

\TeXhacksnote: The result is returned within the \unexpanded primitive (\exp_not:n), which means that the ⟨item⟩ does not expand further when appearing in an e-type or x-type argument expansion.

Selects a pseudo-random item of the ⟨clist var⟩⟨comma list⟩. If the ⟨comma list⟩ has no item, the result is empty.

\TeXhacksnote: The result is returned within the \unexpanded primitive (\exp_not:n), which means that the ⟨item⟩ does not expand further when appearing in an e-type or x-type argument expansion.

23.9 Viewing comma lists

Displays the entries in the ⟨clist var⟩ in the terminal.
23.10 Constant and scratch comma lists

Constant that is always empty.

Scratch comma lists for local assignment. These are never used by the kernel code, and so are safe for use with any \LaTeX3-defined function. However, they may be overwritten by other non-kernel code and so should only be used for short-term storage.

Scratch comma lists for global assignment. These are never used by the kernel code, and so are safe for use with any \LaTeX3-defined function. However, they may be overwritten by other non-kernel code and so should only be used for short-term storage.
Chapter 24

The \texttt{l3token} module

Token manipulation

This module deals with tokens. Now this is perhaps not the most precise description so let's try with a better description: When programming in \TeX, it is often desirable to know just what a certain token is: is it a control sequence or something else. Similarly one often needs to know if a control sequence is expandable or not, a macro or a primitive, how many arguments it takes etc. Another thing of great importance (especially when it comes to document commands) is looking ahead in the token stream to see if a certain character is present and maybe even remove it or disregard other tokens while scanning. This module provides functions for both and as such has two primary function categories: \texttt{\token} for anything that deals with tokens and \texttt{\peek} for looking ahead in the token stream.

Most functions we describe here can be used on control sequences, as those are tokens as well.

It is important to distinguish two aspects of a token: its “shape” (for lack of a better word), which affects the matching of delimited arguments and the comparison of token lists containing this token, and its “meaning”, which affects whether the token expands or what operation it performs. One can have tokens of different shapes with the same meaning, but not the converse.

For instance, \texttt{\if:w}, \texttt{\if_charcode:w}, and \texttt{\tex_if:D} are three names for the same internal operation of \TeX, namely the primitive testing the next two characters for equality of their character code. They have the same meaning hence behave identically in many situations. However, \TeX distinguishes them when searching for a delimited argument. Namely, the example function \texttt{\show_until_if:w} defined below takes everything until \texttt{\if:w} as an argument, despite the presence of other copies of \texttt{\if:w} under different names.

\begin{verbatim}
cs_new:Npn \show_until_if:w #1 \if:w { \tl_show:n {#1} } \show_until_if:w \tex_if:D \if_charcode:w \if:w
\end{verbatim}

A list of all possible shapes and a list of all possible meanings are given in section \ref{sec:l3token}.

195
24.1 Creating character tokens

\char_set_active_eq:NN \char_set_active_eq:Nc \char_gset_active_eq:NN \char_gset_active_eq:Nc
\char_set_active_eq:nN \char_set_active_eq:nc \char_gset_active_eq:nN \char_gset_active_eq:nc
\char_generate:nn {\langle char\rangle} {\langle function\rangle}
\char_generate:nn {\langle charcode\rangle} {\langle catcode\rangle}

Sets the behaviour of the \langle char\rangle in situations where it is active (category code 13) to be equivalent to that of the definition of the \langle function\rangle at the time \char_set_active_eq:NN is used. The category code of the \langle char\rangle is unchanged by this process. The \langle function\rangle may itself be an active character.

\char_generate:nn * \char_generate:nn \{\langle charcode\rangle\} \{\langle catcode\rangle\}

Generates a character token of the given \langle charcode\rangle and \langle catcode\rangle (both of which may be integer expressions). The \langle catcode\rangle may be one of

- 1 (begin group)
- 2 (end group)
- 3 (math toggle)
- 4 (alignment)
- 6 (parameter)
- 7 (math superscript)
- 8 (math subscript)
- 10 (space)
- 11 (letter)
- 12 (other)
- 13 (active)

and other values raise an error. The \langle charcode\rangle may be any one valid for the engine in use, except that for \langle catcode\rangle 10, \langle charcode\rangle 0 is not allowed. Active characters cannot be generated in older versions of Xe\TeX. Another way to build token lists with unusual category codes is \regex_replace:nnN {.*} {\langle replacement\rangle} {\tl var}.

\TeXhacks{Exactly two expansions are needed to produce the character.}

\c_catcode_active_space_tl \c_catcode_active_space_tl
Token list containing one character with category code 13, (“active”), and character code 32 (space).
24.2 Manipulating and interrogating character tokens

Sets the category code of the \texttt{character} to that indicated in the function name. Depending on the current category code of the \texttt{token} the escape token may also be needed:

\begin{verbatim}
\char_set_catcode_other:N \%
\end{verbatim}

The assignment is local.
Sets the category code of the \langle character \rangle which has character code as given by the \langle integer expression \rangle. This version can be used to set up characters which cannot otherwise be given (cf. the N-type variants). The assignment is local.

These functions set the category code of the \langle character \rangle which has character code as given by the \langle integer expression \rangle. The first \langle integer expression \rangle is the character code and the second is the category code to apply. The setting applies within the current \TeX group. In general, the symbolic functions \char_set_catcode\_\langle type \rangle should be preferred, but there are cases where these lower-level functions may be useful.

Expands to the current category code of the \langle character \rangle with character code given by the \langle integer expression \rangle.

Displays the current category code of the \langle character \rangle with character code given by the \langle integer expression \rangle on the terminal.

Sets up the behaviour of the \langle character \rangle when found inside \text_lowercase:n, such that \langle character_1 \rangle will be converted into \langle character_2 \rangle. The two \langle characters \rangle may be specified using an \langle integer expression \rangle for the character code concerned. This may include the \TeX \textbackslash 'character' method for converting a single character into its character code:

\char_set_lccode:nn { '\A ' } { '\a ' } % Standard behaviour
\char_set_lccode:nn { '\A ' } { '\a + 32 ' }
\char_set_lccode:nn { 50 } { 60 }

The setting applies within the current \TeX group.
\char_value_lccode:n \char_value_lccode:n \((\text{integer expression})\)

Expands to the current lower case code of the \textit{(character)} with character code given by the \textit{(integer expression)}.

\char_show_value_lccode:n \char_show_value_lccode:n \((\text{integer expression})\)

Displays the current lower case code of the \textit{(character)} with character code given by the \textit{(integer expression)} on the terminal.

\char_set_uccode:nn \char_set_uccode:nn \((\text{int expr}_1)\) \((\text{int expr}_2)\)

Sets up the behaviour of the \textit{(character)} when found inside \texttt{\textit{text_uppercase:n}}, such that \textit{(character}_1) will be converted into \textit{(character}_2). The two \textit{(characters)} may be specified using an \textit{(integer expression)} for the character code concerned. This may include the \TeX \texttt{\textit{\textbackslash character}} method for converting a single character into its character code:

\begin{verbatim}
\char_set_uccode:nn { \texttt{\char\char\char} } { \texttt{\char\char\char - 32} }
\char_set_uccode:nn { 60 } { 50 }
\end{verbatim}

The setting applies within the current \TeX group.

\char_value_uccode:n \char_value_uccode:n \((\text{integer expression})\)

Expands to the current upper case code of the \textit{(character)} with character code given by the \textit{(integer expression)}.

\char_show_value_uccode:n \char_show_value_uccode:n \((\text{integer expression})\)

Displays the current upper case code of the \textit{(character)} with character code given by the \textit{(integer expression)} on the terminal.

\char_set_mathcode:nn \char_set_mathcode:nn \((\text{int expr}_1)\) \((\text{int expr}_2)\)

This function sets up the math code of \textit{(character)}. The \textit{(character)} is specified as an \textit{(integer expression)} which will be used as the character code of the relevant character. The setting applies within the current \TeX group.

\char_value_mathcode:n \char_value_mathcode:n \((\text{integer expression})\)

Expands to the current math code of the \textit{(character)} with character code given by the \textit{(integer expression)}.

\char_show_value_mathcode:n \char_show_value_mathcode:n \((\text{integer expression})\)

Displays the current math code of the \textit{(character)} with character code given by the \textit{(integer expression)} on the terminal.

\char_set_sfcode:nn \char_set_sfcode:nn \((\text{int expr}_1)\) \((\text{int expr}_2)\)

This function sets up the space factor for the \textit{(character)}. The \textit{(character)} is specified as an \textit{(integer expression)} which will be used as the character code of the relevant character. The setting applies within the current \TeX group.
\char_value_sfcode:n \char_value_sfcode:n \{\text{integer expression}\}

Expands to the current space factor for the \text{character} with character code given by the \text{integer expression}.

\char_show_value_sfcode:n \char_show_value_sfcode:n \{\text{integer expression}\}

Displays the current space factor for the \text{character} with character code given by the \text{integer expression} on the terminal.

\l_char_active_seq
\begin{verbatim}
New: 2012-01-23
Updated: 2015-11-11
\end{verbatim}

Used to track which tokens may require special handling at the document level as they are (or have been at some point) of category \text{active} (catcode 13). Each entry in the sequence consists of a single escaped token, for example \text{"\textbackslash"}. Active tokens should be added to the sequence when they are defined for general document use.

\l_char_special_seq
\begin{verbatim}
New: 2012-01-23
Updated: 2015-11-11
\end{verbatim}

Used to track which tokens will require special handling when working with verbatim-like material at the document level as they are not of categories \text{letter} (catcode 11) or \text{other} (catcode 12). Each entry in the sequence consists of a single escaped token, for example \text{"\textbackslash\textbackslash"} for the backslash or \text{"\textbackslash\{}" for an opening brace. Escaped tokens should be added to the sequence when they are defined for general document use.

### 24.3 Generic tokens

These are implicit tokens which have the category code described by their name. They are used internally for test purposes but are also available to the programmer for other uses.

\text{T\TeXhackers note:} The tokens \text{\c_group_begin_token}, \text{\c_group_end_token}, and \text{\c_\text{-space_token}} are expl3 counterparts of \LaTeXe’s \text{\bgroup}, \text{\egroup}, and \text{\@sptoken}.

\c_group_begin_token \c_group_end_token \c_math_toggle_token \c_alignment_token \c_parameter_token \c_math_superscript_token \c_math_subscript_token \c_space_token

These are implicit tokens which have the category code described by their name. They are used internally for test purposes and should not be used other than for category code tests.

\c_catcode_letter_token \c_catcode_other_token

A token list containing an active token. This is used internally for test purposes and should not be used other than in appropriately-constructed category code tests.
24.4 Converting tokens

\token_to_meaning:N \token_to_meaning:c

Inserts the current meaning of the \token into the input stream as a series of characters of category code 12 (other). This is the primitive \TeX description of the \token, thus for example both functions defined by \cs_set_nopar:Npn and token list variables defined using \tl_new:N are described as macros.

\TeXhackers note: This is the \TeX primitive \meaning. The \token can thus be an explicit space token or an explicit begin-group or end-group character token (\{ or \} when normal \TeX category codes apply) even though these are not valid \N-type arguments.

\token_to_str:N \token_to_str:c

Converts the given \token into a series of characters with category code 12 (other). If the \token is a control sequence, this will start with the current escape character with category code 12 (the escape character is part of the \token). This function requires only a single expansion.

\TeXhackers note: \token_to_str:N is the \TeX primitive \string. The \token can thus be an explicit space tokens or an explicit begin-group or end-group character token (\{ or \} when normal \TeX category codes apply) even though these are not valid \N-type arguments.

\token_to_catcode:N

Converts the given \token into a number describing its category code. If \token is a control sequence this expands to 16. This can’t detect the categories 0 (escape character), 5 (end of line), 9 (ignored character), 14 (comment character), or 15 (invalid character). Control sequences or active characters let to a token of one of the detectable category codes will yield that category.

24.5 Token conditionals

\token_if_group_begin_p:N \token_if_group_begin:NTF

Tests if \token has the category code of a begin group token (\{ when normal \TeX category codes are in force). Note that an explicit begin group token cannot be tested in this way, as it is not a valid \N-type argument.

\token_if_group_end_p:N \token_if_group_end:NTF

Tests if \token has the category code of an end group token (\} when normal \TeX category codes are in force). Note that an explicit end group token cannot be tested in this way, as it is not a valid \N-type argument.
\token_if_math_toggle_p:N = \token_if_math_toggle_p:N \{token\}
\token_if_math_toggle:NTF = \token_if_math_toggle:NTF \{token\} \{true code\} \{false code\}

Tests if \{token\} has the category code of a math shift token ($ when normal \TeX{} category codes are in force).

\token_if_alignment_p:N = \token_if_alignment_p:N \{token\}
\token_if_alignment:NTF = \token_if_alignment:NTF \{token\} \{true code\} \{false code\}

Tests if \{token\} has the category code of an alignment token (& when normal \TeX{} category codes are in force).

\token_if_parameter_p:N = \token_if_parameter_p:N \{token\}
\token_if_parameter:NTF = \token_if_parameter:NTF \{token\} \{true code\} \{false code\}

Tests if \{token\} has the category code of a macro parameter token (# when normal \TeX{} category codes are in force).

\token_if_math_superscript_p:N = \token_if_math_superscript_p:N \{token\}
\token_if_math_superscript:NTF = \token_if_math_superscript:NTF \{token\} \{true code\} \{false code\}

Tests if \{token\} has the category code of a superscript token (^ when normal \TeX{} category codes are in force).

\token_if_math_subscript_p:N = \token_if_math_subscript_p:N \{token\}
\token_if_math_subscript:NTF = \token_if_math_subscript:NTF \{token\} \{true code\} \{false code\}

Tests if \{token\} has the category code of a subscript token (_ when normal \TeX{} category codes are in force).

\token_if_space_p:N = \token_if_space_p:N \{token\}
\token_if_space:NTF = \token_if_space:NTF \{token\} \{true code\} \{false code\}

Tests if \{token\} has the category code of a space token. Note that an explicit space token with character code 32 cannot be tested in this way, as it is not a valid \TeX{}-type argument.

\token_if_letter_p:N = \token_if_letter_p:N \{token\}
\token_if_letter:NTF = \token_if_letter:NTF \{token\} \{true code\} \{false code\}

Tests if \{token\} has the category code of a letter token.

\token_if_other_p:N = \token_if_other_p:N \{token\}
\token_if_other:NTF = \token_if_other:NTF \{token\} \{true code\} \{false code\}

Tests if \{token\} has the category code of an “other” token.

\token_if_active_p:N = \token_if_active_p:N \{token\}
\token_if_active:NTF = \token_if_active:NTF \{token\} \{true code\} \{false code\}

Tests if \{token\} has the category code of an active character.

\token_if_eq_catcode_p:NN = \token_if_eq_catcode_p:NN \{token_1\} \{token_2\}
\token_if_eq_catcode:NN = \token_if_eq_catcode:NN \{token_1\} \{token_2\} \{true code\} \{false code\}

Tests if the two \{tokens\} have the same category code.

202
\token_if_eq_charcode_p:NN \ token_if_eq_charcode_p:NN \ token1 \ token2 \{true code\} \{false code\}

Tests if the two \textit{tokens} have the same character code.

\token_if_eq_meaning_p:NN \ token_if_eq_meaning_p:NN \ token1 \ token2 \{true code\} \{false code\}

Tests if the two \textit{tokens} have the same meaning when expanded.

\token_if_macro_p:N \ token_if_macro_p:N \ token \{true code\} \{false code\}

Tests if the \textit{token} is a \TeX\ macro.

\token_if_cs_p:N \ token_if_cs_p:N \ token \{true code\} \{false code\}

Tests if the \textit{token} is a control sequence.

\token_if_expandable_p:N \ token_if_expandable_p:N \ token \{true code\} \{false code\}

Tests if the \textit{token} is expandable. This test returns \textit{false} for an undefined token.

\token_if_long_macro_p:N \ token_if_long_macro_p:N \ token \{true code\} \{false code\}

Tests if the \textit{token} is a long macro.

\token_if_protected_macro_p:N \ token_if_protected_macro_p:N \ token \{true code\} \{false code\}

Tests if the \textit{token} is a protected macro: for a macro which is both protected and long this returns \textit{false}.

\token_if_protected_long_macro_p:N \ token_if_protected_long_macro_p:N \ token \{true code\} \{false code\}

Tests if the \textit{token} is a protected long macro.

\token_if_chardef_p:N \ token_if_chardef_p:N \ token \{true code\} \{false code\}

Tests if the \textit{token} is defined to be a chardef.

\textbf{\TeX\ hackers note:} Booleans, boxes and small integer constants are implemented as \chardefs.
Tests if the \(\langle\text{token}\rangle\) is defined to be a mathchardef.

Tests if the \(\langle\text{token}\rangle\) is defined to be a font selection command.

Tests if the \(\langle\text{token}\rangle\) is defined to be a dimension register.

Tests if the \(\langle\text{token}\rangle\) is defined to be a integer register.

Tests if the \(\langle\text{token}\rangle\) is defined to be a muskip register.

Tests if the \(\langle\text{token}\rangle\) is defined to be a skip register.

Tests if the \(\langle\text{token}\rangle\) is defined to be a toks register (not used by \LaTeX3).

Tests if the \(\langle\text{token}\rangle\) is an engine primitive. In \LaTeX\ this includes primitive-like commands defined using \texttt{token.set_lua}.

	exttt{TEXhackers note}: Constant integers may be implemented as integer registers, \texttt{chardefs}, or \texttt{mathchardefs} depending on their value.
This function compares the \textit{test token} in turn with each of the \textit{token cases}. If the two are equal (as described for \texttt{token_if_eq_catcode:NNTF}, \texttt{token_if_eq_charcode:NNTF} and \texttt{token_if_eq_meaning:NNTF}, respectively) then the associated \textit{code} is left in the input stream and other cases are discarded. If any of the cases are matched, the \textit{true code} is also inserted into the input stream (after the code for the appropriate case), while if none match then the \textit{false code} is inserted. The functions \texttt{token_case_catcode:Nn}, \texttt{token_case_charcode:Nn}, and \texttt{token_case_meaning:Nn}, which do nothing if there is no match, are also available.

### 24.6 Peeking ahead at the next token

There is often a need to look ahead at the next token in the input stream while leaving it in place. This is handled using the “peek” functions. The generic \texttt{peek_after:Nw} is provided along with a family of predefined tests for common cases. Peeking ahead does not skip spaces: rather, \texttt{peek_remove_spaces:n} should be used. In addition, using \texttt{peek_analysis_map_inline:n}, one can map through the following tokens in the input stream and repeatedly perform some tests.

\begin{itemize}
    \item \texttt{\textbackslash peek_after:Nw (function) \{token\}}
        
        Locally sets the test variable \texttt{l_peek_token} equal to \texttt{\{token\}} (as an implicit token, \textit{not} as a token list), and then expands the \texttt{(function)}. The \texttt{\{token\}} remains in the input stream as the next item after the \texttt{(function)}. The \texttt{\{token\}} here may be \texttt{\{code\}} (assuming normal \TeX{} category codes), \textit{i.e.} it is not necessarily the next argument which would be grabbed by a normal function.

    \item \texttt{\textbackslash peek_gafter:Nw (function) \{token\}}
        
        Globally sets the test variable \texttt{g_peek_token} equal to \texttt{\{token\}} (as an implicit token, \textit{not} as a token list), and then expands the \texttt{(function)}. The \texttt{\{token\}} remains in the input stream as the next item after the \texttt{(function)}. The \texttt{\{token\}} here may be \texttt{\{code\}} (assuming normal \TeX{} category codes), \textit{i.e.} it is not necessarily the next argument which would be grabbed by a normal function.
\end{itemize}

\texttt{l_peek_token} Token set by \texttt{peek_after:Nw} and available for testing as described above.

\texttt{g_peek_token} Token set by \texttt{peek_gafter:Nw} and available for testing as described above.
Tests if the next \{token\} in the input stream has the same category code as the \{test token\} (as defined by the test \texttt{token_if_eq_catcode:NNTF}). Spaces are respected by the test and the \{token\} is left in the input stream after the \{true code\} or \{false code\} (as appropriate to the result of the test).

Tests if the next \{token\} in the input stream has the same character code as the \{test token\} (as defined by the test \texttt{token_if_eq_charcode:NNTF}). Spaces are respected by the test and the \{token\} is removed from the input stream if the test is true. The function then places either the \{true code\} or \{false code\} in the input stream (as appropriate to the result of the test).

Tests if the next \{token\} in the input stream has the same meaning as the \{test token\} (as defined by the test \texttt{token_if_eq_meaning:NNTF}). Spaces are respected by the test and the \{token\} is left in the input stream after the \{true code\} or \{false code\} (as appropriate to the result of the test).

Peeks ahead and detect if the following token is a space (category code 10 and character code 32). If so, removes the token and checks the next token. Once a non-space token is found, the \{code\} will be inserted into the input stream. Typically this will contain a \texttt{peek} operation, but this is not required.
\peek_remove_filler:n \peek_remove_filler:n \langle code \rangle

Peeks ahead and detect if the following token is a space (category code 10) or has meaning equal to \scan_stop:. If so, removes the token and checks the next token. If neither of these cases apply, expands the next token using f-type expansion, then checks the resulting leading token in the same way. If after expansion the next token is neither of the two test cases, the \langle code \rangle will be inserted into the input stream. Typically this will contain a peek operation, but this is not required.

\TeXhackers note: This is essentially a macro-based implementation of how \TeX handles the search for a left brace after for example \everypar, except that any non-expandable token cleanly ends the \langle filler \rangle (i.e. it does not lead to a \TeX error).

In contrast to \TeX's filler removal, a construct \exp_not:N \foo will be treated in the same way as \foo.

\peek_N_type:TF \peek_N_type:TF \langle\langle true code\rangle\rangle \langle\langle false code\rangle\rangle

Tests if the next \langle token \rangle in the input stream can be safely grabbed as an N-type argument. The test is \langle false \rangle if the next \langle token \rangle is either an explicit or implicit begin-group or end-group token (with any character code), or an explicit or implicit space character (with character code 32 and category code 10), or an outer token (never used in \LaTeX3) and \langle true \rangle in all other cases. Note that a \langle true \rangle result ensures that the next \langle token \rangle is a valid N-type argument. However, if the next \langle token \rangle is for instance \c_space_token, the test takes the \langle false \rangle branch, even though the next \langle token \rangle is in fact a valid N-type argument. The \langle token \rangle is left in the input stream after the \langle true code \rangle or \langle false code \rangle (as appropriate to the result of the test).
\peek_analysis_map_inline:n \peek_analysis_map_inline:n \{\textit{inline function}\}\}\\
\l\texttt{\textbackslash peek_analysis_map_inline:n} \{\textit{inline function}\}\}

Repeatedly removes one \textit{token} from the input stream and applies the \textit{inline function} to it, until \texttt{\textbackslash peek_analysis_map_break:} is called. The \textit{inline function} receives three arguments for each \textit{token} in the input stream:

- \texttt{(tokens)}, which both o-expand and e/x-expand to the \textit{token}. The detailed form of \texttt{(tokens)} may change in later releases.
- \texttt{(char code)}, a decimal representation of the character code of the \textit{token}, $-1$ if it is a control sequence.
- \texttt{(catcode)}, a capital hexadecimal digit which denotes the category code of the \textit{token} (0: control sequence, 1: begin-group, 2: end-group, 3: math shift, 4: alignment tab, 6: parameter, 7: superscript, 8: subscript, A: space, B: letter, C: other, D: active). This can be converted to an integer by writing "\texttt{(catcode)}".

These arguments are the same as for \texttt{\textbackslash \texttt{tl_analysis_map_inline:nn}} defined in \texttt{\texttt{l3tl-analysis}}. The \texttt{(char code)} and \texttt{(catcode)} do not take the meaning of a control sequence or active character into account: for instance, upon encountering the token \texttt{\textbackslash c\_group\_begin\_token} in the input stream, \texttt{\textbackslash peek_analysis_map_inline:n} calls the \textit{inline function} with \#1 being \texttt{\textbackslash exp_not:n} \{\texttt{\textbackslash c\_group\_begin\_token}\} (with the current implementation), \#2 being $-1$, and \#3 being 0, as for any other control sequence. In contrast, upon encountering an explicit begin-group token \{, the \textit{inline function} is called with arguments \texttt{\textbackslash exp_after:wN} \{\texttt{\if\_false: \fi:} \} 123 and 1.

The mapping is done at the current group level, \textit{i.e.} any local assignments made by the \textit{inline function} remain in effect after the loop. Within the code, \texttt{\textbackslash l\_peek\_token} is set equal (as a token, not a token list) to the token under consideration.

Peek functions cannot be used within this mapping function (nor other mapping functions) since the input stream contains trailing material necessary for the functioning of the loop.

\textbf{\texttt{\texttt{\LaTeX\textbackslash hackers}}} note: In case the input stream has not yet been tokenized (converted from characters to tokens), characters are tokenized one by one as needed by \texttt{\textbackslash peek_analysis_map-\_inline:n} using the current category code regime.

\texttt{\textbackslash peek_analysis_map_break:} \texttt{\textbackslash peek_analysis_map_break:n} \{\ldots \texttt{\textbackslash peek_analysis_map_break:n} \{\texttt{(code)}\}\}\\
\texttt{\textbackslash peek_analysis_map_break:} \texttt{\textbackslash peek_analysis_map_break:n} \{\ldots \texttt{\textbackslash peek_analysis_map_break:n} \{\texttt{(code)}\}\}

\texttt{\textbackslash peek_analysis_map_break:} \texttt{\textbackslash peek_analysis_map_break:n} \{\ldots \texttt{\textbackslash peek_analysis_map_break:n} \{\texttt{(code)}\}\}

\texttt{\textbackslash peek_analysis_map_break:} \texttt{\textbackslash peek_analysis_map_break:n} \{\ldots \texttt{\textbackslash peek_analysis_map_break:n} \{\texttt{(code)}\}\}

\texttt{\textbackslash \texttt{\textbackslash peek_analysis_map_break:}} stops the \texttt{\textbackslash peek_analysis_map_inline:n} loop from seeking more tokens, and inserts \texttt{(code)} in the input stream (empty for \texttt{\textbackslash peek_analysis_map_break:}).
\peek_regex:nTF \peek_regex:NTF  Tests if the \langle tokens \rangle that follow in the input stream match the \langle regular expression \rangle. Any \langle tokens \rangle that have been read are left in the input stream after the \langle true code \rangle or \langle false code \rangle (as appropriate to the result of the test). See \texttt{l3regex} for documentation of the syntax of regular expressions. The \langle regular expression \rangle is implicitly anchored at the start, so for instance \texttt{\peek_regex:nTF \{ a \}} is essentially equivalent to \texttt{\peek_\charcode:NTF a}.

\textbf{TeXhackers note:} Implicit character tokens are correctly considered by \texttt{\peek_regex:nTF} as control sequences, while functions that inspect individual tokens (for instance \texttt{\peek_\charcode:NTF}) only take into account their meaning.

The \texttt{\peek_regex:nTF} function only inspects as few tokens as necessary to determine whether the regular expression matches. For instance \texttt{\peek_regex:nTF \{ abc | [a-z] \} \{ \} \{ \} abc} will only inspect the first token \texttt{a} even though the first branch \texttt{abc} of the alternative is preferred in functions such as \texttt{\peek_regex_remove_once:nTF}. This may have an effect on tokenization if the input stream has not yet been tokenized and category codes are changed.

\peek_regex_remove_once:nTF \peek_regex_remove_once:NTF  Tests if the \langle tokens \rangle that follow in the input stream match the \langle regex \rangle. If the test is true, the \langle tokens \rangle are removed from the input stream and the \langle true code \rangle is inserted, while if the test is false, the \langle false code \rangle is inserted followed by the \langle tokens \rangle that were originally in the input stream. See \texttt{l3regex} for documentation of the syntax of regular expressions. The \langle regular expression \rangle is implicitly anchored at the start, so for instance \texttt{\peek_regex_remove_once:nTF \{ a \}} is essentially equivalent to \texttt{\peek_\charcode_remove:NTF a}.

\textbf{TeXhackers note:} Implicit character tokens are correctly considered by \texttt{\peek_regex_remove_once:nTF} as control sequences, while functions that inspect individual tokens (for instance \texttt{\peek_\charcode:NTF}) only take into account their meaning.
If the \{tokens\} that follow in the input stream match the \{regex\}, replaces them according to the \{replacement\} as for \regex_replace_once:nnN, and leaves the result in the input stream, after the \{true code\}. Otherwise, leaves \{false code\} followed by the \{tokens\} that were originally in the input stream, with no modifications. See \l3regex for documentation of the syntax of regular expressions and of the \{replacement\}: for instance \0 in the \{replacement\} is replaced by the tokens that were matched in the input stream. The \{regular expression\} is implicitly anchored at the start. In contrast to \regex_replace_once:nnN, no error arises if the \{replacement\} leads to an unbalanced token list: the tokens are inserted into the input stream without issue.

**\TeXhackers note:** Implicit character tokens are correctly considered by \peek_regex_replace_once:nnTF as control sequences, while functions that inspect individual tokens (for instance \peek_charcode:NTF) only take into account their meaning.

## 24.7 Description of all possible tokens

Let us end by reviewing every case that a given token can fall into. This section is quite technical and some details are only meant for completeness. We distinguish the meaning of the token, which controls the expansion of the token and its effect on \TeX’s state, and its shape, which is used when comparing token lists such as for delimited arguments. Two tokens of the same shape must have the same meaning, but the converse does not hold.

A token has one of the following shapes.

- A control sequence, characterized by the sequence of characters that constitute its name: for instance, \use:n is a five-letter control sequence.

- An active character token, characterized by its character code (between 0 and 1114111 for \LuaLaTeX{} and \XeLaTeX{} and less for other engines) and category code 13.

- A character token, characterized by its character code and category code (one of 1, 2, 3, 4, 6, 7, 8, 10, 11 or 12 whose meaning is described below).

There are also a few internal tokens. The following list may be incomplete in some engines.

- Expanding \the:font results in a token that looks identical to the command that was used to select the current font (such as \textit\font) but it differs from it in shape.

- A “frozen” \relax, which differs from the primitive in shape (but has the same meaning), is inserted when the closing \texti of a conditional is encountered before the conditional is evaluated.

- Expanding \noexpand \texti (when the \texti is expandable) results in an internal token, displayed (temporarily) as \notexpanded: \texti, whose shape coincides with the \texti and whose meaning differs from \relax.
An \outer endtemplate: can be encountered when peeking ahead at the next token; this expands to another internal token, end of alignment template.

Tricky programming might access a frozen \endwrite.

Some frozen tokens can only be accessed in interactive sessions: \cr, \right, \endgroup, \fi, \inaccessible.

In LuaTeX, there is also the strange case of “bytes” \xy where x,y are any two lowercase hexadecimal digits, so that the hexadecimal number ranges from 110000 = 1114112 to 1100ff = 1114367. These are used to output individual bytes to files, rather than UTF-8. For the purposes of token comparisons they behave like non-expandable primitive control sequences (not characters) whose \meaning is the character followed by the given byte. If this byte is in the range 80–ff this gives an “invalid utf-8 sequence” error: applying \token_to_str:N or \token_to_meaning:N to these tokens is unsafe. Unfortunately, they don’t seem to be detectable safely by any means except perhaps Lua code.

The meaning of a (non-active) character token is fixed by its category code (and character code) and cannot be changed. We call these tokens explicit character tokens. Category codes that a character token can have are listed below by giving a sample output of the \TeX primitive \meaning, together with their \LaTeX3 names and most common example:

1 begin-group character (\group_begin, often {),
2 end-group character (\group_end, often }),
3 math shift character (\math_toggle, often $),
4 alignment tab character (\alignment, often &),
6 macro parameter character (\parameter, often #),
7 superscript character (\math_superscript, often ^),
8 subscript character (\math_subscript, often _),
10 blank space (\space, often character code 32),
11 the letter (\letter, such as A),
12 the character (\other, such as 0).

Category code 13 (active) is discussed below. Input characters can also have several other category codes which do not lead to character tokens for later processing: 0 (escape), 5 (end_line), 9 (ignore), 14 (comment), and 15 (invalid).

The meaning of a control sequence or active character can be identical to that of any character token listed above (with any character code), and we call such tokens implicit character tokens. The meaning is otherwise in the following list:

- a macro, used in \LaTeX for most functions and some variables (tl, fp, seq, ...),
- a primitive such as \def or \topmark, used in \LaTeX for some functions,
- a register such as \count123, used in \LaTeX for the implementation of some variables (int, dim, ...),
• a constant integer such as \texttt{char"56} or \texttt{mathchar"121},

• a font selection command,

• undefined.

Macros can be \texttt{\textbackslash protected} or not, \texttt{\textbackslash long} or not (the opposite of what \LaTeX3 calls \texttt{nopar}), and \texttt{\textbackslash outer} or not (unused in \LaTeX3). Their \texttt{\textbackslash meaning} takes the form

\begin{verbatim}
⟨prefix⟩ macro:(argument)->⟨replacement⟩
\end{verbatim}

where \(⟨prefix⟩\) is among \texttt{\textbackslash protected\textbackslash long\textbackslash outer}, \(⟨argument⟩\) describes parameters that the macro expects, such as \#1#2#3, and \(⟨replacement⟩\) describes how the parameters are manipulated, such as \texttt{\texttt{\textbackslash int_eval:n(#2+#1*#3)}}.

Now is perhaps a good time to mention some subtleties relating to tokens with category code 10 (space). Any input character with this category code (normally, space and tab characters) becomes a normal space, with character code 32 and category code 10.

When a macro takes an undelimited argument, explicit space characters (with character code 32 and category code 10) are ignored. If the following token is an explicit character token with category code 1 (begin-group) and an arbitrary character code, then \LaTeX scans ahead to obtain an equal number of explicit character tokens with category code 1 (begin-group) and 2 (end-group), and the resulting list of tokens (with outer braces removed) becomes the argument. Otherwise, a single token is taken as the argument for the macro: we call such single tokens “N-type”, as they are suitable to be used as an argument for a function with the signature :N.

When a macro takes a delimited argument \LaTeX scans ahead until finding the delimiter (outside any pairs of begin-group/end-group explicit characters), and the resulting list of tokens (with outer braces removed) becomes the argument. Note that explicit space characters at the start of the argument are \textit{not} ignored in this case (and they prevent brace-stripping).
Chapter 25

The l3prop module

Property lists

expl3 implements a “property list” data type, which contain an unordered list of entries each of which consists of a ⟨key⟩ (string) and an associated ⟨value⟩ (token list). The ⟨key⟩ and ⟨value⟩ may both be given as any balanced text, and the ⟨key⟩ is processed using \t1_to_str:n, meaning that category codes are ignored. Entries can be manipulated individually, as well as collectively by applying a function to every key–value pair within the list.

Each entry in a property list must have a unique ⟨key⟩: if an entry is added to a property list which already contains the ⟨key⟩ then the new entry overwrites the existing one. The ⟨keys⟩ are compared on a string basis, using the same method as \str_if_eq:nnTF.

Property lists are intended for storing key-based information for use within code. They can be converted from and to key–value lists, which are a form of input parsed by the l3keys module. If a key–value list contains a ⟨key⟩ multiple times, only the last ⟨value⟩ associated to it will be kept in the conversion to a property list.

Internally, property lists can use two distinct implementations with different data storage, which are decided when declaring the property list variable using \prop_new:N (“flat” storage) or \prop_new_linked:N (“linked” storage). After a property list is declared with \prop_new:N or \prop_new_linked:N, the type of internal data storage can be changed by \prop_make_flat:N or \prop_make_linked:N, but only at the outermost group level. All other l3prop functions transparently manipulate either storage method and convert as needed.

- The (default) “flat” storage method is suited for a relatively small number of entries, or when the property list is likely to be manipulated (copied, mapped) as a whole rather than entry-wise. It is significantly faster for \prop_set_eq:NN, and only slightly faster for \prop_clear:N, \prop_concat:NNN, and mapping functions \prop_map: . . . .

- The “linked” storage method is meant for property lists with a large numbers of entries. It takes up more of \TeX’s memory during a run, but is significantly faster (for long lists) when accessing or modifying individual entries using functions such as \prop_if_in:Nn, \prop_item:Nn, \prop_put:Nnn, \prop_get:NnN, \prop_pop:NnN, \prop_remove:Nn, as it takes a constant time for these operations (rather
than the number of items for a “flat” property list). A technical drawback is that memory is permanently used\(^7\) by \texttt{⟨keys⟩} stored in a “linked” property list, even after they are removed and the property list is deleted.

### 25.1 Creating and initialising property lists

\begin{verbatim}
\prop_new:N \prop_new:N \prop_list
\prop_clear:N \prop_clear:c \prop_gclear:N \prop_gclear:c
\prop_set_eq:NN \prop_set_eq: \prop_gset_eq: \prop_gset_eq:
\prop_clear_new:N \prop_clear_new:c \prop_gclear_new:N \prop_gclear_new:c
\prop_clear_new_linked:N \prop_clear_new_linked:c \prop_gclear_new_linked:N \prop_gclear_new_linked:c
\end{verbatim}

- \texttt{\prop_new:N \prop_list} creates a new “flat” \texttt{⟨property list⟩} or raises an error if the name is already taken. The declaration is global. The \texttt{⟨property list⟩} initially contains no entries. See also \texttt{\prop_new_linked:N}.

- \texttt{\prop_new_linked:N \prop_new_linked:c} creates a new “linked” \texttt{⟨property list⟩} or raises an error if the name is already taken. The declaration is global. The \texttt{⟨property list⟩} initially contains no entries. The internal data storage differs from that produced by \texttt{\prop_new:N} and it is optimized for property lists with a large number of entries.

- \texttt{\prop_clear:N \prop_clear:c \prop_gclear:N \prop_gclear:c} clears all entries from the \texttt{⟨property list⟩}.

- \texttt{\prop_set_eq:NN \prop_set_eq: \prop_gset_eq: \prop_gset_eq:} sets the content of \texttt{⟨property list1⟩} equal to that of \texttt{⟨property list2⟩}. This converts as needed between the two storage types.

\textbf{TEXhackers note:} If the property list exists and is of “linked” type, it is cleared but not made into a flat property list.

\textbf{TEXhackers note:} If the property list exists and is of “flat” type, it is cleared but not made into a linked property list.

\(^7\)Until the end of the run, that is.
Sets \textit{property list} to contain key–value pairs given in the second argument. If duplicate keys appear only the last of the values is kept. In contrast to most keyval lists (e.g. those in \texttt{l3keys}), each key here \textit{must} be followed with an = sign even to specify an empty \langle value \rangle.

Spaces are trimmed around every \langle key \rangle and every \langle value \rangle, and if the result of trimming spaces consists of a single brace group then a set of outer braces is removed. This enables both the \langle key \rangle and the \langle value \rangle to contain spaces, commas or equal signs. The \langle key \rangle is then processed by \texttt{tl_to_str:n}. This function correctly detects the = and , signs provided they have the standard category code 12 or they are active.

Creates a new constant “flat” \textit{property list} or raises an error if the name is already taken. The \textit{property list} is set globally to contain key–value pairs given in the second argument, processed in the way described for \texttt{prop_set_from_keyval:Nn}. If duplicate keys appear only the last of the values is kept. This function correctly detects the = and , signs provided they have the standard category code 12 or they are active.

Changes the internal storage type of the \textit{property list} to be the same “flat” storage as \texttt{prop_new:N}. The key–value pairs of the \textit{property list} are preserved by the change. If the property list was already flat then nothing is done. This function can only be used at the outermost group level.

Changes the internal storage type of the \textit{property list} to be the same “linked” storage as \texttt{prop_new_linked:N}. The key–value pairs of the \textit{property list} are preserved by the change. If the property list was already linked then nothing is done. This function can only be used at the outermost group level.


---

25.2 Adding and updating property list entries

\prop_put:Nnn
\prop_put:{\property list}{\key}{\value}

Add an entry to the \textit{property list} which may be accessed using the \textit{key} and which has \textit{value}. If the \textit{key} is already present in the \textit{property list}, the existing entry is overwritten by the new \textit{value}. Both the \textit{key} and \textit{value} may contain any \textit{balanced text}. The \textit{key} is stored after processing with \texttt{\tl_to_str:n}, meaning that category codes are ignored.

\prop_put_if_not_in:Nnn
\prop_put_if_not_in:{\property list}{\key}{\value}

If the \textit{key} is present in the \textit{property list} then no action is taken. Otherwise, a new entry is added as described for \texttt{\prop_put:Nnn}.

\prop_gput:Nnn
\prop_gput:{\property list}{\key}{\value}

Combines the key–value pairs of \textit{property list}_1 and \textit{property list}_2, and saves the result in \textit{property list}_1. If a key appears in both \textit{property list}_2 and \textit{property list}_3 then the last value, namely the value in \textit{property list}_3 is kept. This converts as needed between the two storage types.

---

Rev: 2024-05-07
Updated: 2024-05-07

Revised: 2024-03-30
Updated: 2024-05-07

Rev: 2022-07-09
Updated: 2022-07-09
Updates the ⟨property list⟩ by adding entries for each key–value pair given in the second argument. The addition is done through \prop_put:Nnn, hence if the ⟨property list⟩ already contains some of the keys, the corresponding values are discarded and replaced by those given in the key–value list. If duplicate keys appear in the key–value list then only the last of the values is kept.

The function is equivalent to storing the key–value pairs in a temporary property list using \prop_set_from_keyval:Nn, then combining ⟨property list⟩ with the temporary variable using \prop_concat:NNN. In particular, the ⟨keys⟩ and ⟨values⟩ are space-trimmed and unbraced as described in \prop_set_from_keyval:Nn. This function correctly detects the = and , signs provided they have the standard category code 12 or they are active.

### 25.3 Recovering values from property lists

Recovers the ⟨value⟩ stored with ⟨key⟩ from the ⟨property list⟩, and places this in the ⟨token list variable⟩. If the ⟨key⟩ is not found in the ⟨property list⟩ then the ⟨token list variable⟩ is set to the special marker \q_no_value. The ⟨token list variable⟩ is set within the current \TeX group. See also \prop_get:NnNTF.

Recovers the ⟨value⟩ stored with ⟨key⟩ from the ⟨property list⟩, and places this in the ⟨token list variable⟩. If the ⟨key⟩ is not found in the ⟨property list⟩ then the ⟨token list variable⟩ is set to the special marker \q_no_value. The ⟨key⟩ and ⟨value⟩ are then deleted from the property list. Both assignments are local. See also \prop_pop:NnNTF.

Recovers the ⟨value⟩ stored with ⟨key⟩ from the ⟨property list⟩, and places this in the ⟨token list variable⟩. If the ⟨key⟩ is not found in the ⟨property list⟩ then the ⟨token list variable⟩ is set to the special marker \q_no_value. The ⟨property list⟩ is modified globally, while the assignment of the ⟨token list variable⟩ is local. See also \prop_gpop:NnNTF.
\prop_item:Nn \prop_item:(NV|Ne|No|cn|cV|ce|co) *

New: 2014-07-17

Expands to the \langle value\rangle corresponding to the \langle key\rangle in the \langle property list\rangle. If the \langle key\rangle is missing, this has an empty expansion.

\textbf{TeXhackers note:} For “flat” property lists, this expandable function iterates through every key–value pair and is therefore slower than a non-expandable approach based on \prop_get:NnN. (For “linked” property lists both functions are fast.)

The result is returned within the \unexpanded primitive (\exp_not:n), which means that the \langle value\rangle does not expand further when appearing in an e-type or x-type argument expansion.

\prop_count:N \prop_count:c *

Leaves the number of key–value pairs in the \langle property list\rangle in the input stream as an \langle integer denotation\rangle.

\prop_to_keyval:N \prop_to_keyval:N (property list)

Expands to the \langle property list\rangle in a key–value notation. Keep in mind that a \langle property list\rangle is unordered, while key–value interfaces are not necessarily, so this cannot be used for arbitrary interfaces.

\textbf{TeXhackers note:} The result is returned within the \unexpanded primitive (\exp_not:n), which means that the key–value list does not expand further when appearing in an e-type or x-type argument expansion. It also needs exactly two steps of expansion.

### 25.4 Modifying property lists

\prop_remove:Nn \prop_remove:(NV|Ne|No|cn|cV|ce)
\prop_gremove:Nn \prop_gremove:(NV|Ne|No|cn|cV|ce)

New: 2012-05-12

Removes the entry listed under \langle key\rangle from the \langle property list\rangle. If the \langle key\rangle is not found in the \langle property list\rangle no change occurs, \emph{i.e} there is no need to test for the existence of a key before deleting it.

### 25.5 Property list conditionals

\prop_if_exist_p:N \prop_if_exist_p:c \prop_if_exist:N \prop_if_exist:c

New: 2012-05-03

Tests whether the \langle property list\rangle is currently defined. This does not check that the \langle property list\rangle really is a property list variable.
Tests if the \textit{property list} is empty (containing no entries).

Tests if the \textit{key} is present in the \textit{property list}, making the comparison using the method described by \texttt{str_if_eq:nnTF}.

\textbf{TexHackers note:} For “flat” property lists, this expandable function iterates through every key–value pair and is therefore slower than a non-expandable approach based on \texttt{prop-get:NnNTF}. (For “linked” property lists both functions are fast.)

### 25.6 Recovering values from property lists with branching

The functions in this section combine tests for the presence of a key in a property list with recovery of the associated value. This makes them useful for cases where different code follows depending on the presence or absence of a key in a property list. They offer increased readability and performance over separate testing and recovery phases.

If the \textit{key} is not present in the \textit{property list}, leaves the \textit{false code} in the input stream. The value of the \textit{token list variable} is not defined in this case and should not be relied upon. If the \textit{key} is present in the \textit{property list}, stores the corresponding \textit{value} in the \textit{token list variable} without removing it from the \textit{property list}, then leaves the \textit{true code} in the input stream. The \textit{token list variable} is assigned locally.

If the \textit{key} is not present in the \textit{property list}, leaves the \textit{false code} in the input stream. The value of the \textit{token list variable} is not defined in this case and should not be relied upon. If the \textit{key} is present in the \textit{property list}, pops the corresponding \textit{value} in the \textit{token list variable}, i.e. removes the item from the \textit{property list}. Both the \textit{property list} and the \textit{token list variable} are assigned locally.
If the \texttt{key} is not present in the \texttt{property list}, leaves the \texttt{false code} in the input stream. The value of the \texttt{token list variable} is not defined in this case and should not be relied upon. If the \texttt{key} is present in the \texttt{property list}, pops the corresponding \texttt{value} in the \texttt{token list variable}, i.e. removes the item from the \texttt{property list}. The \texttt{property list} is modified globally, while the \texttt{token list variable} is assigned locally.

### 25.7 Mapping over property lists

All mappings are done at the current group level, \textit{i.e.} any local assignments made by the \texttt{function} or \texttt{code} discussed below remain in effect after the loop.

\begin{verbatim}
\prop_map_function:NN \prop_map_function:cn
\end{verbatim}

\texttt{\prop_map_function:NN} \texttt{(property list) (function)}

Applies \texttt{(function)} to every \texttt{(entry)} stored in the \texttt{property list}. The \texttt{(function)} receives two arguments for each iteration: the \texttt{key} and associated \texttt{value}. The order in which \texttt{(entries)} are returned is not defined and should not be relied upon. To pass further arguments to the \texttt{(function)}, see \texttt{\prop_map_inline:Nn} (non-expandable) or \texttt{\prop_map_tokens:Nn}.

\begin{verbatim}
\prop_map_inline:Nn \prop_map_inline:cn
\end{verbatim}

\texttt{\prop_map_inline:Nn} \texttt{(property list) \{\texttt{inline function}\}}

Applies \texttt{(inline function)} to every \texttt{(entry)} stored within the \texttt{property list}. The \texttt{(inline function)} should consist of code which receives the \texttt{key} as \#1 and the \texttt{value} as \#2. The order in which \texttt{(entries)} are returned is not defined and should not be relied upon.

\begin{verbatim}
\prop_map_tokens:Nn \prop_map_tokens:cn
\end{verbatim}

\texttt{\prop_map_tokens:Nn} \texttt{(property list) \{(code\}\}}

Analogue of \texttt{\prop_map_function:NN} which maps several tokens instead of a single function. The \texttt{(code)} receives each key–value pair in the \texttt{property list} as two trailing brace groups. For instance,

\begin{verbatim}
\prop_map_tokens:Nn \_my_prop \{ \str_if_eq:nnT { mykey } \}
\end{verbatim}

expands to the value corresponding to \texttt{mykey}: for each pair in \texttt{\_my_prop} the function \texttt{\str_if_eq:nnT} receives \texttt{mykey}, the \texttt{(key)} and the \texttt{(value)} as its three arguments. For that specific task, \texttt{\prop_item:Nn} is faster.
Used to terminate a \prop_map_inline: function before all entries in the ⟨property list⟩ have been processed. This normally takes place within a conditional statement, for example

\prop_map_inline:Nn \l_my_prop
{\str_if_eq:nnTF { #1 } { bingo }{ \prop_map_break: }
{ % Do something useful
}%}

Use outside of a \prop_map_inline: scenario leads to low level \TeX{} errors.

\textbf{\TeX{}hackers note:} When the mapping is broken, additional tokens may be inserted before further items are taken from the input stream. This depends on the design of the mapping function.

Used to terminate a \prop_map_inline: function before all entries in the ⟨property list⟩ have been processed, inserting the ⟨code⟩ after the mapping has ended. This normally takes place within a conditional statement, for example

\prop_map_inline:Nn \l_my_prop
{\str_if_eq:nnTF { #1 } { bingo }{ \prop_map_break:n { ⟨code⟩ } }
{ % Do something useful
}%}

Use outside of a \prop_map_inline: scenario leads to low level \TeX{} errors.

\textbf{\TeX{}hackers note:} When the mapping is broken, additional tokens may be inserted before the ⟨code⟩ is inserted into the input stream. This depends on the design of the mapping function.

### 25.8 Viewing property lists

\prop_show:N \prop_show:c
Displays the entries in the ⟨property list⟩ in the terminal, and specifies its storage type.
\prop_log:N \prop_log:c
Writes the entries in the \textit{property list} in the log file, and specifies its storage type.

New: 2014-06-12
Updated: 2021-04-29

\section{Scratch property lists}

There is no need to include both flat and linked property lists as scratch variables. We arbitrarily pick the older implementation.

\l_tmpa_prop \l_tmpb_prop
New: 2012-06-23

Scratch “flat” property lists for local assignment. These are never used by the kernel code, and so are safe for use with any \LaTeX3-defined function. However, they may be overwritten by other non-kernel code and so should only be used for short-term storage.

\g_tmpa_prop \g_tmpb_prop
New: 2012-06-23

Scratch “flat” property lists for global assignment. These are never used by the kernel code, and so are safe for use with any \LaTeX3-defined function. However, they may be overwritten by other non-kernel code and so should only be used for short-term storage.

\section{Constants}

\c_empty_prop
A permanently-empty property list used for internal comparisons.

222
Chapter 26

The l3skip module
Dimensions and skips

\LaTeX{} provides two general length variables: \texttt{dim} and \texttt{skip}. Lengths stored as \texttt{dim} variables have a fixed length, whereas \texttt{skip} lengths have a rubber (stretch/shrink) component. In addition, the \texttt{muskip} type is available for use in math mode: this is a special form of \texttt{skip} where the lengths involved are determined by the current math font (in \texttt{mu}). There are common features in the creation and setting of length variables, but for clarity the functions are grouped by variable type.

Many functions take \texttt{dimension expressions (“\texttt{\textbackslash dim expr}”)} or \texttt{skip expressions (“\texttt{\textbackslash skip expr}”)} as arguments.

26.1 Creating and initialising \texttt{dim} variables

\begin{verbatim}
\dim_new:N \dim_new:Nn \dim_new:c \dim_new:cn
\end{verbatim}

Creates a new \texttt{(dimension)} or raises an error if the name is already taken. The declaration is global. The \texttt{(dimension)} is initially equal to 0 pt.

\begin{verbatim}
\dim_const:Nn \dim_const:cn \dim_zero:N \dim_zero:c \dim_gzero:N \dim_gzero:c
\end{verbatim}

Creates a new constant \texttt{(dimension)} or raises an error if the name is already taken. The value of the \texttt{(dimension)} is set globally to the \texttt{(dim expr)}.

Sets \texttt{(dimension)} to 0 pt.

\begin{verbatim}
\dim_zero_new:N \dim_zero_new:c \dim_gzero_new:N \dim_gzero_new:c
\end{verbatim}

Ensures that the \texttt{(dimension)} exists globally by applying \texttt{\dim_new:N} if necessary, then applies \texttt{\dim_(g)zero:N} to leave the \texttt{(dimension)} set to zero.
26.2 Setting dim variables

\[\text{\texttt{\textbackslash dim_add:Nn}}\]
\[\text{\texttt{\textbackslash dim_add:cn}}\]
\[\text{\texttt{\textbackslash dim_gadd:Nn}}\]
\[\text{\texttt{\textbackslash dim_gadd:cn}}\]

Updated: 2011-10-22

\[\text{\texttt{\textbackslash dim_set:Nn}}\]
\[\text{\texttt{\textbackslash dim_set:cn}}\]
\[\text{\texttt{\textbackslash dim_gset:Nn}}\]
\[\text{\texttt{\textbackslash dim_gset:cn}}\]

Updated: 2011-10-22

\[\text{\texttt{\textbackslash dim_set_eq:NN}}\]
\[\text{\texttt{\textbackslash dim_set_eq:(cN|Nc|cc)}}\]
\[\text{\texttt{\textbackslash dim_gset_eq:NN}}\]
\[\text{\texttt{\textbackslash dim_gset_eq:(cN|Nc|cc)}}\]

\[\text{\texttt{\textbackslash dim_sub:Nn}}\]
\[\text{\texttt{\textbackslash dim_sub:cn}}\]
\[\text{\texttt{\textbackslash dim_gsub:Nn}}\]
\[\text{\texttt{\textbackslash dim_gsub:cn}}\]

Updated: 2011-10-22

26.3 Utilities for dimension calculations

\[\text{\texttt{\textbackslash dim_abs:n}}\]

Updated: 2012-09-26

\[\text{\texttt{\textbackslash dim_max:nn}}\]
\[\text{\texttt{\textbackslash dim_min:nn}}\]

Updated: 2012-09-26

\[\text{\texttt{\textbackslash dim_set_eq:NN}}\]
\[\text{\texttt{\textbackslash dim_set_eq:(cN|Nc|cc)}}\]
\[\text{\texttt{\textbackslash dim_gset_eq:NN}}\]
\[\text{\texttt{\textbackslash dim_gset_eq:(cN|Nc|cc)}}\]

\[\text{\texttt{\textbackslash dim_sub:Nn}}\]
\[\text{\texttt{\textbackslash dim_sub:cn}}\]
\[\text{\texttt{\textbackslash dim_gsub:Nn}}\]
\[\text{\texttt{\textbackslash dim_gsub:cn}}\]

Updated: 2011-10-22

\[\text{\texttt{\textbackslash dim_max:nn}}\]
\[\text{\texttt{\textbackslash dim_min:nn}}\]

Updated: 2012-09-26

\[\text{\texttt{\textbackslash dim_set_eq:NN}}\]
\[\text{\texttt{\textbackslash dim_set_eq:(cN|Nc|cc)}}\]
\[\text{\texttt{\textbackslash dim_gset_eq:NN}}\]
\[\text{\texttt{\textbackslash dim_gset_eq:(cN|Nc|cc)}}\]

\[\text{\texttt{\textbackslash dim_sub:Nn}}\]
\[\text{\texttt{\textbackslash dim_sub:cn}}\]
\[\text{\texttt{\textbackslash dim_gsub:Nn}}\]
\[\text{\texttt{\textbackslash dim_gsub:cn}}\]

Updated: 2011-10-22

\[\text{\texttt{\textbackslash dim_max:nn}}\]
\[\text{\texttt{\textbackslash dim_min:nn}}\]

Updated: 2012-09-26

\[\text{\texttt{\textbackslash dim_set_eq:NN}}\]
\[\text{\texttt{\textbackslash dim_set_eq:(cN|Nc|cc)}}\]
\[\text{\texttt{\textbackslash dim_gset_eq:NN}}\]
\[\text{\texttt{\textbackslash dim_gset_eq:(cN|Nc|cc)}}\]

\[\text{\texttt{\textbackslash dim_sub:Nn}}\]
\[\text{\texttt{\textbackslash dim_sub:cn}}\]
\[\text{\texttt{\textbackslash dim_gsub:Nn}}\]
\[\text{\texttt{\textbackslash dim_gsub:cn}}\]

224
\dim_ratio:nn \dim_ratio:nn \{dim expr1\} \{dim expr2\}

Parses the two \emph{dim exprs} and converts the ratio of the two to a form suitable for use inside a \emph{dim expr}. This ratio is then left in the input stream, allowing syntax such as

\begin{align*}
\def\l_my_dim{10 pt \dim_ratio:nn 5 pt 10 pt} \set\l_my_dim
\end{align*}

The output of \dim_ratio:nn on full expansion is a ratio expression between two integers, with all distances converted to scaled points. Thus

\begin{align*}
\tl_set:N \tl_my_tl \dim_ratio:nn 5 pt 10 pt
\tl_show:N \tl_my_tl
\end{align*}

displays $327680/655360$ on the terminal.

\section{Dimension expression conditionals}

\begin{align*}
\texttt{dim_compare_p:nNn} & \texttt{dim_compare_p:nNn} \{dim expr1\} \{relation\} \{dim expr2\} \\
\texttt{dim_compare:nNnTF} & \texttt{dim_compare:nNnTF} \{dim expr1\} \{relation\} \{dim expr2\} \\
& \{true code\} \{false code\}
\end{align*}

This function first evaluates each of the \emph{dim exprs} as described for \dim_eval:n. The two results are then compared using the \emph{relation}:

\begin{align*}
\text{Equal} & = \\
\text{Greater than} & > \\
\text{Less than} & <
\end{align*}

This function is less flexible than \texttt{dim_compare:nTF} but around 5 times faster.
\dim_compare_p:n * \dim_compare:nTF * 
\dim_compare:p:n

\dim_compare:nTF

\dim_compare:p:n

\dim_compare:nTF

\dim_compare:p:n

This function evaluates the \emph{\dim exprs} as described for \texttt{\dim_eval:n} and compares consecutive result using the corresponding \emph{\relation}, namely it compares \texttt{\dim expr_1} and \texttt{\dim expr_2} using the \emph{\relation_1}, then \texttt{\dim expr_2} and \texttt{\dim expr_3} using the \emph{\relation_2}, until finally comparing \texttt{\dim expr_N} and \texttt{\dim expr_{N+1}} using the \emph{\relation_N}. The test yields \texttt{true} if all comparisons are \texttt{true}. Each \texttt{\dim expr} is evaluated only once, and the evaluation is lazy, in the sense that if one comparison is \texttt{false}, then no other \texttt{\dim expr} is evaluated and no other comparison is performed. The \emph{\relations} can be any of the following:

- Equal \hspace{1cm} = \text{ or } ==
- Greater than or equal to \hspace{1cm} >=
- Greater than \hspace{1cm} >
- Less than or equal to \hspace{1cm} <=
- Less than \hspace{1cm} <
- Not equal \hspace{1cm} !=

This function is more flexible than \texttt{\dim_compare:nNnTF} but around 5 times slower.
This function evaluates the \textit{test dim expr} and compares this in turn to each of the \textit{dim expr cases}. If the two are equal then the associated \textit{code} is left in the input stream and other cases are discarded. If any of the cases are matched, the \textit{true code} is also inserted into the input stream (after the code for the appropriate case), while if none match then the \textit{false code} is inserted. The function \texttt{\texttt{dim_case:nn}}, which does nothing if there is no match, is also available. For example

\begin{verbatim}
\dim_set:Nn \l_tmpa_dim { 5 pt }
\dim_case:nnF { 2 \l_tmpa_dim }
{ { 5 pt } { Small } }
{ { 4 pt + 6 pt } { Medium } }
{ { - 10 pt } { Negative } }
{ No idea! }
\end{verbatim}

leaves “Medium” in the input stream.

### 26.5 Dimension expression loops

\texttt{\texttt{dim_case:nnT}} \texttt{\texttt{dim_case:nnTF}} \texttt{\texttt{dim_case:nn}} \texttt{\texttt{dim_case:nnF}}

\texttt{\texttt{dim_case:nnTF}} \texttt{\texttt{dim_case:nnF}}

This function evaluates the \textit{test dim expr} and compares this in turn to each of the \textit{dim expr cases}. If the two are equal then the associated \textit{code} is left in the input stream and other cases are discarded. If any of the cases are matched, the \textit{true code} is also inserted into the input stream (after the code for the appropriate case), while if none match then the \textit{false code} is inserted. The function \texttt{\texttt{dim_case:nn}, which does nothing if there is no match, is also available. For example

\begin{verbatim}
\dim_case:nnT \l_tmpa_dim { 5 pt }
\dim_case:nnF { 2 \l_tmpa_dim }
{ { 5 pt } { Small } }
{ { 4 pt + 6 pt } { Medium } }
{ { - 10 pt } { Negative } }
{ No idea! }
\end{verbatim}

leaves “Medium” in the input stream.

### 26.5 Dimension expression loops

\texttt{\texttt{dim_case:nnT}} \texttt{\texttt{dim_case:nnTF}} \texttt{\texttt{dim_case:nn}} \texttt{\texttt{dim_case:nnF}}

\texttt{\texttt{dim_case:nnTF}} \texttt{\texttt{dim_case:nnF}}

This function evaluates the \textit{test dim expr} and compares this in turn to each of the \textit{dim expr cases}. If the two are equal then the associated \textit{code} is left in the input stream and other cases are discarded. If any of the cases are matched, the \textit{true code} is also inserted into the input stream (after the code for the appropriate case), while if none match then the \textit{false code} is inserted. The function \texttt{\texttt{dim_case:nn}, which does nothing if there is no match, is also available. For example

\begin{verbatim}
\dim_case:nnT \l_tmpa_dim { 5 pt }
\dim_case:nnF { 2 \l_tmpa_dim }
{ { 5 pt } { Small } }
{ { 4 pt + 6 pt } { Medium } }
{ { - 10 pt } { Negative } }
{ No idea! }
\end{verbatim}

leaves “Medium” in the input stream.

### 26.5 Dimension expression loops

\texttt{\texttt{dim_case:nnT}} \texttt{\texttt{dim_case:nnTF}} \texttt{\texttt{dim_case:nn}} \texttt{\texttt{dim_case:nnF}}

\texttt{\texttt{dim_case:nnTF}} \texttt{\texttt{dim_case:nnF}}

This function evaluates the \textit{test dim expr} and compares this in turn to each of the \textit{dim expr cases}. If the two are equal then the associated \textit{code} is left in the input stream and other cases are discarded. If any of the cases are matched, the \textit{true code} is also inserted into the input stream (after the code for the appropriate case), while if none match then the \textit{false code} is inserted. The function \texttt{\texttt{dim_case:nn}, which does nothing if there is no match, is also available. For example

\begin{verbatim}
\dim_case:nnT \l_tmpa_dim { 5 pt }
\dim_case:nnF { 2 \l_tmpa_dim }
{ { 5 pt } { Small } }
{ { 4 pt + 6 pt } { Medium } }
{ { - 10 pt } { Negative } }
{ No idea! }
\end{verbatim}

leaves “Medium” in the input stream.

### 26.5 Dimension expression loops

\texttt{\texttt{dim_case:nnT}} \texttt{\texttt{dim_case:nnTF}} \texttt{\texttt{dim_case:nn}} \texttt{\texttt{dim_case:nnF}}

\texttt{\texttt{dim_case:nnTF}} \texttt{\texttt{dim_case:nnF}}
\textbf{26.6 Dimension step functions}

\begin{enumerate}
\item \texttt{\dim_step_function:nnnN} \hspace{1cm} \texttt{\dim_step_function:nnnN \{initial value\} \{step\} \{final value\} \{function\}}
\end{enumerate}

This function first evaluates the \texttt{initial value}, \texttt{step} and \texttt{final value}, all of which should be dimension expressions. The \texttt{function} is then placed in front of each \texttt{value} from the \texttt{initial value} to the \texttt{final value} in turn (using \texttt{step} between each \texttt{value}). The \texttt{step} must be non-zero. If the \texttt{step} is positive, the loop stops when the \texttt{value} becomes larger than the \texttt{final value}. If the \texttt{step} is negative, the loop stops when the \texttt{value} becomes smaller than the \texttt{final value}. The \texttt{function} should absorb one argument.

\begin{enumerate}
\item \texttt{\dim_step_inline:nnn} \hspace{1cm} \texttt{\dim_step_inline:nnn \{initial value\} \{step\} \{final value\} \{code\}}
\end{enumerate}

This function first evaluates the \texttt{initial value}, \texttt{step} and \texttt{final value}, all of which should be dimension expressions. Then for each \texttt{value} from the \texttt{initial value} to the \texttt{final value} in turn (using \texttt{step} between each \texttt{value}), the \texttt{code} is inserted into the input stream with \texttt{#1} replaced by the current \texttt{value}. Thus the \texttt{code} should define a function of one argument (\texttt{#1}).
This function first evaluates the *(initial value)*, *(step)* and *(final value)*, all of which should be dimension expressions. Then for each *(value)* from the *(initial value)* to the *(final value)* in turn (using *(step)* between each *(value)*) , the *(code)* is inserted into the input stream, with the *(tl var)* defined as the current *(value)*. Thus the *(code)* should make use of the *(tl var)*.

### 26.7 Using dim expressions and variables

\[ \text{\texttt{\textbackslash dim_use:N \{dimension\}}} \]

Recover the content of a *(dimension)* and places it directly in the input stream. An error is raised if the variable does not exist or if it is invalid. Can be omitted in places where a *(dimension)* is required (such as in the argument of \texttt{\textbackslash dim_eval:n}).

**\TeX{}hackers note:** \texttt{\textbackslash dim_use:N} is the \TeX{} primitive \texttt{\textbackslash the}: this is one of several \LaTeX{} names for this primitive.

\[ \text{\texttt{\textbackslash dim_to_decimal:n \{dim expr\}}} \]

Evaluates the *(dim expr)*, and leaves the result, expressed in points (pt) in the input stream, with *no units*. The result is rounded by \TeX{} to at most five decimal places. If the decimal part of the result is zero, it is omitted, together with the decimal marker.

For example

\[ \text{\texttt{\textbackslash dim_to_decimal:n \{1bp\}}} \]

leaves 1.00374 in the input stream, *i.e.* the magnitude of one “big point” when converted to (\TeX{}) points.
\texttt{\textbackslash dim\_to\_decimal\_in\_bp:n \{\textit{dim expr}\}}

Evaluates the \texttt{\{\textit{dim expr}\}}, and leaves the result, expressed in big points (bp) in the input stream, with no units. The result is rounded by \TeX{} to at most five decimal places. If the decimal part of the result is zero, it is omitted, together with the decimal marker.

For example

\texttt{\textbackslash dim\_to\_decimal\_in\_bp:n \{ 1pt \}}

leaves 0.99628 in the input stream, \textit{i.e.} the magnitude of one (\TeX{}) point when converted to big points.

\TeX{}hackers note: The implementation of this function is re-entrant: the result of

\texttt{\textbackslash dim\_compare:nNnTF}

\begin{verbatim}
{ \langle x\rangle bp } =
{ \texttt{\textbackslash dim\_to\_decimal\_in\_bp:n \{ \langle x\rangle bp \} bp } }
\end{verbatim}

will be logically \texttt{true}. The decimal representations may differ provided they produce the same \TeX{} dimension.

\texttt{\textbackslash dim\_to\_decimal\_in\_cc:n
\textbackslash dim\_to\_decimal\_in\_cm:n
\textbackslash dim\_to\_decimal\_in\_dd:n
\textbackslash dim\_to\_decimal\_in\_in:n
\textbackslash dim\_to\_decimal\_in\_mm:n
\textbackslash dim\_to\_decimal\_in\_pc:n}

\texttt{\{\textit{dim expr}\}}

Evaluates the \texttt{\{\textit{dim expr}\}}, and leaves the result, expressed with the appropriate scaling in the input stream, with no units. If the decimal part of the result is zero, it is omitted, together with the decimal marker. The precision of the result is limited to a maximum of five decimal places with trailing zeros omitted.

The maximum \TeX{} allowable dimension value (available as \texttt{\maxdimen} in plain \TeX{} and \LaTeX{} and \texttt{\c_max_dim} in expl3) can only be expressed exactly in the units \texttt{pt}, \texttt{bp} and \texttt{sp}. The maximum allowable input values to five decimal places are

\begin{itemize}
\item 1276.00215 cc
\item 575.83174 cm
\item 15312.02584 dd
\item 226.70540 in
\item 5758.31742 mm
\item 1365.33333 pc
\end{itemize}

(Note that these are not all equal, but rather any larger value will overflow due to the way \TeX{} converts to \texttt{sp}.) Values given to five decimal places larger that these will result in \TeX{} errors; the behavior if additional decimal places are given depends on the \TeX{} internals and thus larger values are not supported by expl3.

\TeX{}hackers note: The implementation of these functions is re-entrant: the result of

\texttt{\textbackslash dim\_compare:nNnTF}

\begin{verbatim}
{ \langle x\rangle<\texttt{unit}> } =
{ \texttt{\textbackslash dim\_to\_decimal\_in\_<\texttt{unit}>:n \{ \langle x\rangle<\texttt{unit}> \} <\texttt{unit}> } }
\end{verbatim}

will be logically \texttt{true}. The decimal representations may differ provided they produce the same \TeX{} dimension.

230
\dim_to_decimal_in_sp:n \dim_to_decimal_in_sp:n \{\dim expr\}

Evaluates the \dim expr, and leaves the result, expressed in scaled points (sp) in the 
input stream, with no units. The result is necessarily an integer.

\dim_to_decimal_in_unit:nn \dim_to_decimal_in_unit:nn \{\dim expr_1\} \{\dim expr_2\}

\dim_to_decimal_in_unit:nn \dim_to_decimal_in_unit:nn \{\dim exprs\}, and leaves the value of \dim expr_1, expressed in a unit 
given by \dim expr_2, in the input stream. If the decimal part of the result is zero, it is 
 omitted, together with the decimal marker. The precisions of the result is limited to a 
maximum of five decimal places with trailing zeros omitted.

For example

\dim_to_decimal_in_unit:nn \{1bp\} \{1mm\}

leaves 0.35278 in the input stream, i.e. the magnitude of one big point when expressed 
in millimetres. The conversions do not guarantee that \TeX would yield identical results 
for the direct input in an equality test, thus for instance

\dim_compare:nNnTF
\{1bp\} =
\{\dim_to_decimal_in_unit:nn \{1bp\} \{1mm\} \text{mm}\}

will take the false branch.

\dim_to_fp:n \dim_to_fp:n \{\dim expr\}

Expands to an internal floating point number equal to the value of the \dim expr in 
pt. Since dimension expressions are evaluated much faster than their floating point 
equivalent, \dim_to_fp:n can be used to speed up parts of a computation where a low 
precision and a smaller range are acceptable.

### 26.8 Viewing \texttt{dim} variables

\dim_show:N \dim_show:N \{\texttt{dimension}\}
\dim_show:C
Displays the value of the \texttt{dimension} on the terminal.

\dim_show:n \dim_show:n \{\dim expr\}
\dim_show:n \{\dim expr\}
Displays the result of evaluating the \dim expr on the terminal.

\dim_log:N \dim_log:N \{\texttt{dimension}\}
\dim_log:C
Writes the value of the \texttt{dimension} in the log file.
\dim_log:n \{\dim expr\}

Writes the result of evaluating the \{\dim expr\} in the log file.

26.9 Constant dimensions

\c_max_dim
The maximum value that can be stored as a dimension. This can also be used as a component of a skip.

\c_zero_dim
A zero length as a dimension. This can also be used as a component of a skip.

26.10 Scratch dimensions

\l_tmpa_dim
\l_tmpb_dim
Scratch dimension for local assignment. These are never used by the kernel code, and so are safe for use with any \LaTeX3-defined function. However, they may be overwritten by other non-kernel code and so should only be used for short-term storage.

\g_tmpa_dim
\g_tmpb_dim
Scratch dimension for global assignment. These are never used by the kernel code, and so are safe for use with any \LaTeX3-defined function. However, they may be overwritten by other non-kernel code and so should only be used for short-term storage.

26.11 Creating and initialising skip variables

\skip_new:N \skip_new:c
\skip: new \{\skip\}

Creates a new \{\skip\} or raises an error if the name is already taken. The declaration is global. The \{\skip\} is initially equal to 0pt.

\skip_zero:N \skip_zero:c
\skip: zero \{\skip\}

Sets \{\skip\} to 0pt.
Ensures that the \texttt{skip} exists globally by applying \texttt{skip \_ new:N} if necessary, then applies \texttt{skip \_ (g)zero:N} to leave the \texttt{skip} set to zero.

Tests whether the \texttt{skip} is currently defined. This does not check that the \texttt{skip} really is a skip variable.

### 26.12 Setting skip variables

\texttt{skip \_ add:Nn} \texttt{\{skip expr\}}

Adds the result of the \texttt{skip expr} to the current content of the \texttt{skip}.

\texttt{skip \_ set:Nn} \texttt{\{skip expr\}}

Sets \texttt{skip} to the value of \texttt{skip expr}, which must evaluate to a length with units and may include a rubber component (for example 1 \texttt{cm} plus 0.5 \texttt{cm}.

\texttt{skip \_ set_eq:NN} \texttt{\{skip \_ 1\} \{skip \_ 2\}}

Sets the content of \texttt{\_ 1} equal to that of \texttt{\_ 2}.

\texttt{skip \_ sub:Nn} \texttt{\{skip expr\}}

Subtracts the result of the \texttt{skip expr} from the current content of the \texttt{skip}. 
26.13 Skip expression conditionals

\texttt{\textbackslash skip\_if\_eq\_p:nn} = \texttt{\{skip expr\_1\}} \{\{skip expr\_2\}\}
\texttt{\textbackslash skip\_if\_eq\_nnTF} = \{\{skip expr\_1\}\} \{\{skip expr\_2\}\}
\texttt{\{true code\}} \{\{false code\}\}

This function first evaluates each of the \langle skip exprs \rangle as described for \texttt{\textbackslash skip\_eval:n}. The two results are then compared for exact equality, i.e. both the fixed and rubber components must be the same for the test to be true.

\texttt{\textbackslash skip\_if\_finite\_p:n} = \texttt{\{skip expr\}}
\texttt{\textbackslash skip\_if\_finite\_nTF} = \{\{skip expr\}\} \{\{true code\}\} \{\{false code\}\}

Evaluates the \langle skip expr \rangle as described for \texttt{\textbackslash skip\_eval:n}, and then tests if all of its components are finite.

26.14 Using skip expressions and variables

\texttt{\textbackslash skip\_eval:n} = \texttt{\{skip expr\}}

Evaluates the \langle skip expr \rangle, expanding any skips and token list variables within the \langle expression \rangle to their content (without requiring \texttt{\textbackslash skip\_use:N/\textbackslash tl\_use:N}) and applying the standard mathematical rules. The result of the calculation is left in the input stream as a \langle glue denotation \rangle after two expansions. This is expressed in points (pt), and requires suitable termination if used in a \TeX\-style assignment as it is not an \langle internal glue \rangle.

\texttt{\textbackslash skip\_use:N} = \texttt{\{skip\}}
\texttt{\textbackslash skip\_use:c} = \texttt{\{skip\}}

Recovers the content of a \langle skip \rangle and places it directly in the input stream. An error is raised if the variable does not exist or if it is invalid. Can be omitted in places where a \langle dimension \rangle or \langle skip \rangle is required (such as in the argument of \texttt{\textbackslash skip\_eval:n}).

\TeX\ hackers note: \texttt{\textbackslash skip\_use:N} is the \TeX\ primitive \texttt{\the}: this is one of several \LaTeX3 names for this primitive.

26.15 Viewing skip variables

\texttt{\textbackslash skip\_show:N} = \texttt{\{skip\}}
\texttt{\textbackslash skip\_show:c} = \texttt{\{skip\}}

Displays the value of the \langle skip \rangle on the terminal.

\texttt{\textbackslash skip\_show:n} = \texttt{\{skip expr\}}

Displays the result of evaluating the \langle skip expr \rangle on the terminal.
\skip_log:N \skip_log:c
New: 2014-08-22
Updated: 2015-08-03
Writes the value of the \(<\text{skip}\) in the log file.

\skip_log:n \skip_log:n \{\langle\text{skip expr}\rangle\}
New: 2014-08-22
Updated: 2015-08-07
Writes the result of evaluating the \(<\text{skip expr}\rangle\) in the log file.

\c_max_skip
Updated: 2012-11-02
The maximum value that can be stored as a skip (equal to \c_max_dim in length), with no stretch nor shrink component.

\c_zero_skip
Updated: 2012-11-01
A zero length as a skip, with no stretch nor shrink component.

\l_tmpa_skip \l_tmpb_skip
Scratch skip for local assignment. These are never used by the kernel code, and so are safe for use with any \texttt{\LaTeX3}-defined function. However, they may be overwritten by other non-kernel code and so should only be used for short-term storage.

\g_tmpa_skip \g_tmpb_skip
Scratch skip for global assignment. These are never used by the kernel code, and so are safe for use with any \texttt{\LaTeX3}-defined function. However, they may be overwritten by other non-kernel code and so should only be used for short-term storage.

\skip_horizontal:N \skip_horizontal:c \skip_horizontal:n \{\langle\text{skip expr}\rangle\}
Updated: 2011-10-22
Inserts a horizontal \(<\text{skip}\rangle\) into the current list. The argument can also be a \(<\text{dim}\rangle\).

\texttt{\TeXhackers note:} \skip_horizontal:N is the \TeX primitive \texttt{\hskip}.

26.18 Inserting skips into the output
\skip_vertical:N \skip_vertical:C \skip_vertical:n
Inserts a vertical \langle skip \rangle into the current list. The argument can also be a \langle dim \rangle.

\TeXhackers note: \skip_vertical:N is the \TeX primitive \vskip.

26.19 Creating and initialising muskip variables

\muskip_new:N \muskip_new:C
Creates a new \langle muskip \rangle or raises an error if the name is already taken. The declaration is global. The \langle muskip \rangle is initially equal to 0 mu.

\muskip_const:Nn \muskip_const:cn
Creates a new constant \langle muskip \rangle or raises an error if the name is already taken. The value of the \langle muskip \rangle is set globally to the \langle muskip expr \rangle.

\muskip_zero:N \muskip_zero:C \muskip_gzero:N \muskip_gzero:C
Sets \langle muskip \rangle to 0 mu.

\muskip_zero_new:N \muskip_zero_new:C \muskip_gzero_new:N \muskip_gzero_new:C
Ensures that the \langle muskip \rangle exists globally by applying \muskip_new:N if necessary, then applies \muskip_(g)zero:N to leave the \langle muskip \rangle set to zero.

\muskip_if_exist_p:N \muskip_if_exist_p:C \muskip_if_exist:NTF \muskip_if_exist:CN
Tests whether the \langle muskip \rangle is currently defined. This does not check that the \langle muskip \rangle really is a muskip variable.

26.20 Setting muskip variables

\muskip_add:Nn \muskip_add:cn \muskip_gadd:Nn \muskip_gadd:cn
Adds the result of the \langle muskip expr \rangle to the current content of the \langle muskip \rangle.
\muskip_set:Nn \muskip_set:cn \muskip_gset:Nn \muskip_gset:cn

\muskip_set_eq:NN \muskip_set_eq:NN \muskip_gset_eq:NN \muskip_gset_eq:NN

\muskip_sub:Nn \muskip_sub:cn \muskip_gsub:Nn \muskip_gsub:cn

\muskip_eval:n \muskip_eval:cn \muskip_use:N \muskip_use:cn

\muskip_show:N \muskip_show:cn

26.21 Using \texttt{muskip} expressions and variables

\muskip_eval:n \muskip_eval:cn \muskip_use:N \muskip_use:cn

\muskip_show:N \muskip_show:cn

26.22 Viewing \texttt{muskip} variables
\muskip\_show:n

Displays the result of evaluating the \texttt{(muskip expr)} on the terminal.

\muskip\_log:N
\muskip\_log:c

Writes the value of the \texttt{(muskip)} in the log file.

\muskip\_log:n

Writes the result of evaluating the \texttt{(muskip expr)} in the log file.

\section*{26.23 Constant muskips}

\c\_max\_muskip

The maximum value that can be stored as a muskip, with no stretch nor shrink component.

\c\_zero\_muskip

A zero length as a muskip, with no stretch nor shrink component.

\section*{26.24 Scratch muskips}

\l\_tmpa\_muskip
\l\_tmpb\_muskip

Scratch muskip for local assignment. These are never used by the kernel code, and so are safe for use with any LaTeX3-defined function. However, they may be overwritten by other non-kernel code and so should only be used for short-term storage.

\g\_tmpa\_muskip
\g\_tmpb\_muskip

Scratch muskip for global assignment. These are never used by the kernel code, and so are safe for use with any LaTeX3-defined function. However, they may be overwritten by other non-kernel code and so should only be used for short-term storage.

\section*{26.25 Primitive conditional}

\ifdim:w \texttt{\ifdim:w (dimen1) (relation) (dimen2)\newline\{true code\} \else:\newline\{false\}\fi:}

Compare two dimensions. The \texttt{(relation)} is one of <, = or > with category code 12.

\TeX{}hackers note: This is the \TeX{} primitive \texttt{ifdim}. 238
Chapter 27

The l3keys module
Key–value interfaces

The key–value method is a popular system for creating large numbers of settings for controlling function or package behaviour. The system normally results in input of the form

```
\MyModuleSetup{
  key-one = value one,
  key-two = value two
}
```
or

```
\MyModuleMacro[
  key-one = value one,
  key-two = value two
]{argument}
```

for the user.

The high level functions here are intended as a method to create key–value controls. Keys are themselves created using a key–value interface, minimising the number of functions and arguments required. Each key is created by setting one or more properties of the key:

```
\keys_define:nn { mymodule }
{
  key-one .code:n = code including parameter #1,
  key-two .tl_set:N = \l_mymodule_store_tl
}
```

These values can then be set as with other key–value approaches:

```
\keys_set:nn { mymodule }
{
  key-one = value one,
  key-two = value two
}
```
As illustrated, keys are created inside a ⟨module⟩: a set of related keys, typically those for a single module/\LaTeX\TeX\ε package. See Section 27.2 for suggestions on how to divide large numbers of keys for a single module.

At a document level, \keys_set:nn is used within a document function, for example

\DeclareDocumentCommand \MyModuleSetup { m }{ \keys_set:nn { mymodule } { #1 } }
\DeclareDocumentCommand \MyModuleMacro { o m }{\group_begin: \keys_set:nn { mymodule } { #1 } \group_end:}

Key names may contain any tokens, as they are handled internally using \tl_to_str:n. As discussed in section 27.2, it is suggested that the character / is reserved for sub-division of keys into different subsets. Functions and variables are not expanded when creating key names, and so

\tl_set:Nn \l_mymodule_tmp_tl { key }
\keys_define:nn { mymodule }{ \l_mymodule_tmp_tl .code:n = code }

creates a key called \l_mymodule_tmp_tl, and not one called key.

### 27.1 Creating keys

\keys_define:nn { ⟨module⟩ } { ⟨keyval list⟩ }
\keys_define:ne

Parses the ⟨keyval list⟩ and defines the keys listed there for ⟨module⟩. The ⟨module⟩ name is treated as a string. In practice the ⟨module⟩ should be chosen to be unique to the module in question (unless deliberately adding keys to an existing module).

The ⟨keyval list⟩ should consist of one or more key names along with an associated key property. The properties of a key determine how it acts. The individual properties are described in the following text; a typical use of \keys_define:nn might read

\keys_define:nn { mymodule }
{ 
  keyname .code:n = Some-code-using-#1,
  keyname .value_required:n = true
}

where the properties of the key begin from the . after the key name.

The various properties available take either no arguments at all, or require one or more arguments. This is indicated in the name of the property using an argument specification. In the following discussion, each property is illustrated attached to an arbitrary ⟨key⟩, which when used may be supplied with a ⟨value⟩. All key definitions are local.

240
Key properties are applied in the reading order and so the ordering is significant. Key properties which define “actions”, such as \texttt{.code:n}, \texttt{.tl_set:N}, etc., override one another. Some other properties are mutually exclusive, notably \texttt{.value_required:n} and \texttt{.value_forbidden:n}, and so they replace one another. However, properties covering non-exclusive behaviours may be given in any order. Thus for example the following definitions are equivalent.

\begin{verbatim}
\keys_define:nn { mymodule }
  {
    keyname .code:n = Some-code-using-#1,
    keyname .value_required:n = true
  }
\keys_define:nn { mymodule }
  {
    keyname .value_required:n = true,
    keyname .code:n = Some-code-using-#1
  }
\end{verbatim}

Note that all key properties define the key within the current \TeX group, with an exception that the special \texttt{.undefine}: property undefines the key within the current \TeX group.

\begin{verbatim}
\bool_set:N \bool_set:c \bool_gset:N \bool_gset:c
\end{verbatim}

\begin{verbatim}
\bool_set_inverse:N \bool_set_inverse:c \bool_gset_inverse:N \bool_gset_inverse:c
\end{verbatim}

\begin{verbatim}
\clist_set:N \clist_set:c \clist_gset:N \clist_gset:c
\end{verbatim}

\begin{verbatim}
\choice: \choices:nn \choices:(Vn|en|on)
\end{verbatim}

\begin{verbatim}
\clist_set:N \clist_set:c \clist_gset:N \clist_gset:c
\end{verbatim}

\begin{verbatim}
\keys_define:nn { mymodule }
  {
    keyname .code:n = Some-code-using-#1,
    keyname .value_required:n = true
  }
\keys_define:nn { mymodule }
  {
    keyname .value_required:n = true,
    keyname .code:n = Some-code-using-#1
  }
\end{verbatim}

Note that all key properties define the key within the current \TeX group, with an exception that the special \texttt{.undefine}: property undefines the key within the current \TeX group.

\begin{verbatim}
\bool_set:N \bool_set:c \bool_gset:N \bool_gset:c
\end{verbatim}

\begin{verbatim}
\bool_set_inverse:N \bool_set_inverse:c \bool_gset_inverse:N \bool_gset_inverse:c
\end{verbatim}

\begin{verbatim}
\clist_set:N \clist_set:c \clist_gset:N \clist_gset:c
\end{verbatim}

\begin{verbatim}
\choice: \choices:nn \choices:(Vn|en|on)
\end{verbatim}

\begin{verbatim}
\clist_set:N \clist_set:c \clist_gset:N \clist_gset:c
\end{verbatim}

\begin{verbatim}
\keys_define:nn { mymodule }
  {
    keyname .code:n = Some-code-using-#1,
    keyname .value_required:n = true
  }
\keys_define:nn { mymodule }
  {
    keyname .value_required:n = true,
    keyname .code:n = Some-code-using-#1
  }
\end{verbatim}

Note that all key properties define the key within the current \TeX group, with an exception that the special \texttt{.undefine}: property undefines the key within the current \TeX group.

\begin{verbatim}
\bool_set:N \bool_set:c \bool_gset:N \bool_gset:c
\end{verbatim}

\begin{verbatim}
\bool_set_inverse:N \bool_set_inverse:c \bool_gset_inverse:N \bool_gset_inverse:c
\end{verbatim}

\begin{verbatim}
\clist_set:N \clist_set:c \clist_gset:N \clist_gset:c
\end{verbatim}

\begin{verbatim}
\choice: \choices:nn \choices:(Vn|en|on)
\end{verbatim}

\begin{verbatim}
\clist_set:N \clist_set:c \clist_gset:N \clist_gset:c
\end{verbatim}

\begin{verbatim}
\keys_define:nn { mymodule }
  {
    keyname .code:n = Some-code-using-#1,
    keyname .value_required:n = true
  }
\keys_define:nn { mymodule }
  {
    keyname .value_required:n = true,
    keyname .code:n = Some-code-using-#1
  }
\end{verbatim}

Note that all key properties define the key within the current \TeX group, with an exception that the special \texttt{.undefine}: property undefines the key within the current \TeX group.

\begin{verbatim}
\bool_set:N \bool_set:c \bool_gset:N \bool_gset:c
\end{verbatim}

\begin{verbatim}
\bool_set_inverse:N \bool_set_inverse:c \bool_gset_inverse:N \bool_gset_inverse:c
\end{verbatim}

\begin{verbatim}
\clist_set:N \clist_set:c \clist_gset:N \clist_gset:c
\end{verbatim}

\begin{verbatim}
\choice: \choices:nn \choices:(Vn|en|on)
\end{verbatim}

\begin{verbatim}
\clist_set:N \clist_set:c \clist_gset:N \clist_gset:c
\end{verbatim}

\begin{verbatim}
\keys_define:nn { mymodule }
  {
    keyname .code:n = Some-code-using-#1,
    keyname .value_required:n = true
  }
\keys_define:nn { mymodule }
  {
    keyname .value_required:n = true,
    keyname .code:n = Some-code-using-#1
  }
\end{verbatim}

Note that all key properties define the key within the current \TeX group, with an exception that the special \texttt{.undefine}: property undefines the key within the current \TeX group.

\begin{verbatim}
\bool_set:N \bool_set:c \bool_gset:N \bool_gset:c
\end{verbatim}

\begin{verbatim}
\bool_set_inverse:N \bool_set_inverse:c \bool_gset_inverse:N \bool_gset_inverse:c
\end{verbatim}

\begin{verbatim}
\clist_set:N \clist_set:c \clist_gset:N \clist_gset:c
\end{verbatim}

\begin{verbatim}
\choice: \choices:nn \choices:(Vn|en|on)
\end{verbatim}

\begin{verbatim}
\clist_set:N \clist_set:c \clist_gset:N \clist_gset:c
\end{verbatim}
Stores the (code) for execution when (key) is used. The (code) can include one parameter (#1), which will be the (value) given for the (key).

(key) .code:n = {(code)}

(key) .cs_set:Np = {control sequence} {arg. spec.}
Defines (key) to set (control sequence) to have (arg. spec.) and replacement text (value).

(key) .dim_set:N = {dimension}
Defines (key) to set (dimension) to (value) (which must a dimension expression). If the variable does not exist, it is created globally at the point that the key is set up. The key will require a value at point-of-use unless a default is set.

(key) .fp_set:N = {floating point}
Defines (key) to set (floating point) to (value) (which must a floating point expression). If the variable does not exist, it is created globally at the point that the key is set up. The key will require a value at point-of-use unless a default is set.
.groups:n (key) .groups:n = {(groups)}

New: 2013-07-14
Defines (key) as belonging to the (groups) (a comma-separated list). Groups provide a “secondary axis” for selectively setting keys, and are described in Section 27.7.

**TeXhackers note:** The (groups) argument is turned into a string then interpreted as a comma-separated list, so group names cannot contain commas nor start or end with a space character.

.inherit:n (key) .inherit:n = {(parents)}

New: 2016-11-22
Specifies that the (key) path should inherit the keys listed as any of the (parents) (a comma list), which can be a module or a sub-division thereof. For example, after setting

\begin{verbatim}
\keys_define:nn { foo } { test .code:n = \tl_show:n {#1} }
\keys_define:nn { } { bar .inherit:n = foo }
\end{verbatim}

setting

\begin{verbatim}
\keys_set:nn { bar } { test = a }
\end{verbatim}

will be equivalent to

\begin{verbatim}
\keys_set:nn { foo } { test = a }
\end{verbatim}

Inheritance applies at point of use, not at definition, thus keys may be added to the (parent) after the use of .inherit:n and will be active. If more than one (parent) is specified, the presence of the (key) will be tested for each in turn, with the first successful hit taking priority.

.initial:n (key) .initial:n = {(value)}

Updated: 2013-07-09
Initialises the (key) with the (value), equivalent to

\begin{verbatim}
\keys_set:nn { (module) } { (key) = (value) }
\end{verbatim}

.int_set:N (key) .int_set:N = (integer)

Updated: 2020-01-17
Defines (key) to set (integer) to (value) (which must be an integer expression). If the variable does not exist, it is created globally at the point that the key is set up. The key will require a value at point-of-use unless a default is set.

.legacy_if_set:n (key) .legacy_if_set:n = (switch)

Updated: 2022-01-15
Defines (key) to set legacy \if\(\text{switch}\) to (value) (which must be either “true” or “false”). The \text{switch} is the name of the switch without the leading if. The inverse versions will set the \text{switch} to the logical opposite of the (value).

.meta:n (key) .meta:n = {(keyval list)}

Updated: 2013-07-10
Makes (key) a meta-key, which will set (keyval list) in one go. The (keyval list) can refer as \#i to the value given at the time the (key) is used (or, if no value is given, the (key)’s default value).
.meta:nn = {\{path\}} {\{keyval list\}}

Makes \langle key \rangle a meta-key, which will set \langle keyval list \rangle in one go using the \langle path \rangle in place of the current one. The \langle keyval list \rangle can refer as \#1 to the value given at the time the \langle key \rangle is used (or, if no value is given, the \langle key \rangle's default value).

.multichoice = \langle key \rangle
default key:

Sets \langle key \rangle to act as a multiple choice key. Each valid choice for \langle key \rangle must then be created, as discussed in section 27.3.

.multichoices = \langle key \rangle \{choices\} {\langle code \rangle}

Sets \langle key \rangle to act as a multiple choice key, and defines a series \langle choices \rangle which are implemented using the \langle code \rangle. Inside \langle code \rangle, \$\{keys_choice_tl\} will be the name of the choice made, and \$\{keys_choice_int\} will be the position of the choice in the list of \langle choices \rangle (indexed from 1). Choices are discussed in detail in section 27.3.

.muskip_set = \langle key \rangle \{muskip\}

Defines \langle key \rangle to set \langle muskip \rangle to \langle value \rangle (which must be a muskip expression). If the variable does not exist, it is created globally at the point that the key is set up. The key will require a value at point-of-use unless a default is set.

.prop_put = \langle key \rangle \{property list\}

Defines \langle key \rangle to put the \langle value \rangle onto the \langle property list \rangle stored under the \langle key \rangle. If the variable does not exist, it is created globally at the point that the key is set up.

.skip_set = \langle key \rangle \{skip\}

Defines \langle key \rangle to set \langle skip \rangle to \langle value \rangle (which must be a skip expression). If the variable does not exist, it is created globally at the point that the key is set up. The key will require a value at point-of-use unless a default is set.

.str_set = \langle key \rangle \{string variable\}

Defines \langle key \rangle to set \langle string variable \rangle to \langle value \rangle. If the variable does not exist, it is created globally at the point that the key is set up.

.str_set_e = \langle key \rangle \{string variable\}

Defines \langle key \rangle to set \langle string variable \rangle to \langle value \rangle, which will be subjected to an \texttt{e}-type expansion (\i.e. using \texttt{\string_set:Ne}). If the variable does not exist, it is created globally at the point that the key is set up.
\tl_set:N \tl_set_c
\tl_gset:N \tl_gset_c
\langle \text{key} \rangle .\tl_set:N = \langle \text{token list variable} \rangle
\langle \text{value} \rangle

Defines \langle \text{key} \rangle to set \langle \text{token list variable} \rangle to \langle \text{value} \rangle. If the variable does not exist, it is created globally at the point that the key is set up.

\tl_set_e:N \tl_set_e:c
\tl_gset_e:N \tl_gset_e:c
\langle \text{key} \rangle .\tl_set_e:N = \langle \text{token list variable} \rangle
\langle \text{value} \rangle
\text{e-type expansion (i.e. using} \tl_set:Ne)
\langle \text{value} \rangle

Defines \langle \text{key} \rangle to set \langle \text{token list variable} \rangle to \langle \text{value} \rangle, which will be subjected to an e-type expansion (i.e. using \tl_set:Ne). If the variable does not exist, it is created globally at the point that the key is set up.

\text{New: 2023-09-18}

\langle \text{value_required:n} = \langle \text{true|false} \rangle \rangle
\langle \text{true|false} \rangle

Specifies that \langle \text{key} \rangle must receive a \langle \text{value} \rangle when used. If a \langle \text{value} \rangle is not given then an error will be issued. Setting the property “false” cancels the restriction.

\text{New: 2015-07-14}

\langle \text{value_forbidden:n} = \langle \text{true|false} \rangle \rangle
\langle \text{true|false} \rangle

Specifies that \langle \text{key} \rangle cannot receive a \langle \text{value} \rangle when used. If a \langle \text{value} \rangle is given then an error will be issued. Setting the property “false” cancels the restriction.

\text{New: 2015-07-14}

\text{Sub-dividing keys}

When creating large numbers of keys, it may be desirable to divide them into several subsets for a given module. This can be achieved either by adding a sub-division to the module name:

\text{\keys_define:nn} { mymodule / subset }
\{ key .\code:n = code \}

or to the key name:

\text{\keys_define:nn} { mymodule }
\{ subset / key .\code:n = code \}

As illustrated, the best choice of token for sub-dividing keys in this way is / . This is because of the method that is used to represent keys internally. Both of the above code fragments set the same key, which has full name mymodule/subset/key.

As illustrated in the next section, this subdivision is particularly relevant to making multiple choices.
27.3 Choice and multiple choice keys

The \l3keys system supports two types of choice key, in which a series of pre-defined input values are linked to varying implementations. Choice keys are usually created so that the various values are mutually-exclusive: only one can apply at any one time. “Multiple” choice keys are also supported: these allow a selection of values to be chosen at the same time.

Mutually-exclusive choices are created by setting the .choice: property:

\keys_define:nn { mymodule }
{ key .choice: }

For keys which are set up as choices, the valid choices are generated by creating sub-keys of the choice key. This can be carried out in two ways.

In many cases, choices execute similar code which is dependent only on the name of the choice or the position of the choice in the list of all possibilities. Here, the keys can share the same code, and can be rapidly created using the .choices:nn property.

\keys_define:nn { mymodule }
{ key .choices:nn =
  { choice-a, choice-b, choice-c }
  { You\-gave\-choice:'\tl_use:N \l_keys_choice_tl',~
    which\-is\-in\-position-\int_use:N \l_keys_choice_int \c_space_tl
    in\-the\-list.
  }
 }

The index \l_keys_choice_int in the list of choices starts at 1.

\l_keys_choice_int \l_keys_choice_tl Inside the code block for a choice generated using .choices:nn, the variables \l_keys_choice_tl and \l_keys_choice_int are available to indicate the name of the current choice, and its position in the comma list. The position is indexed from 1. Note that, as with standard key code generated using .code:n, the value passed to the key (i.e. the choice name) is also available as #1.

On the other hand, it is sometimes useful to create choices which use entirely different code from one another. This can be achieved by setting the .choice: property of a key, then manually defining sub-keys.

\keys_define:nn { mymodule }
{ key .choice:,
  key / choice-a .code:n = code-a,
  key / choice-b .code:n = code-b,
  key / choice-c .code:n = code-c,
 }

It is possible to mix the two methods, but manually-created choices should not use \l_keys_choice_tl or \l_keys_choice_int. These variables do not have defined
behaviour when used outside of code created using \texttt{choices:nn} (\emph{i.e.} anything might happen).

It is possible to allow choice keys to take values which have not previously been defined by adding code for the special \texttt{unknown} choice. The general behavior of the \texttt{unknown} key is described in Section 27.6. A typical example in the case of a choice would be to issue a custom error message:

\begin{verbatim}
\keys_define:nn { mymodule }
{ 
  key .choice:, 
  key / choice-a .code:n = code-a, 
  key / choice-b .code:n = code-b, 
  key / choice-c .code:n = code-c, 
  key / unknown .code:n = \msg_error:nnee { mymodule } { unknown-choice }
  \{ key \} \% Name of choice key 
  \{ choice-a , choice-b , choice-c \} \% Valid choices 
  \{ \exp_not:n \{#1\} \} \% Invalid choice given
% 
%
}
\end{verbatim}

Multiple choices are created in a very similar manner to mutually-exclusive choices, using the properties \texttt{.multichoice:} and \texttt{.multichices:nn}. As with mutually exclusive choices, multiple choices are define as sub-keys. Thus both

\begin{verbatim}
\keys_define:nn { mymodule }
{ 
  key .multichoices:nn =
  \{ choice-a, choice-b, choice-c \}
  \{ You-gave-choice-\t\use:N \l_keys_choice_tl',-\n    which-is-in-position-\n    \\int_use:N \l_keys_choice_int \c_space_tl
    in-the-list.\n  \}
}
\end{verbatim}

and

\begin{verbatim}
\keys_define:nn { mymodule }
{ 
  key .multichoice:, 
  key / choice-a .code:n = code-a, 
  key / choice-b .code:n = code-b, 
  key / choice-c .code:n = code-c,
}
\end{verbatim}

are valid.

When a multiple choice key is set
\keys_set:nn \{ mymodule \}
{
  key = \{ a , b , c \} \% 'key' defined as a multiple choice
}

Each choice is applied in turn, equivalent to a clist mapping or to applying each value individually:

\keys_set:nn \{ mymodule \}
{
  key = a ,
  key = b ,
  key = c ,
}

Thus each separate choice will have passed to it the \l_keys_choice_tl and \l_keys_choice_int in exactly the same way as described for .choices:nn.

### 27.4 Key usage scope

Some keys will be used as settings which have a strictly limited scope of usage. Some will be only available once, others will only be valid until typesetting begins. To allow formats to support this in a structured way, \l3keys allows this information to be specified using the .usage:n property.

\begin{verbatim}
\keys_set:nn \{ mymodule \}
{
  key = \{ a , b , c \} \% 'key' defined as a multiple choice
}
\end{verbatim}

\begin{verbatim}
\keys_set:nn \{ mymodule \}
{
  key = a ,
  key = b ,
  key = c ,
}
\end{verbatim}

Thus each separate choice will have passed to it the \l_keys_choice_tl and \l_keys_choice_int in exactly the same way as described for .choices:nn.

\begin{verbatim}
\keys_set:nn \{ mymodule \}
{
  key = \{ a , b , c \} \% 'key' defined as a multiple choice
}
\end{verbatim}

Define the (key) to have usage within the (scope), which should be one of general, preamble or load.

\begin{verbatim}
\l_keys_usage_load_prop
\l_keys_usage_preamble_prop
\end{verbatim}

\l3keys itself does not attempt to redefine keys based on the usage scope. Rather, this information is made available with these two property lists. These hold an entry for each module (prefix); the value of each entry is a comma-separated list of the usage-restricted key(s).

### 27.5 Setting keys

\begin{verbatim}
\keys_set:nn \{%module\} \{%keyval list\}
\end{verbatim}

Parses the \{keyval list\}, and sets those keys which are defined for \{module\}. The behaviour on finding an unknown key can be set by defining a special unknown key: this is illustrated later.

\begin{verbatim}
\keys_set:nn \{%module\} \{%keyval list\}
\end{verbatim}

Parse the \{keyval list\}, and sets those keys which are defined for \{module\}. The behaviour on finding an unknown key can be set by defining a special unknown key: this is illustrated later.
For each key processed, information of the full path of the key, the name of the key and the value of the key is available within two string and one token list variables. These may be used within the code of the key.

The path of the key is a “full” description of the key, and is unique for each key. It consists of the module and full key name, thus for example

\keys_set:nn { mymodule } { key-a = some-value }

has path mymodule/key-a while

\keys_set:nn { mymodule } { subset / key-a = some-value }

has path mymodule/subset/key-a. This information is stored in \l_keys_path_str.

The name of the key is the part of the path after the last /, and thus is not unique. In the preceding examples, both keys have name key-a despite having different paths. This information is stored in \l_keys_key_str.

The value is everything after the =, which may be empty if no value was given. This is stored in \l_keys_value_tl, and is not processed in any way by \keys_set:nn.

### 27.6 Handling of unknown keys

If a key has not previously been defined (is unknown), \keys_set:nn looks for a special unknown key for the same module, and if this is not defined raises an error indicating that the key name was unknown. This mechanism can be used for example to issue custom error texts. The unknown key also supports the .default:n property.

\keys_define:nn { mymodule }
{  
unknown .code:n =  
You~tried~to~set~key~'\l_keys_key_str'~to~'#1'. ,  
unknown .default:V = \c_novalue_tl 
}

These functions set keys which are known for the ⟨module⟩, and simply ignore other keys. The \keys_set_known:nn function parses the ⟨keyval list⟩, and sets those keys which are defined for ⟨module⟩. Any keys which are unknown are not processed further by the parser. In addition, \keys_set_known:nnN stores the key–value pairs in the ⟨tl⟩ in comma-separated form (i.e. an edited version of the ⟨keyval list⟩). When a ⟨root⟩ is given (\keys_set_known:nnN), the key–value entries are returned relative to this point in the key tree. When it is absent, only the key name and value are provided. The correct list is returned by nested calls.
27.7 Selective key setting

In some cases it may be useful to be able to select only some keys for setting, even though these keys have the same path. For example, with a set of keys defined using

\keys_define:nn { mymodule }
{   
  key-one .code:n = { \my_func:n {#1} } , 
  key-two .tl_set:N = \l_my_a_tl , 
  key-three .tl_set:N = \l_my_b_tl , 
  key-four .fp_set:N = \l_my_a_fp , 
}

the use of \keys_set:nn attempts to set all four keys. However, in some contexts it may only be sensible to set some keys, or to control the order of setting. To do this, keys may be assigned to \textit{groups}: arbitrary sets which are independent of the key tree. Thus modifying the example to read

\keys_define:nn { mymodule }
{   
  key-one .code:n = { \my_func:n {#1} } , 
  key-one .groups:n = { first } , 
  key-two .tl_set:N = \l_my_a_tl , 
  key-two .groups:n = { first } , 
  key-three .tl_set:N = \l_my_b_tl , 
  key-three .groups:n = { second } , 
  key-four .fp_set:N = \l_my_a_fp , 
}

assigns \texttt{key-one} and \texttt{key-two} to group \texttt{first}, \texttt{key-three} to group \texttt{second}, while \texttt{key-four} is not assigned to a group.

Selective key setting may be achieved either by selecting one or more groups to be made “active”, or by marking one or more groups to be ignored in key setting.

\keys_set_exclude_groups:nn
\keys_set_exclude_groups:nnN
\keys_set_exclude_groups:nnnnN
Sets keys by excluding those in the specified \texttt{groups}. The \texttt{groups} are given as a comma-separated list. Unknown keys are not assigned to any group and are thus always set. The key–value pairs for each key which is filtered out are stored in the \texttt{tl} in a comma-separated form (\textit{i.e.} an edited version of the \texttt{keyval list}). The \texttt{\keys_set_\ldots exclude_groups:nn} version skips this stage.

Use of \texttt{\keys_set_exclude_groups:nnN} can be nested, with the correct residual \texttt{keyval list} returned at each stage. In the version which takes a \texttt{root} argument, the key list is returned relative to that point in the key tree. In the cases without a \texttt{root} argument, only the key names and values are returned.
Activates key filtering in an “opt-in” sense: only keys assigned to one or more of the ⟨groups⟩ specified are set. The ⟨groups⟩ are given as a comma-separated list. Unknown keys are not assigned to any group and are thus never set. The key–value pairs for each key which is filtered out are stored in the ⟨tl⟩ in a comma-separated form (i.e. an edited version of the ⟨keyval list⟩). The \keys_set_groups:nnn version skips this stage.

### 27.8 Digesting keys

\keys_precompile:nnN
\keys_precompile:nnN ⟨module⟩ ⟨keyval list⟩ ⟨tl⟩

New: 2022-03-09
Updated: 2022-01-10

Parses the ⟨keyval list⟩ as for \keys_set:nn, placing the resulting code for those which set variables or functions into the ⟨tl⟩. Thus this function “precompiles” the keyval list into a set of results which can be applied rapidly.

### 27.9 Utility functions for keys

\keys_if_exist_p:nn
\keys_if_exist_p:nn ⟨module⟩ ⟨key⟩

\keys_if_exist_p:ne
\keys_if_exist:nnTF ⟨module⟩ ⟨key⟩ ⟨true code⟩ ⟨false code⟩

\keys_if_exist:nnTF

Tests if the ⟨key⟩ exists for ⟨module⟩, i.e. if any code has been defined for ⟨key⟩.

\keys_if_choice_exist_p:nnn
\keys_if_choice_exist_p:nnn ⟨module⟩ ⟨key⟩ ⟨choice⟩

\keys_if_choice_exist:nnnTF ⟨module⟩ ⟨key⟩ ⟨choice⟩ ⟨true code⟩ ⟨false code⟩

Tests if the ⟨choice⟩ is defined for the ⟨key⟩ within the ⟨module⟩, i.e. if any code has been defined for ⟨key⟩/⟨choice⟩. The test is false if the ⟨key⟩ itself is not defined.

\keys_show:nn
\keys_show:nn ⟨module⟩ ⟨key⟩

Updated: 2015-08-09

Displays in the terminal the information associated to the ⟨key⟩ for a ⟨module⟩, including the function which is used to actually implement it.

\keys_log:nn
\keys_log:nn ⟨module⟩ ⟨key⟩

New: 2014-08-22
Updated: 2015-08-09

Writes in the log file the information associated to the ⟨key⟩ for a ⟨module⟩. See also \keys_show:nn which displays the result in the terminal.
27.10 Low-level interface for parsing key–val lists

To re-cap from earlier, a key–value list is input of the form

\[
\begin{align*}
    \text{KeyOne} &= \text{ValueOne} , \\
    \text{KeyTwo} &= \text{ValueTwo} , \\
    \text{KeyThree}
\end{align*}
\]

where each key–value pair is separated by a comma from the rest of the list, and each key–value pair does not necessarily contain an equals sign or a value! Processing this type of input correctly requires a number of careful steps, to correctly account for braces, spaces and the category codes of separators.

While the functions described earlier are used as a high-level interface for processing such input, in special circumstances you may wish to use a lower-level approach. The low-level parsing system converts a \texttt{(key-value list)} into \texttt{(keys)} and associated \texttt{(values)}. After the parsing phase is completed, the resulting keys and values (or keys alone) are available for further processing. This processing is not carried out by the low-level parser itself, and so the parser requires the names of two functions along with the key–value list. One function is needed to process key–value pairs (it receives two arguments), and a second function is required for keys given without any value (it is called with a single argument).

The parser does not double \# tokens or expand any input. Active tokens = and , appearing at the outer level of braces are converted to category “other” (12) so that the parser does not “miss” any due to category code changes. Spaces are removed from the ends of the keys and values. Keys and values which are given in braces have exactly one set removed (after space trimming), thus

\[
\begin{align*}
    \text{key} &= \{\text{value here}\}, \\
    \text{key} &= \text{value here},
\end{align*}
\]

are treated identically.
\keyval_parse:nnn \keyval_parse:(nnV|nnv)  

\keyval_parse:nnn \{(code_1}\} \{(code_2)\} \{(key-value list)\}

Parses the \{(key-value list)\} into a series of \{(keys\} and associated \{(values\}, or keys alone (if no \{(value\) was given). \{(code_1)\} receives each \{key\} (with no \{(value\}) as a trailing brace group, whereas \{(code_2)\} is appended by two brace groups, the \{key\} and \{value\}. The order of the \{keys\} in the \{(key-value list)\} is preserved. Thus

\keyval_parse:nnn
\{ \use_none:nn \{ code 1 \} \}
\{ \use_none:nnn \{ code 2 \} \}
\{ key1 = value1 \, key2 = value2, key3 = , key4 \}

is converted into an input stream

\use_none:nn \{ code 2 \} \{ key1 \} \{ value1 \}
\use_none:nnn \{ code 2 \} \{ key2 \} \{ value2 \}
\use_none:nnn \{ code 2 \} \{ key3 \} \{ \}
\use_none:nn \{ code 1 \} \{ key4 \}

Note that there is a difference between an empty value (an equals sign followed by nothing) and a missing value (no equals sign at all). Spaces are trimmed from the ends of the \{key\} and \{value\}, then one outer set of braces is removed from the \{key\} and \{value\} as part of the processing. If you need exactly the output shown above, you’ll need to either e-type or x-type expand the function.

\TeXhackers note: The result of each list element is returned within \exp_not:n, which means that the converted input stream does not expand further when appearing in an e-type or x-type argument expansion.
\keyval_parse:NNn \keyval_parse:(NNvNNv)

Parses the \texttt{key–value list} into a series of \texttt{keys} and associated \texttt{values}, or keys alone (if no \texttt{value} was given). \texttt{function} should take one argument, while \texttt{function} should absorb two arguments. After \keyval_parse:NNn has parsed the \texttt{key–value list}, \texttt{function} is used to process keys given with no value and \texttt{function} is used to process keys given with a value. The order of the \texttt{keys} in the \texttt{key–value list} is preserved. Thus

\begin{verbatim}
\keyval_parse:NNn \function:n \function:nn
{ key1 = value1, key2 = value2, key3 = , key4 }
\end{verbatim}

is converted into an input stream

\begin{verbatim}
\function:nn { key1 } { value1 }
\function:nn { key2 } { value2 }
\function:nn { key3 } { }
\function:n { key4 }
\end{verbatim}

Note that there is a difference between an empty value (an equals sign followed by nothing) and a missing value (no equals sign at all). Spaces are trimmed from the ends of the \texttt{key} and \texttt{value}, then one outer set of braces is removed from the \texttt{key} and \texttt{value} as part of the processing.

This shares the implementation of \keyval_parse:nnn, the difference is only semantically.

\TeX hackers note: The result is returned within \texttt{exp_not:n}, which means that the converted input stream does not expand further when appearing in an \emph{e}-type or \emph{x}-type argument expansion.
Chapter 28

The \texttt{l3intarray} module

Fast global integer arrays

For applications requiring heavy use of integers, this module provides arrays which can be accessed in constant time (contrast \texttt{l3seq}, where access time is linear). These arrays have several important features

- The size of the array is fixed and must be given at point of initialisation
- The absolute value of each entry has maximum $2^{30} - 1$ (i.e. one power lower than the usual \texttt{c\_max\_int} ceiling of $2^{31} - 1$)

The use of \texttt{intarray} data is therefore recommended for cases where the need for fast access is of paramount importance.

28.1 Creating and initialising integer array variables

\begin{verbatim}
\intarray_new:Nn \intarray_new:cn
\end{verbatim}

Evaluates the integer expression \texttt{(size)} and allocates an \texttt{(integer array variable)} with that number of (zero) entries. The variable name should start with \texttt{\_g} because assignments are always global.

\begin{verbatim}
\intarray_const_from_clist:Nn \intarray_const_from_clist:cn
\end{verbatim}

Creates a new constant \texttt{(integer array variable)} or raises an error if the name is already taken. The \texttt{(integer array variable)} is set (globally) to contain as its items the results of evaluating each \texttt{(integer expression)} in the \texttt{(comma list)}.

\begin{verbatim}
\intarray_gzero:N \intarray_gzero:cn
\end{verbatim}

Sets all entries of the \texttt{(integer array variable)} to zero. Assignments are always global.
28.2 Adding data to integer arrays

\intarray_gset:Nnn ⟨intarray var⟩ {⟨position⟩} {⟨value⟩}

Stores the result of evaluating the integer expression ⟨value⟩ into the ⟨integer array variable⟩ at the (integer expression) ⟨position⟩. If the ⟨position⟩ is not between 1 and the \intarray_count:N, or the ⟨value⟩’s absolute value is bigger than \(2^{30} - 1\), an error occurs. Assignments are always global.

28.3 Counting entries in integer arrays

\intarray_count:N ⟨intarray var⟩

Expands to the number of entries in the ⟨integer array variable⟩. Contrarily to \seq_count:N this is performed in constant time.

28.4 Using a single entry

\intarray_item:Nn ⟨intarray var⟩ {⟨position⟩}

Expands to the integer entry stored at the (integer expression) ⟨position⟩ in the ⟨integer array variable⟩. If the ⟨position⟩ is not between 1 and the \intarray_count:N, an error occurs.

\intarray_rand_item:N ⟨intarray var⟩

Selects a pseudo-random item of the ⟨integer array⟩. If the ⟨integer array⟩ is empty, produce an error.

28.5 Integer array conditional

\intarray_if_exist_p:N ⟨intarray var⟩
\intarray_if_exist_p:c *
\intarray_if_exist:NTF ⟨intarray var⟩ {⟨true code⟩} {⟨false code⟩}
\intarray_if_exist:CTF *

Tests whether the ⟨intarray var⟩ is currently defined. This does not check that the ⟨intarray var⟩ really is an integer array variable.

28.6 Viewing integer arrays

\intarray_show:N ⟨intarray var⟩
\intarray_log:N ⟨intarray var⟩

Displays the items in the ⟨integer array variable⟩ in the terminal or writes them in the log file.
28.7 Implementation notes

It is a wrapper around the \fontdimen primitive, used to store arrays of integers (with a restricted range: absolute value at most $2^{30} - 1$). In contrast to \l3seq sequences the access to individual entries is done in constant time rather than linear time, but only integers can be stored. More precisely, the primitive \fontdimen stores dimensions but the \l3intarray module transparently converts these from/to integers. Assignments are always global.

While LuaTeX’s memory is extensible, other engines can “only” deal with a bit less than $4 \times 10^6$ entries in all \fontdimen arrays combined (with default \TeX Live settings).
Chapter 29

The l3fp module
Floating points

A decimal floating point number is one which is stored as a significand and a separate exponent. The module implements expandably a wide set of arithmetic, trigonometric, and other operations on decimal floating point numbers, to be used within floating point expressions. Floating point expressions ("\langle fp expr \rangle") support the following operations with their usual precedence.

- Basic arithmetic: addition \(x + y\), subtraction \(x - y\), multiplication \(x \ast y\), division \(x/y\), square root \(\sqrt{x}\), and parentheses.
- Comparison operators: \(x < y\), \(x <= y\), \(x >? y\), \(x ! = y\) etc.
- Boolean logic: sign \(\text{sign } x\), negation \(! x\), conjunction \(x \&\& y\), disjunction \(x || y\), ternary operator \(x ? y : z\).
- Exponentials: \(\exp x\), \(\ln x\), \(x^y\), \(\log_b x\).
- Integer factorial: \(\text{fact } x\).
- Trigonometry: \(\sin x\), \(\cos x\), \(\tan x\), \(\cot x\), \(\sec x\), \(\csc x\) expecting their arguments in radians, and \(\sin d x\), \(\cos d x\), \(\tan d x\), \(\cot d x\), \(\sec d x\), \(\csc d x\) expecting their arguments in degrees.
- Inverse trigonometric functions: \(\text{asin } x\), \(\text{acos } x\), \(\text{atan } x\), \(\text{acot } x\), \(\text{asec } x\), \(\text{acsc } x\) giving a result in radians, and \(\text{asind } x\), \(\text{acosd } x\), \(\text{atand } x\), \(\text{acotd } x\), \(\text{asecd } x\), \(\text{acscd } x\) giving a result in degrees.

(not yet) Hyperbolic functions and their inverse functions: \(\sinh x\), \(\cosh x\), \(\tanh x\), \(\coth x\), \(\text{sech } x\), \(\text{csch } x\), and \(\text{asinh } x\), \(\text{acosh } x\), \(\text{atanh } x\), \(\text{acoth } x\), \(\text{asech } x\), \(\text{acsch } x\).

- Extrema: \(\max(x_1, x_2, \ldots)\), \(\min(x_1, x_2, \ldots)\), \(\text{abs}(x)\).
- Rounding functions, controlled by two optional values, \(n\) (number of places, 0 by default) and \(t\) (behavior on a tie, \text{nan} by default):
  - \(\text{trunc}(x, n)\) rounds towards zero,
  - \(\text{floor}(x, n)\) rounds towards \(-\infty\),

258
ceil(x, n) rounds towards $+\infty$,
round(x, n, t) rounds to the closest value, with ties rounded to an even value
by default, towards zero if $t = 0$, towards $+\infty$ if $t > 0$ and towards $-\infty$ if $t < 0$.

And (not yet) modulo, and “quantize”.

- Random numbers: rand(), randint(m, n).
- Constants: pi, deg (one degree in radians).
- Dimensions, automatically expressed in points, e.g., pc is 12.
- Automatic conversion (no need for \texttt{⟨type⟩_use:N}) of integer, dimension, and skip
  variables to floating point numbers, expressing dimensions in points and ignoring
  the stretch and shrink components of skips.
- Tuples: $(x_1, \ldots, x_n)$ that can be stored in variables, added together, multiplied or
divided by a floating point number, and nested.

Floating point numbers can be given either explicitly (in a form such as 1.234e-34, or
−.0001), or as a stored floating point variable, which is automatically replaced by its
current value. A “floating point” is a floating point number or a tuple thereof. See section
29.12.1 for a description of what a floating point is, section 29.12.2 for details about
how an expression is parsed, and section 29.12.3 to know what the various operations do.
Some operations may raise exceptions (error messages), described in section 29.10.

An example of use could be the following.

\LaTeX{} can now compute: $\frac{\sin (3.5)}{2} + 2\cdot 10^{-3}$
= \ExplSyntaxOn \fp_to_decimal:n {\sin(3.5)/2 + 2e-3} \$.\]

The operation round can be used to limit the result’s precision. Adding +0 avoids the
possibly undesirable output −0, replacing it by +0. However, the \texttt{l3fp} module is mostly
meant as an underlying tool for higher-level commands. For example, one could provide
a function to typeset nicely the result of floating point computations.

\documentclass{article}
\usepackage{siunitx}
\ExplSyntaxOn
\NewDocumentCommand { \calcnum } { m } { \num { \fp_to_scientific:n {#1} } }
\ExplSyntaxOff
\begin{document}
\calcnum { 2 \pi * \sin ( 2.3 ^ 5 ) }
\end{document}

See the documentation of siunitx for various options of \texttt{num}.
29.1 Creating and initialising floating point variables

\fp_new:N \langle fp var \rangle
\fp_new:c

Creates a new \langle fp var \rangle or raises an error if the name is already taken. The declaration is global. The \langle fp var \rangle is initially +0.

\fp_const:Nn \langle fp var \rangle \{ \langle fp expr \rangle \}
\fp_const:cn

Creates a new constant \langle fp var \rangle or raises an error if the name is already taken. The \langle fp var \rangle is set globally equal to the result of evaluating the \langle fp expr \rangle.

\fp_zero:N \langle fp var \rangle
\fp_zero:c
\fp_gzero:N
\fp_gzero:c

Sets the \langle fp var \rangle to +0.

\fp_zero_new:N \langle fp var \rangle
\fp_zero_new:c
\fp_gzero_new:N
\fp_gzero_new:c

Ensures that the \langle fp var \rangle exists globally by applying \fp_new:N if necessary, then applies \fp_(g)zero:N to leave the \langle fp var \rangle set to +0.

29.2 Setting floating point variables

\fp_set:Nn \langle fp var \rangle \{ \langle fp expr \rangle \}
\fp_set:cn
\fp_gset:Nn
\fp_gset:cn

Sets \langle fp var \rangle equal to the result of computing the \langle fp expr \rangle.

\fp_set_eq:NN \langle fp var_1 \rangle \langle fp var_2 \rangle
\fp_set_eq:cn(\langle fp var_1 \rangle\langle fp var_2 \rangle)
\fp_gset_eq:NN \langle fp var_1 \rangle \langle fp var_2 \rangle
\fp_gset_eq:cn(\langle fp var_1 \rangle\langle fp var_2 \rangle)

Sets the floating point variable \langle fp var_1 \rangle equal to the current value of \langle fp var_2 \rangle.

\fp_add:Nn \langle fp var \rangle \{ \langle fp expr \rangle \}
\fp_add:cn
\fp_gadd:Nn
\fp_gadd:cn

Adds the result of computing the \langle fp expr \rangle to the \langle fp var \rangle. This also applies if \langle fp var \rangle and \langle floating point expression \rangle evaluate to tuples of the same size.
\(\text{fp_sub:}\text{Nn}\) \(\text{fp_sub:}\text{cn}\)\n\(\text{fp_gsub:}\text{Nn}\) \(\text{fp_gsub:}\text{cn}\)\n
Subtracts the result of computing the \(\text{floating point expression}\) from the \(\text{fp var}\). This also applies if \(\text{fp var}\) and \(\text{floating point expression}\) evaluate to tuples of the same size.

New: 2012-05-08
Updated: 2012-07-08

\subsection*{29.3 Using floating points}

\(\text{fp_eval:}\text{n}\) \(\text{fp_eval:}\text{cn}\)\n\(\text{fp_to_decimal:}\text{N}\) \(\text{fp_to_decimal:}\text{c}\) \(\text{fp_to_decimal:}\text{n}\)\n
Evaluates the \(\text{fp expr}\) and expresses the result as a decimal number with no exponent. Leading or trailing zeros may be inserted to compensate for the exponent. Non-significant trailing zeros are trimmed, and integers are expressed without a decimal separator. The values \(\pm\infty\) and \(\text{nan}\) trigger an “invalid operation” exception. For a tuple, each item is converted using \(\text{fp_eval:}\text{n}\) and they are combined as \((\langle fp_1 \rangle, \langle fp_2 \rangle, \ldots, \langle fp_n \rangle)\) if \(n > 1\) and \((\langle fp_1 \rangle,)\) or (\) for fewer items. This function is identical to \(\text{fp_to_decimal:}\text{n}\).

New: 2012-05-08
Updated: 2012-07-08

\(\text{fp_sign:}\text{n}\) \(\text{fp_sign:}\text{cn}\)\n
Evaluates the \(\text{fp expr}\) and leaves its sign in the input stream using \(\text{fp_eval:}\text{n}\) \(\text{sign(}\langle\text{result}\rangle)\): +1 for positive numbers and for \(+\infty\), -1 for negative numbers and for \(-\infty\), \(\pm0\) for \(\pm0\). If the operand is a tuple or is \(\text{nan}\), then “invalid operation” occurs and the result is 0.

Updated: 2018-11-03

\(\text{fp_to_dim:}\text{N}\) \(\text{fp_to_dim:}\text{c}\) \(\text{fp_to_dim:}\text{n}\)\n
Evaluates the \(\text{fp expr}\) and expresses the result as a dimension (in pt) suitable for use in dimension expressions. The output is identical to \(\text{fp_to_decimal:}\text{n}\), with an additional trailing pt (both letter tokens). In particular, the result may be outside the range \([-2^{14} + 2^{-17}, 2^{14} - 2^{-17}]\) of valid \TeX{} dimensions, leading to overflow errors if used as a dimension. Tuples, as well as the values \(\pm\infty\) and \(\text{nan}\), trigger an “invalid operation” exception.

Updated: 2016-03-22

\(\text{fp_to_int:}\text{N}\) \(\text{fp_to_int:}\text{c}\) \(\text{fp_to_int:}\text{n}\)\n
Evaluates the \(\text{fp expr}\), and rounds the result to the closest integer, rounding exact ties to an even integer. The result may be outside the range \([-2^{31} + 1, 2^{31} - 1]\) of valid \TeX{} integers, leading to overflow errors if used in an integer expression. Tuples, as well as the values \(\pm\infty\) and \(\text{nan}\), trigger an “invalid operation” exception.

Updated: 2012-07-08

261
\fp_to_scientific:N * \fp_to_scientific:N \fp \var \\
\fp_to_scientific:c * \fp_to_scientific:n \{\fp \expr\} \\
\fp_to_scientific:n * 

Evaluates the \(\fp \expr\) and expresses the result in scientific notation: 
\[
\text{optional } - \langle \text{digit} \rangle. (15 \text{ digits}) e \langle \text{optional sign} \rangle \langle \text{exponent} \rangle
\]
The leading \(\langle \text{digit} \rangle\) is non-zero except in the case of \(\pm 0\). The values \(\pm \infty\) and \text{nan} trigger an “invalid operation” exception. Normal category codes apply: thus the \(e\) is category code 11 (a letter). For a tuple, each item is converted using \\fp_to_scientific:n and they are combined as \(\langle \fp_1 \rangle, \langle \fp_2 \rangle, \ldots \langle \fp_n \rangle\) if \(n > 1\) and \(\langle \fp_1 \rangle\) or \(\langle \rangle\) for fewer items.

\fp_to_tl:N * \fp_to_tl:N \fp \var \\
\fp_to_tl:c * \fp_to_tl:n \{\fp \expr\} \\
\fp_to_tl:n * 

Evaluates the \(\fp \expr\) and expresses the result in (almost) the shortest possible form. Numbers in the ranges \((0, 10^{-3})\) and \([10^{16}, \infty)\) are expressed in scientific notation with trailing zeros trimmed and no decimal separator when there is a single significant digit (this differs from \\fp_to_scientific:n). Numbers in the range \([10^{-3}, 10^{16}]\) are expressed in a decimal notation without exponent, with trailing zeros trimmed, and no decimal separator for integer values (see \\fp_to_decimal:n). Negative numbers start with \(-\). The special values \(\pm 0, \pm \infty\) and \text{nan} are rendered as \(0, -0, \text{inf}, -\text{inf},\) and \text{nan} respectively. Normal category codes apply and thus \text{inf} or \text{nan}, if produced, are made up of letters. For a tuple, each item is converted using \\fp_to_tl:n and they are combined as \(\langle \fp_1 \rangle, \langle \fp_2 \rangle, \ldots \langle \fp_n \rangle\) if \(n > 1\) and \(\langle \fp_1 \rangle\) or \(\langle \rangle\) for fewer items. This function is identical to \\fp_to_-decimal:n.

\fp_use:N * \fp_use:N \fp \var \\
\fp_use:c * \\
\fp_use:N 

Inserts the value of the \(\fp \var\) into the input stream as a decimal number with no exponent. Leading or trailing zeros may be inserted to compensate for the exponent. Non-significant trailing zeros are trimmed. Integers are expressed without a decimal separator. The values \(\pm \infty\) and \text{nan} trigger an “invalid operation” exception. For a tuple, each item is converted using \\fp_to_decimal:n and they are combined as \(\langle \fp_1 \rangle, \langle \fp_2 \rangle, \ldots \langle \fp_n \rangle\) if \(n > 1\) and \(\langle \fp_1 \rangle\) or \(\langle \rangle\) for fewer items. This function is identical to \\fp_to_-decimal:n.

29.4 Floating point conditionals

\fp_if_exist_p:N * \fp_if_exist_p:N \fp \var \\
\fp_if_exist_p:c * \fp_if_exist_p:NTF \fp \var \{\langle \text{true code} \rangle\} \{\langle \text{false code} \rangle\} \\
\fp_if_exist:NTF * \fp_if_exist:cTF * 

Tests whether the \(\fp \var\) is currently defined. This does not check that the \(\fp \var\) really is a floating point variable.
Compares the \( \langle \text{fp expr}_1 \rangle \) and the \( \langle \text{fp expr}_2 \rangle \), and returns true if the \( \langle \text{relation} \rangle \) is obeyed. Two floating points \( x \) and \( y \) may obey four mutually exclusive relations: \( x < y \), \( x = y \), \( x > y \), or \( x ? y \) (“not ordered”). The last case occurs exactly if one or both operands is \text{nan} or is a tuple, unless they are equal tuples. Note that a \text{nan} is distinct from any value, even another \text{nan}, hence \( x = x \) is not true for a \text{nan}. To test if a value is \text{nan}, compare it to an arbitrary number with the “not ordered” relation.

\[
\text{\texttt{\textbackslash fp\_compare\_nNnTF \{ \langle value\rangle \} \ ? \ \{ \ 0 \ \}}}
\]
\[
\{ \} \ % \ \text{\langle value\rangle} \ \text{is} \ \text{nan}
\]
\[
\{ \} \ % \ \text{\langle value\rangle} \ \text{is} \ \text{not} \ \text{nan}
\]

Tuples are equal if they have the same number of items and items compare equal (in particular there must be no \text{nan}). At present any other comparison with tuples yields \( ? \) (not ordered). This is experimental.

This function is less flexible than \texttt{\textbackslash fp\_compare\_nTF} but slightly faster. It is provided for consistency with \texttt{\textbackslash int\_compare\_nNnTF} and \texttt{\textbackslash dim\_compare\_nNnTF}. 
\fp_compare_p:n + \fp_compare_p:n \{ \fp expr_1 \langle \text{relation}_1 \rangle 
\ldots 
\fp expr_N \langle \text{relation}_N \rangle 
\fp expr_{N+1} \rangle \}

\fp_compare:nTF \{ \fp expr_1 \langle \text{relation}_1 \rangle 
\ldots 
\fp expr_N \langle \text{relation}_N \rangle 
\fp expr_{N+1} \rangle \}
{ \langle \text{true code} \rangle \} { \langle \text{false code} \rangle \}

Evaluates the \{\text{exprs}\} as described for \fp_eval:n and compares consecutive result using the corresponding \langle \text{relation} \rangle, namely it compares \langle \text{expr}_1 \rangle and \langle \text{expr}_2 \rangle using the \langle \text{relation}_1 \rangle, then \langle \text{expr}_2 \rangle and \langle \text{expr}_3 \rangle using the \langle \text{relation}_2 \rangle, until finally comparing \langle \text{expr}_N \rangle and \langle \text{expr}_{N+1} \rangle using the \langle \text{relation}_N \rangle. The test yields \text{true} if all comparisons are \text{true}. Each \langle \text{floating point expression} \rangle is evaluated only once. Contrarily to \int_compare:nTF, all \langle \text{exprs} \rangle are computed, even if one comparison is \text{false}. Two floating points \text{x} and \text{y} may obey four mutually exclusive relations: \text{x < y}, \text{x = y}, \text{x > y}, or \text{x?y} ("not ordered"). The last case occurs exactly if one or both operands is \text{nan} or is a tuple, unless they are equal tuples. Each \langle \text{relation} \rangle can be any (non-empty) combination of <, =, >, and ?, plus an optional leading ! (which negates the \langle \text{relation} \rangle), with the restriction that the \langle \text{relation} \rangle may not start with ?, as this symbol has a different meaning (in combination with :) within floating point expressions. The comparison \text{x \langle relation \rangle y} is then \text{true} if the \langle \text{relation} \rangle does not start with ! and the actual relation (<, =, >, or ?) between \text{x} and \text{y} appears within the \langle \text{relation} \rangle, or on the contrary if the \langle \text{relation} \rangle starts with ! and the relation between \text{x} and \text{y} does not appear within the \langle \text{relation} \rangle. Common choices of \langle \text{relation} \rangle include >= (greater or equal), != (not equal), !? or <=> (comparable).

This function is more flexible than \fp_compare:nNnTF and only slightly slower.

\fp_if_nan_p:n + \fp_if_nan_p:n \{ \langle \text{expr} \rangle \}
\fp_if_nan:nTF + \fp_if_nan:nTF \{ \langle \text{expr} \rangle \} \{ \langle \text{true code} \rangle \} \{ \langle \text{false code} \rangle \}

Evaluates the \langle \text{expr} \rangle and tests whether the result is exactly \text{nan}. The test returns \text{false} for any other result, even a tuple containing \text{nan}.

\fp_do_until:nNnn + \fp_do_until:nNnn \{ \langle \text{expr}_1 \rangle \langle \text{relation} \rangle \{ \langle \text{expr}_2 \rangle \} \{ \langle \text{code} \rangle \}

Places the \langle \text{code} \rangle in the input stream for \TeX to process, and then evaluates the relationship between the two \langle \text{floating point expressions} \rangle as described for \fp_compare:nNnTF. If the test is \text{false} then the \langle \text{code} \rangle is inserted into the input stream again and a loop occurs until the \langle \text{relation} \rangle is \text{true}.

### 29.5 Floating point expression loops
\fp_do_while:nNn \{\langle code\rangle\} \{\langle relation\rangle\} \{\langle fp\ expr_1\rangle\} \{\langle fp\ expr_2\rangle\}

Places the \langle code\rangle in the input stream for \TeX to process, and then evaluates the relationship between the two \langle floating point expressions\rangle as described for \fp_compare:nNnTF. If the test is \texttt{true} then the \langle code\rangle is inserted into the input stream again and a loop occurs until the \langle relation\rangle is \texttt{false}.

\fp_until_do:nNn \{\langle relation\rangle\} \{\langle code\rangle\} \{\langle fp\ expr_1\rangle\} \{\langle fp\ expr_2\rangle\}

Evaluates the relationship between the two \langle floating point expressions\rangle as described for \fp_compare:nNnTF, and then places the \langle code\rangle in the input stream if the \langle relation\rangle is \texttt{false}. After the \langle code\rangle has been processed by \TeX the test is repeated, and a loop occurs until the test is \texttt{true}.

\fp_do_while:nNn \{\langle code\rangle\} \{\langle relation\rangle\} \{\langle fp\ expr_1\rangle\} \{\langle fp\ expr_2\rangle\}

Evaluates the relationship between the two \langle floating point expressions\rangle as described for \fp_compare:nNnTF, and then places the \langle code\rangle in the input stream if the \langle relation\rangle is \texttt{false}. After the \langle code\rangle has been processed by \TeX the test is repeated, and a loop occurs until the test is \texttt{true}.

\fp_until_do:nn \{\langle relation\rangle\} \{\langle code\rangle\} \{\langle fp\ expr_1\rangle\} \{\langle fp\ expr_2\rangle\}

Evaluates the relationship between the two \langle floating point expressions\rangle as described for \fp_compare:nNnTF, and then places the \langle code\rangle in the input stream if the \langle relation\rangle is \texttt{false}. After the \langle code\rangle has been processed by \TeX the test is repeated, and a loop occurs until the test is \texttt{true}.

\fp_while_do:nn \{\langle relation\rangle\} \{\langle code\rangle\} \{\langle fp\ expr_1\rangle\} \{\langle fp\ expr_2\rangle\}

Evaluates the relationship between the two \langle floating point expressions\rangle as described for \fp_compare:nNnTF, and then places the \langle code\rangle in the input stream if the \langle relation\rangle is \texttt{true}. After the \langle code\rangle has been processed by \TeX the test is repeated, and a loop occurs until the test is \texttt{false}.
This function first evaluates the \( \langle \text{initial value} \rangle \), \( \langle \text{step} \rangle \) and \( \langle \text{final value} \rangle \), each of which should be a floating point expression evaluating to a floating point number, not a tuple. The \( \langle \text{function} \rangle \) is then placed in front of each \( \langle \text{value} \rangle \) from the \( \langle \text{initial value} \rangle \) to the \( \langle \text{final value} \rangle \) in turn (using \( \langle \text{step} \rangle \) between each \( \langle \text{value} \rangle \)). The \( \langle \text{step} \rangle \) must be non-zero. If the \( \langle \text{step} \rangle \) is positive, the loop stops when the \( \langle \text{value} \rangle \) becomes larger than the \( \langle \text{final value} \rangle \). If the \( \langle \text{step} \rangle \) is negative, the loop stops when the \( \langle \text{value} \rangle \) becomes smaller than the \( \langle \text{final value} \rangle \). The \( \langle \text{function} \rangle \) should absorb one numerical argument. For example

\[
\text{\texttt{cs_set:Npn \my_func:n #1 { \{I saw #1\} quad \}}}
\]

\[
\text{\texttt{\fp_step_function:nnnN \{ 1.0 \} \{0.1\} \{1.5\} my_func:n}}
\]

would print

\[
[I \text{ saw } 1.0] \quad [I \text{ saw } 1.1] \quad [I \text{ saw } 1.2] \quad [I \text{ saw } 1.3] \quad [I \text{ saw } 1.4] \quad [I \text{ saw } 1.5]
\]

\TeX\text{Xhacker} \text{ note:} Due to rounding, it may happen that adding the \( \langle \text{step} \rangle \) to the \( \langle \text{value} \rangle \) does not change the \( \langle \text{value} \rangle \); such cases give an error, as they would otherwise lead to an infinite loop.

This function first evaluates the \( \langle \text{initial value} \rangle \), \( \langle \text{step} \rangle \) and \( \langle \text{final value} \rangle \), all of which should be floating point expressions evaluating to a floating point number, not a tuple. Then for each \( \langle \text{value} \rangle \) from the \( \langle \text{initial value} \rangle \) to the \( \langle \text{final value} \rangle \) in turn (using \( \langle \text{step} \rangle \) between each \( \langle \text{value} \rangle \)), the \( \langle \text{code} \rangle \) is inserted into the input stream with \#1 replaced by the current \( \langle \text{value} \rangle \). Thus the \( \langle \text{code} \rangle \) should define a function of one argument (\#1).

This function first evaluates the \( \langle \text{initial value} \rangle \), \( \langle \text{step} \rangle \) and \( \langle \text{final value} \rangle \), all of which should be floating point expressions evaluating to a floating point number, not a tuple. Then for each \( \langle \text{value} \rangle \) from the \( \langle \text{initial value} \rangle \) to the \( \langle \text{final value} \rangle \) in turn (using \( \langle \text{step} \rangle \) between each \( \langle \text{value} \rangle \)), the \( \langle \text{code} \rangle \) is inserted into the input stream, with the \( \langle \text{tl var} \rangle \) defined as the current \( \langle \text{value} \rangle \). Thus the \( \langle \text{code} \rangle \) should make use of the \( \langle \text{tl var} \rangle \).

### 29.6 Symbolic expressions

Floating point expressions support variables: these can only be set locally, so act like standard \( \&\ldots \) variables.

\[
\text{\texttt{\fp_new_variable:n \{ A \}}} \\
\text{\texttt{\fp_set:Nn \_tmpbf_fp \{ 1 * sin(A) + 3**2 \}}} \\
\text{\texttt{\fp_show:n \_tmpbf_fp}} \\
\text{\texttt{\fp_show:N \_tmpbf_fp}} \\
\text{\texttt{\fp_set_variable:nn \{ A \} \{ pi/2 \}}}
\]
defines \( A \) to be a variable, then defines \( \l_{\text{tmpb}} \_fp \) to stand for \( 1 \ast \sin(A) + 9 \) (note that \( 3^{\ast}2 \) is evaluated, but the \( 1 \ast \) product is not simplified away). Until \( \l_{\text{tmpb}} \_fp \) is changed, \( \fp_{\text{show}}:N\ \l_{\text{tmpb}} \_fp \) will show \( ((1 \ast \sin(A)) + 9) \) regardless of the value of \( A \). The next step defines \( A \) to be equal to \( \pi / 2 \): then \( \fp_{\text{show}}:n\ \{ \l_{\text{tmpb}} \_fp \} \) will evaluate \( \l_{\text{tmpb}} \_fp \) and show 10. We then redefine \( A \) to be 0: since \( \l_{\text{tmpb}} \_fp \) still stands for \( 1 \ast \sin(A) + 9 \), the value shown is then 9. Variables can be set with \( \fp_{\text{set variable}}:nn \) to arbitrary floating point expressions including other variables.

\[
\fp_{\text{new variable}}:n \ {\langle \text{identifier} \rangle}
\]

Declares the \( \langle \text{identifier} \rangle \) as a variable, which allows it to be used in floating point expressions. For instance,

\[
\fp_{\text{new variable}}:n \ {\langle A \rangle} \ \\ \fp_{\text{show}}:n \ {\langle A**2 - A + 1 \rangle}
\]

shows \( (((A^2) - A) + 1) \). If the declaration was missing, the parser would complain about an “Unknown fp word ’A’”. The \( \langle \text{identifier} \rangle \) must consist entirely of Latin letters among \([a-zA-Z]\).

\[
\fp_{\text{set variable}}:nn \ {\langle \text{identifier} \rangle} \ {\langle \text{fp expr} \rangle}
\]

Defines the \( \langle \text{identifier} \rangle \) to stand in any further expression for the result of evaluating the \( \langle \text{floating point expression} \rangle \) as much as possible. The result may contain other variables, which are then replaced by their values if they have any. For instance,

\[
\fp_{\text{new variable}}:n \ {\langle A \rangle} \ \\ \fp_{\text{new variable}}:n \ {\langle B \rangle} \ \\ \fp_{\text{new variable}}:n \ {\langle C \rangle} \ \\ \fp_{\text{set variable}}:nn \ {\langle A \rangle} \ {\langle 3 \rangle} \ \\ \fp_{\text{set variable}}:nn \ {\langle C \rangle} \ {\langle A**2 + B + 1 \rangle} \ \\ \fp_{\text{show}}:n\ \{ C + 4 \} \ \\ \fp_{\text{set variable}}:nn \ {\langle A \rangle} \ {\langle 4 \rangle} \ \\ \fp_{\text{show}}:n\ \{ C + 4 \}
\]

shows \( ((9 + (B + 1)) + 4) \) twice: changing the value of \( A \) to 4 does not alter \( C \) because \( A \) was replaced by its value 3 when evaluating \( A**2+B+1 \).
\fp_clear_variable:n \{ \textit{identifier} \} \\
Removes any value given by \fp_set_variable:nn to the variable with this \textit{identifier}.

For instance,
\begin{verbatim}
\fp_new_variable:n \{ A \}
\fp_set_variable:nn \{ A \} \{ 3 \}
\fp_show:n \{ A \^{} 2 \}
\fp_clear_variable:n \{ A \}
\fp_show:n \{ A \^{} 2 \}
\end{verbatim}
shows 9, then \((A^2)\).

29.7 User-defined functions

It is possible to define new user functions which can be used inside the argument to \fp_eval:n, etc. These functions may take one or more named arguments, and should be implemented using expansion methods only.

\fp_new_function:n \{ \textit{identifier} \} \\
Declares the \textit{identifier} as a function, which allows it to be used in floating point expressions. For instance,
\begin{verbatim}
\fp_new_function:n \{ foo \}
\fp_show:n \{ foo \ ( 1 + 2 , foo(3), A ) \^{} 2 \}
\end{verbatim}
shows \((\text{foo}(3, \text{foo}(3), A))^2\). If the declaration was missing, the parser would complain about an “Unknown fp word ‘foo’”. The \textit{identifier} must consist entirely of Latin letters \([a-zA-Z]\).

\fp_set_function:nnn \{ \textit{identifier} \} \{ \textit{vars} \} \{ \textit{fp expr} \} \\
Defines the \textit{identifier} to stand in any further expression for the result of evaluating the \textit{floating point expression}, with the \textit{identifier} accepting the \textit{vars} (a non-empty comma-separated list). The result may contain other functions, which are then replaced by their results if they have any. For instance,
\begin{verbatim}
\fp_new_function:n \{ foo \}
\fp_set_function:nnn \{ npow \} \{ a,b \} \{ a\^{}b \}
\fp_show:n \{ npow(16,0.25) \}
\end{verbatim}
shows 2. The names of the \textit{vars} must consist entirely of Latin letters \([a-zA-Z]\), but are otherwise not restricted: in particular, they are independent of any variables declared by \fp_new_variable:n.

\fp_clear_function:n \{ \textit{identifier} \} \\
Removes any definition given by \fp_set_function:nnn to the function with this \textit{identifier}.
29.8 Some useful constants, and scratch variables

\c_zero_fp
\c_minus_zero_fp
New: 2012-05-08

Zero, with either sign.

\c_one_fp
New: 2012-05-08

One as an \texttt{fp}: useful for comparisons in some places.

\c_inf_fp
\c_minus_inf_fp
New: 2012-05-08

Infinity, with either sign. These can be input directly in a floating point expression as \texttt{inf} and \texttt{-inf}.

\c_nan_fp
New: 2012-05-08

Not a number. This can be input directly in a floating point expression as \texttt{nan}.

\c_e_fp
Updated: 2012-05-08

The value of the base of the natural logarithm, $e = \exp(1)$.

\c_pi_fp
Updated: 2013-11-17

The value of $\pi$. This can be input directly in a floating point expression as \texttt{pi}.

\c_one_degree_fp
New: 2012-05-08
Updated: 2013-11-17

The value of $1^\circ$ in radians. Multiply an angle given in degrees by this value to obtain a result in radians. Note that trigonometric functions expecting an argument in radians or in degrees are both available. Within floating point expressions, this can be accessed as \texttt{deg}.

29.9 Scratch variables

\l_tmpa_fp
\l_tmpb_fp

Scratch floating points for local assignment. These are never used by the kernel code, and so are safe for use with any \LaTeX3-defined function. However, they may be overwritten by other non-kernel code and so should only be used for short-term storage.

\g_tmpa_fp
\g_tmpb_fp

Scratch floating points for global assignment. These are never used by the kernel code, and so are safe for use with any \LaTeX3-defined function. However, they may be overwritten by other non-kernel code and so should only be used for short-term storage.
29.10 Floating point exceptions

The functions defined in this section are experimental, and their functionality may be altered or removed altogether.

"Exceptions" may occur when performing some floating point operations, such as 0 / 0, or 10 ** 1e9999. The relevant IEEE standard defines 5 types of exceptions, of which we implement 4.

- **Overflow** occurs whenever the result of an operation is too large to be represented as a normal floating point number. This results in ±∞.

- **Underflow** occurs whenever the result of an operation is too close to 0 to be represented as a normal floating point number. This results in ±0.

- **Invalid operation** occurs for operations with no defined outcome, for instance 0/0 or sin(∞), and results in a nan. It also occurs for conversion functions whose target type does not have the appropriate infinite or nan value (e.g., \fp_to_dim:n).

- **Division by zero** occurs when dividing a non-zero number by 0, or when evaluating functions at poles, e.g., ln(0) or cot(0). This results in ±∞.

*(not yet) Inexact* occurs whenever the result of a computation is not exact, in other words, almost always. At the moment, this exception is entirely ignored in \TeX3.

To each exception we associate a “flag": \l_fp_overflow_flag, \l_fp_underflow_flag, \l_fp_invalid_operation_flag and \l_fp_division_by_zero_flag. The state of these flags can be tested and modified with commands from \l3flag.

By default, the “invalid operation” exception triggers an (expandable) error, and raises the corresponding flag. Other exceptions raise the flag but do not trigger an error. The behaviour when an exception occurs can be modified (using \fp_trap:nn) to either produce an error and raise the flag, or only raise the flag, or do nothing at all.

\[\text{\texttt{\fp_trap:nn } \langle \text{exception} \rangle \{\langle \text{trap type} \rangle \}}\]

All occurrences of the \langle\text{exception}\rangle (overflow, underflow, invalid_operation or division_by_zero) within the current group are treated as \langle\text{trap type}\rangle, which can be

- **none**: the \langle\text{exception}\rangle will be entirely ignored, and leave no trace;
- **flag**: the \langle\text{exception}\rangle will turn the corresponding flag on when it occurs;
- **error**: additionally, the \langle\text{exception}\rangle will halt the TeX run and display some information about the current operation in the terminal.

This function is experimental, and may be altered or removed.

\l_fp_overflow_flag
\l_fp_underflow_flag
\l_fp_invalid_operation_flag
\l_fp_division_by_zero_flag

Flags denoting the occurrence of various floating-point exceptions.
29.11 Viewing floating points

\texttt{\textbackslash fp\_show:N} \langle \texttt{fp\_var} \rangle \texttt{\textbackslash fp\_show:n \{\texttt{fp\_expr}\}}

Evaluates the \texttt{\langle fp\_expr \rangle} and displays the result in the terminal.

\texttt{\textbackslash fp\_log:N} \langle \texttt{fp\_var} \rangle \texttt{\textbackslash fp\_log:n \{\texttt{fp\_expr}\}}

Evaluates the \texttt{\langle fp\_expr \rangle} and writes the result in the log file.

New: 2012-05-08
Updated: 2021-04-29

29.12 Floating point expressions

29.12.1 Input of floating point numbers

We support four types of floating point numbers:

- \(\pm m \cdot 10^n\), a floating point number, with integer \(1 \leq m \leq 10^{16}\), and \(-10000 \leq n \leq 10000\);
- \(\pm 0\), zero, with a given sign;
- \(\pm \infty\), infinity, with a given sign;
- \texttt{nan}, is “not a number”, and can be either quiet or signalling (\textit{not yet}: this distinction is currently unsupported);

Normal floating point numbers are stored in base 10, with up to 16 significant figures. On input, a normal floating point number consists of:

- \texttt{\langle sign \rangle}: a possibly empty string of \texttt{+} and \texttt{-} characters;
- \texttt{\langle significand \rangle}: a non-empty string of digits together with zero or one dot;
- \texttt{\langle exponent \rangle} optionally: the character e or E, followed by a possibly empty string of \texttt{+} and \texttt{-} tokens, and a non-empty string of digits.

The sign of the resulting number is \texttt{+} if \texttt{\langle sign \rangle} contains an even number of \texttt{-}, and \texttt{-} otherwise, hence, an empty \texttt{\langle sign \rangle} denotes a non-negative input. The stored significand is obtained from \texttt{\langle significand \rangle} by omitting the decimal separator and leading zeros, and rounding to 16 significant digits, filling with trailing zeros if necessary. In particular, the value stored is exact if the input \texttt{\langle significand \rangle} has at most 16 digits. The stored \texttt{\langle exponent \rangle} is obtained by combining the input \texttt{\langle exponent \rangle} (0 if absent) with a shift depending on the position of the significand and the number of leading zeros.

A special case arises if the resulting \texttt{\langle exponent \rangle} is either too large or too small for the floating point number to be represented. This results either in an overflow (the number is then replaced by \(\pm \infty\)), or an underflow (resulting in \(\pm 0\)).
The result is thus ±0 if and only if \(<\textit{significand}\)> contains no non-zero digit (\textit{i.e.}, consists only in characters 0, and an optional period), or if there is an underflow. Note that a single dot is currently a valid floating point number, equal to +0, but that is not guaranteed to remain true.

The \(<\textit{significand}\)> must be non-empty, so e1 and e-1 are not valid floating point numbers. Note that the latter could be mistaken with the difference of “e” and 1. To avoid confusions, the base of natural logarithms cannot be input as e and should be input as exp(1) or \c_e_fp (which is faster).

Special numbers are input as follows:

- \texttt{inf} represents +\infty, and can be preceded by any \textit{sign}, yielding ±\infty as appropriate.
- \texttt{nan} represents a (quiet) non-number. It can be preceded by any sign, but that sign is ignored.
- Any unrecognizable string triggers an error, and produces a \texttt{nan}.
- Note that commands such as \texttt{\infty}, \texttt{\pi}, or \texttt{\sin} do not work in floating point expressions. They may silently be interpreted as completely unexpected numbers, because integer constants (allowed in expressions) are commonly stored as mathematical characters.

\section*{29.12.2 Precedence of operators}

We list here all the operations supported in floating point expressions, in order of decreasing precedence: operations listed earlier bind more tightly than operations listed below them.

- Function calls (\texttt{sin, ln, etc}).
- Binary \texttt{**} and \texttt{^} (right associative).
- Unary +, -, !.
- Implicit multiplication by juxtaposition (2pi) when neither factor is in parentheses.
- Binary * and /, implicit multiplication by juxtaposition with parentheses (for instance 3(4+5)).
- Binary + and -.  
- Comparisons >=, !=, <?, etc.
- Logical \texttt{and}, denoted by \&\&.
- Logical \texttt{or}, denoted by \mid\mid.
- Ternary operator \texttt{?}: (right associative).
- Comma (to build tuples).
The precedence of operations can be overridden using parentheses. In particular, the precedence of juxtaposition implies that

\[ \frac{1}{2\pi} = 1 / (2\pi), \]
\[ \frac{1}{2\pi} (\pi + \pi) = (2\pi)^{-1} (\pi + \pi) \approx 1, \]
\[ \sin 2\pi = \sin (2\pi) \neq 0, \]
\[ 2^{-2} \max (3, 5) = 2^2 \max (3, 5) = 20, \]
\[ 1\text{in}/1\text{cm} = (1\text{in})/(1\text{cm}) = 2.54. \]

Functions are called on the value of their argument, contrarily to \TeX{} macros.

## 29.12.3 Operations

We now present the various operations allowed in floating point expressions, from the lowest precedence to the highest. When used as a truth value, a floating point expression is \textit{false} if it is \( \pm 0 \), and \textit{true} otherwise, including when it is \texttt{nan} or a tuple such as \((0, 0)\). Tuples are only supported to some extent by operations that work with truth values (?:, \&\&, !), by comparisons (!\&\&?), and by +, -, *, /. Unless otherwise specified, providing a tuple as an argument of any other operation yields the “invalid operation” exception and a \texttt{nan} result.

\[
\texttt{\textbackslash fp\_eval:n \{ \langle operand_1 \rangle ? \langle operand_2 \rangle : \langle operand_3 \rangle \}}
\]

The ternary operator ?: results in \langle operand_2 \rangle if \langle operand_1 \rangle is true (not \( \pm 0 \)), and \langle operand_3 \rangle if \langle operand_1 \rangle is false (\( \pm 0 \)). All three \langle operands \rangle are evaluated in all cases; they may be tuples. The operator is right associative, hence

\[
\texttt{\textbackslash fp\_eval:n}
\texttt{\{ 1 + 3 > 4 ? 1 : 2 + 4 > 5 ? 2 : 3 + 5 > 6 ? 3 : 4 \}}
\]

first tests whether \( 1 + 3 > 4 \); since this isn’t true, the branch following : is taken, and \( 2 + 4 > 5 \) is compared; since this is true, the branch before : is taken, and everything else is (evaluated then) ignored. That allows testing for various cases in a concise manner, with the drawback that all computations are made in all cases.

\[
\texttt{\textbackslash fp\_eval:n \{ \langle operand_1 \rangle \&\& \langle operand_2 \rangle \}}
\]

If \langle operand_1 \rangle is true (not \( \pm 0 \)), use that value, otherwise the value of \langle operand_2 \rangle. Both \langle operands \rangle are evaluated in all cases; they may be tuples. In \langle operand_1 \rangle \&\& \langle operand_2 \rangle \&\& \ldots \&\& \langle operand_n \rangle, the first true (nonzero) \langle operand \rangle is used and if all are zero the last one (\( \pm 0 \)) is used.

\[
\texttt{\textbackslash fp\_eval:n \{ \langle operand_1 \rangle \&\&\& \langle operand_2 \rangle \}}
\]

If \langle operand_1 \rangle is false (equal to \( \pm 0 \)), use that value, otherwise the value of \langle operand_2 \rangle. Both \langle operands \rangle are evaluated in all cases; they may be tuples. In \langle operand_1 \rangle \&\&\& \langle operand_2 \rangle \&\&\& \ldots \&\&\& \langle operand_n \rangle, the first false (\( \pm 0 \)) \langle operand \rangle is used and if none is zero the last one is used.
\[ \text{abs} \{ \text{fp expr} \} \]

Computes the absolute value of the \text{fp expr}. If the operand is a tuple, “invalid operation” occurs. This operation does not raise exceptions in other cases. See also \text{fp_abs:n}. 274
\textbf{exp} \texttt{\textbackslash fp\_eval:n \{ exp( \langle fp\ expr\rangle ) \}}

Computes the exponential of the \langle fp\ expr\rangle. “Underflow” and “overflow” occur when appropriate. If the operand is a tuple, “invalid operation” occurs.

\textbf{fact} \texttt{\textbackslash fp\_eval:n \{ fact( \langle fp\ expr\rangle ) \}}

Computes the factorial of the \langle fp\ expr\rangle. If the \langle fp\ expr\rangle is an integer between $-0$ and 3248 included, the result is finite and correctly rounded. Larger positive integers give $+\infty$ with “overflow”, while \texttt{fact}(+\infty) = +\infty and \texttt{fact}(\text{\texttt{nan}}) = \text{\texttt{nan}} with no exception. All other inputs give \text{\texttt{nan}} with the “invalid operation” exception.

\textbf{ln} \texttt{\textbackslash fp\_eval:n \{ ln( \langle fp\ expr\rangle ) \}}

Computes the natural logarithm of the \langle fp\ expr\rangle. Negative numbers have no (real) logarithm, hence the “invalid operation” is raised in that case, including for ln(−0). “Division by zero” occurs when evaluating ln(+0) = −∞. “Underflow” and “overflow” occur when appropriate. If the operand is a tuple, “invalid operation” occurs.

\textbf{logb} \texttt{\textbackslash fp\_eval:n \{ logb( \langle fp\ expr\rangle ) \}}

Determines the exponent of the \langle fp\ expr\rangle, namely the floor of the base-10 logarithm of its absolute value. “Division by zero” occurs when evaluating logb(±0) = −∞. Other special values are logb(±\infty) = +\infty and logb(\text{\texttt{nan}}) = \text{\texttt{nan}}. If the operand is a tuple or is \text{\texttt{nan}}, then “invalid operation” occurs and the result is \text{\texttt{nan}}.

\textbf{max} \texttt{\textbackslash fp\_eval:n \{ max( \langle fp\ expr_1\rangle, \langle fp\ expr_2\rangle, \ldots\ ) \}}

\textbf{min} \texttt{\textbackslash fp\_eval:n \{ min( \langle fp\ expr_1\rangle, \langle fp\ expr_2\rangle, \ldots\ ) \}}

Evaluates each \langle fp\ expr\rangle and computes the largest (smallest) of those. If any of the \langle fp\ expr\rangle is a \texttt{nan} or tuple, the result is \texttt{nan}. If any operand is a tuple, “invalid operation” occurs; these operations do not raise exceptions in other cases.
Only \texttt{round} accepts a third argument. Evaluates \langle \texttt{fp expr}\rangle = x \text{ and } \langle \texttt{fp expr}_2\rangle = n \text{ and } \langle \texttt{fp expr}_3\rangle = t \text{ then rounds } x \text{ to } n \text{ places. If } n \text{ is an integer, this rounds } x \text{ to a multiple of } 10^{-n}; \text{ if } n = +\infty, \text{ this always yields } x; \text{ if } n = -\infty, \text{ this yields one of } \pm 0, \pm \infty, \text{ or } \texttt{nan}; \text{ if } n = \texttt{nan}, \text{ this yields } \texttt{nan}; \text{ if } n \text{ is neither } \pm \infty \text{ nor an integer, then an “invalid operation” exception is raised. When } \langle \texttt{fp expr}_2\rangle \text{ is omitted, } n = 0, \text{ i.e., } \langle \texttt{fp expr}\rangle \text{ is rounded to an integer. The rounding direction depends on the function.}

- \texttt{round} yields the multiple of } 10^{-n} \text{ closest to } x, \text{ with ties (} x \text{ half-way between two such multiples) rounded as follows. If } t \text{ is } \texttt{nan} \text{ (or not given) the even multiple is chosen (“ties to even”), if } t = \pm 0 \text{ the multiple closest to } 0 \text{ is chosen (“ties to zero”), if } t \text{ is positive/negative the multiple closest to } \infty/-\infty \text{ is chosen (“ties towards positive/negative infinity”).}

- \texttt{floor} yields the largest multiple of } 10^{-n} \text{ smaller or equal to } x \text{ (“round towards negative infinity”)};

- \texttt{ceil} yields the smallest multiple of } 10^{-n} \text{ greater or equal to } x \text{ (“round towards positive infinity”)};

- \texttt{trunc} yields a multiple of } 10^{-n} \text{ with the same sign as } x \text{ and with the largest absolute value less than that of } x \text{ (“round towards zero”).}

“Overflow” occurs if } x \text{ is finite and the result is infinite (this can only happen if } \langle \texttt{fp expr}_2\rangle < -9984). \text{ If any operand is a tuple, “invalid operation” occurs.

\textbf{sign} \fp_eval:n \{ \mathrm{sign}( \langle \texttt{fp expr}\rangle ) \}

Evaluates the \langle \texttt{fp expr}\rangle \text{ and determines its sign: +1 for positive numbers and for } +\infty, -1 \text{ for negative numbers and for } -\infty, \pm 0 \text{ for } \pm 0, \text{ and } \texttt{nan} \text{ for } \texttt{nan}. \text{ If the operand is a tuple, “invalid operation” occurs. This operation does not raise exceptions in other cases.}

\textbf{sin} \fp_eval:n \{ \sin( \langle \texttt{fp expr}\rangle ) \}

\textbf{cos} \fp_eval:n \{ \cos( \langle \texttt{fp expr}\rangle ) \}

\textbf{tan} \fp_eval:n \{ \tan( \langle \texttt{fp expr}\rangle ) \}

\textbf{cot} \fp_eval:n \{ \cot( \langle \texttt{fp expr}\rangle ) \}

\textbf{csc} \fp_eval:n \{ \csc( \langle \texttt{fp expr}\rangle ) \}

\textbf{sec} \fp_eval:n \{ \sec( \langle \texttt{fp expr}\rangle ) \}

Computes the sine, cosine, tangent, cotangent, cosecant, or secant of the \langle \texttt{fp expr}\rangle \text{ given in radians. For arguments given in degrees, see } \texttt{sind}, \texttt{cosd}, \text{ etc.} \text{ Note that since } \pi \text{ is irrational, } \sin(8\pi) \text{ is not quite zero, while its analogue } \texttt{sind}(8 \times 180) \text{ is exactly zero. The trigonometric functions are undefined for an argument of } \pm \infty, \text{ leading to the “invalid operation” exception. Additionally, evaluating tangent, cotangent, cosecant, or secant at one of their poles leads to a “division by zero” exception. “Underflow” and “overflow” occur when appropriate. If the operand is a tuple, “invalid operation” occurs.
Computes the sine, cosine, tangent, cotangent, cosecant, or secant of the \( \langle fp \ expr \rangle \) given in degrees. For arguments given in radians, see \( \sin, \cos, etc. \) Note that since \( \pi \) is irrational, \( \sin(8\pi) \) is not quite zero, while its analogue \( \sin(8 \times 180) \) is exactly zero. The trigonometric functions are undefined for an argument of \( \pm \infty \), leading to the “invalid operation” exception. Additionally, evaluating tangent, cotangent, cosecant, or secant at one of their poles leads to a “division by zero” exception. “Underflow” and “overflow” occur when appropriate. If the operand is a tuple, “invalid operation” occurs.

\[ \text{New: 2013-11-02} \]

Computes the arcsine, arccosine, arccosecant, or arcsecant of the \( \langle fp \ expr \rangle \) and returns the result in radians, in the range \([-\pi/2, \pi/2]\) for \( \text{asin} \) and \( \text{acsc} \) and \([0, \pi]\) for \( \text{acos} \) and \( \text{asec} \). For a result in degrees, use \( \text{asind}, etc. \) If the argument of \( \text{asin} \) or \( \text{acos} \) lies outside the range \([-1, 1]\), or the argument of \( \text{acsc} \) or \( \text{asec} \) inside the range \((-1, 1)\), an “invalid operation” exception is raised. “Underflow” and “overflow” occur when appropriate. If the operand is a tuple, “invalid operation” occurs.

\[ \text{New: 2013-11-02} \]

Computes the arcsine, arccosine, arccosecant, or arcsecant of the \( \langle fp \ expr \rangle \) and returns the result in degrees, in the range \([-\pi/2, \pi/2]\) for \( \text{asin} \) and \( \text{acsc} \) and \([0, 180]\) for \( \text{acosd} \) and \( \text{asecd} \). For a result in radians, use \( \text{asind}, etc. \) If the argument of \( \text{asin} \) or \( \text{acosd} \) lies outside the range \([-1, 1]\), or the argument of \( \text{acscd} \) or \( \text{asecd} \) inside the range \((-1, 1)\), an “invalid operation” exception is raised. “Underflow” and “overflow” occur when appropriate. If the operand is a tuple, “invalid operation” occurs.
Computes the square root of the \( \text{fp expr} \). The “invalid operation” is raised when the \( \text{fp expr} \) is negative or is a tuple; no other exception can occur. Special values yield 
\[
\sqrt{-0} = -0, \quad \sqrt{+0} = +0, \quad \sqrt{+\infty} = +\infty \quad \text{and} \quad \sqrt{\text{nan}} = \text{nan}.
\]
\fp_eval:n { rand() }

\fp_eval:n { randint( \langle fp expr \rangle \) }
\fp_eval:n { randint( \langle fp expr \rangle , \langle fp expr \rangle ) }

\sys_rand_seed:
\sys_gset_rand_seed:n

\TeXhackers_note: This is based on pseudo-random numbers provided by the engine's primitive \pdfuniformdeviate in pdf\TeX, \p\TeX, \up\TeX and \uniformdeviate in Lua\TeX and X\e\TeX. The underlying code is based on Metapost, which follows an additive scheme recommended in Section 3.6 of "The Art of Computer Programming, Volume 2".

While we are more careful than \uniformdeviate to preserve uniformity of the underlying stream of 28-bit pseudo-random integers, these pseudo-random numbers should of course not be relied upon for serious numerical computations nor cryptography.

\inf The special values $+\infty$, $-\infty$, and \nan are represented as \inf, -\inf and \nan (see \c_-inf_fp, \c_minus_inf_fp and \c_nan_fp).

\pi The value of \pi (see \c_pi_fp).

\deg The value of 1° in radians (see \c_one_degree_fp).
Those units of measurement are equal to their values in pt, namely

\[
1 \text{ in} = 72.27 \text{ pt}
\]
\[
1 \text{ pt} = 1 \text{ pt}
\]
\[
1 \text{ pc} = 12 \text{ pt}
\]
\[
1 \text{ cm} = \frac{1}{2.54} \text{ in} = 28.45275590551181 \text{ pt}
\]
\[
1 \text{ mm} = \frac{1}{25.4} \text{ in} = 2.845275590551181 \text{ pt}
\]
\[
1 \text{ dd} = 0.376065 \text{ mm} = 1.0700856496063 \text{ pt}
\]
\[
1 \text{ cc} = 12 \text{ dd} = 12.84010277962756 \text{ pt}
\]
\[
1 \text{ nd} = 0.375 \text{ mm} = 1.066978346456693 \text{ pt}
\]
\[
1 \text{ nc} = 12 \text{ nd} = 12.80374015748031 \text{ pt}
\]
\[
1 bp = \frac{1}{72} \text{ in} = 1.00375 \text{ pt}
\]
\[
1 \text{ sp} = 2^{-16} \text{ pt} = 1.52587890625 \times 10^{-5} \text{ pt}.
\]

The values of the (font-dependent) units em and ex are gathered from \TeX{} when the surrounding floating point expression is evaluated.

\begin{Verbatim}
\texttt{true}
\end{Verbatim}
\begin{Verbatim}
\texttt{false}
\end{Verbatim}

% Other names for 1 and +0.

\begin{Verbatim}
\texttt{\textbackslash fp\_abs\_n} \star \texttt{\textbackslash fp\_abs\_n \{\textit{fp expr}\}}
\end{Verbatim}

\texttt{\textbackslash new: 2012-05-14}
\texttt{\textbackslash updated: 2012-07-08}

\begin{Verbatim}
Evaluates the \textit{fp expr} as described for \texttt{\textbackslash fp\_eval\_n} and leaves the absolute value of the result in the input stream. If the argument is ±∞, nan or a tuple, “invalid operation” occurs. Within floating point expressions, \texttt{abs()} can be used; it accepts ±∞ and \texttt{nan} as arguments.
\end{Verbatim}

\begin{Verbatim}
\texttt{\textbackslash fp\_max\_nn} \star \texttt{\textbackslash fp\_max\_nn \{\textit{fp expr}_1\} \{\textit{fp expr}_2\}}
\end{Verbatim}

\texttt{\textbackslash new: 2012-09-26}

\begin{Verbatim}
\texttt{\textbackslash fp\_min\_nn} \star \texttt{\textbackslash fp\_min\_nn \{\textit{fp expr}_1\} \{\textit{fp expr}_2\}}
\end{Verbatim}

\begin{Verbatim}
Evaluates the \textit{fp exprs} as described for \texttt{\textbackslash fp\_eval\_n} and leaves the resulting larger (max) or smaller (min) value in the input stream. If the argument is a tuple, “invalid operation” occurs, but no other case raises exceptions. Within floating point expressions, \texttt{max()} and \texttt{min()} can be used.
\end{Verbatim}

## 29.13 Disclaimer and roadmap

This module may break if the escape character is among 0123456789+, or if it receives a \TeX{} primitive conditional affected by \texttt{\exp\_not\_n}.

The following need to be done. I’ll try to time-order the items.

- Function to count items in a tuple (and to determine if something is a tuple).
- Decide what exponent range to consider.
• Support signalling \texttt{nan}.
• Modulo and remainder, and rounding function \texttt{quantize} (and its friends analogous to \texttt{trunc}, \texttt{ceil}, \texttt{floor}).
• \texttt{\fp_format:nn \{(fp expr)\} \{(format)\}}, but what should \texttt{(format)} be? More general pretty printing?
• Add \texttt{and}, \texttt{or}, \texttt{xor}? Perhaps under the names \texttt{all}, \texttt{any}, and \texttt{xor}?
• Add log\((x,b)\) for logarithm of \(x\) in base \(b\).
• \texttt{hypot} (Euclidean length). Cartesian-to-polar transform.
• Hyperbolic functions \texttt{cosh}, \texttt{sinh}, \texttt{tanh}.
• Inverse hyperbolics.
• Base conversion, input such as \texttt{0xAB.CDEF}.
• Factorial (not with \texttt{!}), gamma function.
• Improve coefficients of the \texttt{sin} and \texttt{tan} series.
• Treat upper and lower case letters identically in identifiers, and ignore underscores.
• Add an \texttt{array(1,2,3)} and \texttt{i=complex(0,1)}.
• Provide an experimental \texttt{map} function? Perhaps easier to implement if it is a single character, \texttt{@sin(1,2)}?
• Provide an \texttt{isnan} function analogue of \texttt{\fp_if_nan:nTF}?
• Support keyword arguments?

\texttt{Pgfmath} also provides box-measurements (depth, height, width), but boxes are not possible expandably.
Bugs, and tests to add.
• Check that functions are monotonic when they should.
• Add exceptions to \texttt{?:, !=>=?, &&, ||, and !}.
• Logarithms of numbers very close to 1 are inaccurate.
• When rounding towards \(-\infty\), \texttt{\dim_to_fp:n \{0pt\}} should return \(-0\), not \(+0\).
• The result of \((\pm0) + (\pm0)\), of \(x + (\pm x)\), and of \((\pm x) + x\) should depend on the rounding mode.
• \texttt{0e9999999999} gives a \TeX{} “number too large” error.
• Subnormals are not implemented.

Possible optimizations/improvements.
• Document that \texttt{l3trial/l3fp-types} introduces tools for adding new types.
• In subsection 29.12.1, write a grammar.
• It would be nice if the `parse` auxiliaries for each operation were set up in the corresponding module, rather than centralizing in `l3fp-parse`.

• Some functions should get an `_o` ending to indicate that they expand after their result.

• More care should be given to distinguish expandable/restricted expandable (auxiliary and internal) functions.

• The code for the `ternary` set of functions is ugly.

• There are many \_ missing in the doc to avoid bad line-breaks.

• The algorithm for computing the logarithm of the significand could be made to use a 5 terms Taylor series instead of 10 terms by taking $c = 2000/(200x + 1) \in [10, 95]$ instead of $c \in [1, 10]$. Also, it would then be possible to simplify the computation of $t$. However, we would then have to hard-code the logarithms of 44 small integers instead of 9.

• Improve notations in the explanations of the division algorithm (`l3fp-basics`).

• Understand and document `\_fp_basics_pack_weird_low:NNNNw` and `\_fp_basics_pack_weird_high:NNNNNNNNw` better. Move the other `basics_pack` auxiliaries to `l3fp-aux` under a better name.

• Find out if underflow can really occur for trigonometric functions, and redoc as appropriate.

• Add bibliography. Some of Kahan’s articles, some previous TeX fp packages, the international standards,\ldots

• Also take into account the “inexact” exception?

• Support multi-character prefix operators (e.g., @$/$ or whatever)?
Chapter 30

The l3fparray module
Fast global floating point arrays

For applications requiring heavy use of floating points, this module provides arrays which can be accessed in constant time (contrast l3seq, where access time is linear). The interface is very close to that of l3intarray. The size of the array is fixed and must be given at point of initialisation.

30.1 Creating and initialising floating point array variables

\fparray_new:Nn \fparray_new:Nn \fparray_new:cn \fparray_gzero:N \fparray_gzero:C

Evaluates the integer expression \textit{size} and allocates an \textit{floating point array variable} with that number of (zero) entries. The variable name should start with \texttt{g} because assignments are always global.

\fparray_gzero:N \fparray_gzero:C

Sets all entries of the \textit{floating point array variable} to +0. Assignments are always global.

30.2 Adding data to floating point arrays

\fparray_gset:Nnn \fparray_gset:Nnn \fparray_gset:cnm

Stores the result of evaluating the floating point expression \textit{value} into the \textit{floating point array variable} at the (integer expression) \textit{position}. If the \textit{position} is not between 1 and the \fparray_count:N, an error occurs. Assignments are always global.
30.3 Counting entries in floating point arrays

\texttt{\textbackslash fparray\_count:N} \texttt{\langle fparray var \rangle}

Expands to the number of entries in the \texttt{\langle floating point array variable \rangle}. This is performed in constant time.

30.4 Using a single entry

\texttt{\textbackslash fparray\_item:Nn} \texttt{\langle fparray var \rangle\{\langle position\rangle\}}

Applies \texttt{\textbackslash fp\_use:N} or \texttt{\textbackslash fp\_to\_tl:N} (respectively) to the floating point entry stored at the \texttt{\langle integer expression \rangle} \texttt{\langle position \rangle} in the \texttt{\langle floating point array variable \rangle}. If the \texttt{\langle position \rangle} is not between 1 and the \texttt{\textbackslash fparray\_count:N} \texttt{\langle fparray var \rangle}, an error occurs.

30.5 Floating point array conditional

\texttt{\textbackslash fparray\_if\_exist\_p:N} \texttt{\langle fparray var \rangle}

Tests whether the \texttt{\langle fparray var \rangle} is currently defined. This does not check that the \texttt{\langle fparray var \rangle} really is a floating point array variable.
Chapter 31

The \texttt{l3bitset} module

Bitsets

This module defines and implements the data type \texttt{bitset}, a vector of bits. The size of the vector may grow dynamically. Individual bits can be set and unset by names pointing to an index position. The names 1, 2, 3, … are predeclared and point to the index positions 1, 2, 3,…. More names can be added and existing names can be changed. The index is like all other indices in expl3 modules \texttt{1-based}. A \texttt{bitset} can be output as binary number or—as needed e.g. in a PDF dictionary—as decimal (arabic) number. Currently only a small subset of the functions provided by the \texttt{bitset} package are implemented here, mainly the functions needed to use bitsets in PDF dictionaries.

The \texttt{bitset} is stored as a string (but one shouldn’t rely on the internal representation) and so the vector size is theoretically unlimited, only restricted by \TeX-memory. But the functions to set and clear bits use integer functions for the index so bitsets can’t be longer than $2^{31} - 1$. The export function \texttt{\bitset_to_arabic:N} can use functions from the \texttt{int} module only if the largest index used for this bitset is smaller than 32, for longer bitsets \texttt{fp} is used and this is slower.
31.1 Creating bitsets

\bitset_new:N \bitset_new:N (bitset var)
\bitset_new:c \bitset_new:N (bitset var)
\bitset_new:NN \bitset_new:cn
\bitset_new:NN \bitset_new:cn
{ \langle \text{name}_1 \rangle = \langle \text{index}_1 \rangle , \\
\langle \text{name}_2 \rangle = \langle \text{index}_2 \rangle , \ldots 
}\text{New: 2023-11-15}

Creates a new \langle \text{bitset var} \rangle or raises an error if the name is already taken. The declaration is global. The \langle \text{bitset var} \rangle is initially 0.

Bitsets are implemented as string variables consisting of 1’s and 0’s. The rightmost number is the index position 1, so the string variable can be viewed directly as the binary number. But one shouldn’t rely on the internal representation, but use the dedicated \bitset_to_bin:N instead to get the binary number.

The name–index pairs given in the second argument of \bitset_new:NN declares names for some indices, which can be used to set and unset bits. The names 1, 2, 3, ... are predeclared and point to the index positions 1, 2, 3, ....

\langle \text{index}... \rangle should be a positive number or an \langle \text{integer expression} \rangle which evaluates to a positive number. The expression is evaluated when the index is used, not at declaration time. The names \langle \text{name}... \rangle should be unique. Using a number as name, e.g. \text{10}=1, is allowed, it then overwrites the predeclared name 10, but the index position 10 can then only be reached if some other name for it exists, e.g. \text{ten}=10. It is not necessary to give every index a name, and an index can have more than one name. The named index can be extended or changed with the next function.

\bitset_addto_named_index:Nn \bitset_addto_named_index:Nn (bitset var)
\bitset_addto_named_index:Nn \bitset_addto_named_index:Nn (bitset var)
{ \langle \text{name}_1 \rangle = \langle \text{index}_1 \rangle , \\
\langle \text{name}_2 \rangle = \langle \text{index}_2 \rangle , \ldots 
}\text{New: 2023-11-15}

This extends or changes the name–index pairs for \langle \text{bitset var} \rangle globally as described for \bitset_new:NN.

For example after these settings

\bitset_new:NN \l\_pdfannot\_F\_bitset
{ 
Invisible = 1, 
Hidden = 2, 
Print = 3, 
NoZoom = 4, 
NoRotate = 5, 
NoView = 6, 
ReadOnly = 7, 
Locked = 8, 
ToggleNoView = 9, 
LockedContents = 10 
}
\bitset_addto_named_index:Nn \l\_pdfannot\_F\_bitset
{ 
286
\begin{verbatim}
print = 3
}

it is possible to set bit 3 by using any of these alternatives:
\bitset_set_true:Nn \l_pdfannot_F_bitset {Print}
\bitset_set_true:Nn \l_pdfannot_F_bitset {print}
\bitset_set_true:Nn \l_pdfannot_F_bitset {3}
\end{verbatim}

\bitset_if_exist_p:N \langle bitset var \rangle
\bitset_if_exist:NTF \langle bitset var \rangle \{\langle true code\rangle\} \{\langle false code\rangle\}
Tests whether the \langle bitset var \rangle exist.

\section*{31.2 Setting and unsetting bits}

\begin{verbatim}
\bitset_set_true:Nn \langle bitset var \rangle \{\langle name\rangle\}
This sets the bit of the index position represented by \langle\langle name\rangle\rangle to 1. \langle name\rangle should be either one of the predeclared names 1, 2, 3, ..., or one of the names added manually. Index position are 1-based. If needed the length of the bit vector is enlarged.

\bitset_set_false:Nn \langle bitset var \rangle \{\langle name\rangle\}
This unsets the bit of the index position represented by \langle\langle name\rangle\rangle (sets it to 0). \langle name\rangle should be either one of the predeclared names 1, 2, 3, ..., or one of the names added manually. The index is 1-based. If the index position is larger than the current length of the bit vector nothing happens. If the leading (left most) bit is unset, zeros are not trimmed but stay in the bit vector and are still shown by \bitset_show:N.

\bitset_clear:N \langle bitset var \rangle
This resets the bitset to the initial state. The declared names are not changed.
\end{verbatim}

\section*{31.3 Using bitsets}

\begin{verbatim}
\bitset_item:Nn \langle bitset var \rangle \{\langle name\rangle\}
\bitset_item:cn \langle bitset var \rangle \{\langle name\rangle\}
\bitset_item:Nn outputs 1 if the bit with the index number represented by \langle name\rangle is set and 0 otherwise. \langle name\rangle is either one of the predeclared names 1, 2, 3, ..., or one of the names added manually.
\end{verbatim}
\bitset_to_bin:N \bitset_to_bin:c * This leaves the current value of the bitset expressed as a binary (string) number in the input stream. If no bit has been set yet, the output is zero.

\bitset_to_arabic:N \bitset_to_arabic:c * This leaves the current value of the bitset expressed as a decimal number in the input stream. If no bit has been set yet, the output is zero. The function uses \int_from_bin:n if the largest index that have been set or unset is smaller than 32, and a slower implementation based on \fp_eval:n otherwise.

\bitset_show:N \bitset_show:c Displays the binary and decimal values of the ⟨bitset var⟩ on the terminal.

\bitset_log:N \bitset_log:c Writes the binary and decimal values of the ⟨bitset var⟩ in the log file.

\bitset_show_named_index:N \bitset_show_named_index:c Displays declared name–index pairs of the ⟨bitset var⟩ on the terminal.

\bitset_log_named_index:N \bitset_log_named_index:c Writes declared name–index pairs of the ⟨bitset var⟩ in the log file.
Chapter 32

The l3cctab module

Category code tables

A category code table enables rapid switching of all category codes in one operation. For Lua\TeX, this is possible over the entire Unicode range. For other engines, only the 8-bit range (0–255) is covered by such tables. The implementation of category code tables in expl3 also saves and restores the \TeX\ \texttt{\endlinechar} primitive value, meaning they could be used for example to implement \texttt{\ExplSyntaxOn}.

32.1 Creating and initialising category code tables

\begin{verbatim}
\cctab_new:N \cctab_new:N \langle\text{category code table}\rangle
\end{verbatim}

Creates a new \langle\text{category code table}\rangle variable or raises an error if the name is already taken. The declaration is global. The \langle\text{category code table}\rangle is initialised with the codes as used by ini\TeX.

\begin{verbatim}
\cctab_const:Nn \cctab_const:cn \langle\text{category code table}\rangle \{\langle\text{category code set up}\rangle\}
\end{verbatim}

Creates a new \langle\text{category code table}\rangle, applies (in a group) the \langle\text{category code set up}\rangle on top of ini\TeX settings, then saves them globally as a constant table. The \langle\text{category code set up}\rangle can include a call to \cctab_select:N.

\begin{verbatim}
\cctab_gset:Nn \cctab_gset:cn \langle\text{category code table}\rangle \{\langle\text{category code set up}\rangle\}
\end{verbatim}

Starting from the ini\TeX category codes, applies (in a group) the \langle\text{category code set up}\rangle, then saves them globally in the \langle\text{category code table}\rangle. The \langle\text{category code set up}\rangle can include a call to \cctab_select:N.

\begin{verbatim}
\cctab_gsave_current:N \cctab_gsave_current:cn
\end{verbatim}

Saves the current prevailing category codes in the \langle\text{category code table}\rangle.
32.2 Using category code tables

\cctab_begin:N \cctab_begin:c
Switches locally the category codes in force to those stored in the \cctab_end:. The prevailing codes before the function is called are added to a stack, for use with \cctab_end:. This function does not start a \TeX{} group.

\cctab_end:
Ends the scope of a \cctab_end: started using \cctab_begin:N, returning the codes to those in force before the matching \cctab_begin:N was used. This must be used within the same \TeX{} group (and at the same \TeX{} group level) as the matching \cctab_begin:N.

\cctab_select:N \cctab_select:c
Selects the \cctab_end: for the scope of the current group. This is in particular useful in the \cctab_select:N \cctab_select:c \cctab_item:Nn \cctab_item:cn \cctab_if_exist_p:N \cctab_if_exist:NTF \cctab_if_exist:N \cctab_if_exist:c \cctab_if_exist_p:c \cctab_if_exist_p:N \cctab_if_exist:NTF \cctab_if_exist:N \cctab_if_exist:c \cctab_if_exist_p:N \cctab_if_exist:NTF \cctab_if_exist:N \cctab_if_exist:c

32.3 Category code table conditionals

\cctab_if_exist_p:N \cctab_if_exist_p:c \cctab_if_exist:NTF \cctab_if_exist:N \cctab_if_exist:c
Tests whether the \cctab_end: is currently defined. This does not check that the \cctab_end: really is a category code table.

32.4 Constant and scratch category code tables

\c_code_cctab \c_document_cctab
Category code table for the expl3 code environment; this does not include $\emptyset$, which is retained as an “other” character. Sets the \endlinechar value to 32 (a space).

Category code table for a standard \LaTeX{} document, as set by the \LaTeX{} kernel. In particular, the upper-half of the 8-bit range will be set to “active” with pdf\LaTeX{} only. No babel shorthands will be activated. Sets the \endlinechar value to 13 (normal line ending).
\c_initeX_cctab

Category code table as set up by \texttt{iniTEx}.

Updated: 2020-07-02

\c_other_cctab

Category code table where all characters have category code 12 (other). Sets the \texttt{endlinechar} value to $-1$.

Updated: 2020-07-02

\c_str_cctab

Category code table where all characters have category code 12 (other) with the exception of spaces, which have category code 10 (space). Sets the \texttt{endlinechar} value to $-1$.

Updated: 2020-07-02

\g_tmpa_cctab
\g_tmpb_cctab

Scratch category code tables.

Rev: 2023-05-26
Part V

Text manipulation
Chapter 33

The l3unicode module
Unicode support functions

This module provides Unicode-specific functions along with loading data from a range of Unicode Consortium files. Most of the code here is internal, but there are a small set of public functions. These work with Unicode (codepoints) and are designed to give usable results with both Unicode-aware and 8-bit engines.
\codepoint_generate:nn \{\codepoint\} \{\catcode\}

Generates one or more character tokens representing the \codepoint. With Unicode engines, exactly one character token will be generated, and this will have the \catcode specified as the second argument:

- 1 (begin group)
- 2 (end group)
- 3 (math toggle)
- 4 (alignment)
- 6 (parameter)
- 7 (math superscript)
- 8 (math subscript)
- 10 (space)
- 11 (letter)
- 12 (other)
- 13 (active)

For 8-bit engines, between one and four character tokens will be produced: these will be the bytes of the UTF-8 representation of the \codepoint. For all codepoints outside of the classical ASCII range, the generated character tokens will be active (category code 13); for codepoints in the ASCII range, the given \catcode will be used. To allow the result of this function to be used inside an expansion context, the result is protected by \exp_not:n.

\TeXhackers note: Users of (u)p\TeX note that these engines are treated as 8-bit in this context. In particular, for up\TeX, irrespective of the \catcode of the \codepoint, any value outside the ASCII range will result in a series of active bytes being generated.

\codepoint_str_generate:n \{\codepoint\}

Generates one or more character tokens representing the \codepoint. With Unicode engines, exactly one character token will be generated. For 8-bit engines, between one and four character tokens will be produced: these will be the bytes of the UTF-8 representation of the \codepoint. All of the generated character tokens will be of category code 12, except any spaces (codepoint 32), which will be category code 10.
Expands to the Unicode general category identifier of the (codepoint). The general category identifier is a string made up of two letter characters, the first uppercase and the second lowercase. The uppercase letters divide codepoints into broader groups, which are then refined by the lowercase letter. For example, codepoints representing letters all have identifiers starting L, for example Lu (uppercase letter), Lt (titlecase letter), etc. Full details are available in the documentation provided by the Unicode Consortium: see https://www.unicode.org/reports/tr44/#General_Category_Values

Converts the (codepoint) to the Unicode Normalization Form Canonical Decomposition. The generated character(s) will have the current category code as they would if typed in directly for Unicode engines; for 8-bit engines, active characters are used for all codepoints outside of the ASCII range.
Chapter 34

The \texttt{l3text} module

Text processing

This module deals with manipulation of (formatted) text; such material is comprised of a restricted set of token list content. The functions provided here concern conversion of textual content for example in case changing, generation of bookmarks and extraction to tags. All of the major functions operate by expansion. Begin-group and end-group tokens in the \texttt{text} are normalized and become \{ and \}, respectively.

34.1 Expanding text

\texttt{text\_expand:n} \texttt{\{text\}}

Takes user input \texttt{text} and expands the content. Protected commands (typically formatting) are left in place, and no processing takes place of math mode material (as delimited by pairs given in \texttt{\l_{text\_math\_delims\_tl}} or as the argument to commands listed in \texttt{\l_{text\_math\_arg\_tl}}). Commands which are neither engine- nor \LaTeX protected are expanded exhaustively. Any commands listed in \texttt{\l_{text\_expand\_exclude\_tl}} are excluded from expansion, as are those in \texttt{\l_{text\_case\_exclude\_arg\_tl}} and \texttt{\l_{text\_math\_arg\_tl}}.

\texttt{text\_declare\_expand\_equivalent:Nn} \texttt{\{cmd\} \{replacement\}}

Declares that the \texttt{replacement} tokens should be used whenever the \texttt{cmd} (a single token) is encountered. The \texttt{replacement} tokens should be expandable. A token can be “replaced” by itself if the defined replacement wraps it in \texttt{\exp_not:n}, for example

\texttt{text\_declare\_expand\_equivalent:Nn \' \{\exp_not:n \{\'\}\}
### 34.2 Case changing

**\text_lowercase:n** *\text_uppercase:n \{\textit{tokens}\}*

**\text_lowercase:nn** *\text_uppercase:nn \{BCP-47\} \{\textit{tokens}\}*

Takes user input \textit{tokens} first applies \textit{text expend:n}, then transforms the case of character tokens as specified by the function name. The category code of letters are not changed by this process when Unicode engines are used; in 8-bit engines, case changed characters in the ASCII range will have the current prevailing category code, while those outside of it will be represented by active characters.

Upper- and lowercase have the obvious meanings. Titlecasing may be regarded informally as converting the first character of the \textit{tokens} to uppercase. However, the process is more complex than this as there are some situations where a single lowercase character maps to a special form, for example \textit{ij} in Dutch which becomes \textit{IJ}. There are two functions available for titlecasing: one which applies the change to each “word” and a second which only applies at the start of the input. (Here, “word” boundaries are spaces: at present, full Unicode word breaking is not attempted.)

Importantly, notice that these functions are intended for working with user text for typesetting. For case changing programmatic data see the \texttt{lst} module and discussion there of \texttt{str_lowercase:n}, \texttt{str_uppercase:n} and \texttt{str_casefold:n}.

Case changing does not take place within math mode material so for example

\begin{verbatim}
\text_uppercase:n { Some-text-$y = mx + c$-with-{Braces} }
\end{verbatim}

becomes

\begin{verbatim}
SOME TEXT $y = mx + c$ WITH {BRACES}
\end{verbatim}

The first mandatory argument of commands listed in \texttt{l_text_case_exclude_arg-tl} is excluded from case changing; the latter are entirely non-textual content (such as labels).

The standard mappings here follow those defined by the Unicode Consortium in UnicodeData.txt and SpecialCasing.txt. For pH\TeX, only the ASCII range is covered as the engine treats input outside of this range as east Asian.

Locale-sensitive conversions are enabled using the \textit{(BCP-47)} argument, and follow Unicode Consortium guidelines. Currently, the locale strings recognized for special handling are as follows.

- **Armenian** (\texttt{hy} and \texttt{hy-x-yiwn}) The setting \texttt{hy} maps the codepoint U+0587, the ligature of letters ech and yiw, to the codepoints for capital ech and vew when uppercasing: this follows the spelling reform which is used in Armenia. The alternative \texttt{hy-x-yiwn} maps U+0587 to capital ech and yiw on uppercasing (also the output if Armenian is not selected at all).

- **Azeri and Turkish** (\texttt{az} and \texttt{tr}). The case pairs I/i-dotless and I-dot/i are activated for these languages. The combining dot mark is removed when lowercasing I-dot and introduced when upper casing i-dotless.

- **German** (\texttt{de-x-eszett}). An alternative mapping for German in which the lowercase \textit{Eszett} maps to a \textit{großes Eszett}. 

297
• Greek (el). Removes accents from Greek letters when uppercasing; titlecasing leaves accents in place. A variant el-x-iota is available which converts the ypogeogrammeni (subscript muted iota) to capital iota when uppercasing: the standard version retains the subscript versions.

• Lithuanian (lt). The lowercase letters i and j should retain a dot above when the accents grave, acute or tilde are present. This is implemented for lowercasing of the relevant uppercase letters both when input as single Unicode codepoints and when using combining accents. The combining dot is removed when uppercasing in these cases. Note that only the accents used in Lithuanian are covered: the behaviour of other accents are not modified.

• Medieval Latin (la-x-medieval). The characters u and V are interchanged on case changing.

• Dutch (nl). Capitalisation of ij at the beginning of titlecased input produces IJ rather than Ij.

Determining whether non-letter characters at the start of text should count as the uppercase element is controllable. When `\l_text_titlecase_check_letter_bool` is true, codepoints which are not letters (Unicode general category L) are not changed, and only the first letter is uppercased. When `\l_text_titlecase_check_letter_bool` is false, the first codepoint is uppercased, irrespective of the general code of the character.

\textdeclarecaseequivalent\{cmd\} {{replacement}}

Declares that the \text{replacement} tokens should be used whenever the \text{cmd} (a single token) is encountered during case changing.

\textdeclarelowercasemapping\{codepoint\} {\text{replacement}}
\textdeclaretitlecasemapping\{codepoint\} {\text{replacement}}
\textdelegateuppercasemapping\{codepoint\} {\text{replacement}}

\text{New: 2023-04-11}
\text{Updated: 2023-04-20}

Declares that the \text{replacement} tokens should be used when case mapping the \text{codepoint}, rather than the standard mapping given in the Unicode data files. The \text{nnn} version takes a BCP-47 tag, which can be used to specify that the customisation only applies to that locale.

\textcaseswitch\{normal\} \{upper\} \{lower\} \{title\}

\text{New: 2022-07-04}

Context-sensitive function which will expand to one of the \text{normal}, \text{upper}, \text{lower} or \text{title} tokens depending on the current case changing operation. Outside of case changing, the \text{normal} tokens are produced. Within case changing, the appropriate mapping tokens are inserted.
34.3 Removing formatting from text

\text_purify:n \{\langle text \rangle}\}
\text_purify:n

Takes user input \langle text \rangle and expands as described for \text_expand:n, then removes all functions from the resulting text. Math mode material (as delimited by pairs given in \l_text_math_delims_tl or as the argument to commands listed in \l_text_math_arg_tl) is left contained in a pair of $ delimiters. Non-expandable functions present in the \langle text \rangle must either have a defined equivalent (see \text_declare_purify_equivalent:Nn) or will be removed from the result. Implicit tokens are converted to their explicit equivalent.

\text_declare_purify_equivalent:Nn \text_declare_purify_equivalent:Nn \text_declare_purify_equivalent:Nn (cmd) \{\langle replacement \rangle\}
\text_declare_purify_equivalent:Nn
\text_declare_purify_equivalent:Nn

Declares that the \langle replacement \rangle tokens should be used whenever the \langle cmd \rangle (a single token) is encountered. The \langle replacement \rangle tokens should be expandable.

34.4 Control variables

\l_text_math_arg_tl

Lists commands present in the \langle text \rangle where the argument of the command should be treated as math mode material. The treatment here is similar to \l_text_math_delims_tl but for a command rather than paired delimiters.

\l_text_math_delims_tl

Lists pairs of tokens which delimit (in-line) math mode content; such content may be excluded from processing.

\l_text_case_exclude_arg_tl

Lists commands where the first mandatory argument is excluded from case changing.

\l_text_expand_exclude_tl

Lists commands which are excluded from expansion. This protection includes everything up to and including their first braced argument.

\l_text_titlecase_check_letter_bool

Controls how the start of titlecasing is handled: when \text{true}, the first letter in text is considered. The standard setting is \text{true}.  

299
34.5 Mapping to graphemes

Grapheme splitting is implemented using the algorithm described in Unicode Standard Annex #29. This includes support for extended grapheme clusters. Text starting with a line feed or carriage return character will drop this due to standard TeX processing. At present extended pictograms are not supported: these may be added in a future release.

\text_map_function:nN (text) {⟨function⟩}

Takes user input ⟨text⟩ and expands as described for \text_expand:n, then maps over the graphemes within the result, passing each grapheme to the ⟨function⟩. Broadly a grapheme is a “user perceived character”: the Unicode Consortium describe the decomposition of input to graphemes in depth, and the approach used here implements that algorithm. The ⟨function⟩ should accept one argument as ⟨balanced text⟩: this may be comprise codepoints or may be a control sequence. With 8-bit engines, the codepoint(s) themselves may of course be made up of multiple bytes: the mapping will pass the correct codepoints independent of the engine in use. See also \text_map_inline:nn.

\text_map_inline:nn (text) {⟨inline function⟩}

Takes user input ⟨text⟩ and expands as described for \text_expand:n, then maps over the graphemes within the result, passing each grapheme to the ⟨inline function⟩. Broadly a grapheme is a “user perceived character”: the Unicode Consortium describe the decomposition of input to graphemes in depth, and the approach used here implements that algorithm. The ⟨inline function⟩ should consist of code which receives the grapheme as ⟨balanced text⟩: this may be comprise codepoints or may be a control sequence. With 8-bit engines, the codepoint(s) themselves may of course be made up of multiple bytes: the mapping will pass the correct codepoints independent of the engine in use. See also \text_map_function:nN.

\text_map_break:
\text_map_break:n ⟨code⟩

Used to terminate a \text_map... function before all entries in the ⟨text⟩ have been processed. This normally takes place within a conditional statement.
Part VI
Typesetting
Chapter 35

The \texttt{l3box} module

Boxes

Box variables contain typeset material that can be inserted on the page or in other boxes. Their contents cannot be converted back to lists of tokens. There are three kinds of box operations: horizontal mode denoted with prefix \texttt{\hbox\_}, vertical mode with prefix \texttt{\vbox\_}, and the generic operations working in both modes with prefix \texttt{\box\_}. For instance, a new box variable containing the words “Hello, world!” (in a horizontal box) can be obtained by the following code.

\begin{verbatim}
\box_new:N \l_hello_box
\hbox_set:Nn \l_hello_box { Hello, - world! }
\end{verbatim}

The argument is typeset inside a \TeX{} group so that any variables assigned during the construction of this box restores its value afterwards.

Box variables from \texttt{l3box} are compatible with those of \LaTeX{} and plain \TeX{} and can be used interchangeably. The \texttt{l3box} commands to construct boxes, such as \texttt{\hbox:n} or \texttt{\hbox_set:Nn}, are “color-safe”, meaning that

\begin{verbatim}
\hbox:n { \color_select:n { blue } Hello, } ~ world!
\end{verbatim}

will result in “Hello,” taking the color blue, but “world!” remaining with the prevailing color outside the box.

35.1 Creating and initialising boxes

\begin{verbatim}
\box_new:N \box_new:c
\box_clear:N \box_clear:c \box_gclear:N \box_gclear:c
\end{verbatim}

\begin{description}
\item[\texttt{\box_new:N \box_new:c}] Creates a new \langle box\rangle or raises an error if the name is already taken. The declaration is global. The \langle box\rangle is initially void.
\item[\texttt{\box_clear:N \box_clear:c \box_gclear:N \box_gclear:c}] Clears the content of the \langle box\rangle by setting the box equal to \texttt{\c_empty_box}.
\end{description}
Ensures that the \langle box \rangle exists globally by applying \box_new:N if necessary, then applies \box_(g)clear:N to leave the \langle box \rangle empty.

\box_set_eq:NN \box_gset_eq:NN
\box_set_eq:NN (cN|Nc|cc)
\box_gset_eq:NN (cN|Nc|cc)

\box_if_exist_p:N \box_if_exist:NTF
\box_if_exist:cTF
\box_if_exist_p:N ⋆ \box_if_exist_p:c ⋆
\box_if_exist:NTF ⋆ \box_if_exist:cTF ⋆

Tests whether the \langle box \rangle is currently defined. This does not check that the \langle box \rangle really is a box.

New: 2012-03-03

35.2 Using boxes

\box_use:N \box_use:c
\box_move_right:nn \box_move_left:nn
\box_move_up:nn \box_move_down:nn

\box_use:N \box_use:c

Inserts the current content of the \langle box \rangle onto the current list for typesetting. An error is raised if the variable does not exist or if it is invalid.

\TeXhackers note: This is the \TeX primitive \copy.

\box_move_right:nn {dim expr} \langle box function \rangle
\box_move_left:nn {dim expr} \langle box function \rangle
\box_move_up:nn {dim expr} \langle box function \rangle
\box_move_down:nn {dim expr} \langle box function \rangle

This function operates in vertical mode, and inserts the material specified by the \langle box function \rangle such that its reference point is displaced horizontally by the given \langle dim expr \rangle from the reference point for typesetting, to the right or left as appropriate. The \langle box function \rangle should be a box operation such as \box_use:N \langle box \rangle or a “raw” box specification such as \vbox:n { xyz }.

This function operates in horizontal mode, and inserts the material specified by the \langle box function \rangle such that its reference point is displaced vertically by the given \langle dim expr \rangle from the reference point for typesetting, up or down as appropriate. The \langle box function \rangle should be a box operation such as \box_use:N \langle box \rangle or a “raw” box specification such as \vbox:n { xyz }.
35.3 Measuring and setting box dimensions

\box_dp:N  \box_dp:N \langle box \rangle
Calculates the depth (below the baseline) of the \langle box \rangle in a form suitable for use in a \langle dim expr \rangle.

\TeXhackers note: This is the \TeX\ primitive \textbackslash dp.

\box_ht:N  \box_ht:N \langle box \rangle
Calculates the height (above the baseline) of the \langle box \rangle in a form suitable for use in a \langle dim expr \rangle.

\TeXhackers note: This is the \TeX\ primitive \textbackslash ht.

\box_wd:N  \box_wd:N \langle box \rangle
Calculates the width of the \langle box \rangle in a form suitable for use in a \langle dim expr \rangle.

\TeXhackers note: This is the \TeX\ primitive \textbackslash wd.

\box_ht_plus_dp:N  \box_ht_plus_dp:N \langle box \rangle
Calculates the total vertical size (height plus depth) of the \langle box \rangle in a form suitable for use in a \langle dim expr \rangle.

\box_set_dp:Nn  \textbackslash box_set_dp:N \langle box \rangle \{ \langle dim expr \rangle \}
Set the depth (below the baseline) of the \langle box \rangle to the value of the \{ \langle dim expr \rangle \}.

\box_set_ht:Nn  \textbackslash box_set_ht:N \langle box \rangle \{ \langle dim expr \rangle \}
Set the height (above the baseline) of the \langle box \rangle to the value of the \{ \langle dim expr \rangle \}.

\box_set_wd:Nn  \textbackslash box_set_wd:N \langle box \rangle \{ \langle dim expr \rangle \}
Set the width of the \langle box \rangle to the value of the \{ \langle dim expr \rangle \}.
### 35.4 Box conditionals

```
def box_if_empty_p:N
    box_if_empty_p:N (box)

def box_if_empty_p:c
    box_if_empty:NTF (box) \{\{true code\}\} \{\{false code\}\}
```

Tests if \(\langle box\rangle\) is a empty (equal to \texttt{\c_empty_box}).

```
def box_if_horizontal_p:N
    box_if_horizontal_p:N (box)

def box_if_horizontal_p:c
    box_if_horizontal:NTF (box) \{\{true code\}\} \{\{false code\}\}
```

Tests if \(\langle box\rangle\) is a horizontal box.

```
def box_if_vertical_p:N
    box_if_vertical_p:N (box)

def box_if_vertical_p:c
    box_if_vertical:NTF (box) \{\{true code\}\} \{\{false code\}\}
```

Tests if \(\langle box\rangle\) is a vertical box.

### 35.5 The last box inserted

```
def box_set_to_last:N
    box_set_to_last:N (box)

def box_set_to_last:c
    box_set_to_last:NTF (box) \{\{true code\}\} \{\{false code\}\}
```

Sets the \(\langle box\rangle\) equal to the last item (box) added to the current partial list, removing the item from the list at the same time. When applied to the main vertical list, the \(\langle box\rangle\) is always void as it is not possible to recover the last added item.

### 35.6 Constant boxes

```
def \c_empty_box
    This is a permanently empty box, which is neither set as horizontal nor vertical.
    \texttt{\TeXhackers note}: At the \TeX{} level this is a void box.
```

### 35.7 Scratch boxes

```
def \l_tmpa_box
\l_tmpb_box
```

Scratch boxes for local assignment. These are never used by the kernel code, and so are safe for use with any \LaTeX{}3-defined function. However, they may be overwritten by other non-kernel code and so should only be used for short-term storage.

```
def \g_tmpa_box
\g_tmpb_box
```

Scratch boxes for global assignment. These are never used by the kernel code, and so are safe for use with any \EPX3-defined function. However, they may be overwritten by other non-kernel code and so should only be used for short-term storage.
35.8 Viewing box contents

\box_show:N \box_show:N \langle box \rangle \show
Shows full details of the content of the \langle box \rangle in the terminal.

\box_show:Nnn \box_show:Nnn \langle box \rangle \{\langle int \ expr_1 \rangle\} \{\langle int \ expr_2 \rangle\}
Display the contents of \langle box \rangle in the terminal, showing the first \langle int \ expr_1 \rangle items of the
box, and descending into \langle int \ expr_2 \rangle group levels.

\box_log:N \box_log:N \langle box \rangle \show
 Writes full details of the content of the \langle box \rangle to the log.

\box_log:Nnn \box_log:Nnn \langle box \rangle \{\langle int \ expr_1 \rangle\} \{\langle int \ expr_2 \rangle\}
Writes the contents of \langle box \rangle to the log, showing the first \langle int \ expr_1 \rangle items of the box,
and descending into \langle int \ expr_2 \rangle group levels.

35.9 Boxes and color

All \LaTeX3 boxes are “color safe”: a color set inside the box stops applying after the end
of the box has occurred.

35.10 Horizontal mode boxes

\hbox:n \hbox:n \langle contents \rangle
Typesets the \langle contents \rangle into a horizontal box of natural width and then includes this
box in the current list for typesetting.

\hbox_to_wd:nn \hbox_to_wd:nn \langle dim \ expr \rangle \langle contents \rangle
Typesets the \langle contents \rangle into a horizontal box of width \langle dim \ expr \rangle and then includes
this box in the current list for typesetting.

\hbox_to_zero:n \hbox_to_zero:n \langle contents \rangle
Typesets the \langle contents \rangle into a horizontal box of zero width and then includes this box
in the current list for typesetting.

\hbox_set:Nn \hbox_set:cn \hbox_gset:Nn \hbox_gset:cn
Typesets the \langle contents \rangle at natural width and then stores the result inside the \langle box \rangle.
\hbox_set_to_wd:Nnn \hbox_set_to_wd:cnm \hbox_gset_to_wd:Nnn \hbox_gset_to_wd:cnm

Typesets the \langle contents \rangle to the width given by the \langle dim expr \rangle and then stores the result inside the \langle box \rangle.

\hbox_overlap_center:n \hbox_overlap_center:n \langle contents \rangle

Typesets the \langle contents \rangle into a horizontal box of zero width such that material protrudes equally to both sides of the insertion point.

\hbox_overlap_right:n \hbox_overlap_right:n \langle contents \rangle

Typesets the \langle contents \rangle into a horizontal box of zero width such that material protrudes to the right of the insertion point.

\hbox_overlap_left:n \hbox_overlap_left:n \langle contents \rangle

Typesets the \langle contents \rangle into a horizontal box of zero width such that material protrudes to the left of the insertion point.

\hbox_set:Nw \hbox_set:cw \hbox_set_end: \hbox_set:Nw \hbox_set:cw \hbox_set_end:

Typesets the \langle contents \rangle at natural width and then stores the result inside the \langle box \rangle.

In contrast to \hbox_set:Nn this function does not absorb the argument when finding the \langle content \rangle, and so can be used in circumstances where the \langle content \rangle may not be a simple argument.

\hbox_set_to_wd:Nnw \hbox_set_to_wd:cnw \hbox_gset_to_wd:Nnw \hbox_gset_to_wd:cnw

Typesets the \langle contents \rangle to the width given by the \langle dim expr \rangle and then stores the result inside the \langle box \rangle. In contrast to \hbox_set_to_wd:Nnn this function does not absorb the argument when finding the \langle content \rangle, and so can be used in circumstances where the \langle content \rangle may not be a simple argument.

\hbox_unpack:N \hbox_unpack:N \langle box \rangle

Unpacks the content of the horizontal \langle box \rangle, retaining any stretching or shrinking applied when the \langle box \rangle was set.

\TeXhackers note: This is the \TeX{} primitive \texttt{\unhcopy}.

\section{35.11 Vertical mode boxes}

Vertical boxes inherit their baseline from their contents. The standard case is that the baseline of the box is at the same position as that of the last item added to the box. This means that the box has no depth unless the last item added to it had depth. As a result most vertical boxes have a large height value and small or zero depth. The exception are
_top boxes, where the reference point is that of the first item added. These tend to have a large depth and small height, although the latter is typically non-zero.

\vbox:n \{\textit{contents}\}

Typesets the \textit{contents} into a vertical box of natural height and includes this box in the current list for typesetting.

\vbox_top:n \{\textit{contents}\}

Typesets the \textit{contents} into a vertical box of natural height and includes this box in the current list for typesetting. The baseline of the box is equal to that of the first item added to the box.

\vbox_to_ht:nn \{\textit{dim expr}\} \{\textit{contents}\}

Typesets the \textit{contents} into a vertical box of height \textit{dim expr} and then includes this box in the current list for typesetting.

\vbox_to_zero:n \{\textit{contents}\}

Typesets the \textit{contents} into a vertical box of zero height and then includes this box in the current list for typesetting.

\vbox_set:Nn \{\textit{box}\} \{\textit{contents}\}

Typesets the \textit{contents} at natural height and then stores the result inside the \textit{box}.

\vbox_set_to_ht:Nnn \{\textit{box}\} \{\textit{dim expr}\} \{\textit{contents}\}

Typesets the \textit{contents} to the height given by the \textit{dim expr} and then stores the result inside the \textit{box}.

\vbox_set:Nw \{\textit{box}\} \{\textit{contents}\} \vbox_set_end:

Typesets the \textit{contents} at natural height and then stores the result inside the \textit{box}. In contrast to \vbox_set:Nn this function does not absorb the argument when finding the \textit{content}, and so can be used in circumstances where the \textit{content} may not be a simple argument.
\vbox_set_to_ht:Nnw \vbox_set_to_ht:cnw \vbox_gset_to_ht:Nnw \vbox_gset_to_ht:cnw
\vbox_set_split_to_ht:NNn \vbox_set_split_to_ht:(cNn|Ncn|ccn) \vbox_gset_split_to_ht:NNn \vbox_gset_split_to_ht:(cNn|Ncn|ccn)

Typesets the \textit{\textless{}contents\textgreater{}} to the height given by the \textit{\textless{}dim expr\textgreater{}} and then stores the result inside the \textit{\textless{}box\textgreater{}}. In contrast to \vbox_set_to_ht:Nnn this function does not absorb the argument when finding the \textit{\textless{}content\textgreater{}}, and so can be used in circumstances where the \textit{\textless{}content\textgreater{}} may not be a simple argument.

\vbox_unpack:N \vbox_unpack:c

Unpacks the content of the vertical \textit{\textless{}box\textgreater{}}, retaining any stretching or shrinking applied when the \textit{\textless{}box\textgreater{}} was set.

\TeXhacksnote{This is the \TeX{} primitive \unvcopy{.}}

### 35.12 Using boxes efficiently

The functions above for using box contents work in exactly the same way as for any other expl3 variable. However, for efficiency reasons, it is also useful to have functions which drop box contents on use. When a box is dropped, the box becomes empty at the group level \textit{where the box was originally set} rather than necessarily \textit{at the current group level}. For example, with

\hbox_set:Nn \l_tmpa_box { A }
\group_begin:
\hbox_set:Nn \l_tmpa_box { B }
\group_begin:
\box_use_drop:N \l_tmpa_box
\group_end:
\box_show:N \l_tmpa_box
\group_end:
\box_show:N \l_tmpa_box

the first use of \texttt{\box{}show:N} will show an entirely cleared (void) box, and the second will show the letter \texttt{A} in the box.

These functions should be preferred when the content of the box is no longer required after use. Note that due to the unusual scoping behaviour of \texttt{\drop} functions they may be applied to both local and global boxes: the latter will naturally be set and thus cleared at a global level.

309
\usebox {box}

Inserts the current content of the \texttt{box} onto the current list for typesetting then drops the box content. An error is raised if the variable does not exist or if it is invalid. This function may be applied to local or global boxes.

\textbf{\TeX{}hackers note:} This is the \TeX{} primitive \texttt{\usebox}.

\setboxeq{box}{box_1}{box_2}

Sets the content of \texttt{box_1} equal to that of \texttt{box_2}, then drops \texttt{box_2}.

\textbf{New: 2019-01-17}

\gsetboxeq{box}{box_1}{box_2}

Sets the content of \texttt{box_1} globally equal to that of \texttt{box_2}, then drops \texttt{box_2}.

\textbf{New: 2019-01-17}

\hboxunpackdrop{box}

Unpacks the content of the horizontal \texttt{box}, retaining any stretching or shrinking applied when the \texttt{box} was set. The original \texttt{box} is then dropped.

\textbf{\TeX{}hackers note:} This is the \TeX{} primitive \texttt{\hboxunpack}.

\textbf{New: 2019-01-17}

\vboxunpackdrop{box}

Unpacks the content of the vertical \texttt{box}, retaining any stretching or shrinking applied when the \texttt{box} was set. The original \texttt{box} is then dropped.

\textbf{\TeX{}hackers note:} This is the \TeX{} primitive \texttt{\vboxunpack}.

\subsection*{35.13 Affine transformations}

Affine transformations are changes which (informally) preserve straight lines. Simple translations are affine transformations, but are better handled in \TeX{} by doing the translation first, then inserting an unmodified box. On the other hand, rotation and resizing of boxed material can best be handled by modifying boxes. These transformations are described here.
Resizes the \(<\text{box}\>\) to fit within the given \(\langle x\text{-size} \rangle\) (horizontally) and \(\langle y\text{-size} \rangle\) (vertically); both of the sizes are dimension expressions. The \(\langle y\text{-size} \rangle\) is the height only: it does not include any depth. The updated \(\langle \text{box} \rangle\) is an hbox, irrespective of the nature of the \(\langle \text{box} \rangle\) before the resizing is applied. The final size of the \(\langle \text{box} \rangle\) is the smaller of \{\langle x\text{-size} \rangle\} and \{\langle y\text{-size} \rangle\}, i.e. the result fits within the dimensions specified. Negative sizes cause the material in the \(\langle \text{box} \rangle\) to be reversed in direction, but the reference point of the \(\langle \text{box} \rangle\) is unchanged. Thus a negative \(\langle y\text{-size} \rangle\) results in the \(\langle \text{box} \rangle\) having a depth dependent on the height of the original and vice versa.

Resizes the \(\langle \text{box} \rangle\) to fit within the given \(\langle x\text{-size} \rangle\) (horizontally) and \(\langle y\text{-size} \rangle\) (vertically); both of the sizes are dimension expressions. The \(\langle y\text{-size} \rangle\) is the total vertical size (height plus depth). The updated \(\langle \text{box} \rangle\) is an hbox, irrespective of the nature of the \(\langle \text{box} \rangle\) before the resizing is applied. The final size of the \(\langle \text{box} \rangle\) is the smaller of \{\langle x\text{-size} \rangle\} and \{\langle y\text{-size} \rangle\}, i.e. the result fits within the dimensions specified. Negative sizes cause the material in the \(\langle \text{box} \rangle\) to be reversed in direction, but the reference point of the \(\langle \text{box} \rangle\) is unchanged. Thus a negative \(\langle y\text{-size} \rangle\) results in the \(\langle \text{box} \rangle\) having a depth dependent on the height of the original and vice versa.

Resizes the \(\langle \text{box} \rangle\) to \(\langle y\text{-size} \rangle\) (vertically), scaling the horizontal size by the same amount; \(\langle y\text{-size} \rangle\) is a dimension expression. The \(\langle y\text{-size} \rangle\) is the height only: it does not include any depth. The updated \(\langle \text{box} \rangle\) is an hbox, irrespective of the nature of the \(\langle \text{box} \rangle\) before the resizing is applied. A negative \(\langle y\text{-size} \rangle\) causes the material in the \(\langle \text{box} \rangle\) to be reversed in direction, but the reference point of the \(\langle \text{box} \rangle\) is unchanged. Thus a negative \(\langle y\text{-size} \rangle\) results in the \(\langle \text{box} \rangle\) having a depth dependent on the height of the original and vice versa.
Resizes the \texttt{box} to \texttt{(y-size)} (vertically), scaling the horizontal size by the same amount; \texttt{(y-size)} is a dimension expression. The \texttt{(y-size)} is the total vertical size (height plus depth). The updated \texttt{box} is an \texttt{hbox}, irrespective of the nature of the \texttt{box} before the resizing is applied. A negative \texttt{(y-size)} causes the material in the \texttt{box} to be reversed in direction, but the reference point of the \texttt{box} is unchanged. Thus a negative \texttt{(y-size)} results in the \texttt{box} having a depth dependent on the height of the original and \textit{vice versa}.

Resizes the \texttt{box} to \texttt{(x-size)} (horizontally), scaling the vertical size by the same amount; \texttt{(x-size)} is a dimension expression. The updated \texttt{box} is an \texttt{hbox}, irrespective of the nature of the \texttt{box} before the resizing is applied. A negative \texttt{(x-size)} causes the material in the \texttt{box} to be reversed in direction, but the reference point of the \texttt{box} is unchanged. Thus a negative \texttt{(x-size)} results in the \texttt{box} having a depth dependent on the height of the original and \textit{vice versa}.

Resizes the \texttt{box} to \texttt{(x-size)} (horizontally) and \texttt{(y-size)} (vertically): both of the sizes are dimension expressions. The \texttt{(y-size)} is the height only and does not include any depth. The updated \texttt{box} is an \texttt{hbox}, irrespective of the nature of the \texttt{box} before the resizing is applied. Negative sizes cause the material in the \texttt{box} to be reversed in direction, but the reference point of the \texttt{box} is unchanged. Thus a negative \texttt{(y-size)} results in the \texttt{box} having a depth dependent on the height of the original and \textit{vice versa}.

Resizes the \texttt{box} to \texttt{(x-size)} (horizontally) and \texttt{(y-size)} (vertically): both of the sizes are dimension expressions. The \texttt{(y-size)} is the total vertical size (height plus depth). The updated \texttt{box} is an \texttt{hbox}, irrespective of the nature of the \texttt{box} before the resizing is applied. Negative sizes cause the material in the \texttt{box} to be reversed in direction, but the reference point of the \texttt{box} is unchanged. Thus a negative \texttt{(y-size)} results in the \texttt{box} having a depth dependent on the height of the original and \textit{vice versa}. 
\box_rotate:Nn \box_rotate:cn \box_grotate:Nn \box_grotate:cn

Rotates the \texttt{<box>} by \texttt{(angle)} (a \texttt{fp expr} in degrees) anti-clockwise about its reference point. The reference point of the updated box is moved horizontally such that it is at the left side of the smallest rectangle enclosing the rotated material. The updated \texttt{<box>} is an \texttt{hbox}, irrespective of the nature of the \texttt{<box>} before the rotation is applied.

\box_scale:Nnn \box_scale:cnn \box_gscale:Nnn \box_gscale:cnn

Scales the \texttt{<box>} by factors \texttt{(x-scale)} and \texttt{(y-scale)} in the horizontal and vertical directions, respectively (both scales are \texttt{fp expr}). The updated \texttt{<box>} is an \texttt{hbox}, irrespective of the nature of the \texttt{<box>} before the scaling is applied. Negative scalings cause the material in the \texttt{<box>} to be reversed in direction, but the reference point of the \texttt{<box>} is unchanged. Thus a negative \texttt{(y-scale)} results in the \texttt{<box>} having a depth dependent on the height of the original and vice versa.

\box_set_clipped:N \box_set_clipped:c \box_gset_clipped:N \box_gset_clipped:c

Clips the \texttt{<box>} in the output so that only material inside the bounding box is displayed in the output. The updated \texttt{<box>} is an \texttt{hbox}, irrespective of the nature of the \texttt{<box>} before the clipping is applied. Additional box levels are also generated by this operation. \textbf{TeXhackers note}: Clipping is implemented by the driver, and as such the full content of the box is placed in the output file. Thus clipping does not remove any information from the raw output, and hidden material can therefore be viewed by direct examination of the file.

\box_set_trim:Nnnnn \box_set_trim:cnnnn \box_gset_trim:Nnnnn \box_gset_trim:cnnnn

Adjusts the bounding box of the \texttt{<box>} \texttt{(left)} is removed from the left-hand edge of the bounding box, \texttt{(right)} from the right-hand edge and so forth. All adjustments are \texttt{dim exprs}. Material outside of the bounding box is still displayed in the output unless \texttt{\box_set_clipped:N} is subsequently applied. The updated \texttt{<box>} is an \texttt{hbox}, irrespective of the nature of the \texttt{<box>} before the trim operation is applied. Additional box levels are also generated by this operation. The behavior of the operation where the trims requested is greater than the size of the box is undefined.

\box_set_viewport:Nnnnn \box_set_viewport:cnnnn \box_gset_viewport:Nnnnn \box_gset_viewport:cnnnn

Adjusts the bounding box of the \texttt{<box>} such that it has lower-left coordinates \texttt{(llx), (lly)} and upper-right coordinates \texttt{(urx), (ury)}. All four coordinate positions are \texttt{dim exprs}. Material outside of the bounding box is still displayed in the output unless \texttt{\box_set_clipped:N} is subsequently applied. The updated \texttt{<box>} is an \texttt{hbox}, irrespective of the nature of the \texttt{<box>} before the viewport operation is applied. Additional box levels are also generated by this operation.
35.15 Primitive box conditionals

\if_hbox:N \if_hbox:N (box)  
\langle true code \rangle 
\else: \langle false code \rangle 
\fi:
Tests is (box) is a horizontal box.

\TeXhackers note: This is the \TeX primitive \ifhbox.

\if_vbox:N \if_vbox:N (box)  
\langle true code \rangle 
\else: \langle false code \rangle 
\fi:
Tests is (box) is a vertical box.

\TeXhackers note: This is the \TeX primitive \ifvbox.

\if_box_empty:N \if_box_empty:N (box)  
\langle true code \rangle 
\else: \langle false code \rangle 
\fi:
Tests is (box) is an empty (void) box.

\TeXhackers note: This is the \TeX primitive \ifvoid.
Chapter 36

The \texttt{l3coffins} module
Coffin code layer

The material in this module provides the low-level support system for coffins. For details about the design concept of a coffin, see the \texttt{xcoffins} module (in the \texttt{l3experimental} bundle).

36.1 Creating and initialising coffins

\begin{verbatim}
\coffin_new:N \coffin
\coffin_new:N \coffin
New: 2011-08-17
\end{verbatim}

Creates a new \texttt{\coffin} or raises an error if the name is already taken. The declaration is global. The \texttt{\coffin} is initially empty.

\begin{verbatim}
\coffin_clear:N \coffin
\coffin_clear:c
\coffin_gclear:N \coffin_gclear:c
\end{verbatim}

New: 2011-08-17
Updated: 2019-01-21

Clears the content of the \texttt{\coffin}.

\begin{verbatim}
\coffin_set_eq:NN \coffin \coffin
\coffin_set_eq:NN \coffin \coffin
\coffin_gset_eq:NN \coffin_gset_eq:NN
\end{verbatim}

New: 2011-08-17
Updated: 2019-01-21

Sets both the content and poles of \texttt{\coffin_1} equal to those of \texttt{\coffin_2}.

\begin{verbatim}
\coffin_if_exist_p:N \coffin
\coffin_if_exist_p:c \coffin
\coffin_if_exist:NTF \coffin \coffin \coffin
\coffin_if_exist:CTF \coffin
\end{verbatim}

Tests whether the \texttt{\coffin} is currently defined.
36.2 Setting coffin content and poles

\hcoffin_set:Nn  ⟨coffin⟩{⟨material⟩}
Typesets the ⟨material⟩ in horizontal mode, storing the result in the ⟨coffin⟩. The standard poles for the ⟨coffin⟩ are then set up based on the size of the typeset material.

\hcoffin_set:Nw  ⟨coffin⟩ (material)  \hcoffin_set_end:
Typesets the ⟨material⟩ in horizontal mode, storing the result in the ⟨coffin⟩. The standard poles for the ⟨coffin⟩ are then set up based on the size of the typeset material. These functions are useful for setting the entire contents of an environment in a coffin.

\vcoffin_set:Nnn  ⟨coffin⟩ {⟨width⟩} {⟨material⟩}
Typesets the ⟨material⟩ in vertical mode constrained to the given ⟨width⟩ and stores the result in the ⟨coffin⟩. The standard poles for the ⟨coffin⟩ are then set up based on the size of the typeset material.

\vcoffin_set:Nnw  ⟨coffin⟩ {⟨width⟩} (material)  \vcoffin_set_end:
Typesets the ⟨material⟩ in vertical mode constrained to the given ⟨width⟩ and stores the result in the ⟨coffin⟩. The standard poles for the ⟨coffin⟩ are then set up based on the size of the typeset material. These functions are useful for setting the entire contents of an environment in a coffin.

\coffin_set_horizontal_pole:Nnn  ⟨coffin⟩ {⟨pole⟩} {⟨offset⟩}
Sets the ⟨pole⟩ to run horizontally through the ⟨coffin⟩. The ⟨pole⟩ is placed at the ⟨offset⟩ from the baseline of the ⟨coffin⟩. The ⟨offset⟩ should be given as a dimension expression.
\c coffin_set_vertical_pole:Nnn \c coffin_set_vertical_pole:cnn 
\c coffin_gset_vertical_pole:Nnn \c coffin_gset_vertical_pole:cnn

Sets the \langle pole \rangle to run vertically through the \langle coffin \rangle. The \langle pole \rangle is placed at the \langle offset \rangle from the left-hand edge of the bounding box of the \langle coffin \rangle. The \langle offset \rangle should be given as a dimension expression.

\c coffin_reset_poles:N \c coffin_greset_poles:N

Resets the poles of the \langle coffin \rangle to the standard set, removing any custom or inherited poles. The poles will therefore be equal to those that would be obtained from \c hcoffin_set:Nn or similar; the bounding box of the coffin is not reset, so any material outside of the formal bounding box will not influence the poles.

36.3 Coffin affine transformations

\c coffin_resize:Nnn \c coffin_resize:cnn \c coffin_gresize:Nnn \c coffin_gresize:cnn

Resized the \langle coffin \rangle to \langle width \rangle and \langle total-height \rangle, both of which should be given as dimension expressions.

\c coffin_rotate:Nn \c coffin_rotate:cn \c coffin_grotate:Nn \c coffin_grotate:cn

Rotates the \langle coffin \rangle by the given \langle angle \rangle (given in degrees counter-clockwise). This process rotates both the coffin content and poles. Multiple rotations do not result in the bounding box of the coffin growing unnecessarily.

\c coffin_scale:Nnn \c coffin_scale:cnn \c coffin_gscale:Nnn \c coffin_gscale:cnn

Scales the \langle coffin \rangle by a factors \langle x-scale \rangle and \langle y-scale \rangle in the horizontal and vertical directions, respectively. The two scale factors should be given as real numbers.
36.4 Joining and using coffins

This function attaches \(\text{coffin}_2\) to \(\text{coffin}_1\) such that the bounding box of \(\text{coffin}_1\) is not altered, i.e. \(\text{coffin}_2\) can protrude outside of the bounding box of the coffin. The alignment is carried out by first calculating \(\text{handle}_1\), the point of intersection of \(\text{coffin}_1\) and \(\text{coffin}_2\), and \(\text{handle}_2\), the point of intersection of \(\text{coffin}_1\) and \(\text{coffin}_2\). \(\text{coffin}_2\) is then attached to \(\text{coffin}_1\) such that the relationship between \(\text{handle}_1\) and \(\text{handle}_2\) is described by the \(\text{x-offset}\) and \(\text{y-offset}\). The two offsets should be given as dimension expressions.

This function joins \(\text{coffin}_2\) to \(\text{coffin}_1\) such that the bounding box of \(\text{coffin}_1\) may expand. The new bounding box covers the area containing the bounding boxes of the two original coffins. The alignment is carried out by first calculating \(\text{handle}_1\), the point of intersection of \(\text{coffin}_1\) and \(\text{coffin}_2\), and \(\text{handle}_2\), the point of intersection of \(\text{coffin}_2\) and \(\text{coffin}_2\). \(\text{coffin}_2\) is then attached to \(\text{coffin}_1\) such that the relationship between \(\text{handle}_1\) and \(\text{handle}_2\) is described by the \(\text{x-offset}\) and \(\text{y-offset}\). The two offsets should be given as dimension expressions.

Typesetting is carried out by first calculating \(\text{handle}\), the point of intersection of \(\text{pole}_1\) and \(\text{pole}_2\). The coffin is then typeset in horizontal mode such that the relationship between the current reference point in the document and the \(\text{handle}\) is described by the \(\text{x-offset}\) and \(\text{y-offset}\). The two offsets should be given as dimension expressions. Typesetting a coffin is therefore analogous to carrying out an alignment where the “parent” coffin is the current insertion point.

36.5 Measuring coffins

Calculates the depth (below the baseline) of the \(\text{coffin}\) in a form suitable for use in a \(\text{dim expr}\).
\texttt{\coffin ht:N} \ (\texttt{coffin})
\texttt{\coffin ht:c}
Calculates the height (above the baseline) of the \texttt{\textit{coffin}} in a form suitable for use in a \texttt{\textit{dim expr}}.

\texttt{\coffin wd:N} \ (\texttt{coffin})
\texttt{\coffin wd:c}
Calculates the width of the \texttt{\textit{coffin}} in a form suitable for use in a \texttt{\textit{dim expr}}.

36.6 Coffin diagnostics

\texttt{\coffin display_handles:Nn} \ (\texttt{coffin}) \ {\texttt{(color)}}
\texttt{\coffin display_handles:cn}
This function first calculates the intersections between all of the \texttt{\textit{poles}} of the \texttt{\textit{coffin}} to give a set of \texttt{\textit{handles}}. It then prints the \texttt{\textit{coffin}} at the current location in the source, with the position of the \texttt{\textit{handles}} marked on the coffin. The \texttt{\textit{handles}} are labelled as part of this process: the locations of the \texttt{\textit{handles}} and the labels are both printed in the \texttt{\textit{color}} specified.

\texttt{\coffin mark_handle:Nnnn} \ (\texttt{coffin}) \ {\texttt{(pole\_1)}} \ {\texttt{(pole\_2)}} \ {\texttt{(color)}}
\texttt{\coffin mark_handle:cn}
This function first calculates the \texttt{\textit{handle}} for the \texttt{\textit{coffin}} as defined by the intersection of \texttt{\textit{pole\_1}} and \texttt{\textit{pole\_2}}. It then marks the position of the \texttt{\textit{handle}} on the \texttt{\textit{coffin}}. The \texttt{\textit{handle}} are labelled as part of this process: the location of the \texttt{\textit{handle}} and the label are both printed in the \texttt{\textit{color}} specified.

\texttt{\coffin show_structure:N} \ (\texttt{coffin})
\texttt{\coffin show_structure:cn}
This function shows the structural information about the \texttt{\textit{coffin}} in the terminal. The width, height and depth of the typeset material are given, along with the location of all of the poles of the coffin.

Notice that the poles of a coffin are defined by four values: the $x$ and $y$ coordinates of a point that the pole passes through and the $x$- and $y$-components of a vector denoting the direction of the pole. It is the ratio between the later, rather than the absolute values, which determines the direction of the pole.

\texttt{\coffin log_structure:N} \ (\texttt{coffin})
\texttt{\coffin log_structure:cn}
This function writes the structural information about the \texttt{\textit{coffin}} in the log file. See also \texttt{\coffin show_structure:N} which displays the result in the terminal.

\texttt{\coffin show:N} \ (\texttt{coffin})
\texttt{\coffin show:c}
\texttt{\coffin log:N} \ (\texttt{coffin})
\texttt{\coffin log:c}
Shows full details of poles and contents of the \texttt{\textit{coffin}} in the terminal or log file. See \texttt{\coffin show_structure:N} and \texttt{\box show:N} to show separately the pole structure and the contents.

319
\coffin_show:Nnn \coffin_show:cnn \coffin_log:Nnn \coffin_log:cnn

Shows poles and contents of the ⟨coffin⟩ in the terminal or log file, showing the first ⟨int expr⟩ items in the coffin, and descending into ⟨int expr⟩ group levels. See \coffin-_show_structure:N and \box_show:Nnn to show separately the pole structure and the contents.

36.7 Constants and variables

\c_empty_coffin A permanently empty coffin.

\l_tmpa_coffin \l_tmpb_coffin Scratch coffins for local assignment. These are never used by the kernel code, and so are safe for use with any \TeX3-defined function. However, they may be overwritten by other non-kernel code and so should only be used for short-term storage.

\g_tmpa_coffin \g_tmpb_coffin Scratch coffins for global assignment. These are never used by the kernel code, and so are safe for use with any \TeX3-defined function. However, they may be overwritten by other non-kernel code and so should only be used for short-term storage.
Chapter 37

The l3color module
Color support

37.1 Color in boxes

Controlling the color of text in boxes requires a small number of control functions, so that the boxed material uses the color at the point where it is set, rather than where it is used.

\color_group_begin: \color_group_begin: \color_group_end: \color_group_end:
\color_ensure_current: \color_ensure_current:

New: 2011-09-03

Creates a color group: one used to “trap” color settings. This grouping is built in to for example \hbox_set:Nn.

\color_ensure_current:
Ensures that material inside a box uses the foreground color at the point where the box is set, rather than that in force when the box is used. This function should usually be used within a \color_group_begin: ... \color_group_end: group.

37.2 Color models

A color model is a way to represent sets of colors. Different models are particularly suitable for different output methods, e.g. screen or print. Parameter-based models can describe a very large number of unique colors, and have a varying number of axes which define a color space. In contrast, various proprietary models are available which define spot colors (more formally separations).

Core models are used to pass color information to output; these are “native” to l3color. Core models use real numbers in the range [0, 1] to represent values. The core models supported here are

- gray Grayscale color, with a single axis running from 0 (fully black) to 1 (fully white)
- rgb Red-green-blue color, with three axes, one for each of the components
• **cmyk** Cyan-magenta-yellow-black color, with four axes, one for each of the components.

There are also interface models: these are convenient for users but have to be manipulated before storing/passing to the backend. Interface models are primarily integer-based: see below for more detail. The supported interface models are

• **Gray** Grayscale color, with a single axis running from 0 (fully black) to 15 (fully white)

• **hsb** Hue-saturation-brightness color, with three axes, all real values in the range 
\([0, 1]\) for hue saturation and brightness

• **Hsb** Hue-saturation-brightness color, with three axes, integer in the range \([0, 360]\) for hue, real values in the range \([0, 1]\) for saturation and brightness

• **HSB** Hue-saturation-brightness color, with three axes, integers in the range \([0, 240]\) for hue, saturation and brightness

• **HTML** HTML format representation of RGB color given as a single six-digit hexadecimal number

• **RGB** Red-green-blue color, with three axes, one for each of the components, values as integers from 0 to 255

• **wave** Light wavelength, a real number in the range 380 to 780 (nanometres)

All interface models are internally stored as **rgb**.

Finally, there are a small number of models which are parsed to allow data transfer from **xcolor** but which should not be used by end-users. These are

• **cmy** Cyan-magenta-yellow color with three axes, one for each of the components; converted to **cmyk**

• **tHsb** “Tuned” hue-saturation-brightness color with three axes, integer in the range 
\([0, 360]\) for hue, real values in the range \([0, 1]\) for saturation and brightness; converted to **rgb** using the standard tuning map defined by **xcolor**

• **&spot** Spot color tint with one value; treated as a gray tint as spot color data is not available for extraction

To allow parsing of data from **xcolor**, any leading model up the first : will be discarded; the approach of selecting an internal form for data is **not** used in **3color**.

Additional models may be created to allow mixing of separation colors with each other or with those from other models. See Section 37.9 for more detail of color support for additional models.

When color is selected by model, the (**values**) given are specified as a comma-separated list. The length of the list will therefore be determined by the detail of the model involved.

Color models (and interconversion) are complex, and more details are given in the manual to the **LaTeX** **2e xcolor** package and in the **PostScript Language Reference Manual**, published by Addison–Wesley.
37.3 Color expressions

In addition to allowing specification of color by model and values, \texttt{13color} also supports color expressions. These are created by combining one or more color names, with the amount of each specified as a value in the range 0–100. The value should be given between ! symbols in the expression. Thus for example

\begin{verbatim}
  red!50!green
\end{verbatim}

is a mixture of 50\% red and 50\% green. A trailing value is interpreted as implicitly followed by \texttt{white}, and so

\begin{verbatim}
  red!25
\end{verbatim}

specifies 25\% red mixed with 75\% white.

Where the models for the mixed colors are different, the model of the first color is used. Thus

\begin{verbatim}
  red!50!cyan
\end{verbatim}

will result in a color specification using the \texttt{rgb} model, made up of 50\% red and 50\% of cyan \textit{expressed in rgb}. This may be important as color model interconversion is not exact.

The one exception to the above is where the first model in an expression is \texttt{gray}. In this case, the order of mixing is “swapped” internally, so that for example

\begin{verbatim}
  black!50!red
\end{verbatim}

has the same result as

\begin{verbatim}
  red!50!black
\end{verbatim}

(the predefined colors \texttt{black} and \texttt{white} use the \texttt{gray} model).

Where more than two colors are mixed in an expression, evaluation takes place in a stepwise fashion. Thus in

\begin{verbatim}
  cyan!50!magenta!10!yellow
\end{verbatim}

the sub-expression

\begin{verbatim}
  cyan!50!magenta
\end{verbatim}

is first evaluated to give an intermediate color specification, before the second step

\begin{verbatim}
  <intermediate>!10!yellow
\end{verbatim}

where \texttt{<intermediate>} represents this transitory calculated value.

Within a color expression, . may be used to represent the color active for typesetting (the current color). This allows for example

\begin{verbatim}
  .!50
\end{verbatim}

to mean a mixture of 50\% of current color with white.

(Color expressions supported here are a subset of those provided by the \LaTeX\ \texttt{xcolor} package. At present, only such features as are clearly useful have been added here.)
37.4 Named colors

Color names are stored in a single namespace, which makes them accessible as part of color expressions. Whilst they are not reserved in a technical sense, the names black, white, red, green, blue, cyan, magenta and yellow have special meaning and should not be redefined. Color names should be made up of letters, numbers and spaces only: other characters are reserved for use in color expressions. In particular, . represents the current color at the start of a color expression.

\color_set:nn \color_set:nn {\langle name\rangle} \lbrace\langle color expression\rangle\rbrace

Evaluates the \langle color expression\rangle and stores the resulting color specification as the \langle name\rangle.

\color_set:nnn \color_set:nn {\langle name\rangle} \lbrace\langle model(s)\rangle\rbrace \lbrace\langle value(s)\rbrace

Stores the color specification equivalent to the \langle model(s)\rangle and \langle values\rangle as the \langle name\rangle.

\color_set_eq:nn \color_set_eq:nn {\langle name1\rangle} {\langle name2\rangle}

Copies the color specification in \langle name2\rangle to \langle name1\rangle. The special name . may be used to represent the current color, allowing it to be saved to a name.

\color_if_exist_p:n \color_if_exist:nTF {\langle name\rangle} \lbrace\langle true code\rangle\rbrace \lbrace\langle false code\rbrace

Tests whether \langle name\rangle is currently defined to provide a color specification.

\color_show:n \color_show:n {\langle name\rangle}
\color_log:n \color_log:n {\langle name\rangle}

Displays the color specification stored in the \langle name\rangle on the terminal or log file.

37.5 Selecting colors

General selection of color is safe when split across pages: a stack is used to ensure that the correct color is re-selected on the new page. These commands set the current color (.): other more specialised functions such as fill and stroke selectors do not adjust this value.

\color_select:n \color_select:n \lbrace\langle color expression\rangle\rbrace

Parses the \langle color expression\rangle and then activates the resulting color specification for typeset material.

\color_select:nn \color_select:nn {\langle model(s)\rangle} \lbrace\langle value(s)\rbrace

Activates the color specification equivalent to the \langle model(s)\rangle and \langle value(s)\rangle for typeset material.

\l_{color\_fixed\_model\_tl} When this is set to a non-empty value, colors will be converted to the specified model when they are selected. Note that included images and similar are not influenced by this setting.
37.6 Colors for fills and strokes

Colors for drawing operations and so forth are split into strokes and fills (the latter may also be referred to as non-stroke color). The fill color is used for text under normal circumstances. Depending on the backend, stroke color may use a stack, in which case it exhibits the same page breaking behavior as general color. However, \texttt{dvips/dvisvgm} do not support this, and so color will need to be contained within a scope, such as \texttt{\draw_begin:/\draw_end:}.

\begin{verbatim}
\color_fill:n \color_stroke:n
\color_fill:nn \color_stroke:nn
\end{verbatim}

Parses the \texttt{⟨color expression⟩} and then activates the resulting color specification for filling or stroking.

\begin{verbatim}
\color_fill:nn \color_stroke:nn
\end{verbatim}

Activates the color specification equivalent to the \texttt{⟨model(s)⟩} and \texttt{⟨value(s)⟩} for filling or stroking.

\begin{verbatim}
\color_math:nn \color_math:nnn
\end{verbatim}

Works as for \texttt{\color_select:n(n)} but applies color only to the math mode \texttt{⟨content⟩}. The function does not generate a group and the \texttt{⟨content⟩} therefore retains its math atom states. Sub/superscripts are also properly handled.

\begin{verbatim}
\l_color_math_active_tl
\end{verbatim}

This list controls which tokens are considered as math active and should therefore be replaced by their definition during searching for sub/superscripts.

37.7 Multiple color models

When selecting or setting a color with an explicit model, it is possible to give values for more than one model at one time. This is particularly useful where automated conversion between models does not give the desired outcome. To do this, the list of models and list of values are both subdivided using / characters (as for the similar function in \texttt{xcolor}). For example, to save a color with explicit \texttt{cmyk} and \texttt{rgb} values, one could use

\begin{verbatim}
\color_set:nnn { foo } { cmyk / rgb }
{ 0.1 , 0.2 , 0.3 , 0.4 / 0.1 , 0.2 , 0.3 }
\end{verbatim}
The manually-specified conversion will be used in preference to automated calculation whenever the model(s) listed are used: both in expressions and when a fixed model is active.

Similarly, the same syntax can be applied to directly selecting a color.

```
\color_select:nn \{ cmyk / rgb \}
\{ 0.1 , 0.2 , 0.3 , 0.4 / 0.1, 0.2, 0.3 \}
```

Again, this list is used when a fixed model is active: the first entry is used unless there is a fixed model matching one of the other entries.

### 37.8 Exporting color specifications

The major use of color expressions is in setting typesetting output, but there are other places in which some form of color information is required. These may need data in a different format or using a different model to the internal representation. Thus a set of functions are available to export colors in different formats.

Valid export targets are

- **backend** Two brace groups: the first containing the model, the second containing space-separated values appropriate for the model; this is the format required by backend functions of expl3
- **comma-sep-cmyk** Comma-separated cyan-magenta-yellow-black values
- **comma-sep-rgb** Comma-separated red-green-blue values suitable for use as a PDF annotation color
- **HTML** Uppercase two-digit hexadecimal values, expressing a red-green-blue color; the digits are *not* separated
- **space-sep-cmyk** Space-separated cyan-magenta-yellow-black values
- **space-sep-rgb** Space-separated red-green-blue values suitable for use as a PDF annotation color

```
\color_export:nn \langle \text{color expression} \rangle \{ \text{format} \} \{ \text{tl} \}
```

Parses the \langle color expression \rangle as described earlier, then converts to the \langle format \rangle specified and assigns the data to the \langle tl \rangle.

```
\color_export:nnn \langle \text{model} \rangle \{ \text{value(s)} \} \{ \text{format} \} \{ \text{tl} \}
```

Expresses the combination of \langle model \rangle and \langle value(s) \rangle in an internal representation, then converts to the \langle format \rangle specified and assigns the data to the \langle tl \rangle.
37.9 Creating new color models

Additional color models are required to support specialist workflows, for example those involving separations (see https://helpx.adobe.com/indesign/using/spot-process-colors.html for details of the use of separations in print). Color models may be split into families; for the standard device-based color models (DeviceCMYK, DeviceRGB, DeviceGray), these are synonymous. This is not generally the case: see the PDF reference for more details. (Note that l3color uses the shorter names cmyk, etc.)

\color_model_new:nnn \color_model_new:nnn \{model\} \{(family)\} \{params\}

Creates a new (model) which is derived from the color model (family). The latter should be one of

- DeviceN
- ICCBased
- Separation

(The (family) may be given in mixed case as-in the PDF reference: internally, case of these strings is folded.) Depending on the (family), one or more (params) are mandatory or optional.

For a Separation space, there are three compulsory keys.

- name The name of the Separation, for example the formal name of a spot color ink. Such a (name) may contain spaces, etc., which are not permitted in the (model).
- alternative-model An alternative device colorspace, one of cmyk, rgb, gray or CIELAB. The three parameter-based models work as described above; see below for details of CIELAB colors.
- alternative-values A comma-separated list of values appropriate to the alternative-model. This information is used by the PDF application if the Separation is not available.

CIELAB color separations are created using the alternative-model = CIELAB setting. These colors must also have an illuminant key, one of a, c, e, d50, d55, d65 or d75. The alternative-values in this case are the three parameters L*, a* and b* of the CIELAB model. Full details of this device-independent color approach are given in the documentation to the colorspace package.

CIELAB colors cannot be converted into other device-dependent color spaces, and as such, mixing can only occur if colors set up using the CIELAB model are also given with an alternative parameter-based model. If that is not the case, l3color will fallback to using black as the colorant in any mixing.

For a DeviceN space, there is one compulsory key.

- names The names of the components of the DeviceN space. Each should be either the (name) of a Separation model, a process color name (cyan, etc.) or the special name none.

For a ICCBased space, there is one compulsory key.

- file The name of the file containing the profile.
37.9.1 Color profiles

Color profiles are used to ensure color accuracy by linking to collaboration. Applying a profile can be used to standardise color which is otherwise device-dependence.

\color_profile_apply:nn \color_profile_apply:nn \{(\textit{profile})\} \{(\textit{model})\}

This function applies a \textit{(profile)} to one of the device \textit{(models)}. The profile will then apply to all color of the selected \textit{(model)}. The \textit{(profile)} should specify an ICC profile file. The \textit{(model)} has to be one the standard device models: \textit{cmyk, gray} or \textit{rgb}.
Chapter 38

The l3pdf module
Core PDF support

38.1 Objects

38.1.1 Named objects

An \textit{object} name should fully expand to tokens suitable for use in a label-like context.

\begin{verbatim}
\pdf_object_new:n \pdf_object_new:n \langle object \rangle

Declares \textit{object} as a PDF object. The object may be referenced from this point on, and written later using \texttt{\pdf_object_write:nnn}.

\pdf_object_write:nnn \pdf_object_write:nnn \{\langle object \rangle\} \{\langle type \rangle\} \{\langle content \rangle\}

\pdf_object_write:nne

Writes the \textit{content} as content of the \textit{object}. Depending on the \textit{type} declared for the object, the format required for the \textit{data} will vary

\begin{itemize}
  \item \texttt{array} A space-separated list of values
  \item \texttt{dict} Key–value pairs in the form /\langle key \rangle \langle value \rangle
  \item \texttt{fstream} Two brace groups: \textit{file name} and \textit{file content}
  \item \texttt{stream} Two brace groups: \textit{attributes (dictionary)} and \textit{stream contents}
\end{itemize}

\begin{verbatim}
\pdf_object_ref:n \pdf_object_ref:n \{\langle object \rangle\}

\pdf_object_ref:n \pdf_object_ref:n \{\langle object \rangle\}

\pdf_object_if_exist_p:n \pdf_object_if_exist:nTF \{\langle object \rangle\} \{\langle true code \rangle\} \{\langle false code \rangle\}

\pdf_object_if_exist_p:n \pdf_object_if_exist:nTF \{\langle object \rangle\} \{\langle true code \rangle\} \{\langle false code \rangle\}
\end{verbatim}

Inserts the appropriate information to reference the \textit{object} in for example page resource allocation. If the \textit{object} does not exist then the function expands to a reference to object zero; no PDF indirect object ever has this number, so this is a marker for error.

\begin{verbatim}
\pdf_object_if_exist_p:n \pdf_object_if_exist:nTF \{\langle object \rangle\} \{\langle true code \rangle\} \{\langle false code \rangle\}
\end{verbatim}

Tests whether an object with name \{\textit{object}\} has been defined.
38.1.2 Indexed objects

Objects can also be created using a pair of ⟨class⟩ and ⟨index⟩; the ⟨class⟩ argument should expand to character tokens, whilst the ⟨index⟩ is an ⟨int expr⟩ and starts at 1. For large families of objects, this approach is more efficient than using individual names.

\pdf_object_new_indexed:nn \pdf_object_new_indexed:nn ⟨class⟩ ⟨index⟩

Declares a PDF object of ⟨class⟩ and ⟨index⟩. The object may be referenced from this point on, and written later using \pdf_object_write_indexed:nnnn.

\pdf_object_write_indexed:nnnn \pdf_object_write_indexed:nnnn ⟨class⟩ ⟨index⟩ ⟨type⟩ ⟨content⟩

Writes the ⟨content⟩ as content of the object of ⟨class⟩ and ⟨index⟩. Depending on the ⟨type⟩ declared for the object, the format required for the ⟨content⟩ will vary

array A space-separated list of values
dict Key–value pairs in the form /⟨key⟩ ⟨value⟩
fstream Two brace groups: ⟨file name⟩ and ⟨file content⟩
stream Two brace groups: ⟨attributes (dictionary)⟩ and ⟨stream contents⟩

\pdf_object_ref_indexed:nn * \pdf_object_ref_indexed:nn ⟨class⟩ ⟨index⟩

Inserts the appropriate information to reference the object of ⟨class⟩ and ⟨index⟩ in for example page resource allocation. If the ⟨class⟩/⟨index⟩ combination does not exist then the function expands to a reference to object zero; no PDF indirect object ever has this number, so this is a marker for error.

38.1.3 General functions

\pdf_object_unnamed_write:nn \pdf_object_unnamed_write:nn ⟨type⟩ ⟨content⟩
\pdf_object_unnamed_write:ne

Writes the ⟨content⟩ as content of an anonymous object. Depending on the ⟨type⟩, the format required for the ⟨data⟩ will vary

array A space-separated list of values
dict Key–value pairs in the form /⟨key⟩ ⟨value⟩
fstream Two brace groups: ⟨attributes (dictionary)⟩ and ⟨file name⟩
stream Two brace groups: ⟨attributes (dictionary)⟩ and ⟨stream contents⟩
Inserts the appropriate information to reference the last <object> created. This is particularly useful for anonymous objects.

Inserts the appropriate information to reference the <abspage>; the latter is expanded fully before further processing.

### 38.2 Version

\pdf_version_compare_p:Nn \pdf_version_compare_p:Nn \pdf_version_compare:NnTF \pdf_version_compare:NnTF

Compares the version of the PDF being created with the <version> string specified, using the <comparator>. Either the <true code> or <false code> will be left in the output stream.

Sets the <version> of the PDF being created. The min version will not alter the output version unless it is currently lower than the <version> requested.

This function may only be used up to the point where the PDF file is initialised. With dvips it sets \pdf_version_major: and \pdf_version_minor: and allows to compare the values with \pdf_version_compare:Nn, but the PDF version itself still has to be set with the command line option -dCompatibilityLevel of ps2pdf.

Expands to the currently-active PDF version.

### 38.3 Page (media) size

\pdf_pagesize_gset:nn \pdf_pagesize_gset:nn

Sets the page size (mediabox) of the PDF being created to the <width> and <height>, both of which are <dimexpr>. The page size can only be set at the start of the output with dvips; with other backends, this can be adjusted on a per-page basis.

### 38.4 Compression

\pdf_uncompress: \pdf_uncompress:

Disables any compression of the PDF, where possible.

This function may only be used up to the point where the PDF file is initialised.
### 38.5 Destinations

Destinations are the places a link jumped too. Unlike the name may suggest they don’t described an exact location in the PDF. Instead a destination contains a reference to a page along with an instruction how to display this page. The normally used “XYZ top left zoom” for example instructs the viewer to show the page with the given zoom and the top left corner at the top left coordinates—which then gives the impression that there is an anchor at this position.

If an instruction takes a coordinate, it is calculated by the following commands relative to the location the command is issued. So to get a specific coordinate one has to move the command to the right place.

```
\pdf_destination:nn \pdf_destination:nn {⟨name⟩} ⟨⟨type or integer⟩⟩
```

This creates a destination. ⟨⟨type or integer⟩⟩ can be one of fit, fith, fitv, fitb, fitbh, fitbv, fitr, xyz or an integer representing a scale factor in percent. fitr here gives only a lightweight version of /FitR: The backend code defines fitr so that it will with pdfLaTeX and LuaLaTeX use the coordinates of the surrounding box, with dvips and dvipdfmx it falls back to fit. For full control use \pdf_destination:nnnn.

The keywords match to the PDF names as described in the following tabular.

<table>
<thead>
<tr>
<th>Keyword</th>
<th>PDF</th>
<th>Remarks</th>
</tr>
</thead>
<tbody>
<tr>
<td>fit</td>
<td>/Fit</td>
<td>Fits the page to the window</td>
</tr>
<tr>
<td>fith</td>
<td>/FitH top</td>
<td>Fits the width of the page to the window</td>
</tr>
<tr>
<td>fitv</td>
<td>/FitV left</td>
<td>Fits the height of the page to the window</td>
</tr>
<tr>
<td>fitb</td>
<td>/FitB</td>
<td>Fits the page bounding box to the window</td>
</tr>
<tr>
<td>fitbh</td>
<td>/FitBH top</td>
<td>Fits the width of the page bounding box to the window.</td>
</tr>
<tr>
<td>fitbv</td>
<td>/FitBV left</td>
<td>Fits the height of the page bounding box to the window.</td>
</tr>
<tr>
<td>fitr</td>
<td>/FitR left bottom right top</td>
<td>Fits the rectangle specified by the four coordinates to the window (see above for the restrictions)</td>
</tr>
<tr>
<td>xyz</td>
<td>/XYZ left top null</td>
<td>Sets a coordinate but doesn’t change the zoom.</td>
</tr>
</tbody>
</table>

```
\pdf_destination:nnnn \pdf_destination:nnnn {⟨name⟩} ⟨⟨width⟩} ⟨⟨height⟩} ⟨⟨depth⟩}
```

This creates a destination with /FitR type with the given dimensions relative to the current location. The destination is in a box of size zero, but it doesn’t switch to horizontal mode.
Part VII
Implementation
Chapter 39

l3bootstrap implementation

39.1 The \pdfstrcmp primitive in X\textTeX

Only pdf\TeX{} has a primitive called \texttt{\pdfstrcmp}. The X\textTeX{} version is just \texttt{\strcmp}, so there is some shuffling to do. As this is still a real primitive, using the pdf\TeX{} name is “safe”.

\verb|\begingroup\expandafter\expandafter\expandafter\endgroup|\verb|
\verb|\expandafter\ifx\csname pdfstrcmp\endcsname\relax|
\verb|\let\pdfstrcmp\strcmp|
\verb|\fi|

39.2 Loading support Lua code

When Lua\TeX{} is used there are various pieces of Lua code which need to be loaded. The code itself is defined in l3luatex and is extracted into a separate file. Thus here the task is to load the Lua code both now and (if required) at the start of each job.

\verb|\begingroup\expandafter\expandafter\expandafter\endgroup|\verb|
\verb|\expandafter\ifx\csname directlua\endcsname\relax|
\verb|\else|
\verb|\ifnum\luatexversion<110 %|
\verb|\input{ltluatex}|
\verb|\fi|
\verb|\else|
\verb|\ifnum\luatexversion<110 %|
\verb|\newluabytecode\@expl@luadata@bytecode|
\verb|\directlua{require("expl3")}|
\verb|\fi|

For Lua\TeX{} we make sure the basic support is loaded: this is only necessary in plain.

\verb|\begingroup\expandafter\expandafter\expandafter\endgroup|\verb|
\verb|\expandafter\ifx\csname newcatcodetable\endcsname\relax|
\verb|\input{ltluatex}%|
\verb|\fi|
\verb|\begingroup\expandafter\expandafter\expandafter\endgroup|\verb|
\verb|\expandafter\ifx\csname newluabytecode\endcsname\relax|
\verb|\envluabytecode\@expl@luadata@bytecode|
\verb|\fi|
\verb|\directlua{require("expl3")}|

334
As the user might be making a custom format, no assumption is made about matching package mode with only loading the Lua code once. Instead, a query to Lua reveals what mode is in operation.

```latex
\directlua{
if status.ini_version then
tex.write("1")
end
>>0 %
}\everyjob\expandafter{\the\everyjob\csname\detokenize{lua_now:n}\endcsname{require("expl3")}}%}
\fi
\fi
\fi

39.3 Engine requirements

The code currently requires \texttt{\varepsilon-\TeX}, the set of \texttt{pdf\Pi\TeX} extensions \emph{including} \texttt{\expanded}, and for Unicode engines the ability to generate arbitrary character tokens by expansion. That is covered by all supported engines since \TeX\ Live 2019, which we therefore use as a baseline for engine and \B\Pi\TeX\ format support. For \LaTeX, we require at least Lua 5.3 and the \texttt{token.setLua} function. This is available at least since Lua\TeX\ 1.10, which again is the one in \TeX\ Live 2019. (u)p\TeX only gained \texttt{\ifincsname} for \TeX\ Live 2020, but at present that primitive is unused in expl3 so for the present it’s not tested. If and when that changes, we will need to revisit the code here.
The code environment is now set up.

\ExplSyntaxOff

Before changing any category codes, in package mode we need to save the situation before loading. Note the set up here means that once applied \ExplSyntaxOff becomes a "do nothing" command until \ExplSyntaxOn is used.

(End of definition for \ExplSyntaxOff. This function is documented on page 9.)

The code environment is now set up.

39.4 The LATEX3 code environment
\l__kernel_expl_bool The status for code syntax: this is on at present.
\global\chardef\l__kernel_expl_bool = 1

(End of definition for \l__kernel_expl_bool.)

\ExplSyntaxOn The idea here is that multiple \ExplSyntaxOn calls are not going to mess up category codes, and that multiple calls to \ExplSyntaxOff are also not wasting time. Applying \ExplSyntaxOn alters the definition of \ExplSyntaxOff and so in package mode this function should not be used until after the end of the loading process!

\protected \def \ExplSyntaxOn
{\bool_if:NF \l__kernel_expl_bool
 {\cs_set_protected:Npe \ExplSyntaxOff
 {\char_set_catcode:nn { 9 } { \char_value_catcode:n { 9 } }
 \char_set_catcode:nn { 32 } { \char_value_catcode:n { 32 } }
 \char_set_catcode:nn { 34 } { \char_value_catcode:n { 34 } }
 \char_set_catcode:nn { 58 } { \char_value_catcode:n { 58 } }
 \char_set_catcode:nn { 94 } { \char_value_catcode:n { 94 } }
 \char_set_catcode:nn { 95 } { \char_value_catcode:n { 95 } }
 \char_set_catcode:nn { 124 } { \char_value_catcode:n { 124 } }
 \char_set_catcode:nn { 126 } { \char_value_catcode:n { 126 } }
 \tex_endlinechar:D = \tex_the:D \tex_endlinechar:D \scan_stop:
 \bool_set_false:N \l__kernel_expl_bool
 \cs_set_protected:Npn \ExplSyntaxOff { }
}
\char_set_catcode-ignore:n { 9 } \% tab
\char_set_catcode-ignore:n { 32 } \% space
\char_set_catcode-other:n { 34 } \% double quote
\char_set_catcode-letter:n { 58 } \% colon
\char_set_catcode_math_superscript:n { 94 } \% circumflex
\char_set_catcode_letter:n { 95 } \% underscore
\char_set_catcode_other:n { 124 } \% pipe
\char_set_catcode_space:n { 126 } \% tilde
\tex_endlinechar:D = 32 \scan_stop:
\bool_set_true:N \l__kernel_expl_bool
}

(End of definition for \ExplSyntaxOn. This function is documented on page 9.)

{/package}
Chapter 40

\texttt{l3names implementation}

The prefix here is \texttt{kernel}. A few places need \texttt{@@} to be left as is; this is obtained as \texttt{@@@}. The code here simply renames all of the primitives to new, internal, names.

The \texttt{\let} primitive is renamed by hand first as it is essential for the entire process to follow. This also uses \texttt{\global}, as that way we avoid leaving an unneeded csname in the hash table.

Everything is inside a (rather long) group, which keeps \texttt{\_kernel\_primitive:NN} trapped.

\begin{verbatim}
\long \def \_kernel\_primitive:NN #1#2
{ \tex\_global:D \tex\_let:D #2 #1 }
\end{verbatim}

To allow extracting "just the names", a bit of DocStrip fiddling.

In the current incarnation of this module, all \TeX\ primitives are given a new name of the form \texttt{\tex\_oldname:D}. But first three special cases which have symbolic original names. These are given modified new names, so that they may be entered without catcode tricks.

Now all the other primitives.
Primitives introduced by \TeX.

\begin{itemize}
  \item \_\_kernel\_primitive:NN \toks \tex_toks:D
  \item \_\_kernel\_primitive:NN \toksdef \tex_toksdef:D
  \item \_\_kernel\_primitive:NN \tolerance \tex_tolerance:D
  \item \_\_kernel\_primitive:NN \topmark \tex_topmark:D
  \item \_\_kernel\_primitive:NN \topskip \tex_topskip:D
  \item \_\_kernel\_primitive:NN \tracingcommands \tex_tracingcommands:D
  \item \_\_kernel\_primitive:NN \tracinglostchars \tex_tracinglostchars:D
  \item \_\_kernel\_primitive:NN \tracingmacros \tex_tracingmacros:D
  \item \_\_kernel\_primitive:NN \tracingonline \tex_tracingonline:D
  \item \_\_kernel\_primitive:NN \tracingoutput \tex_tracingoutput:D
  \item \_\_kernel\_primitive:NN \tracingpages \tex_tracingpages:D
  \item \_\_kernel\_primitive:NN \tracingparagraphs \tex_tracingparagraphs:D
  \item \_\_kernel\_primitive:NN \tracingrestores \tex_tracingrestores:D
  \item \_\_kernel\_primitive:NN \tracingstats \tex_tracingstats:D
  \item \_\_kernel\_primitive:NN \uccode \tex_uccode:D
  \item \_\_kernel\_primitive:NN \uchyph \tex_uchyph:D
  \item \_\_kernel\_primitive:NN \underline \tex_underline:D
  \item \_\_kernel\_primitive:NN \unhbox \tex_unhbox:D
  \item \_\_kernel\_primitive:NN \unhcopy \tex_unhcopy:D
  \item \_\_kernel\_primitive:NN \unkern \tex_unkern:D
  \item \_\_kernel\_primitive:NN \unpenalty \tex_unpenalty:D
  \item \_\_kernel\_primitive:NN \unskip \tex_unskip:D
  \item \_\_kernel\_primitive:NN \unvbox \tex_unvbox:D
  \item \_\_kernel\_primitive:NN \unvcopy \tex_unvcopy:D
  \item \_\_kernel\_primitive:NN \uppercase \tex_uppercase:D
  \item \_\_kernel\_primitive:NN \vadjust \tex_vadjust:D
  \item \_\_kernel\_primitive:NN \valign \tex_valign:D
  \item \_\_kernel\_primitive:NN \vbadness \tex_vbadness:D
  \item \_\_kernel\_primitive:NN \vbox \tex_vbox:D
  \item \_\_kernel\_primitive:NN \vcenter \tex_vcenter:D
  \item \_\_kernel\_primitive:NN \vfil \tex_vfil:D
  \item \_\_kernel\_primitive:NN \vfill \tex_vfill:D
  \item \_\_kernel\_primitive:NN \vfilneg \tex_vfilneg:D
  \item \_\_kernel\_primitive:NN \vfuzz \tex_vfuzz:D
  \item \_\_kernel\_primitive:NN \voffset \tex_voffset:D
  \item \_\_kernel\_primitive:NN \vrule \tex_vrule:D
  \item \_\_kernel\_primitive:NN \vaize \tex_vaize:D
  \item \_\_kernel\_primitive:NN \vakip \tex vakip:D
  \item \_\_kernel\_primitive:NN \vsplit \tex_vsplit:D
  \item \_\_kernel\_primitive:NN \vss \tex_vss:D
  \item \_\_kernel\_primitive:NN \vtop \tex_vtop:D
  \item \_\_kernel\_primitive:NN \wd \tex_wd:D
  \item \_\_kernel\_primitive:NN \widowpenalty \tex Widowpenalty:D
  \item \_\_kernel\_primitive:NN \write \tex_write:D
  \item \_\_kernel\_primitive:NN \xdef \tex_xdef:D
  \item \_\_kernel\_primitive:NN \xleaders \tex_xleaders:D
  \item \_\_kernel\_primitive:NN \xspaceskip \tex_xspaceskip:D
  \item \_\_kernel\_primitive:NN \year \tex_year:D
\end{itemize}
\texttt{\_kernel\_primitive:NN \currentgrouptype \tex\_currentgrouptype:\textit{D}}
\texttt{\_kernel\_primitive:NN \currentifbranch \tex\_currentifbranch:\textit{D}}
\texttt{\_kernel\_primitive:NN \currentiflevel \tex\_currentiflevel:\textit{D}}
\texttt{\_kernel\_primitive:NN \currentiftype \tex\_currentiftype:\textit{D}}
\texttt{\_kernel\_primitive:NN \detokenize \tex\_detokenize:\textit{D}}
\texttt{\_kernel\_primitive:NN \dimexpr \tex\_dimexpr:\textit{D}}
\texttt{\_kernel\_primitive:NN \displaywidowpenalties \tex\_displaywidowpenalties:\textit{D}}
\texttt{\_kernel\_primitive:NN \endL \tex\_endL:\textit{D}}
\texttt{\_kernel\_primitive:NN \endR \tex\_endR:\textit{D}}
\texttt{\_kernel\_primitive:NN \eTeXrevision \tex\_eTeXrevision:\textit{D}}
\texttt{\_kernel\_primitive:NN \eTeXversion \tex\_eTeXversion:\textit{D}}
\texttt{\_kernel\_primitive:NN \everyeof \tex\_everyeof:\textit{D}}
\texttt{\_kernel\_primitive:NN \firstmarks \tex\_firstmarks:\textit{D}}
\texttt{\_kernel\_primitive:NN \everyeof \tex\_everyeof:\textit{D}}
\texttt{\_kernel\_primitive:NN \fontcharht \tex\_fontcharht:\textit{D}}
\texttt{\_kernel\_primitive:NN \fontcharwd \tex\_fontcharwd:\textit{D}}
\texttt{\_kernel\_primitive:NN \fontchardp \tex\_fontchardp:\textit{D}}
\texttt{\_kernel\_primitive:NN \fontcharic \tex\_fontcharic:\textit{D}}
\texttt{\_kernel\_primitive:NN \interactionmode \tex\_interactionmode:\textit{D}}
\texttt{\_kernel\_primitive:NN \interlinepenalties \tex\_interlinepenalties:\textit{D}}
\texttt{\_kernel\_primitive:NN \lastlinefit \tex\_lastlinefit:\textit{D}}
\texttt{\_kernel\_primitive:NN \lastnodetype \tex\_lastnodetype:\textit{D}}
\texttt{\_kernel\_primitive:NN \firstmarks \tex\_firstmarks:\textit{D}}
\texttt{\_kernel\_primitive:NN \middle \tex\_middle:\textit{D}}
\texttt{\_kernel\_primitive:NN \mutoglue \tex\_mutoglue:\textit{D}}
\texttt{\_kernel\_primitive:NN \numexpr \tex\_numexpr:\textit{D}}
\texttt{\_kernel\_primitive:NN \pagediscards \tex\_pagediscards:\textit{D}}
\texttt{\_kernel\_primitive:NN \parshapedimen \tex\_parshapedimen:\textit{D}}
\texttt{\_kernel\_primitive:NN \parshapeindent \tex\_parshapeindent:\textit{D}}
\texttt{\_kernel\_primitive:NN \parshapelength \tex\_parshapelength:\textit{D}}
\texttt{\_kernel\_primitive:NN \pandedition \tex\_pandedition:\textit{D}}
\texttt{\_kernel\_primitive:NN \pandedition \tex\_pandedition:\textit{D}}
\texttt{\_kernel\_primitive:NN \protected \tex\_protected:\textit{D}}
\texttt{\_kernel\_primitive:NN \readline \tex\_readline:\textit{D}}
\texttt{\_kernel\_primitive:NN \savinghyphcodes \tex\_savinghyphcodes:\textit{D}}
\texttt{\_kernel\_primitive:NN \savingvdiscards \tex\_savingvdiscards:\textit{D}}
\texttt{\_kernel\_primitive:NN \scantokens \tex\_scantokens:\textit{D}}
\texttt{\_kernel\_primitive:NN \showgroups \tex\_showgroups:\textit{D}}
\texttt{\_kernel\_primitive:NN \showifs \tex\_showifs:\textit{D}}
\texttt{\_kernel\_primitive:NN \showtokens \tex\_showtokens:\textit{D}}
\texttt{\_kernel\_primitive:NN \splitbotmarks \tex\_splitbotmarks:\textit{D}}
\texttt{\_kernel\_primitive:NN \splitdiscards \tex\_splitdiscards:\textit{D}}
\texttt{\_kernel\_primitive:NN \splitfirstmarks \tex\_splitfirstmarks:\textit{D}}
\texttt{\_kernel\_primitive:NN \splitfirstmarks \tex\_splitfirstmarks:\textit{D}}
\texttt{\_kernel\_primitive:NN \TeXXeTstate \tex\_TeXXeTstate:\textit{D}}
\texttt{\_kernel\_primitive:NN \topmarks \tex\_topmarks:\textit{D}}
\texttt{\_kernel\_primitive:NN \tracingassigns \tex\_tracingassigns:\textit{D}}
Post-ε-TEx primitives do not always end up with the same name in all engines, if indeed they are available cross-engine anyway. We therefore take the approach of preferring the shortest name that makes sense. First, we deal with the primitives introduced by pdfTEx which directly relate to PDF output: these are copied with the names unchanged.
These are not related to PDF output and either already appear in other engines without the \pdf prefix, or might reasonably do so at some future stage. We therefore drop the leading pdf here.
The version primitives are not related to PDF mode but are \pdfTeX-specific, so again are carried forward unchanged.

These ones appear in \pdfTeX but don’t have \pdf in the name at all: no decisions to make.
Post pdf\TeX\ primitive availability gets more complex. Both X\TeX\ and Lua\TeX\ have varying names for some primitives from pdf\TeX. Particularly for Lua\TeX\ tracking all of that would be hard. Instead, we now check that we only save primitives if they actually exist.

Some pdf\TeX\ primitives are handled here because they got dropped in Lua\TeX\ but the corresponding internal names are emulated later. The Lua code is already loaded at this point, so we shouldn’t overwrite them.

X\TeX-specific primitives. Note that X\TeX’s \texttt{\strcmp} is handled earlier and is “rolled up” into \texttt{pdfstrcmp}. A few cross-compatibility names which lack the pdflatex of the original are handled later.
Primitives from pdFTeX that XeFTeX renames: also helps with LuaFTeX.
Primitives from LuaTeX, some of which have been ported back to XeLaTeX.
Primitives from pdfTeX that LuaTeX renames.

\_\_kernel\_primitive:NN \mathscriptcharmode \tex_mathscriptcharmode:D
\_\_kernel\_primitive:NN \mathstyle \tex_mathstyle:D
\_\_kernel\_primitive:NN \mathsurroundmode \tex_mathsurroundmode:D
\_\_kernel\_primitive:NN \mathsurroundskip \tex_mathsurroundskip:D
\_\_kernel\_primitive:NN \nohrule \tex_nohrule:D
\_\_kernel\_primitive:NN \nokerns \tex_nokerns:D
\_\_kernel\_primitive:NN \noligs \tex_noligs:D
\_\_kernel\_primitive:NN \nospaces \tex_nospaces:D
\_\_kernel\_primitive:NN \novrule \tex_novrule:D
\_\_kernel\_primitive:NN \outputbox \tex_outputbox:D
\_\_kernel\_primitive:NN \pagebottomoffset \tex_pagebottomoffset:D
\_\_kernel\_primitive:NN \pagedir \tex_pagedir:D
\_\_kernel\_primitive:NN \pagedirection \tex_pagedirection:D
\_\_kernel\_primitive:NN \pageleftoffset \tex_pageleftoffset:D
\_\_kernel\_primitive:NN \pagerightoffset \tex_pagerightoffset:D
\_\_kernel\_primitive:NN \pagetopoffset \tex_pagetopoffset:D
\_\_kernel\_primitive:NN \pardir \tex_pardir:D
\_\_kernel\_primitive:NN \pardirection \tex_pardirection:D
\_\_kernel\_primitive:NN \pdfextension \tex_pdfextension:D
\_\_kernel\_primitive:NN \pdffeedback \tex_pdffeedback:D
\_\_kernel\_primitive:NN \pdfvariable \tex_pdfvariable:D
\_\_kernel\_primitive:NN \postexhyphenchar \tex_postexhyphenchar:D
\_\_kernel\_primitive:NN \posthyphenchar \tex_posthyphenchar:D
\_\_kernel\_primitive:NN \prebinoppenalty \tex_prebinoppenalty:D
\_\_kernel\_primitive:NN \predisplaygapfactor \tex_predisplaygapfactor:D
\_\_kernel\_primitive:NN \preexhyphenchar \tex_preexhyphenchar:D
\_\_kernel\_primitive:NN \prehyphenchar \tex_prehyphenchar:D
\_\_kernel\_primitive:NN \prerelpenalty \tex_prerelpenalty:D
\_\_kernel\_primitive:NN \rightghost \tex_rightghost:D
\_\_kernel\_primitive:NN \savecatcodetable \tex_savecatcodetable:D
\_\_kernel\_primitive:NN \scantextokens \tex_scantextokens:D
\_\_kernel\_primitive:NN \setfontid \tex_setfontid:D
\_\_kernel\_primitive:NN \shapemode \tex_shapemode:D
\_\_kernel\_primitive:NN \suppresifsnameerror \tex_suppresifsnameerror:D
\_\_kernel\_primitive:NN \suppresslongerror \tex_suppresslongerror:D
\_\_kernel\_primitive:NN \suppressmathparerror \tex_suppressmathparerror:D
\_\_kernel\_primitive:NN \suppressoutererror \tex_suppressoutererror:D
\_\_kernel\_primitive:NN \textdir \tex_textdir:D
\_\_kernel\_primitive:NN \textdirection \tex_textdirection:D
\_\_kernel\_primitive:NN \toksapp \tex_toksapp:D
\_\_kernel\_primitive:NN \tokspre \tex_tokspre:D
\_\_kernel\_primitive:NN \tpack \tex_tpack:D
\_\_kernel\_primitive:NN \variablefam \tex_variablefam:D
\_\_kernel\_primitive:NN \vpack \tex_vpack:D
\_\_kernel\_primitive:NN \wordboundary \tex_wordboundary:D
\_\_kernel\_primitive:NN \xtoksapp \tex_xtoksapp:D
\_\_kernel\_primitive:NN \xtokspre \tex_xtokspre:D
\_\_kernel\_primitive:NN \adjustspacing \tex_adjustspacing:D
\_\_kernel\_primitive:NN \copyfont \tex_copyfont:D
\_\_kernel\_primitive:NN \draftmode \tex_draftmode:D
The set of Unicode math primitives were introduced by \texttt{Xe\LaTeX} and \texttt{Lua\LaTeX} in a somewhat complex fashion: a few first as \texttt{\LaTeX...}, which were then renamed with \texttt{\LaTeX} having a lot more. These names now all start \texttt{U...} and mainly \texttt{Umath...}.
Primitives from \LaTeX. 

\texttt{Primitives from \LaTeX.} 

\texttt{Primitives from \LaTeX.} 

\texttt{Primitives from \LaTeX.} 

\texttt{Primitives from \LaTeX.} 

\texttt{Primitives from \LaTeX.} 

\texttt{Primitives from \LaTeX.} 

\texttt{Primitives from \LaTeX.} 

\texttt{Primitives from \LaTeX.} 

\texttt{Primitives from \LaTeX.} 

\texttt{Primitives from \LaTeX.} 

\texttt{Primitives from \LaTeX.} 

\texttt{Primitives from \LaTeX.} 

\texttt{Primitives from \LaTeX.} 

\texttt{Primitives from \LaTeX.} 

\texttt{Primitives from \LaTeX.} 

\texttt{Primitives from \LaTeX.} 

\texttt{Primitives from \LaTeX.} 

\texttt{Primitives from \LaTeX.} 

\texttt{Primitives from \LaTeX.} 

\texttt{Primitives from \LaTeX.} 

\texttt{Primitives from \LaTeX.} 

\texttt{Primitives from \LaTeX.} 

\texttt{Primitives from \LaTeX.} 

\texttt{Primitives from \LaTeX.} 

\texttt{Primitives from \LaTeX.} 

\texttt{Primitives from \LaTeX.} 

\texttt{Primitives from \LaTeX.} 

\texttt{Primitives from \LaTeX.} 

\texttt{Primitives from \LaTeX.} 

\texttt{Primitives from \LaTeX.} 

\texttt{Primitives from \LaTeX.} 

\texttt{Primitives from \LaTeX.} 

\texttt{Primitives from \LaTeX.} 

\texttt{Primitives from \LaTeX.} 

\texttt{Primitives from \LaTeX.} 

\texttt{Primitives from \LaTeX.} 

\texttt{Primitives from \LaTeX.} 

\texttt{Primitives from \LaTeX.} 

\texttt{Primitives from \LaTeX.} 

\texttt{Primitives from \LaTeX.} 

\texttt{Primitives from \LaTeX.} 

\texttt{Primitives from \LaTeX.} 

\texttt{Primitives from \LaTeX.} 

\texttt{Primitives from \LaTeX.} 

\texttt{Primitives from \LaTeX.} 

\texttt{Primitives from \LaTeX.} 

\texttt{Primitives from \LaTeX.} 

\texttt{Primitives from \LaTeX.} 

\texttt{Primitives from \LaTeX.} 

\texttt{Primitives from \LaTeX.} 

\texttt{Primitives from \LaTeX.} 

\texttt{Primitives from \LaTeX.} 

\texttt{Primitives from \LaTeX.} 

\texttt{Primitives from \LaTeX.} 

\texttt{Primitives from \LaTeX.} 

\texttt{Primitives from \LaTeX.}
Primitives from upTeX.

\__kernel_primitive:NN \currentcjktoken \tex_currentcjktoken:D
\__kernel_primitive:NN \disablecjktoken \tex_disablecjktoken:D
\__kernel_primitive:NN \enablecjktoken \tex_enablecjktoken:D
\__kernel_primitive:NN \forcecjktoken \tex_forcecjktoken:D
\__kernel_primitive:NN \kchar \tex_kchar:D
\__kernel_primitive:NN \kchardef \tex_kchardef:D
\__kernel_primitive:NN \kuten \tex_kuten:D
\__kernel_primitive:NN \uptexrevision \tex_uptexrevision:D
\__kernelPrimitive:NN \uptexversion \tex_uptexversion:D

Omega primitives provided by p\TeX\ (listed separately mainly to allow understanding of their source).

\__kernel_primitive:NN \odelcode \tex_odelcode:D
\__kernel_primitive:NN \odelimiter \tex_odelimiter:D
\__kernel_primitive:NN \omathaccent \tex_omathaccent:D
\__kernel_primitive:NN \omathchar \tex_omathchar:D
\__kernel_primitive:NN \omathchardef \tex_omathchardef:D
\__kernel_primitive:NN \omathcode \tex_omathcode:D
\__kernel_primitive:NN \oradical \tex_oradical:D

Newer cross-engine primitives.

\__kernel_primitive:NN \partokencontext \tex_partokencontext:D
\__kernel_primitive:NN \partokenname \tex_partokenname:D
\__kernel_primitive:NN \showstream \tex_showstream:D
\__kernelPrimitive:NN \tracingstacklevels \tex_tracingstacklevels:D

End of the “just the names” part of the source.

\tex_endgroup:D

\LaTeX\,\epsilon\, moves a few primitives, so these are sorted out. In newer versions of \LaTeX\,\epsilon, expl3 is loaded rather early, so only some primitives are already renamed, so we need two tests here. At the beginning of the \LaTeX\,\epsilon format, the primitives \texttt{\end} and \texttt{\input} are renamed, and only later on the other ones.

\tex_ifdefined:D \@@end
\tex_let:D \tex_end:D \@@end
\tex_let:D \tex_input:D \@@input
\tex_fi:D

If \texttt{\@@hyph} is defined, we are loading expl3 in a pre-2020/10/01 release of \LaTeX\,\epsilon, so a few other primitives have to be tested as well.

\tex_ifdefined:D \@@hyph
\tex_let:D \tex_everydisplay:D \frozen@everydisplay
\tex_let:D \tex_everymath:D \frozen@everymath
\tex_let:D \tex_hyphen:D \@@hyph
\tex_let:D \tex_italiccorrection:D \@@italiccorr
\tex_let:D \tex_underline:D \@@underline

359
The \texttt{shipout} primitive is particularly tricky as a number of packages want to hook in here. First, we see if a sufficiently-new kernel has saved a copy: if it has, just use that. Otherwise, we need to check each of the possible packages/classes that might move it: here, we are looking for those which do not delay action to the \texttt{AtBeginDocument} hook. (We cannot use \texttt{primitive} as that doesn’t allow us to make a direct copy of the primitive \texttt{itself.}) As we know that \LaTeXe is in use, we use it’s \texttt{@for} loop here.

```
\begin{verbatim}
1235 \texttt{tex_ifdefined:D \@@@@shipout}
1236 \texttt{tex_let:D \texttt{tex_shipout:D \@@@@shipout}}
1237 \texttt{tex_fi:D}
1238 \texttt{tex_begingroup:D}
1239 \texttt{tex_edef:D \l_tmpa_tl { \texttt{tex_string:D \shipout} }}
1240 \texttt{tex_edef:D \l_tmpb_tl { \texttt{tex_meaning:D \shipout} }}
1241 \texttt{tex_ifx:D \l_tmpa_tl \l_tmpb_tl}
1242 \texttt{tex_else:D}
1243 \texttt{tex_expandafter:D \@for \texttt{tex_expandafter:D \@tempa \texttt{tex_string:D :=}}}
1244 \texttt{\@ROPshipout}
1245 \texttt{\dup@shipout}
1246 \texttt{\GPTorg@shipout}
1247 \texttt{\LL@shipout}
1248 \texttt{\mem@oldshipout}
1249 \texttt{\opem@shipout}
1250 \texttt{\pgfpages@originalshipout}
1251 \texttt{\pr@shipout}
1252 \texttt{\Shipout}
1253 \texttt{\verso@orig@shipout}
1254 \texttt{\do}
1255 \texttt{\ifx:D \l_tmpa_tl \l_tmpb_tl}
1256 \texttt{\{ \texttt{tex_expandafter:D \texttt{tex_meaning:D \@@shipout} \}}}
1257 \texttt{\tex_ifx:D \l_tmpa_tl \l_tmpb_tl}
1258 \texttt{\tex_global:D \texttt{tex_expandafter:D \texttt{tex_set:D}}}
1259 \texttt{\texttt{tex_expandafter:D \texttt{tex_shipout:D \@@shipout} \}}
1260 \texttt{\tex_fi:D}
1261 \end{verbatim}
```

Some tidying up is needed for \texttt{(pdf)tracingfonts}. Newer Lua\TeX{} has this simply as \texttt{tracingfonts}, but that is overwritten by the \LaTeXe kernel. So any spurious definition has to be removed, then the real version saved either from the pdf\TeX{} name or from Lua\TeX{}. In the latter case, we leave \texttt{@tracingfonts} available: this might be useful and almost all \LaTeXe users will have expl3 loaded by fontspec. (We follow the usual kernel convention that \texttt{@@} is used for saved primitives.)

```
\begin{verbatim}
\texttt{tex_let:D \texttt{tex_tracingfonts:D \texttt{tex_undefined:D}}}
\texttt{tex_ifdefined:D \texttt{pdftracingfonts}}
\texttt{tex_let:D \texttt{tex_tracingfonts:D \pdftracingfonts}}
\texttt{tex_else:D}
\texttt{tex_ifdefined:D \texttt{pdfdirectlua:D}}
\texttt{\pdfdirectlua:D \{ tex.enableprimitives("@@", \{"tracingfonts"\}) \}}
\texttt{\tex_let:D \texttt{tex_tracingfonts:D \@@tracingfonts}}
\texttt{\tex_fi:D}
\texttt{\tex_fi:D}
\end{verbatim}
```
Only pdfTeX and LuaTeX define \texttt{\pdfmapfile} and \texttt{\pdfmapline}: tidy up the fact that some format-building processes leave a couple of questionable decisions about that!

A few packages do unfortunate things to date-related primitives.

cslatex moves a couple of primitives which we recover here; as there is no other marker, we can only work by looking for the names.

For Con\TeXt, two tests are needed. Both Mark II and Mark IV move several primitives: these are all covered by the first test, again using \verb|\end| as a marker. For Mark IV, a few more primitives are moved: they are implemented using some Lua code in the current Con\TeXt.
\text\strcmp \text\Ucharcat

In \LaTeXe, we additionally emulate some primitives using Lua code.

\text\strcmp Compare two strings, expanding to 0 if they are equal, −1 if the first one is smaller and 1 if the second one is smaller. Here “smaller” refers to codepoint order which does not correspond to the user expected order for most non-ASCII strings.

\verbatim
local minus_tok = token_new(string.byte'-', 12)
local zero_tok = token_new(string.byte'0', 12)
local one_tok = token_new(string.byte'1', 12)
luacmd('tex\strcmp', function()
local first = scan_string()
local second = scan_string()
if first < second then
  put_next(minus_tok, one_tok)
else
  put_next(first == second and zero_tok or one_tok)
end
end, 'global')

(End of definition for \text\strcmp.)

\text\Ucharcat Creating arbitrary chars using \text\cprint. The alternative approach using \text\token.new(...) is about 10% slower but needed to create arbitrary space tokens.

\verbatim
local sprint = tex.sprint
local cprint = tex.cprint
luacmd('tex\Ucharcat', function()
local charcode = scan_int()
local catcode = scan_int()
if catcode == 10 then
  sprint(token_new(charcode, 10))
else
  put_next(charcode, one_tok)
end
end

(End of definition for \text\Ucharcat.)
else
    cprint(catcode, utf8_char(charcode))
end
end, 'global')

(End of definition for \textcat:D.)

\filesize:D Wrap the function from \ltxutils.
luacmd('\filesize:D', function()
    local size = filesize(scan_string())
    if size then write(size) end
end, 'global')

(End of definition for \filesize:D.)

\mdfivesum:D There are two cases: Either hash a file or a string. Both are already implemented in \l3luatex or built-in.
luacmd('\mdfivesum:D', function()
    local hash
    if scan_keyword"file" then
        hash = filemd5sum(scan_string())
    else
        hash = md5_HEX(scan_string())
    end
    if hash then write(hash) end
end, 'global')

(End of definition for \mdfivesum:D.)

\filemoddate:D A primitive for getting the modification date of a file.
luacmd('\filemoddate:D', function()
    local date = filemoddate(scan_string())
    if date then write(date) end
end, 'global')

(End of definition for \filemoddate:D.)

\filedump:D An emulated primitive for getting a hexdump from a (partial) file. The length has a default of 0. This is consistent with pdfTEX, but it effectively makes the primitive useless without an explicit length. Therefore we allow the keyword whole to be used instead of a length, indicating that the whole remaining file should be read.
luacmd('\filedump:D', function()
    local offset = scan_keyword'offset' and scan_int() or nil
    local length = scan_keyword'length' and scan_int()
    or not scan_keyword'whole' and 0 or nil
    local data = filedump(scan_string(), offset, length)
    if data then write(data) end
end, 'global')

(End of definition for \filedump:D.)

{/lua}
{/package}
Chapter 41

l3kernel-functions:
kernel-reserved functions

41.1 Internal l3debug kernel functions

These functions are only created if debugging is enabled, hence they are actually defined in l3debug.

\__kernel_chk_var_local:N \__kernel_chk_var_local:N \__kernel_chk_var_global:N
\__kernel_chk_var_global:N

Applies \__kernel_chk_var_exist:N \__kernel_chk_var_exist:N as well as \__kernel_chk_var_scope:NN \__kernel_chk_var_scope:NN, where \scope is l or g.

\__kernel_chk_var_scope:NN \__kernel_chk_var_scope:NN \__kernel_chk_var_scope:NN
\__kernel_chk_var_scope:NN

Checks the \var has the correct \scope, and if not raises a kernel-level error. The \scope is a single letter l, g, c denoting local variables, global variables, or constants. More precisely, if the variable name starts with a letter and an underscore (normal expl3 convention) the function checks that this single letter matches the \scope. Otherwise the function cannot know the scope \var the first time: instead, it defines \__debug_chk_/\var name to store that information for the next call. Thus, if a given \var is subject to assignments of different scopes a kernel error will result.

\__kernel_chk_cs_exist:N \__kernel_chk_cs_exist:N \__kernel_chk_cs_exist:N
\__kernel_chk_cs_exist:N
\__kernel_chk_var_exist:N \__kernel_chk_var_exist:N
\__kernel_chk_var_exist:N

Checks that their argument is defined according to the criteria for \cs_if_exist_p:N, and if not raises a kernel-level error. Error messages are different.

\__kernel_chk_flag_exist:NN \__kernel_chk_flag_exist:NN \__kernel_chk_flag_exist:NN
\__kernel_chk_flag_exist:NN

Checks that the \flag is defined according to the criterion for \flag_if_exist_p:N, and if not raises a kernel-level error and calls the function with the argument \l_tmpa_flag to proceed somehow without producing too many errors.
\__kernel_debug_log:e \__kernel_debug_log:e \{message\ \text{text}\}

If the log-functions option is active, this function writes the \{message\ \text{text}\} to the log file using \text{iow_log:e}. Otherwise, the \{message\ \text{text}\} is ignored using \text{use_none:n}.

41.2 Internal kernel functions

\__kernel_chk_defined:NT \__kernel_chk_defined:NT \{variable\} \{true\ \text{code}\}

If \{variable\} is not defined (according to \text{cs_if_exist:NTF}), this triggers an error, otherwise the \{true\ \text{code}\} is run.

\__kernel_chk_expr:nNnN \__kernel_chk_expr:nNnN \{expr\} \{eval\} \{convert\} \{caller\}

This function is only created if debugging is enabled. By default it is equivalent to \text{use_i:nnnn}. When expression checking is enabled, it leaves in the input stream the result of \text{tex_the:D} \{eval\} \{expr\} \text{tex_relax:D} after checking that no token was left over. If any token was not taken as part of the expression, there is an error message displaying the result of the evaluation as well as the \{caller\}. For instance \{eval\} can be \text{\_int_eval:w} and \{caller\} can be \text{\int_eval:n} or \text{\int_set:Nn}. The argument \{convert\} is empty except for \text{mu} expressions where it is \text{\tex_mutoglue:D}, used for internal purposes.

\__kernel_chk_tl_type:NnnT \__kernel_chk_tl_type:NnnT \{control\ \text{sequence}\} \{specific\ \text{type}\} \{reconstruction\} \{true\ \text{code}\}

Helper to test that the \{control\ \text{sequence}\} is a variable of the given \{specific\ \text{type}\} of token list. Produces suitable error messages if the \{control\ \text{sequence}\} does not exist, or if it is not a token list variable at all, or if the \{control\ \text{sequence}\} differs from the result of e-expanding \{reconstruction\}. If all of these tests succeed then the \{true\ \text{code}\} is run.

\__kernel_codepoint_to_bytes:n \__kernel_codepoint_to_bytes:n \{codepoint\}

Converts the \{codepoint\} to UTF-8 bytes. The expansion of this function comprises four brace groups, each of which will contain a hexadecimal value: the appropriate byte. As UTF-8 is a variable-length, one or more of the groups may be empty: the bytes read in the logical order, such that a two-byte codepoint will have groups #1 and #2 filled and #3 and #4 empty.

\__kernel_cs_parm_from_arg_count:nnF \__kernel_cs_parm_from_arg_count:nnF \{follow-on\} \{args\} \{false\ \text{code}\}

Evaluates the number of \{args\} and leaves the \{follow-on\} code followed by a brace group containing the required number of primitive parameter markers (#1, etc.). If the number of \{args\} is outside the range [0,9], the \{false\ \text{code}\} is inserted instead of the \{follow-on\}.
\__kernel_dependency_version_check:Nn \__kernel_dependency_version_check:Nn \langle\text{date}\rangle \{\langle\text{file}\rangle\}
\__kernel_dependency_version_check:nn \__kernel_dependency_version_check:nn \langle\text{date}\rangle \{\langle\text{file}\rangle\}

Checks if the loaded version of the expl3 kernel is at least \langle\text{date}\rangle, required by \langle\text{file}\rangle. If the kernel date is older than \langle\text{date}\rangle, the loading of \langle\text{file}\rangle is aborted and an error is raised.

\__kernel_deprecation_code:nn \__kernel_deprecation_code:nn \langle\text{error code}\rangle \{\langle\text{working code}\rangle\}

Stores both an \langle\text{error}\rangle and \langle\text{working}\rangle definition for given material such that they can be exchanged by \debug_on:n and \debug_off:n.

\__kernel_exp_not:w \__kernel_exp_not:w \langle\text{expandable tokens}\rangle \{\langle\text{content}\rangle\}

Carries out expansion on the \langle\text{expandable tokens}\rangle before preventing further expansion of the \langle\text{content}\rangle as for \exp_not:n. Typically, the \langle\text{expandable tokens}\rangle will alter the nature of the \langle\text{content}\rangle, i.e. allow it to be generated in some way.

\l__kernel_expl_bool
A boolean which records the current code syntax status: \text{true} if currently inside a code environment. This variable should only be set by \ExplSyntaxOn/\ExplSyntaxOff.

(End of definition for \l__kernel_expl_bool.)

\c__kernel_expl_date_tl
A token list containing the release date of the l3kernel preloaded in \LaTeX used to check if dependencies match.

(End of definition for \c__kernel_expl_date_tl.)

\__kernel_file_missing:n \__kernel_file_missing:n \langle\text{name}\rangle

Expands the \langle\text{name}\rangle as per \__kernel_file_name_sanitize:n then produces an error message indicating that this file was not found.

\__kernel_file_name_sanitize:n \__kernel_file_name_sanitize:n \langle\text{name}\rangle

Updated: 2021-04-17

Expands the file name using a \csname-based approach, and relies on active characters (for example from UTF-8 characters) being properly set up to expand to a expansion-safe version using \ifcsname. This is less conservative than the token-by-token approach used before, but it is much faster.

\__kernel_file_input_push:n \__kernel_file_input_push:n \langle\text{name}\rangle
\__kernel_file_input_pop: \__kernel_file_input_pop:

Used to push and pop data from the internal file stack: needed only in package mode, where interfacing with the \LaTeX kernel is necessary.

\__kernel_int_add:nnn \__kernel_int_add:nnn \langle\text{integer}_1\rangle \{\langle\text{integer}_2\rangle\} \{\langle\text{integer}_3\rangle\}

Expands to the result of adding the three \langle\text{integers}\rangle (which must be suitable input for \int_eval:w), avoiding intermediate overflow. Overflow occurs only if the overall result is outside \([-2^{31} + 1, 2^{31} - 1]\). The \langle\text{integers}\rangle may be of the form \int_eval:w ... \scan_stop: but may be evaluated more than once.
\_\_\_kernel_intarray_gset:Nnn \_\_\_kernel_intarray_gset:Nnn \{intarray\} \{\{index\}\} \{\{value\}\}

New: 2018-03-31

Faster version of \intarray_gset:Nnn. Stores the \{value\} into the \{integer array\} variable at the \{position\}. The \{index\} and \{value\} must be suitable for a direct assignment to a \TeX{} count register, for instance expanding to an integer denotation or obtained through the primitive \numexpr{} (which may be un-terminated). No bound checking is performed: the caller is responsible for ensuring that the \{position\} is between 1 and the \\intarray_count:N, and the \{value\}’s absolute value is at most \(2^{30} - 1\). Assignments are always global.

\_\_\_kernel_intarray_item:Nn \_\_\_kernel_intarray_item:Nn \{intarray\} \{\{index\}\}

New: 2018-03-31

Faster version of \intarray_item:Nn. Expands to the integer entry stored at the \{index\} in the \{integer array variable\}. The \{index\} must be suitable for a direct assignment to a \TeX{} count register and must be between 1 and the \\intarray_count:N, lest a low-level \TeX{} error occur.

\_\_\_kernel_intarray_range_to_clist:Nnn \_\_\_kernel_intarray_range_to_clist:Nnn \{intarray\} \{\{start index\}\} \{\{end index\}\}

New: 2020-07-12

Converts to integer denotations separated by commas the entries of the \{intarray\} from positions \{start index\} to \{end index\} included. The \{start index\} and \{end index\} must be suitable for a direct assignment to a \TeX{} count register, and be suitably ordered. All tokens have category code other.

\_\_\_kernel_intarray_gset_range_from_clist:Nnn \_\_\_kernel_intarray_gset_range_from_clist:Nnn \{intarray\} \{\{start index\}\} \{\{integer\ clist\}\}

New: 2020-07-12

Stores the entries of the \{clist\} as entries of the \{intarray\} starting from the \{start index\}, upwards. This is done without any bound checking. The \{start index\} and all entries of the \{integer comma list\} (which do not undergo space trimming and brace stripping as in normal clist mappings) must be suitable for a direct assignment to a \TeX{} count register. An empty entry may stop the loop.

\_\_\_kernel_ior_open:Nn \_\_\_kernel_ior_open:Nn \{stream\} \{\{file name\}\}

\_\_\_kernel_iow_open:Nn \_\_\_kernel_iow_open:Nn \{stream\} \{\{file name\}\}

This function has identical syntax to the public version. However, does not take precautions against active characters in the \{file name\}, and it does not attempt to add a \{path\} to the \{file name\}: it is therefore intended to be used by higher-level functions which have already fully expanded the \{file name\} and which need to perform multiple open or close operations. See for example the implementation of \ior_shell_open:Nn.

This function has identical syntax to the public version. However, does not take precautions against active characters in the \{file name\}, and it does not attempt to add a \{path\} to the \{file name\}: it is therefore intended to be used by higher-level functions which have already fully expanded the \{file name\} and which need to perform multiple open or close operations. See for example the implementation of \iow_shell_open:Nn.
\_\_kernel\_iow\_with:Nnn \_\_kernel\_iow\_with:Nnn \langle integer \rangle \{ \langle value \rangle \} \{ \langle code \rangle \}

If the \langle integer \rangle is equal to the \langle value \rangle then this function simply runs the \langle code \rangle. Otherwise it saves the current value of the \langle integer \rangle, sets it to the \langle value \rangle, runs the \langle code \rangle, and restores the \langle integer \rangle to its former value. This is used to ensure that the \texttt{\newline} character is 10 when writing to a stream, which lets \texttt{\iow\_newline} work, and that \texttt{\errorcontextlines} is $-1$ when displaying a message.

\_\_kernel\_kern:n \_\_kernel\_kern:n \{ \langle length \rangle \}

Inserts a kern of the specified \langle length \rangle, a dimension expression.

\_\_kernel\_msg\_show\_eval:Nn \_\_kernel\_msg\_show\_eval:Nn \langle function \rangle \{ \langle expression \rangle \}

Shows or logs the \langle expression \rangle (turned into a string), an equal sign, and the result of applying the \langle function \rangle to the \{ \langle expression \rangle \} (with f-expansion). For instance, if the \langle function \rangle is \texttt{\int\_eval:n} and the \langle expression \rangle is $1+2$ then this logs $> 1+2=3$.

\_\_kernel\_pdf\_object\_id:n \_\_kernel\_pdf\_object\_id:n \* \_\_kernel\_pdf\_object\_id:n \{ \langle object \rangle \}

\_\_kernel\_pdf\_object\_id\_indexed:nn \_\_kernel\_pdf\_object\_id\_indexed:nn \* \_\_kernel\_pdf\_object\_id\_indexed:nn \{ \langle object \rangle \} \{ \langle class \rangle \} \{ \langle number \rangle \}

Expands to the ID of \langle object \rangle (or object of \langle number \rangle within the \langle class \rangle), in for example page resource allocation. Depending on the backend, the result may be the same as \texttt{\pdf\_object\_id:n/\pdf\_object\_id\_indexed:nn}.

\g\_\_kernel\_prg\_map\_int

This integer is used by non-expandable mapping functions to track the level of nesting in force. The functions \texttt{\(\langle type\rangle\_map\_1:w\)}, \texttt{\(\langle type\rangle\_map\_2:w\)}, etc., labelled by \texttt{\g\_\_kernel\_prg\_map\_int} hold functions to be mapped over various list datatypes in inline and variable mappings.

(End of definition for \g\_\_kernel\_prg\_map\_int.)
Defines a quark-test function \( \langle \text{arg spec} \rangle \) which tests if its argument is \q__\langle namespace\rangle_recursion_tail, then acts accordingly, as described below for each possible \( \langle \text{arg spec} \rangle \).

The \langle namespace\rangle is determined as the first (nonempty) \_delimited word in \langle name\rangle and is used internally in the definition of auxiliaries. The function \__kernel_quark_new_test:N does not define the \q__\langle namespace\rangle_recursion_tail and \q__\langle namespace\rangle_recursion_stop quarks. They should be manually defined with \quark_new:N.

There are 6 different types of quark-test functions. Which one is defined depends on the \langle arg spec\rangle, which must be one of the options listed now. Four of them are modeled after \quark_if_recursion_tail:(N|n) and \quark_if_recursion_tail_do:(N|n)n.

n defines \langle name\rangle:n such that it checks if \#1 contains only \q__\langle namespace\rangle_recursion_tail, and if so consumes all tokens up to \q__\langle namespace\rangle_recursion_stop (c.f. \quark_if_recursion_tail_stop:n).

nn defines \langle name\rangle:nn such that it checks if \#1 contains only \q__\langle namespace\rangle_recursion_tail, and if so consumes all tokens up to \q__\langle namespace\rangle_recursion_stop, then executes the code \#2 after that (c.f. \quark_if_recursion_tail_stop_do:nn).

N defines \langle name\rangle:N such that it checks if \#1 is \q__\langle namespace\rangle_recursion_tail, and if so consumes all tokens up to \q__\langle namespace\rangle_recursion_stop (c.f. \quark_if_recursion_tail_stop:N).

Nn defines \langle name\rangle:Nn such that it checks if \#1 is \q__\langle namespace\rangle_recursion_tail, and if so consumes all tokens up to \q__\langle namespace\rangle_recursion_stop, then executes the code \#2 after that (c.f. \quark_if_recursion_tail_stop_do:Nn).

The last two are modeled after \quark_if_recursion_tail_break:(n|N)N, and in those cases the quark \q__\langle namespace\rangle_recursion_stop is not used (and thus needs not be defined).

nN defines \langle name\rangle:nN such that it checks if \#1 contains only \q__\langle namespace\rangle_recursion_tail, and if so uses the \langle type\rangle_map_break: function \#2.

NN defines \langle name\rangle:NN such that it checks if \#1 is \q__\langle namespace\rangle_recursion_tail, and if so uses the \langle type\rangle_map_break: function \#2.

Any other signature, as well as a function without signature are errors, and in such case the definition is aborted.
\__kernel_quark_new_conditional:Nn  \__kernel_quark_new_conditional:Nn
\__\langle namespace\rangle_quark_if_{\langle name\rangle}:{\langle arg spec\rangle} \{\langle conditions\rangle\}

Defines a collection of quark conditionals that test if their argument is the quark \_\langle namespace\rangle_\langle name\rangle and perform suitable actions. The \langle conditions\rangle are a comma-separated list of one or more of p, T, F, and TF, and one conditional is defined for each \langle condition\rangle in the list, as described for \prg_new_conditional:Npnn. The conditionals are defined using \prg_new_conditional:Npnn, so that their name is obtained by adding p, T, F, or TF to the base name \_\langle namespace\rangle_quark_if_{\langle name\rangle}:{\langle arg spec\rangle}.

The first argument of \__kernel_quark_new_conditional:Nn must contain \_quark_if_ and ;, as these markers are used to determine the \langle name\rangle of the quark \_\langle namespace\rangle_\langle name\rangle to be tested. This quark should be manually defined with \quark_new:N, as \__kernel_quark_new_conditional:Nn does not define it.

The function \__kernel_quark_new_conditional:Nn can define 2 different types of quark conditionals. Which one is defined depends on the \langle arg spec\rangle, which must be one of the following options, modeled after \quark_if:nil:(N|n)(TF).

n defines \_\langle namespace\rangle_quark_if_{\langle name\rangle}:n(TF) such that it checks if #1 contains only \_\langle namespace\rangle_\langle name\rangle, and executes the proper conditional branch.

N defines \_\langle namespace\rangle_quark_if_{\langle name\rangle}:N(TF) such that it checks if #1 is \_\langle namespace\rangle_\langle name\rangle, and executes the proper conditional branch.

Any other signature, as well as a function without signature are errors, and in such case the definition is aborted.

\__kernel_sys_everyjob:  \__kernel_sys_everyjob:

Inserts the internal token list required at the start of every run (job).

\c__kernel_randint_max_int

Maximal allowed argument to \__kernel_randint:n. Equal to 2^{17} - 1.

(End of definition for \c__kernel_randint:n.)

\__kernel_randint:n  \__kernel_randint:n \{\langle max\rangle\}

Used in an integer expression this gives a pseudo-random number between 1 and \langle max\rangle included. One must have \langle max\rangle \leq 2^{17} - 1. The \langle max\rangle must be suitable for \int_value:w (and any \int_eval:w must be terminated by \scan_stop: or equivalent).

\__kernel_randint:nn  \__kernel_randint:nn \{\langle min\rangle\} \{\langle max\rangle\}

Used in an integer expression this gives a pseudo-random number between \langle min\rangle and \langle max\rangle included. The \langle min\rangle and \langle max\rangle must be suitable for \int_value:w (and any \int_eval:w must be terminated by \scan_stop: or equivalent). For small ranges \(R = \langle max\rangle - \langle min\rangle + 1 \leq 2^{17} - 1\), \langle min\rangle - 1 + \__kernel_randint:n{\langle R\rangle} is faster.

\__kernel_register_show:N  \__kernel_register_show:N \{register\}

Used to show the contents of a \TeX{} register at the terminal, formatted such that internal parts of the mechanism are not visible.

\__kernel_register_log:N  \__kernel_register_log:N \{register\}

Used to write the contents of a \TeX{} register to the log file in a form similar to \__kernel_register_show:N.

370
\_\_kernel\_\_str\_to\_other:n \* \_\_kernel\_\_str\_to\_other:n \{(token list)\}

Converts the (token list) to a (other string), where spaces have category code “other”. This function can be f-expanded without fear of losing a leading space, since spaces do not have category code 10 in its result. It takes a time quadratic in the character count of the string.

\_\_kernel\_\_str\_to\_other\_fast:n \* \_\_kernel\_\_str\_to\_other\_fast:n \{(token list)\}

Same behaviour \_\_kernel\_\_str\_to\_other:n but only restricted-expandable. It takes a time linear in the character count of the string.

\_\_kernel\_tl\_to\_str:w \* \_\_kernel\_tl\_to\_str:w \{expandable tokens\} \{\{tokens\}\}

Carries out expansion on the (expandable tokens) before conversion of the (tokens) to a string as describe for \tl\_to\_str:n. Typically, the (expandable tokens) will alter the nature of the (tokens), i.e. allow it to be generated in some way. This function requires only a single expansion.

\_\_kernel\_tl\_set:Nx \_\_kernel\_tl\_set:Nx \tl\ var \{\{tokens\}\}

\_\_kernel\_tl\_gset:Nx

Fully expands \{tokens\} and assigns the result to \tl\ var. \{tokens\} must be given in braces and there must be no token between \tl\ var and \{tokens\}.

\_\_kernel\_codepoint\_data:nn \* \_\_kernel\_codepoint\_data:nn \{\{type\}\} \{\{codepoint\}\}

Expands to the appropriate value for the \{type\} of data requested for a \{codepoint\}. The current list of \{types\} and results are

- lowercase The single codepoint specified by UnicodeData.txt for lowercase mapping of the codepoint: will be equal to the input \{codepoint\} if there is no mapping specified in UnicodeData.txt
- uppercase The single codepoint specified by UnicodeData.txt for uppercase mapping of the codepoint: will be equal to the input \{codepoint\} if there is no mapping specified in UnicodeData.txt

\_\_kernel\_codepoint\_case:nn \* \_\_kernel\_codepoint\_case:nn \{\{mapping\}\} \{\{codepoint\}\}

Expands to a list of three balanced text, of which at least the first will contain a codepoint. This list of up to three codepoints specifies the full case mapping for the input \{codepoint\}. The \{mapping\} should be one of

- casefold
- lowercase
- titlecase
- uppercase
41.3 Kernel backend functions

These functions are required to pass information to the backend. The nature of these means that they are defined only when the relevant backend is in use.

\__kernel_backend_literal:n \{\langle content\rangle\}
\__kernel_backend_literal:(e|e)

Adds the \langle content\rangle literally to the current vertical list as a whatsit. The nature of the \langle content\rangle will depend on the backend in use.

\__kernel_backend_literal_postscript:n \{\langle PostScript\rangle\}
\__kernel_backend_literal_postscript:e

Adds the \langle PostScript\rangle literally to the current vertical list as a whatsit. No positioning is applied.

\__kernel_backend_literal_pdf:n \{\langle PDF instructions\rangle\}
\__kernel_backend_literal_pdf:e

Adds the \langle PDF instructions\rangle literally to the current vertical list as a whatsit. No positioning is applied.

\__kernel_backend_literal_svg:n \{\langle SVG instructions\rangle\}
\__kernel_backend_literal_svg:e

Adds the \langle SVG instructions\rangle literally to the current vertical list as a whatsit. No positioning is applied.

\__kernel_backend_postscript:n \{\langle PostScript\rangle\}
\__kernel_backend_postscript:e

Adds the \langle PostScript\rangle to the current vertical list as a whatsit. The PostScript reference point is adjusted to match the current position. The PostScript is inserted inside a SDict begin/end pair.

\__kernel_backend_align_begin: \{\langle PostScript literals\rangle\}
\__kernel_backend_align_end:

Arranges to align the PostScript and DVI current positions and scales.

\__kernel_backend_scope_begin: \{\langle content\rangle\}
\__kernel_backend_scope_end:

Creates a scope for instructions at the backend level.

\__kernel_backend_matrix:n \{\langle matrix\rangle\}
\__kernel_backend_matrix:e

Applies the \langle matrix\rangle to the current transformation matrix.

\g__kernel_backend_header_bool

Specifies whether to write headers for the backend.
\l__kernel\_color\_stack\_int  The color stack used in pdf\TeX{} and Lua\TeX{} for the main color.
Chapter 42

\texttt{\textbf{l3basics} implementation}

42.1 Renaming some \TeX\ primitives (again)

Having given all the \TeX\ primitives a consistent name, we need to give sensible names to the ones we actually want to use. These will be defined as needed in the appropriate modules, but we do a few now, just to get started.\footnote{This renaming gets expensive in terms of csname usage, an alternative scheme would be to just use the \texttt{\textbackslash tex,...:D} name in the cases where no good alternative exists.}

\begin{verbatim}
\if_true:\text_global:D \text_let:D \if_true:\text_iftrue:D
\if_false:\text_global:D \text_let:D \if_false:\text_iffalse:D
\or:\text_global:D \text_let:D \or:\text_or:D
\else:\text_global:D \text_let:D \else:\text_else:D
\fi:\text_global:D \text_let:D \fi:\text_fi:D
\reverse_if:N \text_global:D \text_let:D \reverse_if:N \text_unless:D
\if:w \text_global:D \text_let:D \if:w \text_if:D
\if_charcode:w \text_global:D \text_let:D \if_charcode:w \text_if:D
\if_catcode:w \text_global:D \text_let:D \if_catcode:w \text_ifcat:D
\if_meaning:w \text_global:D \text_let:D \if_meaning:w \text_ifx:D
\if_bool:N \text_global:D \text_let:D \if_bool:N \text_ifodd:D
\end{verbatim}

Then some conditionals.

\begin{verbatim}
\if_mode_math: \text_global:D \text_let:D \if_mode_math: \text_ifmmode:D
\if_mode_horizontal: \text_global:D \text_let:D \if_mode_horizontal: \text_ifhmode:D
\if_mode_vertical: \text_global:D \text_let:D \if_mode_vertical: \text_ifvmode:D
\if_mode_inner: \text_global:D \text_let:D \if_mode_inner: \text_ifinner:D
\end{verbatim}

(End of definition for \texttt{\textbackslash if_true:} and others. These functions are documented on page \pageref{page}.)

\begin{verbatim}
\if_cs_exist:N \text_global:D \text_let:D \if_cs_exist:N \text_ifdefined:D
\if_cs_exist:w \text_global:D \text_let:D \if_cs_exist:w \text_ifcsname:D
\cs:w \text_global:D \text_let:D \cs:w \text_csname:D
\cs_end: \text_global:D \text_let:D \cs_end: \text_endcsname:D
\end{verbatim}

\begin{verbatim}
\if_mode_math: \text_global:D \text_let:D \if_mode_math: \text_ifmmode:D
\if_mode_horizontal: \text_global:D \text_let:D \if_mode_horizontal: \text_ifhmode:D
\if_mode_vertical: \text_global:D \text_let:D \if_mode_vertical: \text_ifvmode:D
\if_mode_inner: \text_global:D \text_let:D \if_mode_inner: \text_ifinner:D
\end{verbatim}

(End of definition for \texttt{\textbackslash if_mode_math:} and others. These functions are documented on page \pageref{page}.)

Building csnames and testing if control sequences exist.

\begin{verbatim}
\if_cs_exist:N \text_global:D \text_let:D \if_cs_exist:N \text_ifdefined:D
\if_cs_exist:w \text_global:D \text_let:D \if_cs_exist:w \text_ifcsname:D
\cs:w \text_global:D \text_let:D \cs:w \text_csname:D
\cs_end: \text_global:D \text_let:D \cs_end: \text_endcsname:D
\end{verbatim}

374
The five \exp functions are used in the \l3exp module where they are described.

\exp_after:wN \exp_not:N \exp_not:n

Examing a control sequence or token.

\token_to_meaning:N \cs_meaning:N

Making strings.

\tl_to_str:n \token_to_str:N \__kernel_tl_to_str:w

The next three are basic functions for which there also exist versions that are safe inside alignments. These safe versions are defined in the \l3prg module.

\scan_stop: \group_begin: \group_end:

For integers.

\if_int_compare:w \__int_to_roman:w

Adding material after the end of a group.

\group_insert_after:N

Discussed in \l3exp, but needed much earlier.

\exp_args:Nc \exp_args:cc
A small number of variants defined by hand. Some of the necessary functions (\use_i:nn, \use_ii:nn, and \exp_args:NNc) are not defined at that point yet, but will be defined before those variants are used. The \cs_meaning:c command must check for an undefined control sequence to avoid defining it mistakenly.

\begin{verbatim}
\token_to_str:c \token_to_meaning:c \cs_meaning:c

\token_to_str:c \token_to_meaning:c \cs_meaning:c

\cs_gset_protected_nopar:Npn \cs_gset_protected_nopar:Npe
\cs_gset_protected_nopar:Npx \cs_gset_protected:Npn \cs_gset_protected:Npe
\cs_gset:Npn \cs_gset:Npe \cs_gset:Npx

\token_to_str:c \token_to_meaning:c \cs_meaning:c

\use_i:nn \use_ii:nn \exp_args:NNc

\exp_after:wN \use_i:nn \else:
\exp_after:wN \use_ii:nn \fi:
\{ \exp_args:Nc \cs_meaning:N {#1} \}
\{ \tl_to_str:n {undefined} \}

\cs_gset_nopar:Npn \cs_gset_nopar:Npe \cs_gset_nopar:Npx
\cs_gset:Npn \cs_gset:Npe \cs_gset:Npx

\end{verbatim}

(End of definition for \token_to_meaning:N. This function is documented on page 201.)

\section{Defining some constants}

We need the constant \c_zero_int which is used by some functions in current module. The rest are defined in the l3int module – at least for the ones that can be defined with \tex_chardef:D or \tex_mathchardef:D. For other constants the l3int module is required but it can’t be used until the allocation has been set up properly!

\begin{verbatim}
\c_zero_int
\c_max_register_int

\c_zero_int
\c_max_register_int

\exp_after:wN \use_i:nn \else:
\exp_after:wN \use_ii:nn \fi:
\{ \exp_args:Nc \cs_meaning:N {#1} \}
\{ \tl_to_str:n {undefined} \}

\cs_gset_nopar:Npn \cs_gset_nopar:Npe \cs_gset_nopar:Npx
\cs_gset:Npn \cs_gset:Npe \cs_gset:Npx

\end{verbatim}

(End of definition for \c_zero_int. This variable is documented on page 178.)

\section{Defining functions}

We start by providing functions for the typical definition functions. First the global ones.

All assignment functions in \TeX{} should be naturally protected; after all, the \TeX{} primitives for assignments are and it can be a cause of problems if others aren’t.

\begin{verbatim}
\cs_gset_nopar:Npn \cs_gset_nopar:Npe \cs_gset_nopar:Npx
\cs_gset:Npn \cs_gset:Npe \cs_gset:Npx

\token_to_str:c \token_to_meaning:c \cs_meaning:c

\use_i:nn \use_ii:nn \exp_args:NNc

\exp_after:wN \use_i:nn \else:
\exp_after:wN \use_ii:nn \fi:
\{ \exp_args:Nc \cs_meaning:N {#1} \}
\{ \tl_to_str:n {undefined} \}

\cs_gset_nopar:Npn \cs_gset_nopar:Npe \cs_gset_nopar:Npx
\cs_gset:Npn \cs_gset:Npe \cs_gset:Npx

\end{verbatim}

(End of definition for \c_max_register_int. This variable is documented on page 178.)
42.4 Selecting tokens

\l__exp_internal_tl Scratch token list variable for `\textxpan`, used by `\use:x`, used in defining conditionals. We don’t use t1 methods because `\textbasics` is loaded earlier.
\l__exp_internal_tl \cs_gset_nopar:Npn \l__exp_internal_tl { } (End of definition for `\l__exp_internal_tl`.)

\use:c This macro grabs its argument and returns a csname from it.
\cs_gset:Npn \use:c #1 { \cs:w #1 \cs_end: } (End of definition for `\use:c`. This function is documented on page 21.)
\use:x  Fully expands its argument and passes it to the input stream. Uses the reserved \_\_\_\_exp\_internal\_t1 which we’ve set up above.
\begin{verbatim}
\cs_gset_protected:Npn \use:x #1
{ \cs_set_nopar:Npx \l__exp_internal_tl {#1} \l__exp_internal_tl }
\end{verbatim}

(End of definition for \use:x.)
\begin{verbatim}
\use:e
\cs_gset:Npn \use:e #1 { \tex_expanded:D {#1} }
\end{verbatim}

(End of definition for \use:e. This function is documented on page 26.)
\begin{verbatim}
\use:n \use:nn \use:nnn \use:nnnn \use:_i:nn \use:_ii:nn \use:_iii:nn \use:_iv:nn \use:_i:nnn \use:_ii:nnn \use:_iii:nnn \use:_iv:nnn \use:_v:nnn \use:_i:nnnn \use:_ii:nnnn \use:_iii:nnnn \use:_iv:nnnn \use:_v:nnnn \use:_vi:nnnn \use:_i:nnnnn \use:_ii:nnnnn \use:_iii:nnnnn \use:_iv:nnnnn \use:_v:nnnnn \use:_vi:nnnnn \use:_i:nnnnn \use:_ii:nnnnn \use:_iii:nnnnn \use:_iv:nnnnn \use:_v:nnnnn \use:_vi:nnnnn \use:_i:nnnnnn \use:_ii:nnnnnn \use:_iii:nnnnnn \use:_iv:nnnnnn \use:_v:nnnnnn \use:_vi:nnnnnn \use:_i:nnnnnnn \use:_ii:nnnnnnn \use:_iii:nnnnnnn \use:_iv:nnnnnnn \use:_v:nnnnnnn \use:_vi:nnnnnnn \use:_i:nnnnnnnn \use:_ii:nnnnnnnn \use:_iii:nnnnnnnn \use:_iv:nnnnnnnn \use:_v:nnnnnnnn \use:_vi:nnnnnnnn
\end{verbatim}

We also need something for picking up arguments from a longer list.
\begin{verbatim}
\use_i:nnn #1#2#3 {#1}
\cs_gset:Npn \use_i:nnn #1#2#3 {#1}
\cs_gset:Npn \use_ii:nnn #1#2#3 {#2}
\cs_gset:Npn \use_iii:nnn #1#2#3 {#3}
\cs_gset:Npn \use_i:nnnn #1#2#3#4 {#1}
\cs_gset:Npn \use_ii:nnnn #1#2#3#4 {#2}
\cs_gset:Npn \use_iii:nnnn #1#2#3#4 {#3}
\cs_gset:Npn \use_iv:nnnn #1#2#3#4 {#4}
\cs_gset:Npn \use_i:nnnnn #1#2#3#4#5 {#1}
\cs_gset:Npn \use_ii:nnnnn #1#2#3#4#5 {#2}
\cs_gset:Npn \use_iii:nnnnn #1#2#3#4#5 {#3}
\cs_gset:Npn \use_iv:nnnnn #1#2#3#4#5 {#4}
\cs_gset:Npn \use_v:nnnnn #1#2#3#4#5 {#5}
\cs_gset:Npn \use_i:nnnnnn #1#2#3#4#5#6 {#1}
\cs_gset:Npn \use_ii:nnnnnn #1#2#3#4#5#6 {#2}
\cs_gset:Npn \use_iii:nnnnnn #1#2#3#4#5#6 {#3}
\cs_gset:Npn \use_iv:nnnnnn #1#2#3#4#5#6 {#4}
\cs_gset:Npn \use_v:nnnnnn #1#2#3#4#5#6 {#5}
\cs_gset:Npn \use_i:nnnnnnn #1#2#3#4#5#6#7 {#1}
\cs_gset:Npn \use_ii:nnnnnnn #1#2#3#4#5#6#7 {#2}
\cs_gset:Npn \use_iii:nnnnnnn #1#2#3#4#5#6#7 {#3}
\cs_gset:Npn \use_iv:nnnnnnn #1#2#3#4#5#6#7 {#4}
\cs_gset:Npn \use_v:nnnnnnn #1#2#3#4#5#6#7 {#5}
\cs_gset:Npn \use_i:nnnnnnnn #1#2#3#4#5#6#7 {#1}
\cs_gset:Npn \use_ii:nnnnnnnn #1#2#3#4#5#6#7 {#2}
\cs_gset:Npn \use_iii:nnnnnnnn #1#2#3#4#5#6#7 {#3}
\end{verbatim}

378
\use_i:nnn

\cs_gset:Npn \use_i:nnn #1#2#3 {#1#2}

(End of definition for \use_i:nnn. This function is documented on page 26.)

\use_ii_i:nn

\cs_gset:Npn \use_ii_i:nn #1#2 { #2 #1 }

(End of definition for \use_ii_i:nn. This function is documented on page 26.)

Functions that gobble everything until they see either \texttt{\q_nil}, \texttt{\q_stop}, or \texttt{\q_recursion_stop}, respectively.

\cs_gset:Npn \use_none_delimit_by_q_nil:w #1 \q_nil { #1 } \q_nil \}
\cs_gset:Npn \use_none_delimit_by_q_stop:w #1 \q_stop { #1 } \q_stop \}
\cs_gset:Npn \use_none_delimit_by_q_recursion_stop:w #1 \q_recursion_stop { #1 } \q_recursion_stop \}

(End of definition for \use_none_delimit_by_q_nil:w, \use_none_delimit_by_q_stop:w, and \use_none_delimit_by_q_recursion_stop:w. These functions are documented on page 26.)

Same as above but execute first argument after gobbling. Very useful when you need to skip the rest of a mapping sequence but want an easy way to control what should be expanded next.

\cs_gset:Npn \use_i_delimit_by_q_nil:nw #1 \q_nil { #1 } \q_nil \}
\cs_gset:Npn \use_i_delimit_by_q_stop:nw #1 \q_stop { #1 } \q_stop \}
\cs_gset:Npn \use_i_delimit_by_q_recursion_stop:nw #1 \q_recursion_stop { #1 } \q_recursion_stop \}

(End of definition for \use_i_delimit_by_q_nil:nw, \use_i_delimit_by_q_stop:nw, and \use_i_delimit_by_q_recursion_stop:nw. These functions are documented on page 27.)
42.5 Gobbling tokens from input

To gobble tokens from the input we use a standard naming convention: the number of tokens gobbled is given by the number of n’s following the : in the name. Although we could define functions to remove ten arguments or more using separate calls of \use_none:nnnnnnnnnn, this is very non-intuitive to the programmer who will assume that expanding such a function once takes care of gobbling all the tokens in one go.

\begin{verbatim}
\cs_gset:Npn \use_none:n { }  
\cs_gset:Npn \use_none:nn { }  
\cs_gset:Npn \use_none:nnn { }  
\cs_gset:Npn \use_none:nnnn { }  
\cs_gset:Npn \use_none:nnnnn { }  
\cs_gset:Npn \use_none:nnnnnn { }  
\cs_gset:Npn \use_none:nnnnnnn { }  
\cs_gset:Npn \use_none:nnnnnnnn { }  
\end{verbatim}

(End of definition for \use_none:n and others. These functions are documented on page 26.)

42.6 Debugging and patching later definitions

\begin{verbatim}
\cs_gset_protected:Npn \__kernel_if_debug:TF #1#2 {#2}  
\cs_gset_protected:Npn \debug_on:n #1  { \sys_load_debug: \cs_if_exist:NT \__debug_all_on: { \debug_on:n {#1} } }  
\cs_gset_protected:Npn \debug_off:n #1 { \sys_load_debug: \cs_if_exist:NT \__debug_all_on: { \debug_off:n {#1} } }  
\end{verbatim}

(End of definition for \__kernel_if_debug:TF.)

\begin{verbatim}
\debug_on:n \cs_gset_protected:Npn \debug_on:n #1 { \sys_load_debug: \cs_if_exist:NT \__debug_all_on: { \debug_on:n {#1} } }  
\end{verbatim}

(End of definition for \debug_on:n and \debug_off:n. These functions are documented on page 30.)

\begin{verbatim}
\debug_suspend: \cs_gset_protected:Npn \debug_suspend: { }  
\cs_gset_protected:Npn \debug_resume: { }  
\end{verbatim}

(End of definition for \debug_suspend: and \debug_resume:. These functions are documented on page 30.)
Make deprecated commands throw errors if the user requests it. This relies on two token lists, filled up in \texttt{l3deprecation}.

\verbatimex{\texttt{cs_gset_nopar:Npn \__kernel_deprecation_code:nn #1#2}}{\begin{verbatim}
\tl_gput_right:Nn \g__debug_deprecation_on_tl {#1}
\tl_gput_right:Nn \g__debug_deprecation_off_tl {#2}
\end{verbatim}}

(End of definition for \__kernel_deprecation_code:nn, \g__debug_deprecation_on_tl, \texttt{and} \g__debug_deprecation_off_tl.)

\section{Conditional processing and definitions}

Underneath any predicate function (\_p) or other conditional forms (TF, etc.) is a built-in logic saying that it after all of the testing and processing must return the \texttt{state} this leaves \LaTeX{} in. Therefore, a simple user interface could be something like

\verbatimex{\texttt{\if_meaning:w #1\prg_return_true:\else:\if_meaning:w #1\prg_return_true:\else:\prg_return_false:\fi:\fi:}}{\begin{verbatim}
\if_meaning:w #1\prg_return_true:
\else:
  \if_meaning:w #1\prg_return_true:
  \else:
    \prg_return_false:
  \fi:
\fi:
\end{verbatim}}

The idea here is that \texttt{\exp:w} expands fully any \texttt{\else:} and \texttt{\fi:} that are waiting to be discarded, before reaching the \texttt{\exp_end:} which leaves an empty expansion. The code can then leave either the first or second argument in the input stream. This means that all of the branching code has to contain at least two tokens: see how the logical tests are actually implemented to see this.

\verbatimex{\texttt{\cs_gset:Npn \prg_return_true:\cs_gset:Npn \prg_return_false:}}{\begin{verbatim}
\cs_gset:Npn \prg_return_true: { \exp_after:wN \use_i:nn \exp:w }
\cs_gset:Npn \prg_return_false: { \exp_after:wN \use_ii:nn \exp:w }
\end{verbatim}}

An extended state space could be implemented by including a more elaborate function in place of \texttt{use_i:nn/use_ii:nn}. Provided two arguments are absorbed then the code would work.

(End of definition for \texttt{\prg_return_true:} \texttt{and} \texttt{\prg_return_false:}. These functions are documented on page 65.)
Private version of \texttt{use\_none\_delimit\_by\_q\_recursion\_stop:w}.

\begin{verbatim}
\cs_gset:Npn \_\_prg_use_none_delimit_by_q_recursion_stop:w
  \#1 \_\_prg_recursion_stop \{ \}
\end{verbatim}

(End of definition for \_\_prg_use_none_delimit_by_q_recursion_stop:w.)

The user functions for the types using parameter text from the programmer. The various functions only differ by which function is used for the assignment. For those \texttt{Npnn} type functions, we must grab the parameter text, reading everything up to a left brace before continuing. Then split the base function into name and signature, and feed \{\texttt{(name)}\} \{\texttt{(signature)}\} \{\texttt{(set or new)}\} \{\texttt{(maybe protected)}\} \{\texttt{(parameters)}\} \{\texttt{TF,...}\} \{\texttt{(code)}\} to the auxiliary function responsible for defining all conditionals. Note that \texttt{e} stands for expandable and \texttt{p} for protected.

\begin{verbatim}
\cs_gset_protected:Npn \prg_set_conditional:Npnn
  \{ \_\_prg_generate_conditional_parm:NNNpnn \cs_set:Npn e \}
\cs_gset_protected:Npn \prg_gset_conditional:Npnn
  \{ \_\_prg_generate_conditional_parm:NNNpnn \cs_gset:Npn e \}
\cs_gset_protected:Npn \prg_new_conditional:Npnn
  \{ \_\_prg_generate_conditional_parm:NNNpnn \cs_new:Npn e \}
\cs_gset_protected:Npn \prg_set_protected_conditional:Npnn
  \{ \_\_prg_generate_conditional_parm:NNNpnn \cs_set_protected:Npn p \}
\cs_gset_protected:Npn \prg_gset_protected_conditional:Npnn
  \{ \_\_prg_generate_conditional_parm:NNNpnn \cs_gset_protected:Npn p \}
\cs_gset_protected:Npn \prg_new_protected_conditional:Npnn
  \{ \_\_prg_generate_conditional_parm:NNNpnn \cs_new_protected:Npn p \}
\cs_gset_protected:Npn \__prg_generate_conditional_count:NNNnnn
  \{ \_\_prg_generate_conditional_count:nnNNNnnn \cs_split_function:N #3 \}
  \#1 \#2 \{\#4\}
\end{verbatim}

(End of definition for \texttt{prg\_set\_conditional:Npnn} and others. These functions are documented on page 63.)

The user functions for the types automatically inserting the correct parameter text based on the signature. The various functions only differ by which function is used for the assignment. Split the base function into name and signature. The second auxiliary generates the parameter text from the number of letters in the signature. Then feed \{\texttt{(name)}\} \{\texttt{(signature)}\} \{\texttt{(set or new)}\} \{\texttt{(maybe protected)}\} \{\texttt{(parameters)}\} \{\texttt{TF,...}\} \{\texttt{(code)}\} to the auxiliary function responsible for defining all conditionals. If the \texttt{(signature)} has more than 9 letters, the definition is aborted since \TeX macros have at most 9 arguments. The erroneous case where the function name contains no colon is captured later.

\begin{verbatim}
\cs_gset_protected:Npn \prg_set_conditional:Nnn
  \{ \_\_prg_generate_conditional_count:NNNnnn \cs_set:Npn e \}
\cs_gset_protected:Npn \prg_gset_conditional:Nnn
  \{ \_\_prg_generate_conditional_count:NNNnnn \cs_gset:Npn e \}
\cs_gset_protected:Npn \prg_new_conditional:Nnn
  \{ \_\_prg_generate_conditional_count:NNNnnn \cs_new:Npn e \}
\end{verbatim}
The workhorse here is going through a list of desired forms, \textit{i.e.}, \texttt{p}, \texttt{TF}, \texttt{T} and \texttt{F}. The first three arguments come from splitting up the base form of the conditional, which gives the name, signature and a boolean to signal whether or not there was a colon in the name. In the absence of a colon, we throw an error and don’t define any conditional. The fourth and fifth arguments build up the defining function. The sixth is the parameters to use (possibly empty), the seventh is the list of forms to define, the eighth is the replacement text which we will augment when defining the forms. The use of \texttt{\tl_to_str:n} makes the later loop more robust.

A large number of our low-level conditionals look like \langle\texttt{code}\rangle \prg_return_true: \else: \prg_return_false: \fi: so we optimize this special case by calling \texttt{\__prg_generate_conditional_fast:nw \langle\texttt{code}\rangle}. This passes \texttt{\use_i:nn} instead of \texttt{\use_-i_ii:nnn} to functions such as \texttt{\__prg_generate_p_form:WNNnnnnN}.
Looping through the list of desired forms. First are six arguments and seventh is the form. Use the form to call the correct type. If the form does not exist, the \use:c construction results in \relax, and the error message is displayed (unless the form is empty, to allow for \{T, , F\}), then \use:none:nnnnnnnnn cleans up. Otherwise, the error message is removed by the variant form.

_END of definition for \__prg_generate_conditional:nnNNNnnn and others._

How to generate the various forms. Those functions take the following arguments: 1: junk, 2: \cs_set:Npn or similar, 3: p (for protected conditionals) or e, 4: function name, 5: signature, 6: parameter text, 7: replacement (possibly trimmed by \__prg_generate_conditional_fast:nw), 8: \use_i_ii:nnn or \use_i:nn (for “fast” conditionals). Remember that the logic-returning functions expect two arguments to be present after \exp_end:. notice the construction of the different variants relies on this, and that the TF and F variants will be slightly faster than the T version. The p form is only valid for expandable tests, we check for that by making sure that the second argument is empty. For “fast” conditionals, 7 has an extra \if_. To optimize a bit further we don’t use \exp_after:wN \use_i_ii:nn and similar but instead use \__prg_TP_true:w and similar to swap out the macro after \fi:. It would be a tiny bit faster if we directly
grabbed the T and F arguments there, but if those are actually missing, the recovery from the runaway argument would not insert \fi: back, messing up nesting of conditionals.

\begin{verbatim}
\cs_gset_protected:Npn \_prg_generate_p_form:wNNnnnnN #1 \s__prg_stop #2#3#4#5#6#7#8
{ \if_meaning:w e #3 \exp_after:wN \use_i:nn \else:
 \exp_after:wN \use_ii:nn \fi:
{ \exp_args:Nc #2 { #4 _p: #5 } #6 }
{ \#7 \exp_end: \c_true_bool \c_false_bool } \fi: \c_false_bool
}
\cs_gset_protected:Npn \_prg_generate_T_form:wNNnnnnN #1 \s__prg_stop #2#3#4#5#6#7#8
{ #8
{ \exp_args:Nc #2 { #4 : #5 T } #6 }
{ \#7 \exp_end: \use:n \use_none:n } \fi: \use_none:n
}
\cs_gset_protected:Npn \_prg_generate_F_form:wNNnnnnN #1 \s__prg_stop #2#3#4#5#6#7#8
{ #8
{ \exp_args:Nc #2 { #4 : #5 F } #6 }
{ \#7 \exp_end: { } } \fi: \use:nn
}
\cs_gset_protected:Npn \_prg_generate_TF_form:wNNnnnnN #1 \s__prg_stop #2#3#4#5#6#7#8
{ #8
{ \exp_args:Nc #2 { #4 : #5 TF } #6 }
{ \#7 \exp_end: { } } \fi: \use_ii:nn
}
\cs_gset:Npn \_prg_p_true:w \fi: \c_false_bool \fi: \c_true_bool
\cs_gset:Npn \_prg_T_true:w \fi: \use_none:n \fi: \use:n
\cs_gset:Npn \_prg_F_true:w \fi: \use:n \fi: \use_none:n
\cs_gset:Npn \_prg_TF_true:w \fi: \use_ii:nn
(End of definition for \_prg_generate_p_form:wNNnnnnN and others.)
\end{verbatim}

The setting-equal functions. Split both functions and feed \{⟨name1⟩\} \{⟨signature1⟩\} ⟨boolean1⟩ \{⟨name2⟩\} \{⟨signature2⟩\} ⟨boolean2⟩ ⟨copying function⟩ ⟨conditions⟩,
\_prg\_recursion\_tail , \_prg\_recursion\_stop to a first auxiliary.

\_cs\_gset\_protected:Npn \_\_prg\_set\_eq\_conditional:NNn 
\_\_prg\_set\_eq\_conditional:NNn \_\_set\_eq:cc 
\_cs\_gset\_protected:Npn \_\_prg\_set\_eq\_conditional:NNn 
\_\_prg\_set\_eq\_conditional:NNn \_\_set\_eq:cc 
\_cs\_gset\_protected:Npn \_\_prg\_set\_eq\_conditional:NNn 
\_\_prg\_set\_eq\_conditional:NNn \_\_set\_eq:cc 
\_cs\_gset\_protected:Npn \_\_prg\_set\_eq\_conditional:NNn \_\_prg\_new\_eq\_conditional:NNn 
\_\_prg\_set\_eq\_conditional:NNn \_\_new\_eq:cc 
\_cs\_gset\_protected:Npn \_\_prg\_set\_eq\_conditional:NNn \_\_prg\_set\_eq\_conditional:NNn #1#2#3#4 
\{ 
\use:e 
\{ 
\exp_not:N \_\_prg\_set\_eq\_conditional:nnNnnNNw 
\_cs\_split\_function:N #2 
\_cs\_split\_function:N #3 
\exp_not:N #1 
\tl\_to\_str:n \{ #4 \} 
\exp_not:n \{ , \_\_prg\_recursion\_tail , \_\_prg\_recursion\_stop \} 
\}
\}
(End of definition for \_\_prg\_set\_eq\_conditional:NNn and others. These functions are documented on page 65.)

Split the function to be defined, and setup a manual clist loop over argument #6 of the first auxiliary. The second auxiliary receives twice three arguments coming from splitting the function to be defined and the function to copy. Make sure that both functions contained a colon, otherwise we don’t know how to build conditionals, hence abort. Call the looping macro, with arguments \{⟨name1⟩\} \{⟨signature1⟩\} \{⟨name2⟩\} \{⟨signature2⟩\} \{copying function⟩ and followed by the comma list. At each step in the loop, make sure that the conditional form we copy is defined, and copy it, otherwise abort.

\_cs\_gset\_protected:Npn \_\_prg\_set\_eq\_conditional:nnNnnNNw #1#2#3#4#5#6 
\{ 
\_if\_meaning:w \_\_false\_bool #3 
\_msg\_error:nne \{ kernel \} \{ missing-colon \} 
\{ \token\_to\_str:c \{ #1 \} \} 
\exp\_after:wN \_\_prg\_use\_none\_delimit\_by_q\_recursion\_stop:w 
\_fi: 
\_if\_meaning:w \_\_false\_bool #6 
\_msg\_error:nne \{ kernel \} \{ missing-colon \} 
\{ \token\_to\_str:c \{ #4 \} \} 
\exp\_after:wN \_\_prg\_use\_none\_delimit\_by_q\_recursion\_stop:w 
\_fi: 
\_\_prg\_set\_eq\_conditional\_loop:nnnnNw \{ #1 \} \{ #2 \} \{ #4 \} \{ #5 \} 
\}
\cs\_gset\_protected:Npn \_\_prg\_set\_eq\_conditional\_loop:nnnnNw #1#2#3#4#5#6 , 
\{ 
\_if\_meaning:w \_\_prg\_recursion\_tail #6 
\exp\_after:wN \_\_prg\_use\_none\_delimit\_by_q\_recursion\_stop:w 
\_fi: 
\use:c \{ \_\_prg\_set\_eq\_conditional\_ #6 \_\_form:wNnnn \} 
\tl\_if\_empty:nF \{ #6 \} 
\{ 
\_msg\_error:nnee 
\{ kernel \} \{ conditional-form-unknown \} 
\}

386
All that is left is to define the canonical boolean true and false. I think Michael originated the idea of expandable boolean tests. At first these were supposed to expand into either TT or TF to be tested using \if:w but this was later changed to 00 and 01, so they could be used in logical operations. Later again they were changed to being numerical constants with values of 1 for true and 0 for false. We need this from the get-go.

\c_true_bool \c_false_bool
Here are the canonical boolean values.

(End of definition for \c_true_bool and \c_false_bool. These variables are documented on page 67.)

42.8 Dissecting a control sequence

\__cs_count_signature:N \__cs_count_signature:N ⟨function⟩
Splits the ⟨function⟩ into the ⟨name⟩ (i.e. the part before the colon) and the ⟨signature⟩ (i.e. after the colon). The ⟨number⟩ of tokens in the ⟨signature⟩ is then left in the input stream. If there was no ⟨signature⟩ then the result is the marker value −1.

Function used for various short-term usages, for instance defining functions whose definition involves tokens which are hard to insert normally (spaces, characters with category other).

\cs_to_str:N \cs_to_str:N \cs_to_str:w
This converts a control sequence into the character string of its name, removing the leading escape character. This turns out to be a non-trivial matter as there a different cases:

- The usual case of a printable escape character;
- the case of a non-printable escape characters, e.g., when the value of the \escapechar is negative;
• when the escape character is a space.

One approach to solve this is to test how many tokens result from \texttt{token_to_str:N \a}. If there are two tokens, then the escape character is printable, while if it is non-printable then only one is present.

However, there is an additional complication: the control sequence itself may start with a space. Clearly that should not be lost in the process of converting to a string. So the approach adopted is a little more intricate still. When the escape character is printable, \texttt{token_to_str:N \a} yields the escape character itself and a space. The character codes are different, thus the \texttt{if:w} test is false, and \TeX reads \texttt{__cs_to_str:N} after turning the following control sequence into a string; this auxiliary removes the escape character, and stops the expansion of the initial \texttt{tex_romannumeral:D}. The second case is that the escape character is not printable. Then the \texttt{if:w} test is unfinished after reading a space from \texttt{token_to_str:N \a}, and the auxiliary \texttt{__cs_to_str:w} is expanded, feeding – as a second character for the test; the test is false, and \TeX skips to \texttt{fi:}, then performs \texttt{token_to_str:N}, and stops the \texttt{tex_romannumeral:D} with \texttt{c_zero_int}. The last case is that the escape character is itself a space. In this case, the \texttt{if:w} test is true, and the auxiliary \texttt{__cs_to_str:w} comes into play, inserting \texttt{-\int_value:w}, which expands \texttt{c_zero_int} to the character \texttt{0}. The initial \texttt{tex_romannumeral:D} then sees \texttt{0}, which is not a terminated number, followed by the escape character, a space, which is removed, terminating the expansion of \texttt{tex_romannumeral:D}. In all three cases, \texttt{cs_to_str:N} takes two expansion steps to be fully expanded.

We implement the expansion scheme using \texttt{tex_romannumeral:D} terminating it with \texttt{c_zero_int} rather than using \texttt{exp:w} and \texttt{exp_end}: as we normally do. The reason is that the code heavily depends on terminating the expansion with \texttt{c_zero_int} so we make this dependency explicit.

If speed is a concern we could use \texttt{csstring} in \LaTeX. For the empty csname that primitive gives an empty result while the current \texttt{cs_to_str:N} gives incorrect results in all engines (this is impossible to fix without huge performance hit).

(End of definition for \texttt{cs_to_str:N}, \texttt{__cs_to_str:N}, and \texttt{__cs_to_str:w}. This function is documented on page 22.)

\begin{verbatim}
\cs_gset:Npn \cs_to_str:N
{ \if:w \token_to_str:N \__cs_to_str:w \fi:
\exp_after:wN \__cs_to_str:N \token_to_str:N \token_to_str:N\}
\cs_gset:Npn \__cs_to_str:N #1 { \c_zero_int }
\cs_gset:Npn \__cs_to_str:w #1 \__cs_to_str:N\}
\cs_gset:Npn \__cs_to_str:w #1 \__cs_to_str:N
{ - \int_value:w \fi: \exp_after:wN \c_zero_int }
\end{verbatim}

This function takes a function name and splits it into name with the escape char removed and argument specification. In addition to this, a third argument, a boolean \langle \texttt{true} \rangle or \langle \texttt{false} \rangle is returned with \langle \texttt{true} \rangle for when there is a colon in the function and \langle \texttt{false} \rangle if there is not.

First ensure that we actually get a properly evaluated string by expanding \texttt{cs_to_str:N} twice. If the function contained a colon, the auxiliary takes as \#1 the function name, delimited by the first colon, then the signature \#2, delimited by \texttt{s__cs_mark}, then \texttt{c_true_bool} as \#3, and \#4 cleans up until \texttt{s__cs_stop}. Otherwise, the \#1 contains the function name and \texttt{s__cs_mark} \texttt{c_true_bool}, \#2 is empty, \#3 is \texttt{c_false_bool},
and #4 cleans up. The second auxiliary trims the trailing \_\_cs\_mark from the function name if present (that is, if the original function had no colon).

1810 \cs_gset_protected:Npn \_\_cs\_tmp:w #1
1811 \{
1812 \cs_gset:Npn \cs_split_function:N ##1
1813 \{
1814 \exp_after:wN \exp_after:wN \exp_after:wN
1815 \_\_cs_split_function_auxi:w
1816 \cs_to_str:N \_\_cs\_mark \c\_true\_bool
1817 \#1 \_\_cs\_mark \c\_false\_bool \_\_cs\_stop
1818 \}
1819 \cs_gset:Npn \_\_cs_split_function_auxi:w
1820 \#1 \#2 \_\_cs\_mark \#3 \_\_cs\_stop
1821 \{ \_\_cs_split_function_auxii:w \#1 \_\_cs\_mark \_\_cs\_stop {\#2} \#3 \}
1822 \cs_gset:Npn \_\_cs_split_function_auxii:w \#1 \_\_cs\_mark \#2 \_\_cs\_stop
1823 \{ \{\#1\} \}
1824 \}
1825 \exp_after:wN \_\_cs\_tmp:w \token_to_str:N :

(End of definition for \cs_split_function:N, \_\_cs_split_function_auxi:w, and \_\_cs_split_function_auxii:w. This function is documented on page 22.)

42.9 Exist or free

A control sequence is said to exist (to be used) if has an entry in the hash table and its meaning is different from the primitive \relax token. A control sequence is said to be free (to be defined) if it does not already exist.

Two versions for checking existence. For the \texttt{N} form we firstly check for \scan_stop: and then if it is in the hash table. There is no problem when inputting something like \texttt{\else:} or \texttt{\fi:} as \TeX{} will only ever skip input in case the token tested against is \scan_stop:. In both the \texttt{N} and \texttt{c} form we use the way \prg_set_conditional:Npnn optimizes the conditionals to negate the tests using \texttt{\else:} (the \texttt{\else:} in the top level functions will be removed by the optimization, and this usage of \texttt{\else:} will be fine).

For the \texttt{c} form we firstly check if it is in the hash table and then for \scan_stop: so that we do not add it to the hash table unless it was already there. Here we have to be careful as the text to be skipped if the first test is false may contain tokens that disturb the scanner. Therefore, we ensure that the second test is performed after the first one has concluded completely.
\cs_if_exist:NTF \tex_lastnamedcs:D
  { \prg_gset_conditional:Npnn \cs_if_exist:c #1 { p , T , F , TF }
    { \if_cs_exist:w #1 \cs_end: \\
        \__cs_if_exist_c_aux:
        \prg_return_true:
        \else: \\
        \prg_return_false:
        \fi: 
    }
  \cs_gset:Npn \__cs_if_exist_c_aux:
    { \fi: \exp_after:wN \if_meaning:w \tex_lastnamedcs:D \scan_stop: \else: }
}

\cs_if_exist:NTF \tex_lastnamedcs:D
  { \prg_gset_conditional:Npnn \cs_if_exist:c #1 { p , T , F , TF }
    { \if_cs_exist:w #1 \cs_end: \\
        \__cs_if_exist_c_aux:w
        \fi: \\
        \use_none:n \if_false: \\
        \prg_return_true:
        \else: \\
        \prg_return_false:
        \fi: 
    }
  \cs_gset:Npn \__cs_if_exist_c_aux:w
    { \fi: \exp_after:wN \if_meaning:w \cs:w #1 \cs_end: \scan_stop: \else: }
}

(End of definition for \cs_if_exist:NTF, \__cs_if_exist_c_aux:, and \__cs_if_exist_c_aux:w. This function is documented on page 28.)

\cs_if_free_p:N
\cs_if_free_p:c
\cs_if_free:N
\cs_if_free:cTF

The logical reversal of the above.

\cs_if_free_p:N
\cs_if_free_p:c
\cs_if_free:N
\cs_if_free:cTF

\prg_gset_conditional:Npnn \cs_if_free:N #1 { p , T , F , TF }
{ \if_cs_exist:N #1
  \else: \\
  \use_none:nnnn
  \fi: \\
  \if_meaning:w #1 \scan_stop: \\
  \prg_return_true:
  \else: \\
  \prg_return_false:
  \fi: }
\cs_if_exist:NTF \tex_lastnamedcs:D
{ \prg_gset_conditional:Npnn \cs_if_free:c #1 { p , T , F , TF }
  { \if_cs_exist:w #1 \cs_end: \\
      \__cs_if_free_c_aux:w
      \fi: 
  }
}

390
The \texttt{cs_if_exist_use:...} functions cannot be implemented as conditionals because the true branch must leave both the control sequence itself and the true code in the input stream. For the \texttt{c} variants, we are careful not to put the control sequence in the hash table if it does not exist. If available we use the \texttt{lastnamedcs} primitive.
42.10 Preliminaries for new functions

We provide two kinds of functions that can be used to define control sequences. On the one hand we have functions that check if their argument doesn’t already exist, they are called \..._new. The second type of defining functions doesn’t check if the argument is already defined.

Before we can define them, we need some auxiliary macros that allow us to generate error messages. The next few definitions here are only temporary, they will be redefined later on.

\msg_error:nnee
\msg_error:nne
\msg_error:nn

If an internal error occurs before \LaTeX{} has loaded \l3msg then the code should issue a usable if terse error message and halt. This can only happen if a coding error is made by the team, so this is a reasonable response. Setting the \newlinechar is needed, to turn \^^J into a proper line break in plain \TeX{}.
\__kernel_chk_if_free_cs:N
\__kernel_chk_if_free_cs:c

This command is called by \cs_new:nopar:Npn and \cs_new:eq:NN etc. to make sure that the argument sequence is not already in use. If it is, an error is signalled. It checks if ⟨csname⟩ is undefined or \scan_stop:. Otherwise an error message is issued. We have to make sure we don’t put the argument into the conditional processing since it may be an \if... type function!

\cs_gset_protected:Npn \__kernel_chk_if_free_cs:N #1 1
\cs_gset_protected:Npn \__new_kernel_chk_if_free_cs:N #1
\cs_gset_protected:Npn \__cs_tmp:w
\cs_gset_protected:Npn \__cs_tmp:w #1#2
\cs_gset_protected:Npn \__cs_tmp:w #1#2
\cs_gset_protected:Npn \__cs_tmp:w #1#2
\cs_gset_protected:Npn \__cs_tmp:w #1#2
\cs_gset_protected:Npn \__cs_tmp:w #1#2

42.11 Defining new functions

\cs_new_nopar:Npn \cs_new_nopar:Npe \cs_new:nopar
\cs_gset_nopar:Npn \cs_gset:nopar
\cs_new_protected_nopar:Npn \cs_new_protected:nopar
\cs_new_protected:nopar
\cs_new_protected:nopar
\cs_new_protected:nopar
\cs_new_protected:nopar
\cs_new_protected:nopar
\cs_new_protected:nopar
\cs_new_protected:nopar
\cs_new_protected:nopar
\cs_new_protected:nopar
\cs_new_protected:nopar
\cs_new_protected:nopar
\cs_new_protected:nopar

Function which check that the control sequence is free before defining it.
\__cs_tmp:w \cs_new_nopar:Npe \cs_gset_nopar:Npe
\__cs_tmp:w \cs_new_nopar:Npx \cs_gset_nopar:Npx
\__cs_tmp:w \cs_new:Npn \cs_gset:Npn
\__cs_tmp:w \cs_new:Npe \cs_gset:Npe
\__cs_tmp:w \cs_new:Npx \cs_gset:Npx
\__cs_tmp:w \cs_new_protected_nopar:Npn \cs_gset_protected_nopar:Npn
\__cs_tmp:w \cs_new_protected_nopar:Npe \cs_gset_protected_nopar:Npe
\__cs_tmp:w \cs_new_protected_nopar:Npx \cs_gset_protected_nopar:Npx
\__cs_tmp:w \cs_new_protected_nopar:Npp \cs_gset_protected_nopar:Npp
\__cs_tmp:w \cs_new_protected:Npn \cs_gset_protected:Npn
\__cs_tmp:w \cs_new_protected:Npe \cs_gset_protected:Npe
\__cs_tmp:w \cs_new_protected:Npx \cs_gset_protected:Npx
\__cs_tmp:w \cs_set_nopar:cpn \cs_set_nopar:Npn
\__cs_tmp:w \cs_set_nopar:cpe \cs_set_nopar:Npe
\__cs_tmp:w \cs_set_nopar:cpx \cs_set_nopar:Npx
\__cs_tmp:w \cs_gset_nopar:cpn \cs_gset_nopar:Npn
\__cs_tmp:w \cs_gset_nopar:cpe \cs_gset_nopar:Npe
\__cs_tmp:w \cs_gset_nopar:cpx \cs_gset_nopar:Npx
\__cs_tmp:w \cs_new_nopar:cpn \cs_new_nopar:Npn
\__cs_tmp:w \cs_new_nopar:cpe \cs_new_nopar:Npe
\__cs_tmp:w \cs_new_nopar:cpx \cs_new_nopar:Npx
\__cs_set_nopar:cpn \cs_set_nopar:Npn and others. These functions are documented on page 15.)
\__cs_set_nopar:cpe \cs_set_nopar:Npe
\__cs_set_nopar:cpx \cs_set_nopar:Npx
\__cs_gset_nopar:cpn \cs_gset_nopar:Npn
\__cs_gset_nopar:cpe \cs_gset_nopar:Npe
\__cs_gset_nopar:cpx \cs_gset_nopar:Npx
\__cs_gset_nopar:Npp \cs_gset_nopar:Npp
\__cs_new_nopar:cpn \cs_new_nopar:Npn
\__cs_new_nopar:cpe \cs_new_nopar:Npe
\__cs_new_nopar:cpx \cs_new_nopar:Npx
\cs_set_nopar:cpn \cs_set_nopar:Npn
\cs_set_nopar:cpe \cs_set_nopar:Npe
\cs_set_nopar:cpx \cs_set_nopar:Npx
\cs_gset_nopar:cpn \cs_gset_nopar:Npn
\cs_gset_nopar:cpe \cs_gset_nopar:Npe
\cs_gset_nopar:cpx \cs_gset_nopar:Npx
\cs_new_nopar:cpn \cs_new_nopar:Npn
\cs_new_nopar:cpe \cs_new_nopar:Npe
\cs_new_nopar:cpx \cs_new_nopar:Npx
\cs_set:cpn \cs_set_nopar:Npn
\cs_set:cpx \cs_set_nopar:Npx
\cs_gset:cpn \cs_gset_nopar:Npn
\cs_gset:cpx \cs_gset_nopar:Npx
\cs_new:cpn \cs_new_nopar:Npn
\cs_new:cpx \cs_new_nopar:Npx
\cs_set_protected_nopar:cpn \cs_set_protected_nopar:Npn
\cs_set_protected_nopar:cpe \cs_set_protected_nopar:Npe
\cs_set_protected_nopar:cpx \cs_set_protected_nopar:Npx
\cs_gset_protected_nopar:cpn \cs_gset_protected_nopar:Npn
\cs_gset_protected_nopar:cpe \cs_gset_protected_nopar:Npe
\cs_gset_protected_nopar:cpx \cs_gset_protected_nopar:Npx
\cs_new_protected_nopar:cpn \cs_new_protected_nopar:Npn
\cs_new_protected_nopar:cpe \cs_new_protected_nopar:Npe
\cs_new_protected_nopar:cpx \cs_new_protected_nopar:Npx
\cs_set_protected_nopar:Npp \cs_set_protected_nopar:Npp
\cs_gset_protected_nopar:Npp \cs_gset_protected_nopar:Npp
\cs_new_protected_nopar:Npp \cs_new_protected_nopar:Npp
\cs_set_protected_nopar:cpn \cs_set_protected_nopar:Npn
\cs_set_protected_nopar:cpe \cs_set_protected_nopar:Npe
\cs_set_protected_nopar:cpx \cs_set_protected_nopar:Npx
\cs_gset_protected_nopar:cpn \cs_gset_protected_nopar:Npn
\cs_gset_protected_nopar:cpe \cs_gset_protected_nopar:Npe
\cs_gset_protected_nopar:cpx \cs_gset_protected_nopar:Npx
\cs_new_protected_nopar:cpn \cs_new_protected_nopar:Npn
\cs_new_protected_nopar:cpe \cs_new_protected_nopar:Npe
\cs_new_protected_nopar:cpx \cs_new_protected_nopar:Npx
\End of definition for \cs_set_nopar:Npp. This function is documented on page 16.)
\End of definition for \cs_set_nopar:Npp. This function is documented on page 16.)
\End of definition for \cs_new_nopar:Npp. This function is documented on page 16.)
\End of definition for \cs_new_nopar:Npp. This function is documented on page 16.)

Variants of the \cs_set_protected_nopar:Npp versions which make a csname out of the first arguments. We may also do this globally.
\cs_set_protected_nopar:cpn \cs_set_protected_nopar:Npn
\cs_set_protected_nopar:cpe \cs_set_protected_nopar:Npe
\cs_set_protected_nopar:cpx \cs_set_protected_nopar:Npx
\cs_gset_protected_nopar:cpn \cs_gset_protected_nopar:Npn
\cs_gset_protected_nopar:cpe \cs_gset_protected_nopar:Npe
\cs_gset_protected_nopar:cpx \cs_gset_protected_nopar:Npx
\cs_new_protected_nopar:cpn \cs_new_protected_nopar:Npn
\cs_new_protected_nopar:cpe \cs_new_protected_nopar:Npe
\cs_new_protected_nopar:cpx \cs_new_protected_nopar:Npx

Variants of the \cs_set_nopar:Npp versions which make a csname out of the first arguments. We may also do this globally.
\cs_set_nopar:cpn \cs_set_nopar:Npn
\cs_set_nopar:cpe \cs_set_nopar:Npe
\cs_set_nopar:cpx \cs_set_nopar:Npx
\cs_gset_nopar:cpn \cs_gset_nopar:Npn
\cs_gset_nopar:cpe \cs_gset_nopar:Npe
\cs_gset_nopar:cpx \cs_gset_nopar:Npx
\cs_new_nopar:cpn \cs_new_nopar:Npn
\cs_new_nopar:cpe \cs_new_nopar:Npe
\cs_new_nopar:cpx \cs_new_nopar:Npx

Variants of the \cs_set_nopar:Npp versions which make a csname out of the first arguments. We may also do this globally.
\cs_set:cpn \cs_set_nopar:Npn
\cs_set:cpx \cs_set_nopar:Npx
\cs_gset:cpn \cs_gset_nopar:Npn
\cs_gset:cpx \cs_gset_nopar:Npx
\cs_new:cpn \cs_new_nopar:Npn
\cs_new:cpx \cs_new_nopar:Npx

Like \cs_set:no par:Npp and \cs_new:nopar:Npn, except that the first argument consists of the sequence of characters that should be used to form the name of the desired control sequence (the c stands for csname argument, see the expansion module). Global versions are also provided.
\__cs_set_nopar:cpn \cs_set_nopar:Npn
\__cs_set_nopar:cpe \cs_set_nopar:Npe
\__cs_set_nopar:cpx \cs_set_nopar:Npx
\__cs_gset_nopar:cpn \cs_gset_nopar:Npn
\__cs_gset_nopar:cpe \cs_gset_nopar:Npe
\__cs_gset_nopar:cpx \cs_gset_nopar:Npx
\__cs_new_nopar:cpn \cs_new_nopar:Npn
\__cs_new_nopar:cpe \cs_new_nopar:Npe
\__cs_new_nopar:cpx \cs_new_nopar:Npx

Variants of the \cs_set:Npn versions which make a csname out of the first arguments. We may also do this globally.
\__cs_set:Npn \cs_set_nopar:Npn
\__cs_set:Npe \cs_set_nopar:Npe
\__cs_set:Npx \cs_set_nopar:Npx
\__cs_gset:Npn \cs_gset_nopar:Npn
\__cs_gset:Npe \cs_gset_nopar:Npe
\__cs_gset:Npx \cs_gset_nopar:Npx
\__cs_new:Npn \cs_new_nopar:Npn
\__cs_new:Npe \cs_new_nopar:Npe
\__cs_new:Npx \cs_new_nopar:Npx

\cs_set_protected_nopar:cpn\cs_set_protected_nopar:cpe\cs_set_protected_nopar:cpx
\cs_gset_protected_nopar:cpn \cs_gset_protected_nopar:cpe \cs_gset_protected_nopar:cpx
\cs_new_protected_nopar:cpn \cs_new_protected_nopar:cpe \cs_new_protected_nopar:cpx
Variants of the `\cs_set_protected:Npn` versions which make a csname out of the first arguments. We may also do this globally.

\begin{verbatim}
\cs_set_protected:cpn \cs_set_protected:cpe \cs_set_protected:cpx
\cs_gset_protected:cpn \cs_gset_protected:cpe \cs_gset_protected:cpx
\cs_new_protected:cpn \cs_new_protected:cpe \cs_new_protected:cpx
\end{verbatim}

(End of definition for `\cs_set_protected:Npn`. This function is documented on page 16.)

### 42.12 Copying definitions

These macros allow us to copy the definition of a control sequence to another control sequence.

\begin{verbatim}
\cs_set_eq:NN \cs_set_eq:cN \cs_set_eq:Nc \cs_set_eq:cc
\cs_gset_eq:NN \cs_gset_eq:cN \cs_gset_eq:Nc \cs_gset_eq:cc
\cs_new_eq:NN \cs_new_eq:cN \cs_new_eq:Nc \cs_new_eq:cc
\end{verbatim}

The `=` sign allows us to define funny char tokens like `=` itself or `␣` with this function. For the definition of `\c_space_char{~}` to work we need the `~` after the `=`.

`\cs_set_eq:NN` is long to avoid problems with a literal argument of `\par`. While `\cs_new_eq:NN` will probably never be correct with a first argument of `\par`, define it long in order to throw an “already defined” error rather than “runaway argument”.

\begin{verbatim}
\cs_new_protected:Npn \cs_set_eq:NN \cs_new_eq:NN \cs_new_eq:cN \cs_new_eq:Nc \cs_new_eq:cc
\end{verbatim}

(End of definition for `\cs_set_protected:Npn`, `\cs_gset_protected:Npn`, and `\cs_new_protected:Npn`. These functions are documented on page 20.)
42.13 Undefining functions

\cs_undefine:N
\cs_undefine:c

The following function is used to free the main memory from the definition of some function that isn’t in use any longer. The \texttt{c} variant is careful not to add the control sequence to the hash table if it isn’t there yet, and it also avoids nesting \TeX conditional in case \texttt{#1} is unbalanced in this matter. We optimize the case where the command exists by reducing as much as possible the tokens in the conditional.

\begin{verbatim}
2064 \cs_new_protected:Npn \cs_undefine:N #1
2065 { \cs_gset_eq:NN #1 \tex_undefined:D }
2066 \cs_new_protected:Npn \cs_undefine:c #1
2067 { \if_cs_exist:w #1 \cs_end:
2068 \else:
2069 \use_i:nnnn
2070 \fi:
2071 \exp_args:Nc \cs_undefine:N {#1}
2072 }
\end{verbatim}

(End of definition for \texttt{\cs_undefine:N}. This function is documented on page 20.)

42.14 Generating parameter text from argument count

\texttt{\__kernel_cs_parm_from_arg_count:nnF}
\texttt{\__cs_parm_from_arg_count_test:nnF}

\LaTeX{} provides shorthands to define control sequences and conditionals with a simple parameter text, derived directly from the signature, or more generally from knowing the number of arguments, between 0 and 9. This function expands to its first argument, untouched, followed by a brace group containing the parameter text \texttt{{\#1...\#n}}, where \texttt{n} is the result of evaluating the second argument (as described in \texttt{\int_eval:n}). If the second argument gives a result outside the range [0, 9], the third argument is returned instead, normally an error message. Some of the functions use here are not defined yet, but will be defined before this function is called.

\begin{verbatim}
2075 \cs_new_protected:Npn \__kernel_cs_parm_from_arg_count:nnF #1#2
2076 { \exp_args:Ne \__cs_parm_from_arg_count_test:nnF
2077 { \if_case:w \int_eval:n {#2}
2078 { }
2079 \or: { \#1 } \or: { \#1\#2 } \or: { \#1\#2\#3 } \or: { \#1\#2\#3\#4 } \or: { \#1\#2\#3\#4\#5 } \or: { \#1\#2\#3\#4\#5\#6 } \or: { \#1\#2\#3\#4\#5\#6\#7 } \or: { \#1\#2\#3\#4\#5\#6\#7\#8 } \or: { \#1\#2\#3\#4\#5\#6\#7\#8\#9 } \else: { \c_false_bool } \fi:
2081 }{#1}
\end{verbatim}

396
\begin{quote}

Counting the number of tokens in the signature, \textit{i.e.}, the number of arguments the function should take. Since this is not used in any time-critical function, we simply use \texttt{\_\_count:n} if there is a signature, otherwise \texttt{-1} arguments to signal an error. We need a variant form right away.

\[\text{\texttt{\_\_kernel_cs_parm_from_arg_count:nnF}}\text{\texttt{\_\_cs_parm_from_arg_count_\texttt{-test}:nnF}}\]
\end{quote}

\section{Defining functions from a given number of arguments}

We provide a constructor function for defining functions with a given number of arguments. For this we need to choose the correct parameter text and then use that when defining. Since \TeX{} supports from zero to nine arguments, we use a simple switch to choose the correct parameter text, ensuring the result is returned after finishing the conditional. If it is not between zero and nine, we throw an error.

1: function to define, 2: with what to define it, 3: the number of args it requires and 4: the replacement text

\begin{quote}
\[\text{\texttt{\_\_kernel_cs_parm_from_arg_count:nnF}}\text{\texttt{\_\_cs_parm_from_arg_count_\texttt{-test}:nnF}}\]
\end{quote}
A variant form we need right away, plus one which is used elsewhere but which is most logically created here.

\use_none:n
\{#4\}
\}

\end{definition}

42.16 Using the signature to define functions

We can now combine some of the tools we have to provide a simple interface for defining functions, where the number of arguments is read from the signature. For instance, \cs_set:Nn \foo_bar:nn {#1,#2}. We want to define \cs_set:Nn as

\cs_set:Npn \cs_set:Nn #1#2
\{\cs_generate_from_arg_count:NNnn #1 \cs_set:Npn \{\@@_count_signature:N #1\} {#2}\}

In short, to define \cs_set:Nn we need just use \cs_set:Npn, everything else is the same for each variant. Therefore, we can make it simpler by temporarily defining a function to do this for us.

\cs_set:Npn \__cs_tmp:w #1#2#3
\{\cs_new_protected:cpx { cs_ #1 : #2 } \exp_not:N \__cs_generate_from_signature:NNn \exp_after:wN \exp_not:N \cs:w cs_ #1 : #3 \cs_end:\}

\cs_new_protected:Npn \__cs_generate_from_signature:nnNNNn #1#2#3#4#5#6
\{\bool_if:NTF #3
\{\cs_set_nopar:Npx \__cs_tmp:w \{\@@_count_signature:N \cs_generate_from_signature:nnNNn \exp_after:wN \exp_not:N \cs:w cs_ \#1 : \#2 \cs_end:\}\}
\}

\cs_new_protected:Npm \cs_generate_from_arg_count:NNnn
\{\exp_args:Nc \cs_generate_from_arg_count:NNnn \}
\cs_new_protected:Npm \cs_generate_from_arg_count:Ncnn
\{\exp_args:Ncc \cs_generate_from_arg_count:NNnn \}

(End of definition for \cs_generate_from_arg_count:NNnn. This function is documented on page 19.)
Then we define the 24 variants beginning with $N$.

\begin{verbatim}
\cs_new:Npn \__cs_generate_from_signature:n #1
\begin{verbatim}
\__cs_tmp:w \set \{ Nn \} \{ Npn \}
\__cs_tmp:w \set \{ Ne \} \{ Npe \}
\__cs_tmp:w \set \{ Nx \} \{ Npx \}
\__cs_tmp:w \set \{ set \} \{ Nn \} \{ Npn \}
\__cs_tmp:w \set \{ set \} \{ Ne \} \{ Npe \}
\__cs_tmp:w \set \{ set \} \{ Nx \} \{ Npx \}
\__cs_tmp:w \set \{ set_nopar \} \{ Nn \} \{ Npn \}
\__cs_tmp:w \set \{ set_nopar \} \{ Ne \} \{ Npe \}
\__cs_tmp:w \set \{ set_nopar \} \{ Nx \} \{ Npx \}
\__cs_tmp:w \set \{ set_protected \} \{ Nn \} \{ Npn \}
\__cs_tmp:w \set \{ set_protected \} \{ Ne \} \{ Npe \}
\__cs_tmp:w \set \{ set_protected \} \{ Nx \} \{ Npx \}
\__cs_tmp:w \set \{ set_protected_nopar \} \{ Nn \} \{ Npn \}
\__cs_tmp:w \set \{ set_protected_nopar \} \{ Ne \} \{ Npe \}
\__cs_tmp:w \set \{ set_protected_nopar \} \{ Nx \} \{ Npx \}
\__cs_tmp:w \set \{ gset \} \{ Nn \} \{ Npn \}
\__cs_tmp:w \set \{ gset \} \{ Ne \} \{ Npe \}
\__cs_tmp:w \set \{ gset \} \{ Nx \} \{ Npx \}
\__cs_tmp:w \set \{ gset_nopar \} \{ Nn \} \{ Npn \}
\__cs_tmp:w \set \{ gset_nopar \} \{ Ne \} \{ Npe \}
\__cs_tmp:w \set \{ gset_nopar \} \{ Nx \} \{ Npx \}
\__cs_tmp:w \set \{ gset_protected \} \{ Nn \} \{ Npn \}
\__cs_tmp:w \set \{ gset_protected \} \{ Ne \} \{ Npe \}
\__cs_tmp:w \set \{ gset_protected \} \{ Nx \} \{ Npx \}
\__cs_tmp:w \set \{ gset_protected_nopar \} \{ Nn \} \{ Npn \}
\__cs_tmp:w \set \{ gset_protected_nopar \} \{ Ne \} \{ Npe \}
\__cs_tmp:w \set \{ gset_protected_nopar \} \{ Nx \} \{ Npx \}
\__cs_tmp:w \set \{ new \} \{ Nn \} \{ Npn \}
\__cs_tmp:w \set \{ new \} \{ Ne \} \{ Npe \}
\__cs_tmp:w \set \{ new \} \{ Nx \} \{ Npx \}
\end{verbatim}
\end{verbatim}

\begin{verbatim}
\__cs_tmp:w \set \{ new_nopar \} \{ Nn \} \{ Npn \}
\__cs_tmp:w \set \{ new_nopar \} \{ Ne \} \{ Npe \}
\__cs_tmp:w \set \{ new_nopar \} \{ Nx \} \{ Npx \}
\__cs_tmp:w \set \{ new_protected \} \{ Nn \} \{ Npn \}
\__cs_tmp:w \set \{ new_protected \} \{ Ne \} \{ Npe \}
\__cs_tmp:w \set \{ new_protected \} \{ Nx \} \{ Npx \}
\__cs_tmp:w \set \{ new_protected_nopar \} \{ Nn \} \{ Npn \}
\__cs_tmp:w \set \{ new_protected_nopar \} \{ Ne \} \{ Npe \}
\__cs_tmp:w \set \{ new_protected_nopar \} \{ Nx \} \{ Npx \}
\end{verbatim}
\end{verbatim}

\end{verbatim}
The 24 \texttt{c} variants simply use \texttt{exp_args:Nc}.

\begin{verbatim}
\cs_set:cn \cs_set:ce \cs_set:nopar:cn \cs_set:nopar:ce
\cs_set_protected:cn \cs_set_protected:ce \cs_set_protected:nopar:cn \cs_set_protected:nopar:ce
\cs_set_protected_protected:cn \cs_set_protected_protected:ce \cs_set_protected_protected:nopar:cn \cs_set_protected_protected:nopar:ce
\cs_gset:cn \cs_gset:ce \cs_gset:nopar:cn \cs_gset:nopar:ce
\cs_gset_protected:cn \cs_gset_protected:ce \cs_gset_protected:nopar:cn \cs_gset_protected:nopar:ce
\cs_gset_protected_protected:cn \cs_gset_protected_protected:ce \cs_gset_protected_protected:nopar:cn \cs_gset_protected_protected:nopar:ce
\cs_new:cn \cs_new:ce \cs_new:nopar:cn \cs_new:nopar:ce
\cs_new_protected:cn \cs_new_protected:ce \cs_new_protected:nopar:cn \cs_new_protected:nopar:ce
\end{verbatim}

(End of definition for \cs_set:Nn and others. These functions are documented on page 18.)
42.17 Checking control sequence equality

\cs_if_eq_p:NN \cs_if_eq_p:cN \cs_if_eq_p:Nc \cs_if_eq_p:cc \cs_if_eq:NN \cs_if_eq:cN \cs_if_eq:Nc \cs_if_eq:cc

Check if two control sequences are identical.

\prg_new_conditional:Npnn \cs_if_eq:NN #1#2 { p , T , F , TF }
{ \if_meaning:w #1#2 \prg_return_true: \else: \prg_return_false: \fi:
}

\cs_new:Npn \cs_if_eq_p:cN { \exp_args:Nc \cs_if_eq_p:NN }
\cs_new:Npn \cs_if_eq:cNTF { \exp_args:Nc \cs_if_eq:NNTF }
\cs_new:Npn \cs_if_eq:ccT { \exp_args:Ncc \cs_if_eq:NNT }
\cs_new:Npn \cs_if_eq:ccF { \exp_args:Ncc \cs_if_eq:NNF }

(End of definition for \cs_if_eq:NNTF. This function is documented on page 28.)

42.18 Diagnostic functions

\__kernel_chk_defined:NT Error if the variable #1 is not defined.
\cs_new_protected:Npn \__kernel_chk_defined:NT #1#2
{ \cs_if_exist:NTF #1 {#2} {\msg_error:nne { kernel } { variable-not-defined } { 	oken_to_str:N #1 } }
}

(End of definition for \__kernel_chk_defined:NT.)

\__kernel_register_show:N \__kernel_register_show:c \__kernel_register_log:N \__kernel_register_log:c \__kernel_register_show_aux:N \__kernel_register_show_aux:NN

Simply using the \showthe primitive does not allow for line-wrapping, so instead use \tl_show:n and \tl_log:n (defined in l3tl and that performs line-wrapping). This displays \texttt{\textless variable\textgreater}=\texttt{ value}. We expand the value beforehand as otherwise some integers (such as \currentgrouplevel or \currentgrouptype) altered by the line-wrapping code would show wrong values.

\cs_new_protected:Npn \__kernel_register_show:N
{ \__kernel_register_show_aux:NN \tl_show:n }
\cs_new_protected:Npn \__kernel_register_show:c
{ \exp_args:Nc \__kernel_register_show:N }
\cs_new_protected:Npn \__kernel_register_log:N
{ \__kernel_register_show_aux:NN \tl_log:n }
\cs_new_protected:Npn \__kernel_register_log:c
{ \exp_args:Nc \__kernel_register_log:N }

401
Some control sequences have a very long name or meaning. Thus, simply using \TeX’s primitive \show could lead to overlong lines. The output of this primitive is mimicked to some extent, then the re-built string is given to \tl_show:n or \tl_log:n for line-wrapping. We must expand the meaning before passing it to the wrapping code as otherwise we would wrongly see the definitions that are in place there. To get correct escape characters, set the \escapechar in a group; this also localizes the assignment performed by e-expansion. The \cs_show:N and \cs_log:N commands convert their argument to a control sequence within a group to avoid showing \relax for undefined control sequences.

Wrapper around showgroups. Getting \TeX to write to the log without interruption the run is done by altering the interaction mode.
42.19 Decomposing a macro definition

We sometimes want to test if a control sequence can be expanded to reveal a hidden value. However, we cannot just expand the macro blindly as it may have arguments and none might be present. Therefore we define these functions to pick either the prefix(es), the parameter specification, or the replacement text from a macro. All of this information is returned as characters with catcode 12. If the token in question isn’t a macro, the token \scan_stop: is returned instead.

```latex
\token_if_macro:NTF \cs_prefix_spec:N #1
  \exp_after:wN \__kernel_prefix_arg_replacement:wN \token_to_meaning:N #1 \s__kernel_stop \use_i:nnn
\token_stop: }
\cs_new:Npn \cs_prefix_spec:N #1
  \token_if_macro:NTF #1
  \exp_after:wN \__kernel_prefix_arg_replacement:wN \token_to_meaning:N \exp_not:n \{ \cs_new:Npn \__kernel_prefix_arg_replacement:wN \exp_not:n \#2 \exp_not:n \#3 \s__kernel_stop \#4 \}
\cs_new:Npn \cs_replacement_spec:N #1
  \token_if_macro:NTF \__kernel_prefix_arg_replacement:wN \token_to_meaning:N \exp_not:n \{ \cs_new:Npn \__kernel_prefix_arg_replacement:wN \exp_not:n \#1 \exp_not:n \#2 \exp_not:n \#3 \exp_not:n \#4 \s__kernel_stop \use_i:nnn
\token_stop: }
\cs_new:Npn \cs_replacement_spec:N #1
  \token_if_macro:NTF \__kernel_prefix_arg_replacement:wN \token_to_meaning:N \exp_not:n \{ \cs_new:Npn \__kernel_prefix_arg_replacement:wN \exp_not:n \#1 \exp_not:n \#2 \exp_not:n \#3 \exp_not:n \#4 \s__kernel_stop \use_ii:nnn
\token_stop: }
\cs_new:Npn \cs_parameter_spec:N #1
  \token_if_macro:NTF \__kernel_prefix_arg_replacement:wN \token_to_meaning:N \exp_not:n \{ \cs_new:Npn \__kernel_prefix_arg_replacement:wN \exp_not:n \#1 \exp_not:n \#2 \exp_not:n \#3 \exp_not:n \#4 \s__kernel_stop \use_iii:nnn
\token_stop: }
```

(End of definition for \group_show_list:, \group_log_list:, and \__kernel_group_show:NN. These functions are documented on page 14.)
42.20 Doing nothing functions

\prg_do_nothing: This does not fit anywhere else!

\cs_new:Npn \prg_do_nothing: { }

42.21 Breaking out of mapping functions

\prg_break_point:Nn \prg_map_break:Nn In inline mappings, the nesting level must be reset at the end of the mapping, even when
the user decides to break out. This is done by putting the code that must be performed
as an argument of \_\_prg_break_point:Nn. The breaking functions are then defined to
jump to that point and perform the argument of \_\_prg_break_point:Nn, before the
user’s code (if any). There is a check that we close the correct loop, otherwise we continue
breaking.

\cs_new_eq:NN \prg_break_point:Nn \use_ii:nn
\cs_new:Npn \prg_map_break:Nn #1#2#3 \prg_break_point:Nn #4#5
{ #5 \if_meaning:w #1 #4 \exp_after:wN \use_ii:nn \fi:
 \prg_map_break:Nn #1 {#2}
}

\prg_break_point: \prg_break: \prg_break:n Very simple analogues of \prg_break_point:Nn and \prg_map_break:Nn, for use in fast
short-term recursions which are not mappings, do not need to support nesting, and in
which nothing has to be done at the end of the loop.

\cs_new_eq:NN \prg_break_point: \prg_do_nothing:
\cs_new:Npn \prg_break: #1 \prg_break_point: { }
\cs_new:Npn \prg_break:n #1\prg_break_point: {#1}

42.22 Starting a paragraph

\mode_leave_vertical: The approach here is different to that used by \TeX or plain \TeX, which unbox a
void box to force horizontal mode. That inserts the \everypar tokens before the re-
inserted unboxing tokens. The approach here uses a protected macro, equivalent to the
\quitvmode primitive. In vertical mode, the \indent primitive is inserted: this will
switch to horizontal mode and insert \everypar tokens and nothing else. Unlike the
\LaTeX \varepsilon version, the availability of \varepsilon-\TeX means using a mode test can be done at for example the start of an \halign.

\begin{verbatim}
\cs_new_protected:Npn \mode_leave_vertical:
  {\if_mode_vertical:\exp_after:wN \tex_indent:D \fi:}
\end{verbatim}

(End of definition for \mode_leave_vertical:. This function is documented on page 30.)

\end{verbatim}
Chapter 43

\textit{l3expan} implementation

\verb|\_exp_internal_tl| The \texttt{exp} module has its private variable to temporarily store the result of \texttt{x}-type argument expansion. This is done to avoid interference with other functions using temporary variables.

\textit{End of definition for \_exp_internal_tl.}

\verb|\exp_after:wN| These are defined in \textit{l3basics}, as they are needed “early”. This is just a reminder of that fact!

\textit{End of definition for \exp_after:wN, \exp_not:N, and \exp_not:n. These functions are documented on page 39.}

43.1 General expansion

In this section a general mechanism for defining functions that handle arguments is defined. These general expansion functions are expandable unless \texttt{x} is used. (Any version of \texttt{x} is going to have to use one of the \LaTeX{} names for \texttt{cs_set:Np} at some point, and so is never going to be expandable.)

The definition of expansion functions with this technique happens in section 43.7. In section 43.2 some common cases are coded by a more direct method for efficiency, typically using calls to \texttt{\exp_after:wN}.

\verb|\_exp_internal_tl| This scratch token list variable is defined in \textit{l3basics}.

\textit{End of definition for \_exp_internal_tl.}

This code uses internal functions with names that start with \texttt{\::} to perform the expansions. All macros are \texttt{long} since the tokens undergoing expansion may be arbitrary user input.

An argument manipulator \texttt{\::(2)} always has signature \texttt{\#1\::\#2\#3} where \texttt{\#1} holds the remaining argument manipulations to be performed, \texttt{\::} serves as an end marker for the list of manipulations, \texttt{\#2} is the carried over result of the previous expansion steps and \texttt{\#3} is the argument about to be processed. One exception to this rule is \texttt{\::p}, which has to grab an argument delimited by a left brace.
\_\_exp_arg_next::nnn
\_\_exp_arg_next::Nnn
#1 is the result of an expansion step, #2 is the remaining argument manipulations and
#3 is the current result of the expansion chain. This auxiliary function moves #1 back
after #3 in the input stream and checks if any expansion is left to be done by calling
#2. In by far the most cases we need to add a set of braces to the result of an argument
manipulation so it is more effective to do it directly here. Actually, so far only the \c of
the final argument manipulation variants does not require a set of braces.

\begin{verbatim}
\cs_new:Npn \_\_exp_arg_next::nnn #1#2#3 { #2 ::: { #3 {#1} } }
\cs_new:Npn \_\_exp_arg_next::Nnn #1#2#3 { #2 ::: { #3 #1 } }
\end{verbatim}

\textit{(End of definition for \_\_exp_arg_next::nnn and \_\_exp_arg_next::Nnn.)}

\::: The end marker is just another name for the identity function.
\begin{verbatim}
\cs_new:Npn \::: #1 {#1}
\end{verbatim}

\textit{(End of definition for \:::. \textit{This function is documented on page 43.})}

\::n This function is used to skip an argument that doesn’t need to be expanded.
\begin{verbatim}
\cs_new:Npn \::n #1 \::: #2#3 { #1 \::: {#2 {#3} } }
\end{verbatim}

\textit{(End of definition for \::n. This function is documented on page 43.)}

\::N This function is used to skip an argument that consists of a single token and doesn’t
need to be expanded. It is not wrapped in braces in the result.
\begin{verbatim}
\cs_new:Npn \::N #1 \::: #2#3 { #1 \::: {#2#3} }
\end{verbatim}

\textit{(End of definition for \::N. This function is documented on page 43.)}

\::p This function is used to skip an argument that is delimited by a left brace and doesn’t
need to be expanded. It is not wrapped in braces in the result.
\begin{verbatim}
\cs_new:Npn \::p #1 \::: #2#3# { #1 \::: {#2#3} }
\end{verbatim}

\textit{(End of definition for \::p. This function is documented on page 43.)}

\::c This function is used to skip an argument that is turned into a control sequence without
expansion.
\begin{verbatim}
\cs_new:Npn \::c #1 \::: #2#3
\{ \exp_after:wN \_\_exp_arg_next::Nnn \cs:w #3 \cs_end: {#1} {#2} \}
\end{verbatim}

\textit{(End of definition for \::c. This function is documented on page 43.)}

\::o This function is used to expand an argument once.
\begin{verbatim}
\cs_new:Npn \::o #1 \::: #2#3
\{ \exp_after:wN \_\_exp_arg_next::nnn \exp_after:wN \{#3\} {#1} {#2} \}
\end{verbatim}

\textit{(End of definition for \::o. This function is documented on page 43.)}

\::e With the \texttt{\textbackslash \texttt{\texttt{\texttt{\textbackslash expanded}}} primitive available, just expand.
\begin{verbatim}
\cs_new:Npn \::e #1 \::: #2#3
\{ \tex_expanded:D \{ \exp_not:n \{ #1 \::: \} \{ \exp_not:n \{#2\} \{#3\} \} \}
\end{verbatim}

\textit{(End of definition for \::e. This function is documented on page 43.)}
This function is used to expand a token list until the first unexpandable token is found. This is achieved through \exp:w \exp_end_continue_f:w that expands everything in its way following it. This scanning procedure is terminated once the expansion hits something non-expandable (if that is a space it is removed). We introduce \exp_stop_f: to mark such an end-of-expansion marker. For example, f-expanding \cs_set_eq:Nc \aaa { b \l_tmpa_tl b } where \l_tmpa_tl contains the characters lur gives \tex_let:D \aaa = \blurb which then turns out to start with the non-expandable token \tex_let:D. Since the expansion of \exp:w \exp_end_continue_f:w is empty, we wind up with a fully expanded list, only \TeX has not tried to execute any of the non-expandable tokens. This is what differentiates this function from the e and x argument type.

\cs_new:Npn \::f #1 \::: #2#3
\exp_after:wN \__exp_arg_next:nnn
\exp_after:wN { \exp:w \exp_end_continue_f:w #3 }
{#1} {#2}
\exp_end_continue_f:**
\use:nn { \cs_new_eq:NN \exp_stop_f: } { ~ }

(End of definition for \::f and \exp_stop_f:. These functions are documented on page 43.)

This function is used to expand an argument fully. We build in the expansion of \__exp_arg_next:nnn.
\cs_new_protected:Npn \::x #1 \::: #2#3
\exp_after:wN \__exp_arg_next:nnn
\exp_after:wN { \exp:w \__exp_eval_register:N #3 }
{#1} {#2}
\__exp_eval_register:
\l__exp_internal_tl
\cs_set_nopar:Npe \l__exp_internal_tl
{ \exp_not:n { #1 \::: } { \exp_not:n {#2} {#3} } }
\l__exp_internal_tl
\l__exp_internal_tl

(End of definition for \::x. This function is documented on page 43.)

These functions return the value of a register, i.e., one of \tl, \clist, \int, \skip, \dim, \muskip, or built-in \TeX register. The V version expects a single token whereas v like c creates a csname from its argument given in braces and then evaluates it as if it was a V. The \exp:w sets off an expansion similar to an f-type expansion, which we terminate using \exp_end:. The argument is returned in braces.
\cs_new:Npn \::v #1 \::: #2#3
\exp_after:wN \__exp_arg_next:nnn
\exp_after:wN { \exp:w \__exp_eval_register:c {#3} }
{#1} {#2}
\__exp_eval_register:
\l__exp_internal_tl
\l__exp_internal_tl
\l__exp_internal_tl

(End of definition for \::v and \::V. These functions are documented on page 43.)
This function evaluates a register. Now a register might exist as one of two things: A parameter-less macro or a built-in TeX register such as \count. For the TeX registers we have to utilize a \the whereas for the macros we merely have to expand them once. The trick is to find out when to use \the and when not to. What we want here is to find out whether the token expands to something else when hit with \exp_after:wN. The technique is to compare the meaning of the token in question when it has been prefixed with \exp_not:N and the token itself. If it is a macro, the prefixed \exp_not:N temporarily turns it into the primitive \scan_stop:

\cs_new:Npn \__exp_eval_register:N #1
\exp_after:wN \if_meaning:w \exp_not:N #1 #1
\__exp_eval_error_msg:w
\fi:

If the token was not a macro it may be a malformed variable from a c expansion in which case it is equal to the primitive \scan_stop:. In that case we throw an error. We could let TeX do it for us but that would result in the rather obscure

! You can’t use ‘\relax’ after \the.

which while quite true doesn’t give many hints as to what actually went wrong. We provide something more sensible.

\cs_new:Npn \__exp_eval_register:c #1
\exp_after:wN \__exp_eval_register:N \cs:w #1 \cs_end:

The next bit requires some explanation. The function must be initiated by \exp:w and we want to terminate this expansion chain by inserting the \exp_end: token. However, we have to expand the register \#1 before we do that. If it is a TeX register, we need to execute the sequence \exp_after:wN \exp_end: \tex_the:D #1 and if it is a macro we need to execute \exp_after:wN \exp_end: \#1. We therefore issue the longer of the two sequences and if the register is a macro, we remove the \tex_the:D.

\cs_new:Npn \__exp_eval_error_msg:w #1 \tex_the:D #2
\fi:
\fi:
\msg_expandable_error:nnn { kernel } { bad-variable } {#2}
\exp_end:

Clean up nicely, then call the undefined control sequence. The result is an error message looking like this:

! Undefined control sequence.
<argument> \LaTeX3 error: Erroneous variable used!
1.55 \tl_set:Nv \l_tmpa_tl {undefined_tl}

(End of definition for \__exp_eval_register:N and \__exp_eval_error_msg:w.)
43.2 Hand-tuned definitions

One of the most important features of these functions is that they are fully expandable.

Here are the functions that turn their argument into csnames but are expandable.

\begin{Verbatim}
\cs_new:Npn \exp_args:NNc #1#2#3
  { \exp_after:wN #1 \exp_after:wN #2 \cs:w #3 \cs_end: }
\cs_new:Npn \exp_args:Ncc #1#2#3
  { \exp_after:wN #1 \cs:w #2 \exp_after:wN \cs_end: \cs:w #3 \cs_end: }
\cs_new:Npn \exp_args:Nccc #1#2#3#4
  { \exp_after:wN #1 \exp_after:wN \cs:w #2 \exp_after:wN \cs_end:
    \exp_after:wN \cs:w #3 \exp_after:wN \cs_end:
    \cs:w #4 \cs_end: }
\end{Verbatim}

Those lovely runs of expansion!

\begin{Verbatim}
\cs_new:Npn \exp_args:No #1#2
  { \exp_after:wN #1 \exp_after:wN {#2} }
\cs_new:Npn \exp_args:NNo #1#2#3
  { \exp_after:wN #1 \exp_after:wN #2 \exp_after:wN {#3} }
\cs_new:Npn \exp_args:NNNo #1#2#3#4
  { \exp_after:wN #1 \exp_after:wN #2 \exp_after:wN #3 \exp_after:wN {#4} }
\end{Verbatim}

When the \texttt{\expanded} primitive is available, use it.

\begin{Verbatim}
\cs_new:Npn \exp_args:Ne #1#2
  { \exp_after:wN #1 \tex_expanded:D { {#2} } }
\end{Verbatim}

\begin{Verbatim}
\cs_new:Npn \exp_args:Nf #1#2
  { \exp_after:wN #1 \exp_after:wN { \exp:w \exp_end_continue_f:w #2 } }
\cs_new:Npn \exp_args:Nv #1#2
  { \exp_after:wN #1 \exp_after:wN \exp:w \__exp_eval_register:c {#2} }
\cs_new:Npn \exp_args:NV #1#2
  { \exp_after:wN #1 \exp_after:wN \exp:w \__exp_eval_register:N #2 }
\end{Verbatim}

When the \texttt{\expanded} primitive is available, use it.

\begin{Verbatim}
\cs_new:Npn \exp_args:Nf #1#2
  { \exp_after:wN #1 \exp_after:wN { \exp:w \exp_end_continue_f:w #2 } }
\cs_new:Npn \exp_args:Nv #1#2
  { \exp_after:wN #1 \exp_after:wN \exp:w \__exp_eval_register:c {#2} }
\cs_new:Npn \exp_args:NV #1#2
  { \exp_after:wN #1 \exp_after:wN \exp:w \__exp_eval_register:N #2 }
\end{Verbatim}
Some more hand-tuned function with three arguments. If we forced that an \texttt{o} argument always has braces, we could implement \texttt{\exp_args:Nco} with less tokens and only two arguments.

\begin{verbatim}
\newcommand{void_document}{
\newcommand{\exp_args:NNV} #1#2#3\exp_args:NNv #1#2#3\exp_args:NNf #1#2#3
\exp_args:Nco #1#2#3\exp_args:NcV #1#2#3\exp_args:Ncv #1#2#3\exp_args:Ncf #1#2#3\exp_args:NVV #1#2#3
\exp_args:NNv #1#2#3
\exp_args:NNf #1#2#3
\exp_args:Nco #1#2#3\exp_args:NcV #1#2#3\exp_args:Ncv #1#2#3\exp_args:Ncf #1#2#3
\exp_args:NVV #1#2#3
\cs_new:Npn #1\exp_args:NNV #1#2#3 #1\exp_after:wN #1\exp_after:wN { \__exp_eval_register:N #3 }
\cs_new:Npn #1\exp_args:NNv #1#2#3 #1\exp_after:wN #1\exp_after:wN { \exp:w \__exp_eval_register:c {#3} }
\cs_new:Npn #1\exp_args:NNe #1#2#3 #1\exp_after:wN #1\exp_after:wN \tex_expanded:D { {#3} }
\cs_new:Npn #1\exp_args:NNf #1#2#3 #1\exp_after:wN #1\exp_after:wN { \exp:w \exp_end_continue_f:w #3 }
\cs_new:Npn #1\exp_args:Nco #1#2#3 #1\exp_after:wN #1\cs:w #2 \exp_after:wN \cs_end: \exp_after:wN {#3}
\cs_new:Npn #1\exp_args:NcV #1#2#3 #1\exp_after:wN #1\cs:w #2 \exp_after:wN \cs_end: \exp_after:wN { \exp:w \__exp_eval_register:N #3 }
\cs_new:Npn #1\exp_args:Ncv #1#2#3 #1\exp_after:wN #1\cs:w #2 \exp_after:wN \cs_end: \exp_after:wN { \exp:w \__exp_eval_register:c {#3} }
\cs_new:Npn #1\exp_args:Ncf #1#2#3 #1\exp_after:wN #1\cs:w #2 \exp_after:wN \cs_end: \exp_after:wN { \exp:w \exp_end_continue_f:w #3 }
\end{verbatim}

411
\exp_after:wN \exp_after:wN \__exp_eval_register:N \exp_after:wN #2 \exp_after:wN \\
\exp_after:wN { \exp:w \__exp_eval_register:N #3 }
}

(End of definition for \exp_args:NNNV and others. These functions are documented on page 37.)

\exp_args:NNNV
\exp_args:NNNv
\exp_args:NNNe
\exp_args:NcNc
\exp_args:NcNo
\exp_args:Ncco

A few more that we can hand-tune.

\cs_new:Npn \exp_args:NNNV #1#2#3#4
\cs_after:wN \__exp_eval_register:N \exp_after:wN #2 \exp_after:wN #3 \\
\exp_after:wN { \exp:w \__exp_eval_register:N #4 }
}

\cs_new:Npn \exp_args:NNNv #1#2#3#4
\cs_after:wN \__exp_eval_register:c {#4} \exp_after:wN \__exp_eval_register:N \exp_after:wN #2 \exp_after:wN #3 \\
\exp_after:wN { \exp:w \__exp_eval_register:N #4 }
}

\cs_new:Npn \exp_args:NNNe #1#2#3#4
\tex_expanded:D { {#4} }
}

\cs_new:Npn \exp_args:NcNc #1#2#3#4
\cs:w #2 \exp_after:wN \cs_end: \cs:w #3 \exp_after:wN \cs_end: \\
\exp_after:wN {#4 }
}

(End of definition for \exp_args:NNNV and others. These functions are documented on page 37.)

\exp_args:Nx
43.3 Last-unbraced versions

There are a few places where the last argument needs to be available unbraced. First some helper macros.

\cs_new:Npn \__exp_arg_last_unbraced:nn #1#2 { #2#1 }
\cs_new:Npn \::o_unbraced \::: #1#2 { \exp_after:wN \__exp_arg_last_unbraced:nn \exp_after:wN {#2} {#1} }
\cs_new:Npn \::V_unbraced \::: #1#2 { \exp_after:wN \__exp_arg_last_unbraced:nn \exp_after:wN { \exp:w \__exp_eval_register:N #2 } {#1} }
\cs_new:Npn \::v_unbraced \::: #1#2 { \exp_after:wN \__exp_arg_last_unbraced:nn \exp_after:wN { \exp:w \exp_end_continue_f:w #2 } {#1} }
\cs_new:Npn \::e_unbraced \::: #1#2 { \tex_expanded:D { \exp_not:n {#1} #2 } }
\cs_new:Npn \::f_unbraced \::: #1#2 { \cs_set_nopar:Npe \l__exp_internal_tl { \exp_not:n {#1} #2 } \l__exp_internal_tl }

(End of definition for \__exp_arg_last_unbraced:nn and others. These functions are documented on page 43.)

Now the business end: most of these are hand-tuned for speed, but the general system is in place.

\cs_new:Npn \exp_last_unbraced:No \exp_last_unbraced:NV \exp_last_unbraced:Nv \exp_last_unbraced:Nf \exp_last_unbraced:NNo \exp_last_unbraced:NNv \exp_last_unbraced:NNV \exp_last_unbraced:NfNo \exp_last_unbraced:NfNo \exp_last_unbraced:NfNNv \exp_last_unbraced:NfNNv \exp_last_unbraced:NfNfNv \exp_last_unbraced:NfNfNfNf \exp_last_unbraced:NfNfNfNfNfNf \exp_last_unbraced:NfNfNfNfNfNfNf

(End of definition for \__exp_arg_last_unbraced:nn and others. These functions are documented on page 43.)
End of definition for \texttt{\_\_exp_last_unbraced:no} and others. These functions are documented on page 39.

If \#2 is a single token then this can be implemented as

\texttt{\_\_exp_last_two_unbraced:no}\texttt{N}
However, for robustness this is not suitable. Instead, a bit of a shuffle is used to ensure that #2 can be multiple tokens.

(End of definition for \exp_last_two_unbraced:Noo and \__exp_last_two_unbraced:nO. This function is documented on page 39.)

### 43.4 Preventing expansion

\exp_not:w

At the kernel level, we need the primitive behaviour to allow expansion before the brace group.

(End of definition for \__kernel_exp_not:w.)

\exp_not:c All these except \exp_not:c call the kernel-internal \__kernel_exp_not:w namely \tex_unexpanded:D.

\exp_not:o \exp_not:e \exp_not:f \exp_not:v

(End of definition for \exp_not:c and others. These functions are documented on page 40.)

### 43.5 Controlled expansion

To trigger a sequence of "arbitrarily" many expansions we need a method to invoke \TeX's expansion mechanism in such a way that (a) we are able to stop it in a controlled manner and (b) the result of what triggered the expansion in the first place is null, i.e., that we do not get any unwanted side effects. There aren't that many possibilities in \TeX; in fact the one explained below might well be the only one (as normally the result of expansion is not null).

The trick here is to make use of the fact that \tex_romannumeral:D expands the tokens following it when looking for a number and that its expansion is null if that number...
turns out to be zero or negative. So we use that to start the expansion sequence: \texttt{\exp:w} is set equal to \texttt{\tex_romannumeral:D} in \texttt{l3basics}. To stop the expansion sequence in a controlled way all we need to provide is a constant integer zero as part of expanded tokens. As this is an integer constant it immediately stops \texttt{\tex_romannumeral:D}'s search for a number. Again, the definition of \texttt{\exp_end:} as the integer constant zero is in \texttt{l3basics}. (Note that according to our specification all tokens we expand initiated by \texttt{\exp:w} are supposed to be expandable (as well as their replacement text in the expansion) so we will not encounter a "number" that actually result in a roman numeral being generated. Or if we do then the programmer made a mistake.)

If on the other hand we want to stop the initial expansion sequence but continue with an \texttt{f}-type expansion we provide the alphabetic constant \texttt{'^'^@} that also represents \texttt{0} but this time \TeX{}'s syntax for a \langle \texttt{number} \rangle continues searching for an optional space (and it continues expansion doing that) — see \TeX{}book page 269 for details.

\begin{verbatim}
\group_begin:
\textcatcode:D '\^^@ = 13
\cs_new_protected:Npn \exp_end_continue_f:w { '^^@ }
\ifcs_exist:N ^^@
\else:
\cs_new:Npn ^^@ { \msg_expandable_error:nn { kernel } { bad-exp-end-f } }
\fi:
\end{verbatim}

The same but grabbing an argument to remove spaces and braces.

\begin{verbatim}
\cs_new:Npn \exp_end_continue_f:nw #1 { '^^@ #1 }
\group_end:
\end{verbatim}

\texttt{(End of definition for \exp:w and others. These functions are documented on page 42.)}

### 43.6 Defining function variants

\begin{verbatim}
\__cs_use_none_delimit_by_s_stop:w \__cs_use_i_delimit_by_s_stop:nw \__cs_use_none_delimit_by_q_recursion_stop:w
\end{verbatim}

\texttt{Internal scan marks. No \texttt{l3quark} yet, so do things by hand.}

\begin{verbatim}
\cs_new_eq:NN \s__cs_mark \scan_stop:
\cs_new_eq:NN \s__cs_stop \scan_stop:
\end{verbatim}

\texttt{(End of definition for \s__cs_mark and \s__cs_stop.)}

\begin{verbatim}
\cs_new:Npn \q__cs_recursion_stop { \q__cs_recursion_stop }
\end{verbatim}

\texttt{Internal recursion quarks. No \texttt{l3quark} yet, so do things by hand.}

\begin{verbatim}
\cs_new:Npn \q__cs_recursion_stop { \q__cs_recursion_stop }
\end{verbatim}

\texttt{(End of definition for \q__cs_recursion_stop.)}
\cs_generate_variant:Nn \cs_generate_variant:cn

#1: Base form of a function; e.g., \tl_set:Nn #2: One or more variant argument specifiers; e.g., \{Nx,c,cx\}

After making sure that the base form exists, test whether it is protected or not and define \cs_tmp:w as either \cs_new:Npe or \cs_new_protected:Npe, which is then used to define all the variants (except those involving x-expansion, always protected). Split up the original base function only once, to grab its name and signature. Then we wish to iterate through the comma list of variant argument specifiers, which we first convert to a string: the reason is explained later.

\begin{verbatim}
\cs_new_protected:Npn \cs_generate_variant:Nn #1#2
\__cs_generate_variant:N #1 \use:e\{\__cs_generate_variant:nnNN \cs_split_function:N #1 \exp_not:N #1 \tl_to_str:n {#2} \exp_not:N \scan_stop: \exp_not:N \q__cs_recursion_stop\}
\end{verbatim}

\begin{verbatim}
\cs_new_protected:Npn \cs_generate_variant:cn \exp_args:Nc \cs_generate_variant:Nn
\end{verbatim}

The goal here is to pick up protected parent functions. There are four cases: the parent function can be a primitive or a macro, and can be expandable or not. For non-expandable primitives, all variants should be protected; skipping the \else: branch is safe because non-expandable primitives cannot be \TeX{} conditionals.

The other case where variants should be protected is when the parent function is a protected macro: then \texttt{protected} appears in the meaning before the first occurrence of \texttt{macro}. The \texttt{ww} auxiliary removes everything in the meaning string after the first \texttt{ma}. We use \texttt{ma} rather than the full \texttt{macro} because the meaning of the \texttt{firstmark} primitive (and four others) can contain an arbitrary string after a leading \texttt{firstmark}: Then, look for \texttt{pr} in the part we extracted: no need to look for anything longer: the only strings we can have are an empty string, \texttt{\long␣}, \texttt{\protected␣}, \texttt{\protected\long␣}, \texttt{\first}, \texttt{\top}, \texttt{\bot}, \texttt{\splittop}, or \texttt{\splitbot}, with \texttt{\protect} replaced by the appropriate escape character. If \texttt{pr} appears in the part before \texttt{ma}, the first \texttt{\s__cs_mark} is taken as an argument of the \texttt{wwNw} auxiliary, and \texttt{#3} is \texttt{\cs_new_protected:Npe}, otherwise it is \texttt{\cs_new:Npe}.

\begin{verbatim}
\cs_new_protected:Npe \__cs_generate_variant:N \#1
\exp_not:N \exp_after:wN \exp_not:N \if_meaning:w \exp_not:N \exp_not:N \cs_set_eq:NN \exp_not:N \__cs_tmp:w \cs_new_protected:Npe \exp_not:N \else:
\exp_not:N \exp_after:wN \exp_not:N \__cs_generate_variant:ww
\exp_not:N \tl_to_str:n \{ \exp_not:N \token_to Meaning:N #1 \tl_to_str:n \{ ma \} \s__cs_mark
\end{verbatim}

(End of definition for \_cs_use_none_delimit_by_s_stop:w, \_cs_use_i_delimit_by_s_stop:nw, and \_cs_use_none_delimit_by_q_recursion_stop:w.)
\s__cs_mark \cs_new_protected:Npe
\tl_to_str:n { pr }
\s__cs_mark \cs_new:Npe
\s__cs_stop
\exp_not:N \fi:
\exp_last_unbraced:NNNNo
\cs_new_protected:Npn \__cs_generate_variant:ww
  #1 { \tl_to_str:n { ma } } #2 \s__cs_mark
  \__cs_generate_variant:wwNw #1
\exp_last_unbraced:NNNNo
\cs_new_protected:Npn \__cs_generate_variant:wwwNw
  #1 { \tl_to_str:n { pr } } #2 \s__cs_stop
  \cs_set_eq:NN \__cs_tmp:w #3 #4 \s__cs_stop
  { \__cs_generate_variant:wwNw #1 }
\__cs_generate_variant:nnNN
#1 : Base name.
#2 : Base signature.
#3 : Boolean.
#4 : Base function.
If the boolean is \c_false_bool, the base function has no colon and we abort with
an error; otherwise, set off a loop through the desired variant forms. The original function
is retained as #4 for efficiency.
\cs_new_protected:Npm \__cs_generate_variant:nnNN #1#2#3#4
{ \if_meaning:w \c_false_bool #3
  \msg_error:nne { kernel } { missing-colon }
  { \token_to_str:c {#1} }
  \exp_after:wN \__cs_use_none_delimit_by_q_recursion_stop:w
  \fi:
  \__cs_generate_variant:Nnnw #4 {#1}{#2}
}(End of definition for \__cs_generate_variant:nnNN, \__cs_generate_variant:w, and \__cs_generate_variant:wNw.)
\__cs_generate_variant:Nnnw
#1 : Base function.
#2 : Base name.
#3 : Base signature.
#4 : Beginning of variant signature.
First check whether to terminate the loop over variant forms. Then, for each variant
form, construct a new function name using the original base name, the variant signature
consisting of \l letters and the last \k \l letters of the base signature (of length \k). For
example, for a base function \prop_put:Nnn which needs a \cv variant form, we want the
new signature to be \cvn.
There are further subtleties:
\begin{itemize}
\item In \cs_generate_variant:Nn \foo:nnTF \xxTF, we must define \foo:xxTF using
  \exp_args:Nxx, rather than a hypothetical \exp_args:NxxTF. Thus, we wish to
  trim a common trailing part from the base signature and the variant signature.
\item In \cs_generate_variant:Nn \foo:on \ox, the function \foo:ox must be defined
  using \exp_args:Nnx, not \exp_args:Nox, to avoid double \o expansion.
\end{itemize}
Lastly, \texttt{cs\_generate\_variant:Nn \foo:on \{xn\}} must trigger an error, because we do not have a means to replace o-expansion by x-expansion. More generally, we can only convert N to c, or convert n to V, v, o, e, f, or x.

All this boils down to a few rules. Only n and N-type arguments can be replaced by \texttt{cs\_generate\_variant:Nn}. Other argument types are allowed to be passed unchanged from the base form to the variant: in the process they are changed to n except for N and p-type arguments. A common trailing part is ignored.

We compare the base and variant signatures one character at a time within e-expansion. The result is given to \texttt{\_\_cs\_generate\_variant:WNWW} (defined later) in the form \texttt{(processed variant signature) \_\_cs\_mark \_\_cs\_stop base function \_\_cs\_stop \_\_cs\_mark \_\_cs\_stop new function}. If all went well, \texttt{\_\_cs\_mark \_\_cs\_stop} is empty; otherwise, it is a kernel error message and some clean-up code.

Note the space after \texttt{#3} and after the following brace group. Those are ignored by \TeX\ when fetching the last argument for \texttt{\_\_cs\_generate\_variant\_loop:nWWN}, but can be used as a delimiter for \texttt{\_\_cs\_generate\_variant\_loop\_end:nWWNNnn}.

\begin{verbatim}
\cs_new_protected:Npn \__cs_generate_variant:Nnnw #1#2#3#4 ,
\__cs_generate_variant_loop:nNwN { } #4
\__cs_generate_variant_loop_end:nwwwNNnn \s__cs_mark #3 ~
\__cs_generate_variant_loop_long:wNNnn } ~ \s__cs_stop
\exp_not:N #1 {#2} {#4}
\__cs_generate_variant:Nnnw #1 {#2} {#3}
\end{verbatim}

(End of definition for \texttt{\_\_cs\_generate\_variant\_Nnnw}.)

\#1: Last few consecutive letters common between the base and variant (more precisely, \texttt{\_\_cs\_generate\_variant\_same:N \{letter\}} for each letter).
\#2: Next variant letter.
\#3: Remainder of variant form.
\#4: Next base letter.

The first argument is populated by \texttt{\_\_cs\_generate\_variant\_loop\_same:w} when a variant letter and a base letter match. It is flushed into the input stream whenever the two letters are different: if the loop ends before, the argument is dropped, which means that trailing common letters are ignored.

The case where the two letters are different is only allowed if the base is N and the variant is c, or when the base is n and the variant is V, v, o, e, f, or x. Otherwise, call \texttt{\_\_cs\_generate\_variant\_loop\_invalid:NNWWNNnn} to remove the end of the loop, get arguments at the end of the loop, and place an appropriate error message as a second
argument of \__cs_generate_variant:wwNN. If the letters are distinct and the base letter is indeed n or N, leave in the input stream whatever argument #1 was collected, and the next variant letter #2, then loop by calling \__cs_generate_variant_loop:nNwN.

The loop can stop in three ways.

- If the end of the variant form is encountered first, #2 is \__cs_generate_variant_loop_end:nwwwNNnn (expanded by the conditional \if:w), which inserts some tokens to end the conditional; grabs the ⟨base name⟩ as #7, the ⟨variant signature⟩ #8, the ⟨next base letter⟩ #1 and the part #3 of the base signature that wasn’t read yet; and combines those into the ⟨new function⟩ to be defined.

- If the end of the base form is encountered first, #4 is \if:w\fi:, which ends the conditional (with an empty expansion), followed by \__cs_generate_variant_loop_long:wnNNnn, which places an error as the second argument of \__cs_generate_variant:wwNN.

- The loop can be interrupted early if the requested expansion is unavailable, namely when the variant and base letters differ and the base is not the right one (n or N to support the variant). In that case too an error is placed as the second argument of \__cs_generate_variant:wwNN.

Note that if the variant form has the same length as the base form, #2 is as described in the first point, and #4 as described in the second point above. The \__cs_generate_variant_loop_end:nwwwNNnn breaking function takes the empty brace group in #4 as its first argument: this empty brace group produces the correct signature for the full variant.

\begin{verbatim}
2773 \cs_new:Npn \__cs_generate_variant_loop:nNwN #1#2#3 \s__cs_mark #4 {    
    \if:w #2 #4 
    \exp_after:wN \__cs_generate_variant_loop_same:w 
    \else: 
    \if:w #4 \__cs_generate_variant_loop_base:N #2 \else:   
        \if:w 0 
            \if:w N #4 \else: \if:w n #4 \else: 1 \fi: \fi: 
            \if:w \scan_stop: \__cs_generate_variant_loop_base:N #2 1 \fi: 0 
        \__cs_generate_variant_loop_special:NNwNNnn #4#2 
    \else:   
        \__cs_generate_variant_loop_invalid:NNwNNnn #4#2 
    \fi: 
    \fi: 
    \fi: 
    \prg_do_nothing: 
    \__cs_generate_variant_loop:nNwN { } \s__cs_mark #3 \s__cs_mark #4 } 
2793 \cs_new:Npn \__cs_generate_variant_loop_base:N #1 
2794 {    
    \if:w c #1 N \else: 
        \if:w o #1 n \else:   
            \if:w V #1 n \else:   
                \if:w v #1 n \else:   

420
\end{verbatim}
\cs_new:Npn \__cs_generate_variant_loop_special:NNwNNnn #1#2#3 \s__cs_stop #4#5#6#7 
\exp_not:n
\msg_error:nnee{ \token_to_str:N #5 } {#1} {#2} 
\__cs_stop #3 {#6} {#7}

(End of definition for \__cs_generate_variant_loop_special:NNwNNnn and others.)

\__cs_generate_variant_same:N
When the base and variant letters are identical, don’t do any expansion. For most argument types, we can use the \_type no-expansion, but the N and p types require a slightly different behaviour with respect to braces. For V-type this function could output N to avoid adding useless braces but that is not a problem.

\cs_new:Npn \__cs_generate_variant_same:N #1
\if:w N #1 #1 \else:
\if:w p #1 #1 \else:
\token_to_str:N n \if:w n #1 
\__cs_generate_variant_loop_special:NNwNNnn #1#1
\fi:
\fi:
\fi:
\fi:

(End of definition for \__cs_generate_variant_same:N.)

\__cs_generate_variant:wwNN
If the variant form has already been defined, log its existence (provided log-functions is active). Otherwise, make sure that the \exp_args:N #3 form is defined, and if it contains x, change \__cs_tmp:w locally to \cs_new_protected:Npe. Then define the variant by combining the \exp_args:N #3 variant and the base function.

\cs_new_protected:Npn \__cs_generate_variant:wwNN #1 \s__cs_mark #2 \s__cs_stop #3#4
\cs_if_free:NT #4
\group_begin:
\__cs_generate_internal_variant:n {#1}
\__cs_generate_internal_variant_loop:n
\exp_args:N #1 \exp_args:N #3
\exp_not:c
\exp_not:N #3
\group_end:

(End of definition for \__cs_generate_variant:wwNN.)

\__cs_generate_internal_variant:n
\__cs_generate_internal_variant_loop:n
First test for the presence of x (this is where working with strings makes our lives easier), as the result should be protected, and the next variant to be defined using that internal variant should be protected (done by setting \__cs_tmp:w). Then
call \__cs_generate_internal_variant:NNn with arguments \cs_new_protected:cpn \use:x (for protected) or \cs_new:cpn \tex_expanded:D (expandable) and the signature. If p appears in the signature, or if the function to be defined is expandable and the primitive \expanded is not available, or if there are more than 8 arguments, call some fall-back code that just puts the appropriate :: commands. Otherwise, call \__cs_generate_internal_one_go:NNn to construct the \exp_args:N... function as a macro taking up to 9 arguments and expanding them using \use:x or \tex_expanded:D.

\begin{verbatim}
2888 \cs_new_protected:Npe \__cs_generate_internal_variant:n #1 { \exp_not:N \__cs_generate_internal_variant:wwnNwn #1 \s__cs_mark { \cs_set_eq:NN \exp_not:N \__cs_tmp:w \cs_new_protected:Npe } \cs_new_protected:cpn \use:x \token_to_str:N x \s__cs_mark { } \cs_new:cpn \exp_not:N \tex_expanded:D \s__cs_stop \s__cs_mark {#1} \exp_last_unbraced:NNNNo \cs_new_protected:Npn \__cs_generate_internal_variant:wwnNwn #1 { \token_to_str:N x } #2 \s__cs_mark #3#4#5#6 \s__cs_stop #7 { \exp_not:N \__cs_generate_internal_variant:wwnNwn #4 #5 \cs_new_protected:Npn \__cs_generate_internal_variant:NNn ##1##2##3 { \if_catcode:w X \use_none:nnnnnnnn ##3 \prg_do_nothing: \prg_do_nothing: \prg_do_nothing: \prg_do_nothing: \prg_do_nothing: X \exp_after:wN \__cs_generate_internal_test:Nw \exp_after:wN ##2 \else: \exp_after:wN \__cs_generate_internal_test_aux:w \exp_after:wN #1 \fi: \s__cs_mark } \s__cs_mark { \use:e { \exp_args:N #3 } { \__cs_generate_internal_variant_loop:n #3 { \use_i:nn } } } \s__cs_mark { \exp_not:n \__cs_generate_internal_one_go:NNn #1 #2 {#3} } }
\end{verbatim}

423
This command grabs char by char outputting `\:` #1 (not expanded further). We avoid tests by putting a trailing `\use_i:nn`, which leaves `\cs_end:` and removes the looping macro. The colon is in fact also turned into `\:::`, so that the required structure for `\exp_args:N...` commands is correctly terminated.

(End of definition for `\__cs_generate_internal_variant:n` and `\__cs_generate_internal_variant_loop:n`.)
\exp_args_generate:n

This function is not used in the kernel hence we can use functions that are defined in later modules. It also does not need to be fast so use inline mappings. For each requested variant we check that there are no characters besides \texttt{NnpcofVvx}, in particular that there are no spaces. Then we just call the internal function.

\exp_args:Nnc
\exp_args:Nno
\exp_args:NnV
\exp_args:Nnv
\exp_args:Nne
\exp_args:Nnf
\exp_args:Noc
\exp_args:Noo
\exp_args:NNx
\exp_args:Ncx
\exp_args:Nnx
\exp_args:Nxx
\exp_args:Nxx

(End of definition for \texttt{\prg_generate_conditional_variant:Nnn} and others. This function is documented on page 65.)

43.7 Definitions with the automated technique

Some of these could be done more efficiently, but the complexity of coding then becomes an issue. Notice that the auto-generated functions actually take no arguments themselves.

Here are the actual function definitions, using the helper functions above. The group is used because \texttt{\_\_cs_generate_internal_variant:n} redefines \texttt{\_\_cs_tmp:w} locally.
43.8 Held-over variant generation

A couple of variants that are from early functions.

\cs_generate_from_arg_count:NNno \cs_replacement_spec:c
\cs_generate_variant:Nn \cs_generate_from_arg_count:NNnn { NNno }
\cs_generate_variant:Nn \cs_replacement_spec:N { c }
(End of definition for \cs_generate_from_arg_count:N\nnn and \cs_replacement_spec:N. These functions are documented on page 19.)

318 ⟨/package⟩
Chapter 44

l3sort implementation

44.1 Variables

\g__sort_internal_seq
\g__sort_internal_tl

Sorting happens in a group; the result is stored in those global variables before being copied outside the group to the proper places. For seq and tl this is more efficient than using \use:e (or some \exp_args:NNe) to smuggle the definition outside the group since \TeX{} does not need to re-read tokens. For clist we don’t gain anything since the result is converted from seq to clist anyways.

\seq_new:N \g__sort_internal_seq
\tl_new:N \g__sort_internal_tl

(\textit{End of definition for }\g__sort_internal_seq \textit{and }\g__sort_internal_tl.)

\l__sort_length_int
\l__sort_min_int
\l__sort_top_int
\l__sort_max_int
\l__sort_true_max_int

The sequence has \l__sort_length_int items and is stored from \l__sort_min_int to \l__sort_top_int − 1. While reading the sequence in memory, we check that \l__sort_top_int remains at most \l__sort_max_int, precomputed by \_sort:-compute_range:. That bound is such that the merge sort only uses \toks registers less than \l__sort_true_max_int, namely those that have not been allocated for use in other code: the user’s comparison code could alter these.

\int_new:N \l__sort_length_int
\int_new:N \l__sort_min_int
\int_new:N \l__sort_top_int
\int_new:N \l__sort_max_int
\int_new:N \l__sort_true_max_int

(\textit{End of definition for }\l__sort_length_int \textit{and others}.)

\l__sort_block_int

Merge sort is done in several passes. In each pass, blocks of size \l__sort_block_int are merged in pairs. The block size starts at 1, and, for a length in the range \([2^k + 1, 2^{k+1})\), reaches \(2^k\) in the last pass.

\int_new:N \l__sort_block_int

(\textit{End of definition for }\l__sort_block_int.)
When merging two blocks, \texttt{\_\_sort_begin_int} marks the lowest index in the two blocks, and \texttt{\_\_sort_end_int} marks the highest index, plus 1.

\begin{verbatim}
\int_new:N \l__sort_begin_int
\int_new:N \l__sort_end_int
(End of definition for \texttt{\_\_sort_begin_int} and \texttt{\_\_sort_end_int}.)
\end{verbatim}

When merging two blocks (whose end-points are \texttt{beg} and \texttt{end}), \texttt{A} starts from the high end of the low block, and decreases until reaching \texttt{beg}. The index \texttt{B} starts from the top of the range and marks the register in which a sorted item should be put. Finally, \texttt{C} points to the copy of the high block in the interval of registers starting at \texttt{\_\_sort_length_int}, upwards. \texttt{C} starts from the upper limit of that range.

\begin{verbatim}
\int_new:N \l__sort_A_int
\int_new:N \l__sort_B_int
\int_new:N \l__sort_C_int
(End of definition for \texttt{\_\_sort_A_int}, \texttt{\_\_sort_B_int}, and \texttt{\_\_sort_C_int}.)
\end{verbatim}

Internal scan marks.

\begin{verbatim}
\scan_new:N \s__sort_mark
\scan_new:N \s__sort_stop
(End of definition for \texttt{s__sort_mark} and \texttt{s__sort_stop}.)
\end{verbatim}

\section*{Finding available \texttt{toks} registers}

After \texttt{\_\_sort_compute_range}: (defined below) determines that \texttt{toks} registers between \texttt{\_\_sort_min_int} (included) and \texttt{\_\_sort_true_max_int} (excluded) have not yet been assigned, \texttt{\_\_sort_shrink_range}: computes \texttt{\_\_sort_max_int} to reflect the need for a buffer when merging blocks in the merge sort. Given $2^n \leq A \leq 2^n + 2^{n-1}$ registers we can sort $\lfloor A/2 \rfloor + 2^{n-2}$ items while if we have $2^n + 2^{n-1} \leq A \leq 2^{n+1}$ registers we can sort $A - 2^{n-1}$ items. We first find out a power $2^n$ such that $2^n \leq A \leq 2^{n+1}$ by repeatedly halving \texttt{\_\_sort_block_int}, starting at $2^{15}$ or $2^{14}$ namely half the total number of registers, then we use the formulas and set \texttt{\_\_sort_max_int}.

\begin{verbatim}
\cs_new_protected:Npn \_\_sort_shrink_range:
\int_set:Nn \l__sort_A_int { \l__sort_true_max_int - \l__sort_min_int + 1 }
\int_set:Nn \l__sort_block_int { \c_max_register_int / 2 }
\_\_sort_shrink_range_loop:
\int_set:Nn \l__sort_max_int \l__sort_block_int
{
\int_compare:nNnTF { \l__sort_block_int * 3 / 2 } > \l__sort_A_int
{
\l__sort_min_int
+ ( \l__sort_A_int - 1 ) / 2 + \l__sort_block_int / 4 - 1

\l__sort_true_max_int - \l__sort_block_int / 2
}\}
}
\end{verbatim}

430
First find out what \toks have not yet been assigned. There are many cases. In \LaTeX\2\epsilon\ with no package, available \toks range from \texttt{\count 15 + 1} to \texttt{\max_register_int} included (this was not altered despite the 2015 changes). When \texttt{\loctoks} is defined, namely in plain (e)\TeX, or when the package \texttt{etex} is loaded in \LaTeX\2\epsilon, redefine \texttt{\sort_compute_range:} to use the range \texttt{\count 265} to \texttt{\count 275 − 1}. The \texttt{elocalloc} package also defines \texttt{\loctoks} but uses yet another number for the upper bound, namely \texttt{\alloc@top} (minus one). We must check for \texttt{\loctoks} every time a sorting function is called, as \texttt{etex} or \texttt{elocalloc} could be loaded.

In Con\TeX\ MkIV the range is from \texttt{\syst_last_allocated_toks+1} to \texttt{\max_register_int}, and in MkII it is from \texttt{\lastallocatedtoks+1} to \texttt{\max_register_int}. In all these cases, call \texttt{\sort_shrink_range:}.

\begin{verbatim}
3146 \cs_new_protected:Npn \__sort_shrink_range_loop:
3147 \{\fi:\}

(End of definition for \texttt{\sort_shrink_range:} and \texttt{\sort_shrink_range_loop:})
\end{verbatim}
\cs_if_exist:NT \loctoks { \__sort_redefine_compute_range: }
\tl_map_inline:nn { \lastallocatedtoks \c_syst_last_allocated_toks }
{
\cs_if_exist:NT #1
{
\cs_gset_protected:Npn \__sort_compute_range:
{
\int_set:Nn \l__sort_min_int { #1 + 1 }
\int_set:Nn \l__sort_true_max_int { \c_max_register_int + 1 }
\__sort_shrink_range:
\}
}
(End of definition for \__sort_compute_range:, \__sort_redefine_compute_range:, and \c__sort_max_length_int.)

\__sort_main:NNNn

Sorting happens in three steps. First store items in \toks registers ranging from \l__sort_min_int to \l__sort_top_int − 1, while checking that the list is not too long. If we reach the maximum length, that’s an error; exit the group. Secondly, sort the array of \toks registers, using the user-defined sorting function: \__sort_level: calls \__sort_compare:nn as needed. Finally, unpack the \toks registers (now sorted) into the target \tl, or into \g__sort_internal_seq for seq and clist. This is done by \__sort_seq:NNNNn and \__sort_tl:NNn.

\cs_new_protected:Npn \__sort_main:NNNn #1#2#3#4
{
\__sort_disable_toksdef:
\__sort_compute_range:
\int_set_eq:NN \l__sort_top_int \l__sort_min_int
#1 #3
{\if_int_compare:w \l__sort_top_int = \l__sort_max_int
\__sort_too_long_error:NNw #2 #3
\fi:
\tex_toks:D \l__sort_top_int {##1}
\int_incr:N \l__sort_top_int
}
\int_set:Nn \l__sort_length_int
{ \l__sort_top_int - \l__sort_min_int }
\cs_set:Npn \__sort_compare:nn {#1} {#4}
\int_set:Nn \l__sort_block_int { 1 }
\__sort_level:
\}
(End of definition for \__sort_main:NNNn.)

\tl_sort:Nn \tl_sort:cn \tl_gsort:Nn \tl_gsort:cn
\__sort_tl:NNn \__sort_tl_toks:w

Call the main sorting function then unpack \toks registers outside the group into the target token list. The unpacking is done by \__sort_tl_toks:w; registers are numbered from \l__sort_min_int to \l__sort_top_int − 1. For expansion behaviour we need
a couple of primitives. The \texttt{\tl_gclear:N} reduces memory usage. The \texttt{\prg_break_point} is used by \texttt{\__sort_main:NNNn} when the list is too long.

\begin{verbatim}
cs_new_protected:Npn \tl_sort:Nn { \__sort_tl:NNn \tl_set_eq:NN }
cs_generate_variant:Nn \tl_sort:Nn { c }
cs_new_protected:Npn \tl_gsort:Nn { \__sort_tl:NNn \tl_gset_eq:NN }
cs_generate_variant:Nn \tl_gsort:Nn { c }
cs_new_protected:Npn \__sort_tl:NNn #1#2#3
\group_begin:
\__sort_main:NNNn \tl_map_inline:Nn \tl_map_break:n #2 {#3}
\__kernel_tl_gset:Nx \g__sort_internal_tl
\{ \__sort_tl_toks:w \_\_sort_min_int ; \}
\group_end:
#1 #2 \g__sort_internal_tl
\tl_gclear:N \g__sort_internal_tl
\prg_break_point:
\cs_new:Npn \__sort_tl_toks:w #1 ;
\{ 
\if_int_compare:w #1 < \_\_sort_top_int
\{ \tex_the:D \tex_toks:D #1 \}
\exp_after:wN \__sort_tl_toks:w
\int_value:w \int_eval:n { #1 + 1 } \exp_after:wN ;
\fi:
\}
\end{verbatim}
(End of definition for \texttt{\tl_sort:Nn} and others. These functions are documented on page \pageref{page}.)

Use the same general framework for \texttt{\seq_sort:Nn} and others. Apply the general sorting code, then unpack \texttt{\toks} into \texttt{\_\_sort_internal_seq}. Outside the group copy or convert (for clist) the data to the target variable. The \texttt{\seq_gclear:N} reduces memory usage. The \texttt{\prg_break_point} is used by \texttt{\__sort_main:NNNn} when the list is too long.

\begin{verbatim}
cs_new_protected:Npn \seq_sort:Nn { \__sort_seq:NNNNn \seq_map_inline:Nn \seq_map_break:n \seq_set_eq:NN }
cs_generate_variant:Nn \seq_sort:Nn { c }
cs_new_protected:Npn \seq_gsort:Nn { \__sort_seq:NNNNn \seq_map_inline:Nn \seq_map_break:n \seq_gset_eq:NN }
cs_generate_variant:Nn \seq_gsort:Nn { c }
cs_new_protected:Npn \clist_sort:Nn
\{ \__sort_seq:NNNNn \clist_map_inline:Nn \clist_map_break:n
\clist_set_from_seq:NN
\}
cs_generate_variant:Nn \clist_sort:Nn { c }
cs_new_protected:Npm \clist_gsort:Nn
\{ \__sort_seq:NNNNn \clist_map_inline:Nn \clist_map_break:n
\clist_gset_from_seq:NN
\}
cs_generate_variant:Nn \clist_gsort:Nn { c }
cs_new_protected:Npm \_\_sort_seq:NNNNn #1#2#3#4#5
\group_begin:
\_\_sort_main:NNNn #1 #2 #4 #5
\end{verbatim}

433
44.4 Merge sort

\__sort_level: This function is called once blocks of size \l__sort_block_int (initially 1) are each sorted. If the whole list fits in one block, then we are done (this also takes care of the case of an empty list or a list with one item). Otherwise, go through pairs of blocks starting from 0, then double the block size, and repeat.

\cs_new_protected:Npn \__sort_level:
{
\if_int_compare:w \l__sort_block_int < \l__sort_length_int
\l__sort_end_int \l__sort_min_int
\__sort_merge_blocks:
\tex_advance:D \l__sort_end_int \l__sort_block_int
\exp_after:wN \__sort_level:
\fi:
}

(End of definition for \seq_sort:Nn and others. These functions are documented on page 157.)

\__sort_merge_blocks: This function is called to merge a pair of blocks, starting at the last value of \l__sort_end_int (end-point of the previous pair of blocks). If shifting by one block to the right we reach the end of the list, then this pass has ended: the end of the list is sorted already. Otherwise, store the result of that shift in \A, which indexes the first block starting from the top end. Then locate the end-point (maximum) of the second block: shift end upwards by one more block, but keeping it \leq top. Copy this upper block of \toks registers in registers above length, indexed by \C: this is covered by \__sort_copy_block:. Once this is done we are ready to do the actual merger using \__sort_merge_blocks_aux:. After shifting \A, \B and \C so that they point to the largest index in their respective ranges rather than pointing just beyond those ranges. Of course, once that pair of blocks is merged, move on to the next pair.

\cs_new_protected:Npn \__sort_merge_blocks:
{
\l__sort_begin_int \l__sort_end_int
\tex_advance:D \l__sort_end_int \l__sort_block_int
\if_int_compare:w \l__sort_end_int < \l__sort_top_int
\l__sort_A_int \l__sort_end_int
\tex_advance:D \l__sort_end_int \l__sort_block_int
\if_int_compare:w \l__sort_end_int > \l__sort_top_int
}
\_\_sort_end_int \_\_sort_top_int
\fi:
\_\_sort_B_int \_\_sort_A_int
\_\_sort_C_int \_\_sort_top_int
\_\_sort_copy_block:
\int_decr:N \_\_sort_A_int
\int_decr:N \_\_sort_B_int
\int_decr:N \_\_sort_C_int
\exp_after:wN \_\_sort_merge_blocks_aux:
\exp_after:wN \_\_sort_merge_blocks:
\fi:
\}

\_\_sort_copy_block: We wish to store a copy of the “upper” block of \toks registers, ranging between the initial value of \_\_sort_B_int (included) and \_\_sort_end_int (excluded) into a new range starting at the initial value of \_\_sort_C_int, namely \_\_sort_top_int.
\cs_new_protected:Npn \_\_sort_copy_block:
\{ \tex_toks:D \_\_sort_C_int \tex_toks:D \_\_sort_B_int
\int_incr:N \_\_sort_C_int
\int_incr:N \_\_sort_B_int
\if_int_compare:w \_\_sort_B_int = \_\_sort_end_int \fi:
\_\_sort_copy_block:
\}

\_\_sort_merge_blocks_aux: At this stage, the first block starts at \_\_sort_begin_int, and ends at \_\_sort_A_int, and the second block starts at \_\_sort_top_int and ends at \_\_sort_C_int. The result of the merger is stored at positions indexed by \_\_sort_B_int, which starts at \_\_sort_end_int – 1 and decreases down to \_\_sort_begin_int, covering the full range of the two blocks. In other words, we are building the merger starting with the largest values. The comparison function is defined to return either \textit{swapped} or \textit{same}. Of course, this means the arguments need to be given in the order they appear originally in the list.
\cs_new_protected:Npn \_\_sort_merge_blocks_aux:
\{ \exp_after:wN \_\_sort_compare:nn \exp_after:wN
\{ \tex_the:D \tex_toks:D \exp_after:wN \_\_sort_A_int \exp_after:wN \}
\exp_after:wN \{ \tex_the:D \tex_toks:D \_\_sort_C_int \}
\prg_do_nothing:
\_\_sort_return_mark:w
\_\_sort_return_mark:w
\_\_sort_return_none_error:
\}

(End of definition for \_\_sort_merge_blocks_aux:)
Each comparison should call \sort_return_same: or \sort_return_swapped: exactly once. If neither is called, \__sort_return_none_error: is called, since the return_mark removes tokens until \s__sort_mark. If one is called, the return_mark auxiliary removes everything except \__sort_return_same:w (or its swapped analogue) followed by \__sort_return_none_error:. Finally if two or more are called, \__sort_return_two_error: ends up before any \__sort_return_mark:w, so that it produces an error.

\sort_return_same:
\sort_return_swapped:
\__sort_return_mark:w
\__sort_return_none_error:
\__sort_return_two_error:

If the comparison function returns same, then the second argument fed to \__sort_compare:nn should remain to the right of the other one. Since we build the merger starting from the right, we copy that \toks register into the allotted range, then shift the pointers \texttt{B} and \texttt{C}, and go on to do one more step in the merger, unless the second block has been exhausted: then the remainder of the first block is already in the correct registers and we are done with merging those two blocks.

\__sort_return_same:w
3363 \int_decr:N \l__sort_C_int
3364 \if_int_compare:w \l__sort_C_int < \l__sort_top_int
3365 \use_i:nn
3366 \fi:
3367 \__sort_merge_blocks_aux:
3368 )

(End of definition for \__sort_return_same:w)

\__sort_return_swapped:w If the comparison function returns swapped, then the next item to add to the merger is
the first argument, contents of the \toks register \texttt{A}. Then shift the pointers \texttt{A} and \texttt{B}
to the left, and go for one more step for the merger, unless the left block was exhausted
(\texttt{A} goes below the threshold). In that case, all remaining \toks registers in the second
block, indexed by \texttt{C}, are copied to the merger by \__sort_merge_blocks_end:.
3379 \cs_new_protected:Npn \__sort_return_swapped:w #1 \__sort_return_none_error:
3380 { \text_toks:D \l__sort_B_int \text_toks:D \l__sort_A_int
3381 \int_decr:N \l__sort_B_int
3382 \int_decr:N \l__sort_A_int
3383 \if_int_compare:w \l__sort_A_int < \l__sort_begin_int
3384 \__sort_merge_blocks_end: \use_i:nn
3385 \fi:
3386 \__sort_merge_blocks_aux:
3387 )

(End of definition for \__sort_return_swapped:w)

\__sort_merge_blocks_end: This function’s task is to copy the \toks registers in the block indexed by \texttt{C}
to the merger indexed by \texttt{B}. The end can equally be detected by checking when \texttt{B}
reaches the threshold \texttt{begin}, or when \texttt{C} reaches \texttt{top}.
3370 \cs_new_protected:Npm \__sort_merge_blocks_end:
3371 { \text_toks:D \l__sort_B_int \text_toks:D \l__sort_A_int
3372 \int_decr:N \l__sort_B_int
3373 \int_decr:N \l__sort_A_int
3374 \if_int_compare:w \l__sort_A_int < \l__sort_begin_int
3375 \__sort_merge_blocks_end: \use_i:nn
3376 \fi:
3377 \__sort_merge_blocks_aux:
3378 )

(End of definition for \__sort_merge_blocks_end:)

44.5 Expandable sorting

Sorting expandably is very different from sorting and assigning to a variable. Since tokens
cannot be stored, they must remain in the input stream, and be read through at every
step. It is thus necessarily much slower (at best $O(n^2 \ln n)$) than non-expandable sorting
functions ($O(n \ln n)$).

A prototypical version of expandable quicksort is as follows. If the argument has no
item, return nothing, otherwise partition, using the first item as a pivot (argument \#4 of
\__sort:nnNnn). The arguments of \__sort:nnNnn are 1. items less than \#4, 2. items
greater or equal to \#4, 3. comparison, 4. pivot, 5. next item to test. If \#5 is the tail of
the list, call \texttt{\_\_sort:nn} on \#1 and on \#2, placing \#4 in between; \texttt{use:ff} expands the parts to make \texttt{\_\_sort:nn} f-expandable. Otherwise, compare \#4 and \#5 using \#3. If they are ordered, place \#5 amongst the “greater” items, otherwise amongst the “lesser” items, and continue partitioning.

\begin{verbatim}
\cs_new:Npn \_\_sort:nn #1#2
{\__sort:nnNnn { } { } #2
 #1 \q__sort_recursion_tail \q__sort_recursion_stop}
\end{verbatim}

There are quite a few optimizations available here: the code below is less legible, but more than twice as fast. In the simple version of the code, \texttt{\_\_sort:nnNnn} is called $O(n \ln n)$ times on average (the number of comparisons required by the quicksort algorithm). Hence most of our focus is on optimizing that function.

The first speed up is to avoid testing for the end of the list at every call to \texttt{\_\_sort:nnNnn}. For this, the list is prepared by changing each \langle item \rangle of the original token list into \langle command \rangle \langle item \rangle, just like sequences are stored. We arrange things such that the \langle command \rangle is the \langle conditional \rangle provided by the user: the loop over the \langle prepared tokens \rangle then looks like

\begin{verbatim}
\cs_new:Npn \_\_sort_loop:wNn ... #6#7
{ #6 \langle pivot \rangle \{ #7 \langle loop big \rangle \langle loop small \rangle \langle extra arguments \rangle \}
 \_\_sort_loop:wNn ... \langle prepared tokens \rangle
 \end{loop} \{ \} \s__sort_stop
\end{verbatim}

In this example, which matches the structure of \texttt{\_\_sort_quick_split_i:NnnnnNn} and a few other functions below, the \texttt{\_\_sort_loop:wNn} auxiliary normally receives the user’s \langle conditional \rangle as \#6 and an \langle item \rangle as \#7. This is compared to the \langle pivot \rangle (the argument \#5, not shown here), and the \langle conditional \rangle leaves the \langle loop big \rangle or \langle loop small \rangle auxiliary, which both have the same form as \texttt{\_\_sort_loop:wNn}, receiving the next pair \langle conditional \rangle \langle item \rangle as \#6 and \#7. At the end, \#6 is the \langle end-loop \rangle function, which terminates the loop.

The second speed up is to minimize the duplicated tokens between the \texttt{true} and \texttt{false} branches of the conditional. For this, we introduce two versions of \texttt{\_\_sort:nnNnn},
which receive the new item as #1 and place it either into the list #2 of items less than
the pivot #4 or into the list #3 of items greater or equal to the pivot.

\cs_new:Npn \__sort_i:nnnnNn #1#2#3#4#5#6
{  
  #5 {#4} {#6} \__sort_ii:nnnnNn \__sort_i:nnnnNn  
  {#6} { #2 {#1} } {#3} {#4}  
}
\cs_new:Npn \__sort_ii:nnnnNn #1#2#3#4#5#6
{  
  #5 {#4} {#6} \__sort_ii:nnnnNn \__sort_i:nnnnNn  
  {#6} {#2} { #3 {#1} } {#4}  
}

Note that the two functions have the form of \__sort_loop:wNn above, receiving as #5
the conditional or a function to end the loop. In fact, the lists #2 and #3 must be made
of pairs ⟨conditional⟩{⟨item⟩}, so we have to replace {#6} above by { #5 {#6} }, and
{#1} by #1. The actual functions have one more argument, so all argument numbers are
shifted compared to this code.

The third speed up is to avoid \use:ff using a continuation-passing style:
\__sort_quick_split:NnNn expects a list followed by \s__sort_mark {⟨code⟩}, and
expands to ⟨code⟩⟨sorted list⟩. Sorting the two parts of the list around the pivot is
done with

\__sort_quick_split:NnNn #2 ... \s__sort_mark  
{  
\__sort_quick_split:NnNn #1 ... \s__sort_mark {⟨code⟩}  
{⟨pivot⟩}  
}

Items which are larger than the ⟨pivot⟩ are sorted, then placed after code that sorts the
smaller items, and after the (braced) ⟨pivot⟩.

The fourth speed up is avoid the recursive call to \tl_sort:nN with an empty first
argument. For this, we introduce functions similar to the \__sort_i:nnnnNn of the last
example, but aware of whether the list of ⟨conditional⟩{⟨item⟩} read so far that are
less than the pivot, and the list of those greater or equal, are empty or not: see \__sort-
quick_split:NnNn and functions defined below. Knowing whether the lists are empty
or not is useless if we do not use distinct ending codes as appropriate. The splitting
auxiliaries communicate to the ⟨end-loop⟩ function (that is initially placed after the
“prepared” list) by placing a specific ending function, ignored when looping, but useful
at the end. In fact, the ⟨end-loop⟩ function does nothing but place the appropriate
ending function in front of all its arguments. The ending functions take care of sorting
non-empty sublists, placing the pivot in between, and the continuation before.

The final change in fact slows down the code a little, but is required to avoid memory
issues: schematically, when \TeX{} encounters

\use:n \use:n \use:n { \use:n { ... } ... } ...  

the argument of the first \use:n is not completely read by the second \use:n, hence
must remain in memory; then the argument of the second \use:n is not completely read
when grabbing the argument of the third \use:n, hence must remain in memory, and so
on. The memory consumption grows quadratically with the number of nested \use:n. In
practice, this means that we must read everything until a trailing \texttt{\_sort_stop} once in a while, otherwise sorting lists of more than a few thousand items would exhaust a typical \TeX's memory.

The code within the \texttt{\exp_not:f} sorts the list, leaving in most cases a leading \texttt{\exp_not:n}, which stops the expansion, letting the result be return within \texttt{\exp_not:n}. We filter out the case of a list with no item, which would otherwise cause problems. Then prepare the token list \texttt{\_sort_stop} by inserting the conditional \texttt{\_sort_prepare:Nnnn} before each item. The \texttt{\prepare} auxiliary receives the conditional as \texttt{\_sort_prepare:Nnnn}, the prepared token list so far as \texttt{\_sort_prepare_end:NNNw}, and the item after that as \texttt{\_sort_stop}. The loop ends when \texttt{\_sort_stop} contains \texttt{\prg_break_point:}, then the \texttt{\prepare_end} auxiliary finds the prepared token list as \texttt{\_sort_stop}. The scene is then set up for \texttt{\_sort_split:NnNn}, which sorts the prepared list and perform the post action placed after \texttt{\_sort_mark}, namely removing the trailing \texttt{\_sort_stop} and leaving \texttt{\exp_stop_f:} to stop f-expansion.

\begin{verbatim}
\cs_new:Npn \tl_sort:nN #1#2
{ \exp_not:f
  { \tl_if_blank:nF {#1}
    { \_sort_form:nNnnn #2 { } { } #1
      { \prg_break_point: \_sort_form_end:NNNw } \_sort_form_stop
    } \_sort_form_stop
  }
}
\cs_new:Npn \_sort_form:nNnnn #1#2#3#4
{ \prg_break: #4 \prg_break_point:
  \_sort_form:nNnnn #1 { #2 #3 } { #1 {#4} } #1
}
\cs_new:Npn \_sort_form_end:NNNw #1#2#3
{ \_sort_form_split:NnnN \_sort_form_end:wnnNn { #2 \_sort_form_clean_up: \_sort_stop_f: }
  \_sort_form_clean_up: \_sort_stop #1
}
\cs_new:Npn \_sort_form_clean_up:w #1 \_sort_form_clean_up \_sort_stop {#1}
\end{verbatim}

The \texttt{\_sort_form:} and \texttt{\_sort_form_split:} auxiliaries receive a useless first argument, the new item \texttt{#2} (that they append to either one of the next two arguments), the list \texttt{\_sort_form:} of items less than the pivot, bigger items \texttt{\_sort_form:}, the pivot \texttt{\_sort_form:}, and an item \texttt{\_sort_form:}. The \texttt{\_sort_form:} is the user's \texttt{\_sort_form:} except at the end of the list where it is \texttt{\_sort_form:} and \texttt{\_sort_form:}. The comparison is applied to the \texttt{\_sort_form:} and the \texttt{\_sort_form:}, and calls the \texttt{\_sort_form:} or \texttt{\_sort_form:} auxiliaries if the \texttt{\_sort_form:} is smaller, and the \texttt{\_sort_form:} or \texttt{\_sort_form:} auxiliaries otherwise. In both cases, the next auxiliary goes to work right away, with no intermediate expansion that would slow down operations. Note that the argument \texttt{\_sort_form:} left for the next call has the form \texttt{\_sort_form:}, so that the lists \texttt{\_sort_form:} and \texttt{\_sort_form:} keep the right form to be fed to the next sorting function. The \texttt{\_sort_form:} differs from these in that it is missing three of the arguments, which would be

\texttt{\_sort_form:}
empty, and its first argument is always the user’s \&lt;\textit{conditional} \&gt; rather than an ending function.

\begin{verbatim}
\cs_new:Npn \__sort_quick_split:NnNn #1#2#3#4
\{ #3 {#2} {#4} \__sort_quick_only_ii:NnnnNn \__sort_quick_single_end:nnnwnw
 \{ #3 {#4} \} \{ \} \{ \} #2 \}
\cs_new:Npn \__sort_quick_only_i:NnnnNn #1#2#3#4#5#6#7
\{ #6 {#5} {#7} \__sort_quick_split_ii:NnnnNn \__sort_quick_only_i_end:nnnwnw
 \{ #6 {#7} \} \{ #3 #2 \} \{ \} #5 \}
\cs_new:Npn \__sort_quick_only_ii:NnnnNn #1#2#3#4#5#6#7
\{ #6 {#5} {#7} \__sort_quick_split_i:NnnnNn \__sort_quick_only_ii_end:nnnwnw
 \{ #6 {#7} \} \{ \} #4 #2 \} #5 \}
\cs_new:Npn \__sort_quick_split_i:NnnnNn #1#2#3#4#5#6#7
\{ #6 {#5} {#7} \__sort_quick_split_ii:NnnnNn \__sort_quick_split_i_end:nnnwnw
 \{ #6 {#7} \} \{ #3 #2 \} {#4} \} #5 \}
\cs_new:Npn \__sort_quick_split_ii:NnnnNn #1#2#3#4#5#6#7
\{ #6 {#5} {#7} \__sort_quick_split_i:NnnnNn \__sort_quick_split_end:nnnwnw
 \{ #6 {#7} \} \{ #3 \} \{ #4 #2 \} \{ #5 \}
\end{verbatim}

\textit{(End of definition for \__sort_quick_split:NnNn and others.)}

The \__sort_quick_end:nnTFNn appears instead of the user’s conditional, and receives as its arguments the pivot \#1, a fake item \#2, a true and a false branches \#3 and \#4, followed by an ending function \#5 (one of the four auxiliaries here) and another copy \#6 of the fake item. All those are discarded except the function \#5. This function receives lists \#1 and \#2 of items less than or greater than the pivot \#3, then a continuation code \#5 just after \s__sort_mark. To avoid a memory problem described earlier, all of the ending functions read \#6 until \s__sort_stop and place \#6 back into the input stream. When the lists \#1 and \#2 are empty, the single auxiliary simply places the continuation \#5 before the pivot \{#3}. When \#2 is empty, \#1 is sorted and placed before the pivot \{#3), taking care to feed the continuation \#5 as a continuation for the function sorting \#1. When \#1 is empty, \#2 is sorted, and the continuation argument is used to place the continuation \#5 and the pivot \{#3\} before the sorted result. Finally, when both
lists are non-empty, items larger than the pivot are sorted, then items less than the pivot, and the continuations are done in such a way to place the pivot in between.

\[ \text{End of definition for } \texttt{\_\_sort\_quick\_end:nnTFNn} \text{ and others.} \]

\subsection*{44.6 Messages}

\texttt{\_\_sort\_error}: Bailing out of the sorting code is a bit tricky. It may not be safe to use a delimited argument, so instead we redefine many \texttt{l3sort} commands to be trivial, with \texttt{\_\_sort\_level}: jumping to the break point. This error recovery won’t work in a group.

\[ \text{End of definition for } \texttt{\_\_sort\_error:} \]

\texttt{\_\_sort\_disable\_toksdef}: While sorting, \texttt{\toksdef} is locally disabled to prevent users from using \texttt{\newtoks} or similar commands in their comparison code: the \texttt{\toks} registers that would be assigned are in use by \texttt{l3sort}. In format mode, none of this is needed since there is no \texttt{\toks} allocator.

\[ \text{End of definition for } \texttt{\_\_sort\_disable\_toksdef:} \]
\msg_error:nne { sort } { toksdef }
\{ \token_to_str:N #1 \}
\__sort_error:
\tex_toksdef:D #1
\}
\msg_new:nnnn { sort } { toksdef }
\{ Allocation-of-\iow_char:N\toks-registers-impossible-while-sorting. \}
\{ The-comparison-code-used-for-sorting-a-list-has-attempted-to-
define-#1-as-a-new-\iow_char:N\toks-register-using-
\iow_char:N\newtoks-
or-a-similar-command.-The-list-will-not-be-sorted.
\}
\msg_new:nnnn { sort } { too-large }
\{ The-list-#1-is-too-long-to-be-sorted-by-TeX. \}
\{ TeX-has-#2-toks-registers-still-available:-
this-only-allows-to-sort-with-up-to-#3-
items.-The-list-will-not-be-sorted.
\}
\msg_new:nnnn { sort } { return-none }
\{ The-comparison-code-did-not-return. \}
\{ When-sorting-a-list,-the-code-to-compare-items-#1-and-#2-
did-not-call-
\iow_char:N\sort_return_same: -nor-
\iow_char:N\sort_return_swapped: .-
Exactly-one-of-these-should-be-called.
\}
\msg_new:nnnn { sort } { return-two }
\{ The-comparison-code-returned-multiple-times. \}
\{ When-sorting-a-list,-the-code-to-compare-items-#1-and-#2-called-
\iow_char:N\sort_return_same: -or-
\iow_char:N\sort_return_swapped: -multiple-times.-
Exactly-one-of-these-should-be-called.
\}
\prop_gput:Nnn \g_msg_module_name_prop { sort } { LaTeX }
\prop_gput:Nnn \g_msg_module_type_prop { sort } { }
Chapter 45

l3tl-analysis implementation

45.1 Internal functions

\s__tl  The format used to store token lists internally uses the scan mark \s__tl as a delimiter.  
(End of definition for \s__tl.)

45.2 Internal format

The task of the l3tl-analysis module is to convert token lists to an internal format which allows us to extract all the relevant information about individual tokens (category code, character code), as well as reconstruct the token list quickly. This internal format is used in l3regex where we need to support arbitrary tokens, and it is used in conversion functions in l3str-convert, where we wish to support clusters of characters instead of single tokens.

We thus need a way to encode any ⟨token⟩ (even begin-group and end-group character tokens) in a way amenable to manipulating tokens individually. The best we can do is to find ⟨tokens⟩ which both o-expand and e/x-expand to the given ⟨token⟩. Collecting more information about the category code and character code is also useful for regular expressions, since most regexes are catcode-agnostic. The internal format thus takes the form of a succession of items of the form

⟨tokens⟩ \s__tl ⟨catcode⟩ ⟨char code⟩ \s__tl

The ⟨tokens⟩ o- and e/x-expand to the original token in the token list or to the cluster of tokens corresponding to one Unicode character in the given encoding (for l3str-convert). The ⟨catcode⟩ is given as a single hexadecimal digit, 0 for control sequences. The ⟨char code⟩ is given as a decimal number, −1 for control sequences.

Using delimited arguments lets us build the ⟨tokens⟩ progressively when doing an encoding conversion in l3str-convert. On the other hand, the delimiter \s__tl may not appear unbraced in ⟨tokens⟩. This is not a problem because we are careful to wrap control sequences in braces (as an argument to \exp_not:n) when converting from a general token list to the internal format.

The current rule for converting a ⟨token⟩ to a balanced set of ⟨tokens⟩ which both o-expands and e/x-expands to it is the following.
A control sequence \textbackslash cs becomes \texttt{\exp_not:n \{ \textbackslash cs \} \_\_tl 0 -1 \_\_tl}.

A begin-group character \{ becomes \texttt{\exp_after:wN \{ \texttt{\if_false: \fi: \_\_tl 1 \langle \texttt{char code} \rangle \_\_tl}}.

An end-group character } becomes \texttt{\texttt{\if_false: \fi: \_\_tl 2 \langle \texttt{char code} \rangle \_\_tl}}.

A character with any other category code becomes \texttt{\exp_not:n \{ \langle \texttt{character} \rangle \} \_\_tl \langle \texttt{hex catcode} \rangle \langle \texttt{char code} \rangle \_\_tl}.

In contrast, for \texttt{\peek_analysis_map_inline:n} we must allow for an input stream containing \texttt{\outer} macros, so that wrapping all control sequences in \texttt{\exp_not:n} is unsafe. Instead, we write the more elaborate \texttt{\_\_kernel_exp_not:w \exp_after:wN \{ \exp_not:N \cs \}}. (On the other hand we make a better effort by avoiding \texttt{\exp_not:n} for characters other than active and macro parameters.)

\section{Variables and helper functions}

\texttt{\_\_tl} The scan mark \texttt{\_\_tl} is used as a delimiter in the internal format. This is more practical than using a quark, because we would then need to control expansion much more carefully: compare \texttt{\int_value:w '#1 \_\_tl} with \texttt{\int_value:w '#1 \exp_stop_f: \exp_not:N \q_mark} to extract a character code followed by the delimiter in an e-expansion.

\texttt{\_\_tl_analysis_token} The tokens in the token list are probed with the \TeX{} primitive \texttt{\futurelet}. We use \texttt{\_\_tl_analysis_token} in that construction. In some cases, we convert the following token to a string before probing it: then the token variable used is \texttt{\_\_tl_analysis_char_token}.

\texttt{\_\_tl_peek_code_tl} Holds some code to be run once the next token has been fully analysed in \texttt{\peek_analysis_map_inline:n}.

\texttt{\_\_tl_catcodes_tl} A token list containing the character number 32 (space) with all possible category codes except 1 and 2 (begin-group and end-group). Why 32? Because some Lua\TeX{} versions only allow creation of catcode 10 (space) tokens with this character code, so that we decided to make \texttt{\char_generate:nn} refuse to create such weird spaces as well. We do not include the macro parameter case (catcode 6) because it cannot be used as a macro delimiter.
\l__tl_analysis_normal_int  The number of normal (N-type argument) tokens since the last special token.
\l__tl_analysis_index_int  During the first pass, this is the index in the array being built. During the second pass, it is equal to the maximum index in the array from the first pass.
\l__tl_analysis_nesting_int  Nesting depth of explicit begin-group and end-group characters during the first pass. This lets us detect the end of the token list without a reserved end-marker.
\l__tl_analysis_type_int  When encountering special characters, we record their “type” in this integer.
\g__tl_analysis_result_tl  The result of the conversion is stored in this token list, with a succession of items of the form
\tl_extract_charcode:  Extracting the character code from the meaning of \l__tl_analysis_token. This has no error checking, and should only be assumed to work for begin-group and end-group character tokens. It produces a number in the form ‘\langle char\rangle’.
Counts the number of spaces in the string representation of its second argument, as well as the number of characters following the last space in that representation, and feeds the two numbers as semicolon-delimited arguments to the first argument. When this function is used, the escape character is printable and non-space.

\cs_new:Npn \__tl_analysis_cs_space_count:NN \#1 \#2
\exp_after:wN \#1
\int_value:w \int_eval:w 0
\exp_after:wN \__tl_analysis_cs_space_count:w
\token_to_str:N \#2
\fi: \__tl_analysis_cs_space_count_end:w ; ~ !
\cs_new:Npn \__tl_analysis_cs_space_count:w \#1 ~
\if_false: \#1 \#1 \fi: + 1
\__tl_analysis_cs_space_count:w
\cs_new:Npn \__tl_analysis_cs_space_count_end:w ; \#1 \fi: \#2 !
\exp_after:wN ; \int_value:w \str_count_ignore_spaces:n \#1 \} \}
\cs_new:Npn \__tl_analysis_cs_space_count:NN, \__tl_analysis_cs_space_count:w, and \__tl_analysis_cs_space_count_end:w\

45.4 Plan of attack

Our goal is to produce a token list of the form roughly

\langle \text{token 1} \rangle \text{$_{\text{s__tl}}$} \langle \text{catcode 1} \rangle \langle \text{char code 1} \rangle \text{$_{\text{s__tl}}$} \\
\langle \text{token 2} \rangle \text{$_{\text{s__tl}}$} \langle \text{catcode 2} \rangle \langle \text{char code 2} \rangle \text{$_{\text{s__tl}}$} \\
\cdots \langle \text{token N} \rangle \text{$_{\text{s__tl}}$} \langle \text{catcode N} \rangle \langle \text{char code N} \rangle \text{$_{\text{s__tl}}$} \\

Most but not all tokens can be grabbed as an undelimited (N-type) argument by \TeX. The plan is to have a two pass system. In the first pass, locate special tokens, and store them in various \toks registers. In the second pass, which is done within an e-expanding assignment, normal tokens are taken in as N-type arguments, and special tokens are retrieved from the \toks registers, and removed from the input stream by some means. The whole process takes linear time, because we avoid building the result one item at a time.

We make the escape character printable (backslash, but this later oscillates between slash and backslash): this allows us to distinguish characters from control sequences.

A token has two characteristics: its \text{\textbackslash meaning}, and what it looks like for \TeX when it is in scanning mode (e.g., when capturing parameters for a macro). For our purposes, we distinguish the following meanings:

- begin-group token (category code 1), either space (character code 32), or non-space;
- end-group token (category code 2), either space (character code 32), or non-space;
- space token (category code 10, character code 32);
• anything else (then the token is always an N-type argument).

The token itself can “look like” one of the following:

• a non-active character, in which case its meaning is automatically that associated to its character code and category code, we call it “true” character;
• an active character;
• a control sequence.

The only tokens which are not valid N-type arguments are true begin-group characters, true end-group characters, and true spaces. We detect those characters by scanning ahead with \futurelet, then distinguishing true characters from control sequences set equal to them using the \string representation.

The second pass is a simple exercise in expandable loops.

\__tl_analysis:n
Everything is done within a group, and all definitions are local. We use \group_align_safe_begin/end: to avoid problems in case \_tl_analysis:n is used within an alignment and its argument contains alignment tab tokens.

\cs_new_protected:Npn \__tl_analysis:n #1
\group_begin:
\group_align_safe_begin:
\_tl_analysis_a:n {#1}
\_tl_analysis_b:n {#1}
\group_align_safe_end:
\group_end:
\}

(End of definition for \_tl_analysis:n.)

45.5 Disabling active characters

Active characters can cause problems later on in the processing, so we provide a way to disable them, by setting them to undefined. Since Unicode contains too many characters to loop over all of them, we instead do this whenever we encounter a character. For \pdfTeX and \upTeX we skip characters beyond [0,255] because \lccode only allows those values.

\cs_gset_protected:Npn \_tl_analysis_disable:n #1
\{\if_int_compare:w 256 > #1 \exp_stop_f:
\tex_lccode:D 0 = #1 \exp_stop_f:
\tex_lowercase:D { \tex_let:D ^^@ } \tex_undefined:D
\}
Similar to \_\_tl_analysis_disable:n, but it receives a normal character token, tests if that token is active (by turning it into a space: the active space has been undefined at this point), and if so, disables it. Even if the character is active and set equal to a primitive conditional, nothing blows up. Again, in \TeX{} and \upTeX{} we skip characters beyond [0, 255], which cannot be active anyways.

\__tl_analysis_disable_char:N

The goal of this pass is to detect special (non-N-type) tokens, and count how many N-type tokens lie between special tokens. Also, we wish to store some representation of each special token in a \toks register.

We have 11 types of tokens:

1. a true non-space begin-group character;
2. a true space begin-group character;
3. a true non-space end-group character;
4. a true space end-group character;
5. a true space blank space character;
6. an active character;
7. any other true character;
8. a control sequence equal to a begin-group token (category code 1);
9. a control sequence equal to an end-group token (category code 2);
10. a control sequence equal to a space token (character code 32, category code 10);
11. any other control sequence.

Our first tool is \futurelet. This cannot distinguish case 8 from 1 or 2, nor case 9 from 3 or 4, nor case 10 from case 5. Those cases are later distinguished by applying the \string primitive to the following token, after possibly changing the escape character to ensure that a control sequence’s string representation cannot be mistaken for the true character.

In cases 6, 7, and 11, the following token is a valid N-type argument, so we grab it and distinguish the case of a character from a control sequence: in the latter case, \str_tail:n \{\langle token\rangle\} is non-empty, because the escape character is printable.

\tl_analysis_a:n
We read tokens one by one using \futurelet. While performing the loop, we keep track of the number of true begin-group characters minus the number of true end-group characters in \l__tl_analysis_nesting_int. This reaches −1 when we read the closing brace.

\tl_analysis_a_loop:w
Read one character and check its type.

\tl_analysis_a_type:w
At this point, \l__tl_analysis_token holds the meaning of the following token. We store in \l__tl_analysis_type_int information about the meaning of the token ahead:

- 0 space token;
- 1 begin-group token;
- -1 end-group token;
- 2 other.
The values 0, 1, −1 correspond to how much a true such character changes the nesting level (2 is used only here, and is irrelevant later). Then call the auxiliary for each case. Note that nesting conditionals here is safe because we only skip over \l__tl_analysis_token if it matches with one of the character tokens (hence is not a primitive conditional).

\cs_new_protected:Npn \__tl_analysis_a_type:w
  \l__tl_analysis_type_int =
  \if_meaning:w \l__tl_analysis_token \c_space_token
  0
  \else:
    \if_catcode:w \exp_not:N \l__tl_analysis_token \c_group_begin_token
      1
    \else:
      \if_catcode:w \exp_not:N \l__tl_analysis_token \c_group_end_token
        - 1
      \else:
        2
      \fi:
    \fi:
  \fi:
  \exp_stop_f:
  \if_case:w \l__tl_analysis_type_int
    \exp_after:wN \__tl_analysis_a_space:w
    \or: \exp_after:wN \__tl_analysis_a_bgroup:w
    \or: \exp_after:wN \__tl_analysis_a_safe:N
    \else: \exp_after:wN \__tl_analysis_a_egroup:w
  \fi:
\endtlcase:w

In this branch, the following token’s meaning is a blank space. Apply \string to that token: a true blank space gives a space, a control sequence gives a result starting with the escape character, an active character gives something else than a space since we disabled the space. We grab as \l__tl_analysis_char_token the first character of the string representation then test it in \__tl_analysis_a_space_test:w. Also, since \__tl_analysis_a_store: expects the special token to be stored in the relevant \toks register, we do that. The extra \exp_not:n is unnecessary of course, but it makes the treatment of all tokens more homogeneous. If we discover that the next token was actually a control sequence or an active character instead of a true space, then we stop the counter of normal tokens. We now have in front of us the whole string representation of the control sequence, including potential spaces; those will appear to be true spaces later in this pass. Hence, all other branches of the code in this first pass need to consider the string representation, so that the second pass does not need to test the meaning of tokens, only strings.

\cs_new_protected:Npn \__tl_analysis_a_space:w
  \__tl_analysis_a_space_test:w
\endtlcase:w
The token is most likely a true character token with catcode 1 or 2, but it might be a control sequence, or an active character. Optimizing for the first case, we store in a toks register some code that expands to that token. Since we will turn what follows into a string, we make sure the escape character is different from the current character code (by switching between solidus and backslash). To detect the special case of an active character let to the catcode 1 or 2 character with the same character code, we disable the active character with that character code and re-test: if the following token has become undefined we can in fact safely grab it. We are finally ready to turn what follows to a string and test it. This is one place where we need \__tl_analysis_char_token to be a separate control sequence from \__tl_analysis_token, to compare them.
This function is called each time we meet a special token; at this point, the \toks register \l__tl_analysis_index_int holds a token list which expands to the given special token. Also, the value of \l__tl_analysis_type_int indicates which case we are in:

- -1 end-group character;
- 0 space character;
- 1 begin-group character.

We need to distinguish further the case of a space character (code 32) from other character codes, because those behave differently in the second pass. Namely, after testing the \lccode of 0 (which holds the present character code) we change the cases above to

- -2 space end-group character;
- -1 non-space end-group character;
- 0 space blank space character;
- 1 non-space begin-group character;
- 2 space begin-group character.

This has the property that non-space characters correspond to odd values of \l__tl_analysis_type_int. The number of normal tokens until here and the type of special token are packed into a \skip register. Finally, we check whether we reached the last closing brace, in which case we stop by disabling the looping function (locally).
This should be the simplest case: since the upcoming token is safe, we can simply grab it in a second pass. If the token is a single character (including space), the `if_charcode:w` test yields true; we disable a potentially active character (that could otherwise masquerade as the true character in the next pass) and we count one “normal” token. On the other hand, if the token is a control sequence, we should replace it by its string representation for compatibility with other code branches. Instead of slowly looping through the characters with the main code, we use the knowledge of how the second pass works: if the control sequence name contains no space, count that token as a number of normal tokens equal to its string length. If the control sequence contains spaces, they should be registered as special characters by increasing `\l__tl_analysis_index_int` (no need to carefully count character between each space), and all characters after the last space should be counted in the following sequence of “normal” tokens.

\begin{verbatim}
\cs_new_protected:Npn \__tl_analysis_a_safe:N #1
\__tl_analysis_a_loop:w
\end{verbatim}

End of definition for `\__tl_analysis_a_safe:N` and `\__tl_analysis_a_cs:ww`.

45.7 Second pass

The second pass is an exercise in expandable loops. All the necessary information is stored in `\skip` and `\toks` registers.

\begin{verbatim}
\__tl_analysis_b:n \__tl_analysis_b_loop:w
\end{verbatim}

Start the loop with the index 0. No need for an end-marker: the loop stops by itself when the last index is read. We repeatedly oscillate between reading long stretches of normal tokens, and reading special tokens.

455
The first argument is the number of normal tokens which remain to be read, and the second argument is the index in the array produced in the first step. A character’s string representation is always one character long, while a control sequence is always longer (we have set the escape character to a printable value). In both cases, we leave $\exp_not:n$ \langle \text{token} \rangle \ s\_tl in the input stream (after e-expansion). Here, $\exp_not:n$ is used rather than $\exp_not:N$ because #3 could be a macro parameter character or could be \s__tl (which must be hidden behind braces in the result).

This function is called here with arguments $\__tl_analysis_b_char:n$ and a normal character, while in the peek analysis code it is called with $\use_none:n$ and possibly a space character, which is why the function has signature $Nn$. If the normal token we grab is a character, leave \langle \text{catcode} \rangle \langle \text{charcode} \rangle followed by $\s__tl$ in the input stream, and call $\__tl_analysis_b_normals:ww$ with its first argument decremented.

(End of definition for $\__tl_analysis_b:n$ and $\__tl_analysis_b_loop:w$.)
\exp_not:N \if_meaning:w #2 \exp_not:N \tex_undefined:D
\token_to_str:N D \exp_not:N \else:\n\exp_not:N \if_catcode:w #2 \c_catcode_other_token
\token_to_str:N C \exp_not:N \else:\n\exp_not:N \if_catcode:w #2 \c_catcode_letter_token
\token_to_str:N B \exp_not:N \else:\n\exp_not:N \if_catcode:w #2 \c_math_toggle_token
3 \exp_not:N \else:\n\exp_not:N \if_catcode:w #2 \c_alignment_token
4 \exp_not:N \else:\n\exp_not:N \if_catcode:w #2 \c_math_superscript_token
7 \exp_not:N \else:\n\exp_not:N \if_catcode:w #2 \c_math_subscript_token
8 \exp_not:N \else:\n\exp_not:N \if_catcode:w #2 \c_space_token
\token_to_str:N A \exp_not:N \else:\n6 \exp_not:n { \fi: \fi: \fi: \fi: \fi: \fi: \fi: \fi: } \fi: \fi: \fi: \fi: \fi: \fi: \fi: \fi: } \fi:}
\cs_new:Npn \__tl_analysis_b_char_aux:nww #1
{ \int_value:w '#1 \s__tl \exp_after:wN \__tl_analysis_b_normals:ww \int_value:w \int_eval:w - 1 +}
(End of definition for \__tl_analysis_b_char:Nn and \__tl_analysis_b_char_aux:nww.)

\__tl_analysis_b_cs:Nww \__tl_analysis_b_cs_test:ww
If the token we grab is a control sequence, leave 0 -1 (as category code and character code) in the input stream, followed by \s__tl, and call \__tl_analysis_b_normals:ww
with updated arguments.
\cs_new:Npn \__tl_analysis_b_cs:Nww #1
{ 0 -1 \s__tl \__tl_analysis_cs_space_count:NN \__tl_analysis_b_cs_test:ww #1 \int_value:w \int_eval:w
\exp_after:wN \__tl_analysis_b_normals:ww \int_value:w \int_eval:w - 1 +}
\__tl_analysis_b_special:w \__tl_analysis_b_special_char:wN \__tl_analysis_b_special_space:w
Here, #1 is the current index in the array built in the first pass. Check now whether
we reached the end (we shouldn’t keep the trailing end-group character that marked the
end of the token list in the first pass). Unpack the \toks register: when e/x-expanding again, we will get the special token. Then leave the category code in the input stream, followed by the character code, and call \__tl_analysis_b_loop:w with the next index.

3852 \group_begin:
3853 \char_set_catcode_other:N A
3854 \cs_new:Npn \__tl_analysis_b_special:w
3855 \fi: \__tl_analysis_b_normal:wwN 0 ; #1 ;
3856 \exp_after:wN \prg_break:
3857 \fi:
3858 \if_int_compare:w #1 = \l__tl_analysis_index_int
3859 \exp_after:wN \prg_break:
3860 \fi:
3861 \tex_the:D \tex_toks:D #1 \s__tl
3862 \if_case:w \tex_gluestretch:D \tex_skip:D #1 \exp_stop_f:
3863 \token_to_str:N A
3864 \or: 1
3865 \or: 1
3866 \else: 2
3867 \fi:
3868 \if_int_odd:w \tex_gluestretch:D \tex_skip:D #1 \exp_stop_f:
3869 \exp_after:wN \__tl_analysis_b_special_char:wN \int_value:w
3870 \else:
3871 \exp_after:wN \__tl_analysis_b_special_space:w \int_value:w
3872 \fi:
3873 \int_eval:n { 1 + #1 } \exp_after:wN ;
3874 \token_to_str:N }
3875 \group_end:
3876 \cs_new:Npn \__tl_analysis_b_special_char:wN #1 ; #2
3877 \exp_after:wN \__tl_analysis_b_special_space:w #1 ;
3878 \fi:
3879 \if_int_compare:w #1 = \l__tl_analysis_index_int
3880 \exp_after:wN \prg_break:
3881 \fi:
3882 \if_case:w \tex_gluestretch:D \tex_skip:D #1 \exp_stop_f:
3883 \token_to_str:N A
3884 \or: 1
3885 \or: 1
3886 \else: 2
3887 \fi:
3888 \if_int_odd:w \tex_gluestretch:D \tex_skip:D #1 \exp_stop_f:
3889 \exp_after:wN \__tl_analysis_b_special_char:wN \int_value:w
3890 \else:
3891 \exp_after:wN \__tl_analysis_b_special_space:w \int_value:w
3892 \fi:
3893 \int_eval:n { 1 + #1 } \exp_after:wN ;
3894 \token_to_str:N }

45.8 Mapping through the analysis

First obtain the analysis of the token list into \g__tl_analysis_result_tl. To allow nested mappings, increase the nesting depth \g_kernel_prg_map_int (shared between all modules), then define the payload macro, which runs the user code and has a name specific to that nesting depth. The looping macro grabs the \langle tokens \rangle, \langle catcode \rangle and \langle char code \rangle; it checks for the end of the loop with \use_none:n ##2, normally empty, but which becomes \tl_map_break: at the end; it then calls the payload macro with the arguments in the correct order (this is the reason why we cannot directly use the
same macro for looping and payload), and loops by calling itself. When the loop ends,
remember to decrease the nesting depth.

\begin{verbatim}
\cs_new_protected:Npn \tl_analysis_map_inline:Nn #1
\exp_args:No \tl_analysis_map_inline:nn \#1
\cs_new_protected:Npn \tl_analysis_map_inline:nn #1 #2
\__tl_analysis:n {#1}
\int_gincr:N \g__kernel_prg_map_int
\exp_args:Nc \__tl_analysis_map:Nn
\{ __tl_analysis_map_inline_ \int_use:N \g__kernel_prg_map_int :wNw \}
\cs_new_protected:Npn \__tl_analysis_map:Nn #1 #2 #3 #4 \s__tl
\cs_gset_protected:Npn #1 ##1##2##3 {#2}
\exp_after:wN \__tl_analysis_map:NwNw \exp_after:wN #1
\g__tl_analysis_result_tl \s__tl { ? \tl_map_break: } \s__tl
\prg_break_point:Nn \tl_map_break:
\{ \int_gdecr:N \g__kernel_prg_map_int \}
\cs_new_protected:Npn \tl_analysis_map:Nn \__tl_analysis_map:NNN #1 #2 #3
\{ \tl_if_exist:NTF #3
\exp_args:No \__tl_analysis:n {#3}
\#1 \{ tl \} \{ show-analysis \}
\{ \token_to_str:N \#3 \} \{ \__tl_analysis_show: \} \{ \}
\}
\cs_new_protected:Npn \__tl_analysis_map:NNN #1 #2 #3 #4 \s__tl
\{ \use_none:n \#3 \#1 \{#2} \{#4\} \{#3\}
\cs_new_protected:Npn \__tl_analysis_map:NNN #1 #2 #3
\{ __tl_analysis_map:NNN #1 #2 #3 #4 \s__tl
\cs_new_protected:Npn \__tl_analysis_map:NNN #1 #2 #3 #4 \s__tl
\}
\end{verbatim}

(End of definition for \tl_analysis_map_inline:Nn and others. These functions are documented on page 46.)

\section{Showing the results}

\texttt{\tl_analysis_show:N, \tl_analysis_log:N, and \tl_analysis_shown:NNN. These functions are documented on page 46.)

Add to \__tl_analysis:N a third pass to display tokens to the terminal. If the token list variable is not defined, throw the same error as \tl_show:N by simply calling that function.

\begin{verbatim}
\cs_new_protected:Npn \tl_analysis_show:N \tl_analysis_log:N \tl_analysis_shown:NNN
\{ \tl_if_exist:NTF #3
\exp_args:No \__tl_analysis:n {#3}
\#1 \{ tl \} \{ show-analysis \}
\{ \token_to_str:N \#3 \} \{ \__tl_analysis_show: \} \{ \}
\}
\end{verbatim}

(End of definition for \tl_analysis_show:N, \tl_analysis_log:N, and \tl_analysis_shown:NNN. These functions are documented on page 46.)
\tl_analysis_show:n
\tl_analysis_log:n
\__tl_analysis_show:Nn
No existence test needed here.
\cs_new_protected:Npn \tl_analysis_show:n
{ \__tl_analysis_show:Nn \msg_show:nneee }
\cs_new_protected:Npn \tl_analysis_log:n
{ \__tl_analysis_show:Nn \msg_log:nneee }
\cs_new_protected:Npn \__tl_analysis_show:Nn #1#2
{ \__tl_analysis:n {#2} #1 { tl } { show-analysis } { } { \__tl_analysis_show: } { } { }
}
\__tl_analysis_show:
\__tl_analysis_show_loop:wNw
Here, #1-o- and e/x-expands to the token; #2 is the category code (one uppercase hexadecimal digit), 0 for control sequences; #3 is the character code, which we ignore. In the cases of control sequences and active characters, the meaning may overflow one line, and we want to truncate it. Those cases are thus separated out.
\cs_new:Npn \__tl_analysis_show:
{ \exp_after:wN \__tl_analysis_show_loop:wNw \g__tl_analysis_result_tl
  \s__tl { ? \prg_break: } \s__tl
  \prg_break_point:
  }
\cs_new:Npn \__tl_analysis_show_loop:wNw #1 \s__tl #2 #3 \s__tl
{ \use_none:n #2
  \iow_newline: > \use:nn { - } { - }
  \if_int_compare:w "#2 = \c_zero_int
  \exp_after:wN \__tl_analysis_show_cs:n
  \else:
  \if_int_compare:w "#2 = 13 \exp_stop_f:
    \exp_after:wN \exp_after:wN
    \exp_after:wN \__tl_analysis_show_active:n
  \else:
    \exp_after:wN \exp_after:wN
    \exp_after:wN \__tl_analysis_show_normal:n
  \fi:
  \fi:
  {#1}
  \__tl_analysis_show_loop:wNw
  }
\__tl_analysis_show_normal:n
Non-active characters are a simple matter of printing the character, and its meaning. Our test suite checks that begin-group and end-group characters do not mess up \TeX's alignment status.
\cs_new:Npn \__tl_analysis_show_normal:n #1
{ \exp_after:wN \token_to_str:N #1 - 
  ( \exp_after:wN \token_to_meaning:N #1 )
} 460
This expands to the value of \#1 if it has any.

\cs_new:Npn \__tl_analysis_show_value:N \#1
{\token_if_expandable:NF \#1
{\token_if_chardef:NTF \#1 \prg_break: { }
\token_if_mathchardef:NTF \#1 \prg_break: { }
\token_if_dim_register:NTF \#1 \prg_break: { }
\token_if_int_register:NTF \#1 \prg_break: { }
\token_if_skip_register:NTF \#1 \prg_break: { }
\token_if_toks_register:NTF \#1 \prg_break: { }
\use_none:nnn
\prg_break_point:
\use:n { \exp_after:wN = \tex_the:D \#1 } }
}

Control sequences and active characters are printed in the same way, making sure not to go beyond the \l_iow_line_count_int. In case of an overflow, we replace the last characters by \c__tl_analysis_show_etc_str.

\cs_new:Npn \__tl_analysis_show_cs:n \#1
{\exp_args:No \__tl_analysis_show_long:nn {\token_to_str:N \#1} { control~sequence= } }
\cs_new:Npn \__tl_analysis_show_active:n \#1
{\exp_args:No \__tl_analysis_show_long:nn {\token_to_meaning:N \#1} { active~character= } }
\cs_new:Npn \__tl_analysis_show_long:nn \#1
{\__tl_analysis_show_long_aux:oofn {\token_to_str:N \#1} { \token_to_meaning:N \#1} { \__tl_analysis_show_value:N \#1} }
\cs_new:Npn \__tl_analysis_show_long_aux:nnnn \#1#2#3#4
{\int_compare:nNnTF \str_count:n { \token_to_str:N \#1 (~ \token_to_meaning:N \#1 ( \token_to_str:N \#1) } > \l_iow_line_count_int - 3
{\str_range:nnn { \token_to_str:N \#1 (~ \token_to_meaning:N \#1 ( \token_to_str:N \#1) } { 1 }
{\l_iow_line_count_int - 3 - \str_count:N \c__tl_analysis_show_etc_str
\c__tl_analysis_show_etc_str
{ \token_to_str:N \#1 (~ \token_to_meaning:N \#1 ( \token_to_str:N \#1) } }
\cs_generate_variant:Nn \__tl_analysis_show_long_aux:nnnn { oof }
}

(End of definition for \__tl_analysis_show_value:n)
(End of definition for \__tl_analysis_show_value:N)
45.10 Peeking ahead

The break statements use the general `\prg_map_break:Nn`.

\begin{verbatim}
\peek_analysis_map_break:
\peek_analysis_map_break:n
\end{verbatim}

The break statements use the general `\prg_map_break:Nn`.

\begin{verbatim}
\cs_new:Npn \peek_analysis_map_break: \{ \}
\cs_new:Npn \peek_analysis_map_break:n \{ \}
\end{verbatim}

(End of definition for `\peek_analysis_map_break:` and `\peek_analysis_map_break:n`. These functions are documented on page 208.)

\begin{verbatim}
\l__tl_peek_charcode_int
\end{verbatim}

After a call to `\futurelet \l__tl_analysis_token` followed by a stringified character token (either explicit space or catcode other character), grab the argument and pass it to \#1. We only need to do anything in the case of a space.

\begin{verbatim}
\cs_new:Npn \__tl_analysis_char_arg:Nw \__tl_analysis_char_arg_aux:Nw #1 \{ #1 \}
\end{verbatim}

(End of definition for `\__tl_analysis_char_arg:Nw` and `\__tl_analysis_char_arg_aux:Nw`.)

\begin{verbatim}
\peek_analysis_map_inline:n \__tl_peek_analysis_loop:NNn
\__tl_peek_analysis_test:
\__tl_peek_analysis_exp:N
\__tl_peek_analysis_exp_aux:N
\__tl_peek_analysis_nonexp:N
\__tl_peek_analysis_char:N
\__tl_peek_analysis_char:w
\__tl_peek_analysis_special:
\__tl_peek_analysis_retest:
\__tl_peek_analysis_str:n
\__tl_peek_analysis_collect:n
\__tl_peek_analysis_collect_loop:
\__tl_peek_analysis_collect_test:
\__tl_peek_analysis_active_str:n
\__tl_peek_analysis_explicit:n
\__tl_peek_analysis_escape:
\__tl_peek_analysis_collect:w
\__tl_peek_analysis_collect:n
\__tl_peek_analysis_collect_loop:
\__tl_peek_analysis_collect_test:
\__tl_peek_analysis_collect_end:NNNN
\end{verbatim}

Save the user’s code in a control sequence that is suitable for nested maps. We may wish to pass to this function an `\outer` control sequence or active character; for this we will undefine any expandable token (testing if it is `\outer` is much slower) within a group, closed immediately after the function reads its arguments to avoid affecting the user’s code or even our peek code (there is no risk of undefining `\group_end:` itself since that is not expandable). This user’s code function also calls the loop auxiliary, and includes the trailing `\prg_break_point:Nn` for when the user wants to stop the loop. The loop auxiliary must remove that break point because it must look at the input stream.

\begin{verbatim}
\cs_new_protected:Npn \peek_analysis_map_inline:n \#1
\end{verbatim}

(End of definition for `\__tl_analysis_char_arg:Nw` and `\__tl_analysis_char_arg_aux:Nw`.)

462
The loop starts a group (closed by the user-code function defined above) with a normalized escape character, and checks if the next token is special or $N$-type (distinguishing expandable from non-expandable tokens). The test for nonexpandable tokens in \_\_tl\_peek\_analysis\_test must be done after the tests for begin-group, end-group, and space tokens, in case \l_\_peek\_token is either $\outer$ or is a primitive \TeX\ conditional, as such tokens cannot be skipped over correctly by conditional code.

\cs\_new\_protected:Npn \_\_tl\_peek\_analysis\_loop:NNn #1#2#3

\group\_begin:
\tl\_set:Nn \l\_\_tl\_peek\_code\_tl
{
    \exp\_not:c
\{ \_\_tl\_analysis\_map_ \int\_\_use:N \g\_\_kernel\_prg\_map\_int :nnN \}
\int\_set:Nn \tex\_escape\_char:D { \char\_max \_int } \}
\peek\_after:Nw \_\_tl\_peek\_analysis\_test:
\group\_end:
\cs\_new\_protected:Npn \_\_tl\_peek\_analysis\_test:
{
    \if\_case:w
        \if\_cat\_code:w \exp\_not:N \l\_\_peek\_token \c\_max\_int \fi:
        \if\_cat\_code:w \exp\_not:N \l\_\_peek\_token \c\_max\_int \fi:
        \if\_meaning:w \l\_\_peek\_token \c\_space\_token \c\_max\_int \fi:
        \exp\_after:wN \if\_meaning:w \exp\_not:N \l\_\_peek\_token \l\_\_peek\_token \c\_one\_int
        \fi:
        \c\_zero\_int
        \exp\_after:wN \exp\_after:wN
        \exp\_after:wN \_\_tl\_peek\_analysis\_exp:N
        \exp\_after:wN \exp\_not:N
        \or:
        \exp\_after:wN \_\_tl\_peek\_analysis\_nonexp:N
        \else:
        \exp\_after:wN \_\_tl\_peek\_analysis\_special:
        \fi:
    \}

Expandable tokens (which are automatically $N$-type) can be $\outer$ macros, hence the need for \exp\_after:wN and \exp\_not:N in the code above, which allows the next function to safely grab the token as an argument. We run some code that is expanded using the primitive \cs\_set\_nopar:Npe rather than \tl\_set:Nn to avoid grabbing it as an argument as $\#1$ may be $\outer$. To allow $\#1$ as an argument of the user’s function (stored in \l\_\_tl\_peek\_code\_tl), we set it equal to \scan\_stop:, but we do it at the last minute because $\#1$ may be some pretty important function such as \exp\_after:wN. Then we put the user’s function and the elaborate first argument \_\_kernel\_exp\_not:w \exp\_after:wN \{ \exp\_not:N \#1 \}: indeed we cannot use \exp\_not:n \{\#1\} as this breaks for an $\outer$ macro and we cannot use \exp\_not:N \#1, as o-expanding this yields a “notexpanded” token equal to (a weird) $\relax$, which would have the wrong value for primitive \TeX\ conditionals such as $\if\_meaning:w$.

Then we must add $(-1)\relax$ if the token is a control sequence and $\langle$\char\_code$\rangle\relax$ otherwise. Distinguishing the two cases is easy: since we have made the escape character
\begin{verbatim}
\cs_new_protected:Npn \__tl_peek_analysis_exp:N #1
\{
  \cs_set_nopar:Npe \l__tl_peek_code_tl
  \{
    \tex_let:D \exp_not:N #1 \scan_stop:
    \exp_not:o \l__tl_peek_code_tl
    \{
      \exp_not:n { \__kernel_exp_not:w \exp_after:wN }
      { \exp_not:N \exp_not:N \exp_not:N #1 }
    }
  \exp_after:wN \__tl_peek_analysis_exp_aux:Nw
  \token_to_str:N #1 \prg_do_nothing: \s__tl
\}
\l__tl_peek_code_tl
\}
\cs_new:Npe \__tl_peek_analysis_exp_aux:Nw #1#2 \s__tl
\{
  \exp_not:N \if:w \scan_stop: #2 \scan_stop:
  \{ \exp_not:N \int_value:w \':#1 } \token_to_str:N D
  \exp_not:N \else:
  \{ -1 \} 0
  \exp_not:N \fi:
\}
\cs_new_protected:Npn \__tl_peek_analysis_nonexp:N #1
\{
  \exp_not:N \if_charcode:w \scan_stop: #1 \\use_none:n \token_to_str:N #1 \prg_do_nothing: \\scan_stop:
  \exp_after:wN \__tl_peek_analysis_char:N
  \\exp_after:wN \__tl_peek_analysis_exp_aux:Nw
  \\token_to_str:N #1 \prg_do_nothing: \s__tl
\}
\cs_new_protected:Npn \__tl_peek_analysis_char:N #1
\{
  \\exp_after:wN \__tl_peek_analysis_nonexp:N \#1
\}
\cs_new_protected:Npn \__tl_peek_analysis_cs:N #1
\{
  \exp_not:n { \__kernel_exp_not:w \exp_after:wN }
  { \exp_not:N \exp_not:N \exp_not:N #1 }
\exp_after:wN \__tl_peek_analysis_char:N
\if:w \scan_stop: #1 \scan_stop:
  \{ \exp_not:N \int_value:w \':#1 } \token_to_str:N D
  \exp_not:N \else:
  \{ -1 \} 0
  \exp_not:N \fi:
\}
\end{verbatim}

For normal non-expandable tokens we must distinguish characters (including active ones and macro parameter characters) from control sequences (whose string representation is more than one character because we made the escape character printable). For a control sequence call the user code with suitable arguments, wrapping \texttt{#1} within \texttt{\exp_not:n} just in case it happens to be equal to a macro parameter character. We do not skip \texttt{\exp_not:n} when unnecessary, because this auxiliary is also called in \texttt{\__tl_peek_analysis_retest:} where we have changed some control sequences or active characters to \texttt{\scan_stop:} temporarily.

For normal characters we must determine their category. The main difficulty is that the character may be an active character masquerading as (i.e., set equal to) itself with a different category. Two approaches based on \texttt{\lowercase} can detect this. One could make an active character with the same category as \texttt{#1} and change its definition before testing the
catcode of \#1, but in some Unicode engine this fills up the hash table uselessly. Instead, we lowercase \#1 itself, changing its character code to 32, namely space (because LuaTeX cannot turn catcode 10 characters to anything else than character code 32), then we apply \_\_tl_analysis_b_char:Nn, which detects active characters by comparing them to \tex_undefined:D, and we must have undefined the active space (locally) for this test to work. To define \_\_tl_peek_analysis_char:N itself we use an e-expanding assignment to get the active space in the right place after making it (just for this definition) unexpandable. Finally \_\_tl_peek_analysis_char:w receives the (charcode), (user function), (catcode), and (token), and places the arguments in the correct order. It keeps \exp_not:n for macro parameter characters and active characters (the latter could be macro parameter characters, and it seems more uniform to always put \exp_not:n), and otherwise eliminates it by expanding once with \exp_args:NNNo.

\begin{verbatim}
\group_begin:
\char_set_active_eq:NN \scan_stop:
\cs_new_protected:Npe \_\_tl_peek_analysis_char:N #1
{
  \char_set_eq:NN \char_generate:nn { 32 } { 13 }
  \exp_not:N \tex_undefined:D
  \tex_lccode:D '#1 = 32 \exp_stop_f:
  \tex_lowercase:D
    \tl_put_right:Ne \exp_not:N \l__tl_peek_code_tl
    { \exp_not:n { \__tl_analysis_b_char:Nn \use_none:n } {#1} }
  \exp_not:n
    { \exp_after:wN \__tl_peek_analysis_char:w
      \int_value:w
    } '#1
  \exp_not:n
    { \exp_after:wN \s__tl \l__tl_peek_code_tl }
}{#1}
\group_end:
\cs_new_protected:Npm \_\_tl_peek_analysis_char:w #1 \s__tl #2#3#4
{
  \if_charcode:w 6 #3
  \else:
    \if_charcode:w D #3
    \else:
      \exp_args:NNNo
      \fi:
  \fi:
  #2 { \exp_not:n {#4} } {#1} #3
}
\end{verbatim}

For special characters the idea is to eventually act with \token_to_str:N, then pick up one by one the characters of this string representation until hitting the token that follows. First determine the character code of (the meaning of) the (token) (which we know is a special token), make sure the escape character is different from it, normalize the meanings of two active characters and the empty control sequence, and filter out these cases in \_\_tl_peek_analysis_retest:
At this point we know the meaning of the \textlangle token\texrangle in the input stream is \textlbrace \textlangle token \textrbrace token, either a space (32, 10) or a begin-group or end-group token (catcode 1 or 2), and we excluded a few cases that would be difficult later (empty control sequence, active character with the same character code as its meaning or as the escape character). The idea is to apply \textlangle token \textrangle token to \textlangle token \textrangle token then grab characters (of category code 12 except for spaces that have category code 10) to reconstruct it. In earlier versions of the code we would peek at the \langle next token \rangle next token that lies after \langle token \rangle token in the input stream, which would help us be more accurate in reconstructing the \langle token \rangle token case in edge cases (mentioned below), but this had the side-effect of tokenizing the input stream (turning characters into tokens) farther ahead than needed.

We hit the \langle token \rangle token with \textlangle token \textrangle token to \textlangle token \textrangle token and start grabbing characters. More precisely, by looking at the first character in the string representation of the \langle token \rangle token we distinguish three cases: a stringified control sequence starts with the escape character; for an explicit character we find that same character; for an active character we find anything else (we made sure to exclude the case of an active character whose string representation coincides with the other two cases).
When \#1 is a stringified active character we pass appropriate arguments to the user's code; thankfully \char_generate:nn can make active characters.

```
cs_new_protected:Npn \__tl_peek_analysis_active_str:n #1
{
  \tl_put_right:Ne \l__tl_peek_code_tl
  { \char_generate:nn { '#1 } { 13 } }
  \int_value:w '#1
  \token_to_str:N D
  \l__tl_peek_code_tl
}
```

When \#1 matches the character we had extracted from the meaning of \l_peek_token, the token was an explicit character, which can be a standard space, or a begin-group or end-group character with some character code. In the latter two cases we call \char_generate:nn with suitable arguments and put suitable \if_false: \fi: constructions to make the result balanced and such that o-expanding or e/x-expanding gives back a single (unbalanced) begin-group or end-group character.

```
cs_new_protected:Npn \__tl_peek_analysis_explicit:n #1
{
  \tl_put_right:Ne \l__tl_peek_code_tl
  { \if_meaning:w \l_peek_token \c_space_token
    { ~ } { 32 } \token_to_str:N A
  }
  \else:
    \if_catcode:w \l_peek_token \c_space_token
      { - } { 32 } \token_to_str:N A
    \else:
      \if_catcode:w \l_peek_token \c_group_begin_token
        { \exp_not:N \exp_after:wN
          \char_generate:nn { '#1 } { 1 }
          \exp_not:N \if_false:
          \if_false: \fi: 
          \exp_not:N \fi:
        }
        { \int_value:w '#1 }
        1
      \else:
        { \exp_not:N \if_false:
          \exp_not:N \if_false: \fi:
          \exp_not:N \fi:
          \char_generate:nn { '#1 } { 2 }
        }
        { \int_value:w '#1 }
        2
      \fi:
      \fi:
    }
}
```
Finally there is the case of a special token whose string representation starts with an escape character, namely the token was a control sequence. In that case we could have grabbed the token directly as an \texttt{N}-type argument, but of course we couldn’t know that until we had run all the various tests including stringifying the token. We are thus left with the hard work of picking up one by one the characters in the csname (being careful about spaces), until the constructed csname has the expected meaning. This fails if someone defines a token like \texttt{bgroup@my} whose string representation starts the same as another token with the same meaning being an implicit character token of category code 1, 2, or 10.

As in all other cases, end by calling the user code with suitable arguments (here \texttt{#1} is \texttt{\fi}).
45.11 Messages

When a control sequence (or active character) and its meaning are too long to fit in one line of the terminal, the end is replaced by this token list.

\c__tl_analysis_show_etc_str
\tl_const:Ne \c__tl_analysis_show_etc_str % ( \token_to_str:N \ETC.)

(End of definition for \c__tl_analysis_show_etc_str.)

\msg_new:nnn { tl } { show-analysis }
\tl_if_empty:nF {#1} { #1 ~ }
\tl_if_empty:nTF {#2} { is-empty }
\tl_if_empty:nTF {#2} { contains-the-tokens: #2 }

}
Chapter 46

l3regex implementation

46.1 Plan of attack

Most regex engines use backtracking. This allows to provide very powerful features (back-references come to mind first), but it is costly, and raises the problem of catastrophic backtracking. Since \TeX{} is not first and foremost a programming language, complicated code tends to run slowly, and we must use faster, albeit slightly more restrictive, techniques, coming from automata theory.

Given a regular expression of \(n\) characters, we do the following:

- (Compiling.) Analyse the regex, finding invalid input, and convert it to an internal representation.
- (Building.) Convert the compiled regex to a non-deterministic finite automaton (NFA) with \(O(n)\) states which accepts precisely token lists matching that regex.
- (Matching.) Loop through the query token list one token (one “position”) at a time, exploring in parallel every possible path (“active thread”) through the NFA, considering active threads in an order determined by the quantifiers’ greediness.

We use the following vocabulary in the code comments (and in variable names).

- **Group**: index of the capturing group, \(-1\) for non-capturing groups.
- **Position**: each token in the query is labelled by an integer \(\langle \text{position} \rangle\), with \(\text{min\_pos} - 1 \leq \langle \text{position} \rangle \leq \text{max\_pos}\). The lowest and highest positions \(\text{min\_pos} - 1\) and \(\text{max\_pos}\) correspond to imaginary begin and end markers (with non-existent category code and character code). \text{max\_pos} is only set quite late in the processing.
- **Query**: the token list to which we apply the regular expression.
- **State**: each state of the NFA is labelled by an integer \(\langle \text{state} \rangle\) with \(\text{min\_state} \leq \langle \text{state} \rangle < \text{max\_state}\).
- **Active thread**: state of the NFA that is reached when reading the query token list for the matching. Those threads are ordered according to the greediness of quantifiers.
• **Step**: used when matching, starts at 0, incremented every time a character is read, and is not reset when searching for repeated matches. The integer \l__regex_step_int is a unique id for all the steps of the matching algorithm.

We use \l3intarray to manipulate arrays of integers. We also abuse \TeX’s \toks registers, by accessing them directly by number rather than tying them to control sequence using the \nevtoks allocation functions. Specifically, these arrays and \toks are used as follows. When building, \toks{state} holds the tests and actions to perform in the ⟨state⟩ of the NFA. When matching,

- \g__regex_state_active_intarray holds the last ⟨step⟩ in which each ⟨state⟩ was active.
- \g__regex_thread_info_intarray consists of blocks for each ⟨thread⟩ (with \texttt{min_thread} ≤ ⟨thread⟩ < \texttt{max_thread}). Each block has \texttt{1+2\l__regex_capturing_group_int} entries: the ⟨state⟩ in which the ⟨thread⟩ currently is, followed by the beginnings of all submatches, and then the ends of all submatches. The ⟨threads⟩ are ordered starting from the best to the least preferred.
- \g__regex_submatch_prev_intarray, \g__regex_submatch_begin_intarray and \g__regex_submatch_end_intarray hold, for each submatch (as would be extracted by \regex_extract_all:nN), the place where the submatch started to be looked for and its two end-points. For historical reasons, the minimum index is twice \texttt{max_state}, and the used registers go up to \l__regex_submatch_int. They are organized in blocks of \l__regex_capturing_group_int entries, each block corresponding to one match with all its submatches stored in consecutive entries.

When actually building the result,

- \toks{position} holds ⟨tokens⟩ which o- and e-expand to the ⟨position⟩-th token in the query.
- \g__regex_balance_intarray holds the balance of begin-group and end-group character tokens which appear before that point in the token list.

The code is structured as follows. Variables are introduced in the relevant section. First we present some generic helper functions. Then comes the code for compiling a regular expression, and for showing the result of the compilation. The building phase converts a compiled regex to NFA states, and the automaton is run by the code in the following section. The only remaining brick is parsing the replacement text and performing the replacement. We are then ready for all the user functions. Finally, messages, and a little bit of tracing code.

### 46.2 Helpers

\__regex_int_eval:w Access the primitive: performance is key here, so we do not use the slower route via \int_eval:n.

4271 \cs_new_eq:NN \__regex_int_eval:w \tex_numexpr:D

(\textit{End of definition for }\__regex_int_eval:w.)

\regex_standard_escapechar: Make the \\texttt{escapechar} into the standard backslash.

4272 \cs_new_protected:Npn \regex_standard_escapechar:

4273 \{ \int_set:Nn \tex_escapechar:D \{ ‘\\ } \}
Unpack a \toks given its number.
\cs_new:Npn \__regex_toks_use:w { \tex_the:D \tex_toks:D }

Empty a \toks or set it to a value, given its number.
\cs_new_protected:Npn \__regex_toks_clear:N \__regex_toks_set:Nn \__regex_toks_set:No
\cs_new_eq:NN \__regex_toks_clear:N \tex_toks:D
\cs_new_protected:Npn \__regex_toks_set:Nn \__regex_toks_set:No
\cs_if_exist:NTF \tex_etokspre:D { \cs_new_eq:NN \__regex_toks_put_left:Ne \tex_etokspre:D }
\cs_if_exist:NTF \tex_etoksapp:D { \cs_new_eq:NN \__regex_toks_put_right:Ne \tex_etoksapp:D }
\cs_if_exist:NTF \tex_toksapp:D { \cs_new_eq:NN \__regex_toks_put_right:Nn \tex_toksapp:D }

Copy \#3 \toks registers from \#2 onwards to \#1 onwards, like C’s memcpy.
\cs_new_protected:Npn \__regex_toks_memcpy:NNn
\prg_replicate:nn {#3}
\tex_toks:D #1 = \tex_toks:D #2 \int_incr:N #1 \int_incr:N #2

During the building phase we wish to add e-expanded material to \toks, either to the left or to the right. The expansion is done “by hand” for optimization (these operations are used quite a lot). The \#n version of \__regex_toks_put_right:Ne is provided because it is more efficient than e-expanding with \exp_not:n.
\cs_if_exist:NTF \tex_etokspre:D { \cs_new_eq:NN \__regex_toks_put_left:Ne \tex_etokspre:D }
\cs_if_exist:NTF \tex_etoksapp:D { \cs_new_eq:NN \__regex_toks_put_right:Ne \tex_etoksapp:D }
\cs_if_exist:NTF \tex_toksapp:D { \cs_new_eq:NN \__regex_toks_put_right:Nn \tex_toksapp:D }

(End of definition for \__regex_standard_escapechar:)
\__regex_curr_cs_to_str: \ Expands to the string representation of the token (known to be a control sequence) at the current position \l__regex_curr_token_tl. It should only be used in e/x-expansion to avoid losing a leading space.

\cs_new:Npn \__regex_curr_cs_to_str: \{
\exp_after:wN \exp_after:wN \exp_after:wN \cs_to_str:N \l__regex_curr_token_tl 
\}

(End of definition for \__regex_curr_cs_to_str:)

\__regex_intarray_item:NnF \__regex_intarray_item_aux:nNF \ Expands to an item of intarray, with a default value.

\cs_new:Npn \__regex_intarray_item:NnF #1#2 \{
\exp_args:No \__regex_intarray_item_aux:nNF \{ \tex_the:D \__regex_int_eval:w #2 \} #1 \}

\cs_new:Npn \__regex_intarray_item_aux:nNF #1#2 \{
\if_int_compare:w #1 > \c_zero_int \exp_after:wN \use_ii:nnn \fi:
\use_ii:nn { \__kernel_intarray_item:Nn #2 {#1} } 
\}

(End of definition for \__regex_intarray_item:NnF and \__regex_intarray_item_aux:nNF.)

\__regex_maplike_break: \ An analogous to \tl_map_break:, this correctly exits \tl_map_inline:nn and similar constructions and jumps to the matching \prg_break_point:Nn \__regex_maplike_break: \{ \}.

\cs_new:Npn \__regex_maplike_break: \{
\prg_map_break:Nn \__regex_maplike_break: { } \}

(End of definition for \__regex_maplike_break:)

\__regex_tl_odd_items:n \__regex_tl_even_items:n \\__regex_tl_even_items_loop:nn \ Map through a token list one pair at a time, leaving the odd-numbered or even-numbered items (the first item is numbered 1).

\cs_new:Npn \__regex_tl_odd_items:n \__regex_tl_even_items:n \\__regex_tl_even_items_loop:nn \#1 \{ \__regex_tl_odd_items:n \#1 \} \cs_new:Npn \__regex_tl_even_items:n \#1 \{
\__regex_tl_even_items_loop:nn \#1 \q__regex_nil \q__regex_nil 
\prg_break_point:
\}
\cs_new:Npn \__regex_tl_even_items_loop:nn \#1\#2 \{
\__regex_use_none_delimit_by_q_nil:w \#2 \prg_break: \q__regex_nil
\{ \exp_not:n \{#2\} \}
\__regex_tl_even_items_loop:nn 
\}

(End of definition for \__regex_tl_odd_items:n, \__regex_tl_even_items:n, and \__regex_tl_even_items_loop:nn.)
46.2.1 Constants and variables

\__regex_tmp:w Temporary function used for various short-term purposes.
\cs_new:Npn \__regex_tmp:w { }
(End of definition for \__regex_tmp:w.)
\l__regex_internal_a_tl \l__regex_internal_b_tl \l__regex_internal_a_int \l__regex_internal_b_int
\l__regex_internal_c_int \l__regex_internal_bool \l__regex_internal_seq \g__regex_internal_tl
Temporary variables used for various purposes.
\tl_new:N \l__regex_internal_a_tl \tl_new:N \l__regex_internal_b_tl
\tl_new:N \l__regex_internal_a_int \tl_new:N \l__regex_internal_b_int
\tl_new:N \l__regex_internal_c_int \tl_new:N \l__regex_internal_bool
\seq_new:N \l__regex_internal_seq \tl_new:N \g__regex_internal_tl
(End of definition for \l__regex_internal_a_tl and others.)
\l__regex_build_tl This temporary variable is specifically for use with the \tl_build machinery.
\tl_new:N \l__regex_build_tl (End of definition for \l__regex_build_tl.)
\c__regex_no_match_regex This regular expression matches nothing, but is still a valid regular expression. We could use a failing assertion, but I went for an empty class. It is used as the initial value for regular expressions declared using \regex_new:N.
\tl_const:Nn \c__regex_no_match_regex { \__regex_branch:n { \__regex_class:NnnnN \c_true_bool { } { 1 } { 0 } \c_true_bool } }
(End of definition for \c__regex_no_match_regex.)
\l__regex_balance_int During this phase, \l__regex_balance_int counts the balance of begin-group and end-group character tokens which appear before a given point in the token list. This variable is also used to keep track of the balance in the replacement text.
\int_new:N \l__regex_balance_int (End of definition for \l__regex_balance_int.)

46.2.2 Testing characters
\c__regex_ascii_min_int \c__regex_ascii_max_control_int \c__regex_ascii_max_int
\int_const:Nn \c__regex_ascii_min_int { 0 }
\int_const:Nn \c__regex_ascii_max_control_int { 31 }
\int_const:Nn \c__regex_ascii_max_int { 127 }
(End of definition for \c__regex_ascii_min_int, \c__regex_ascii_max_control_int, and \c__regex_ascii_max_int.)
\c__regex_ascii_lower_int
\int_const:Nn \c__regex_ascii_lower_int { 'a - 'A }
(End of definition for \c__regex_ascii_lower_int.)
46.2.3 Internal auxiliaries

\_\_regex_recursion_stop

Internal recursion quarks.

\_\_regex_recursion_stop

(End of definition for \_\_regex_recursion_stop.)

\_\_regex_nil

Internal quarks.

\_\_regex_nil

(End of definition for \_\_regex_nil.)

Functions to gobble up to a quark.

\_\_regex_use_none_delimit_by_q_recursion_stop:w

\_\_regex_use_i_delimit_by_q_recursion_stop:nw

\_\_regex_use_none_delimit_by_q_nil:w

Branching quark conditional.

\_\_regex_quark_if_nil:p:n

\_\_regex_quark_if_nil:nTF

(End of definition for \_\_regex_quark_if-nil:p:n, \_\_regex_quark_if-nil:n TF.)

\_\_regex_break_point:TF

\_\_regex_break_true:w

When testing whether a character of the query token list matches a given character class in the regular expression, we often have to test it against several ranges of characters, checking if any one of those matches. This is done with a structure like

\langle test1 \rangle ... \langle testn \rangle

\_\_regex_break_point:TF \{ (true code) \} \{ (false code) \}

If any of the tests succeeds, it calls \_\_regex_break_true:w, which cleans up and leaves \langle true code \rangle in the input stream. Otherwise, \_\_regex_break_point:TF leaves the \langle false code \rangle in the input stream.

\_\_regex_item_reverse:n

This function makes showing regular expressions easier, and lets us define \textbackslash D in terms of \textbackslash d for instance. There is a subtlety: the end of the query is marked by \textbackslash -2, and thus matches \textbackslash D and other negated properties; this case is caught by another part of the code.

(End of definition for \_\_regex_item_reverse:n.)
Simple comparisons triggering `__regex_break_true:w` when true.

```latex
\cs_new_protected:Npn \__regex_item_caseful_equal:n #1
\{\exp_after:wN \__regex_break_true:w \fi:\}
\cs_new_protected:Npn \__regex_item_caseful_range:nn #1 #2
\{\reverse_if:N \if_int_compare:w #1 > \l__regex_curr_char_int \reverse_if:N \if_int_compare:w #2 < \l__regex_curr_char_int \exp_after:wN \exp_after:wN \exp_after:wN \__regex_break_true:w \fi:\fi:\}
```

(End of definition for `\__regex_item_caseful_equal:n` and `\__regex_item_caseful_range:nn`.)

For caseless matching, we perform the test both on the `curr_char` and on the `case_changed_char`. Before doing the second set of tests, we make sure that `case_changed_char` has been computed.

```latex
\cs_new_protected:Npn \__regex_item_caseless_equal:n #1
\{\if_int_compare:w #1 = \l__regex_curr_char_int \exp_after:wN \__regex_break_true:w \fi:\__regex_maybe_compute_ccc:\if_int_compare:w #1 = \l__regex_case_changed_char_int \exp_after:wN \__regex_break_true:w \fi:\}
\cs_new_protected:Npn \__regex_item_caseless_range:nn #1 #2
\{\reverse_if:N \if_int_compare:w #1 > \l__regex_curr_char_int \reverse_if:N \if_int_compare:w #2 < \l__regex_curr_char_int \exp_after:wN \exp_after:wN \exp_after:wN \__regex_break_true:w \fi:\fi:\__regex_maybe_compute_ccc:\reverse_if:N \if_int_compare:w #1 > \l__regex_case_changed_char_int \reverse_if:N \if_int_compare:w #2 < \l__regex_case_changed_char_int \exp_after:wN \exp_after:wN \exp_after:wN \__regex_break_true:w \fi:\fi:\}
```

(End of definition for `\__regex_item_caseless_equal:n` and `\__regex_item_caseless_range:nn`.)

This function is called when `\l__regex_case_changed_char_int` has not yet been computed. If the current character code is in the range \[65, 90\] (upper-case), then add 32, making it lowercase. If it is in the lower-case letter range \[97, 122\], subtract 32.

```latex
\cs_new_protected:Npn \__regex_compute_case_changed_char:
\{\int_set_eq:NN \l__regex_case_changed_char_int \l__regex_curr_char_int\}
```

476
\if_int_compare:w \l__regex_curr_char_int > `'Z \exp_stop_f:
\if_int_compare:w \l__regex_curr_char_int > `'z \exp_stop_f: \else:
  \if_int_compare:w \l__regex_curr_char_int < `'a \exp_stop_f: \else:
    \int_sub:Nn \l__regex_case_changed_char_int \c__regex_ascii_lower_int
  \fi:
\fi:
\fi:
\else:
  \if_int_compare:w \l__regex_curr_char_int < `'A \exp_stop_f: \else:
    \int_add:Nn \l__regex_case_changed_char_int \c__regex_ascii_lower_int
  \fi:
\fi:
\fi:
\cs_set_eq:NN \__regex_maybe_compute_ccc: \prg_do_nothing:
\cs_new_eq:NN \__regex_maybe_compute_ccc: \__regex_compute_case_changed_char:
(End of definition for \__regex_compute_case_changed_char:)

\__regex_item_equal:n
\__regex_item_range:nn
Those must always be defined to expand to a caseful (default) or caseless version, and not be protected: they must expand when compiling, to hard-code which tests are caseless or caseful.
\cs_new_eq:NN \__regex_item_equal:n ?
\cs_new_eq:NN \__regex_item_range:nn ?
(End of definition for \__regex_item_equal:n and \__regex_item_range:nn)

\__regex_item_catcode:nT
\__regex_item_catcode_reverse:nT
\__regex_item_catcode:
The argument is a sum of powers of 4 with exponents given by the allowed category codes (between 0 and 13). Dividing by a given power of 4 gives an odd result if and only if that category code is allowed. If the catcode does not match, then skip the character code tests which follow.
\cs_new_protected:Npn \__regex_item_catcode:
  {
    "
    \if_case:w \l__regex_curr_catcode_int
    1 \or: 4 \or: 10 \or: 40
    \or: 100 \or: 1000 \or: 4000
    \or: 10000 \or: 100000 \or: 400000
    \or: 1000000 \or: 4000000 \else: 1*0
    \fi:
  }
\prg_new_protected_conditional:Npp \__regex_item_catcode:n #1 \{ T \}
{
  \if_int_odd:w \__regex_int_eval:w #1 / \__regex_item_catcode: \scan_stop:
    \prg_return_true:
  \else:
    \prg_return_false:
  \fi:
\}
\cs_new_protected:Npn \__regex_item_catcode_reverse:nT \{ #1 \} \{ \__regex_item_reverse:n {#1} \}
(End of definition for \__regex_item_catcode:nT, \__regex_item_catcode_reverse:nT, and \__regex_item_catcode:)

477
This matches an exact ⟨category⟩-⟨character code⟩ pair, or an exact control sequence, more precisely one of several possible control sequences, separated by \scan_stop:.

\cs_new_protected:Npn \_regex_item_exact:nn #1#2 
{ 
  \if_int_compare:w #1 = \l__regex_curr_catcode_int 
  \if_int_compare:w #2 = \l__regex_curr_char_int 
  \exp_after:wN \exp_after:wN \exp_after:wN \_regex_break_true:w 
  \fi: 
  \fi: 
} 
\cs_new_protected:Npn \_regex_item_exact_cs:n #1 
{ 
  \int_compare:nNnTF \l__regex_curr_catcode_int = \c_zero_int 
  { 
    \_kernel_tl_set:Nx \l__regex_internal_a_tl { \scan_stop: \_regex_curr_cs_to_str: \scan_stop: } 
    \tl_if_in:noTF { \scan_stop: #1 \scan_stop: } \l__regex_internal_a_tl 
    { \_regex_break_true:w } { } 
  } 
} 

(End of definition for \_regex_item_exact:nn and \_regex_item_exact_cs:n)

\_regex_item_cs:n Match a control sequence (the argument is a compiled regex). First test the catcode of the current token to be zero. Then perform the matching test, and break if the csname indeed matches.

\cs_new_protected:Npn \_regex_item_cs:n #1 
{ 
  \int_compare:nNnT \l__regex_curr_catcode_int = \c_zero_int 
  { 
    \_regex_single_match: \_regex_disable_submatches: \_regex_build_for_cs:n {#1} 
    \bool_set_eq:NN \l__regex_saved_success_bool \_regex_success_bool 
    \exp_args:Ne \_regex_match_cs:n { \_regex_curr_cs_to_str: } 
    \if_meaning:w \c_true_bool \_regex_success_bool 
      \group_insert_after:N \_regex_break_true:w 
    \fi: 
    \bool_gset_eq:NN \l__regex_saved_success_bool 
    \_regex_success_bool 
    \_regex_saved_success_bool 
    \_regex_saved_success_bool 
    \group_end: 
  } 
} 

(End of definition for \_regex_item_cs:n)

46.2.4 Character property tests

Character property tests for \d, \W, etc. These character properties are not affected by the (?i) option. The characters recognized by each one are as follows: \d=[0-9],
\w=[0-9A-Z_a-z], \s=[\n\t\v\f\r\^I\^J\^L\^M], \h=[\n\t], \v=[\v\f\r\^M], and the upper case counterparts match anything that the lower case does not match. The order in which the various tests appear is optimized for usual mostly lower case letter text.

4487 \cs_new_protected:Npn \__regex_prop_d: \__regex_prop_h: \__regex_prop_s: \__regex_prop_v: \__regex_prop_w: \__regex_prop_N:


POSIX properties. No surprise.
46.2.5 Simple character escape

Before actually parsing the regular expression or the replacement text, we go through them once, converting `\n` to the character 10, etc. In this pass, we also convert any special character (`*`, `?`, `{`, etc.) or escaped alphanumerical character into a marker indicating that this was a special sequence, and replace escaped special characters and non-escaped alphanumerical characters by markers indicating that those were “raw” characters. The rest of the code can then avoid caring about escaping issues (those can become quite complex to handle in combination with ranges in character classes).

Usage: \texttt{\_regex_escape_use:nnnn (inline 1) (inline 2) (inline 3) ((token list))} The (token list) is converted to a string, then read from left to right, interpreting backslashes as escaping the next character. Unescaped characters are fed to the function (inline 1), and escaped characters are fed to the function (inline 2) within an e-expansion context (typically those functions perform some tests on their argument to decide how to output them). The escape sequences `\a`, `\e`, `\f`, `\n`, `\r`, `\t` and `\x` are recognized, and those are replaced by the corresponding character, then fed to (inline 3). The result is then left in the input stream. Spaces are ignored unless escaped.

The conversion is done within an e-expanding assignment.

\texttt{\_regex_escape_use:nnnn} The result is built in `\_regex_internal_a_t`l, which is then left in the input stream. Tracing code is added as appropriate inside this token list. Go through #4 once, applying
#1, #2, or #3 as relevant to each character (after de-escaping it).

```latex
\cs_new_protected:Npn \__regex_escape:nnnn \#1\#2\#3\#4
\group_begin:
\tl_clear:N \l__regex_internal_a_tl
\cs_set:Npn \__regex_escape_unescaped:N ##1 { #1 }
\cs_set:Npn \__regex_escape_escaped:N ##1 { #2 }
\cs_set:Npn \__regex_escape_raw:N ##1 { #3 }
\__regex_standard_escapechar:
\__kernel_tl_gset:Nx \g__regex_internal_tl { \__kernel_str_to_other_fast:n {#4} }
\tl_put_right:Ne \l__regex_internal_a_tl { \exp_after:wN \__regex_escape_loop:N \g__regex_internal_tl
\scan_stop: \prg_break_point: }
\exp_after:wN \group_end:
\l__regex_internal_a_tl
\EndOfDefinitionFor __regex_escape:nnnn
```

\_\_regex_escape_loop:N \_\_regex_escape\_\:\w

\_\_regex_escape_loop:N reads one character: if it is special (space, backslash, or end-marker), perform the associated action, otherwise it is simply an unescaped character. After a backslash, the same is done, but unknown characters are “escaped”.

```
\cs_new:Npn \__regex_escape_loop:N #1
\cs_if_exist_use:cF { __regex_escape_\token_to_str:N #1:w }
\__regex_escape_unescaped:N #1:w
\__regex_escape_loop:N
}
\cs_new:cpn { __regex_escape_\c_backslash_str:w }
\__regex_escape_loop:N #1
\cs_if_exist_use:cF { __regex_escape_\token_to_str:N #1:w }
\__regex_escape_escaped:N #1:w
\__regex_escape_loop:N
}
\EndOfDefinitionFor __regex_escape_loop:N and __regex_escape\_\:\w:
```

\_\_regex_escape_unescaped:N \_\_regex_escape_escaped:N \_\_regex_escape_raw:N

Those functions are never called before being given a new meaning, so their definitions here don’t matter.

```
\cs_new_eq:NN \__regex_escape_unescaped:N ?
\cs_new_eq:NN \__regex_escape_escaped:N ?
\cs_new_eq:NN \__regex_escape_raw:N ?
\EndOfDefinitionFor __regex_escape_unescaped:N, __regex_escape_escaped:N, and __regex_escape_raw:N:
```

\_\_regex_escape\_\scan_stop: v
\_\_regex_escape\_\scan_stop: v
\_\_regex_escape/\a:w
\_\_regex_escape/\e:w
\_\_regex_escape/\f:w
\_\_regex_escape/\n:w
\_\_regex_escape/\r:w
\_\_regex_escape/\t:w
\_\_regex_escape/\u:w

The loop is ended upon seeing the end-marker “break”, with an error if the string ended in a backslash. Spaces are ignored, and \a, \e, \f, \n, \r, \t take their meaning here.
When \textbackslash x is encountered, \texttt{\_regex_escape\_x:\textbackslash w} is responsible for grabbing some hexadecimal digits, and feeding the result to \texttt{\_regex_escape\_x\_end:\textbackslash w}. If the number is too big interrupt the assignment and produce an error, otherwise call \texttt{\_regex_escape\_raw:\textbackslash N} on the corresponding character token.

\begin{verbatim}
\cs_new:cpn { \_regex_escape\_x:\textbackslash w } \_regex_escape\_loop:\textbackslash N
\{
  \exp_after:wN \_regex_escape\_x\_end:\textbackslash w
  \int_value:w "0 \_regex_escape\_x\_test:N
\}
\cs_new:Npn \_regex_escape\_x\_end:\textbackslash w #1 \\
{ \int_compare:nNnTF {#1} > \c_max_char_int
  { \msg_expandable_error:nnff { regex } { x-overflow } {#1} { \int_to.Hex:n {#1} } }
  { \exp_last_unbraced:Nf \_regex_escape\_raw:\textbackslash N
    \char_generate:nn {#1} { 12 } }
}
\end{verbatim}

(End of definition for \_regex_escape\_x:\textbackslash w, \_regex_escape\_x\_end:\textbackslash w, and \_regex_escape\_x\_large:n.)

Find out whether the first character is a left brace (allowing any number of hexadecimal digits), or not (allowing up to two hexadecimal digits). We need to check for the end-of-string marker. Eventually, call either \_regex_escape\_x:\textbackslash loop:N or \_regex_escape\_x:\textbackslash N.

\begin{verbatim}
\cs_new:Npn \_regex_escape\_x\_test:N \_regex_escape\_x\_testii:N
{ \if_meaning:w \scan_stop: \_regex_escape\_x\_test:N \_regex_escape\_x\_testii:N
  \exp_after:wN \use_i:nnn \exp_after:wN ;
\end{verbatim}

482
(End of definition for \__regex_escape_x_test:N and \__regex_escape_x_testii:N.)

\__regex_escape_x:N
This looks for the second digit in the unbraced case.

(End of definition for \__regex_escape_x:N.)

\__regex_escape_x_loop:N
\__regex_escape_x_loop_error:
Grab hexadecimal digits, skip spaces, and at the end, check that there is a right brace, otherwise raise an error outside the assignment.

483
\textdef{\__regex_escape_x_loop:N}
{
  \exp_after:wN \token_if_eq_charcode:NNTF \c_right_brace_str #1
  \{ \__regex_escape_loop:N \}
  \{ \__regex_escape_x_loop_error:n \{#1\} \}
}
\cs_new:Npn \__regex_escape_x_loop_error:n #1
{
  \msg_expandable_error:nnn { regex } { x-missing-rbrace } {#1}
  \__regex_escape_loop:N #1
}

(End of definition for \__regex_escape_x_loop:N and \__regex_escape_x_loop_error::)

\__regex_hexadecimal_use:NTF \textdef{\__regex_hexadecimal_use:NTF}
\cs_new:Npn \__regex_hexadecimal_use:NTF #1
{
  \if_int_compare:w \c_one_int < "1 \token_to_str:N #1 \exp_stop_f:
    \else:n
      \if_case:w
        \__regex_int_eval:w \exp_after:wN ' \token_to_str:N #1 - 'a \scan_stop:
        \else:n
          \exp_after:wN \exp_after:wN \exp_after:wN \use_iii:nnn
          \fi:n
    \fi:n
  \use_i:nn
}

(End of definition for \__regex_hexadecimal_use:NTF.)

\__regex_char_if_alphanumeric:NTF \__regex_char_if_special:NTF
These two tests are used in the first pass when parsing a regular expression. That pass is responsible for finding escaped and non-escaped characters, and recognizing which ones have special meanings and which should be interpreted as “raw” characters. Namely,

- alphanumerics are “raw” if they are not escaped, and may have a special meaning when escaped;
- non-alphanumeric printable ascii characters are “raw” if they are escaped, and may have a special meaning when not escaped;
- characters other than printable ascii are always “raw”.

484
The code is ugly, and highly based on magic numbers and the ascii codes of characters. This is mostly unavoidable for performance reasons. Maybe the tests can be optimized a little bit more. Here, “alphanumeric” means 0–9, A–Z, a–z; “special” character means non-alphanumeric but printable ascii, from space (hex 20) to del (hex 7E).

\prg_new_conditional:Npn \__regex_char_if_special:N \#1 { \TF }
\prg_new_conditional:Npn \__regex_char_if_alphanumeric:N \#1 { \TF }

\prg_return_true:
\prg_return_false:

\prg_new_conditional:Npn \__regex_char_if_special:N \#1 { \TF }
\prg_new_conditional:Npn \__regex_char_if_alphanumeric:N \#1 { \TF }
46.3 Compiling

A regular expression starts its life as a string of characters. In this section, we convert it to internal instructions, resulting in a “compiled” regular expression. This compiled expression is then turned into states of an automaton in the building phase. Compiled regular expressions consist of the following:

- \_regex_class:NnnnN (boolean) {tests} {min} {more} {laziness}
- \_regex_group:nnnN {branches} {min} {more} {laziness}, also \_regex_group_no_capture:nnnN and \_regex_group_resetting:nnnN with the same syntax.
- \_regex_branch:n {contents}
- \_regex_command_K:
  - \_regex_assertion:Nn (boolean) {assertion test}, where the \{assertion test\} is \_regex_b_test: or \_regex_Z_test: or \_regex_A_test: or \_regex_G_test:

Tests can be the following:

- \_regex_item_caseful_equal:n {char code}
- \_regex_item_caseless_equal:n {char code}
- \_regex_item_caseful_range:nn {min} {max}
- \_regex_item_caseless_range:nn {min} {max}
- \_regex_item_catcode:nT {catcode bitmap} {tests}
- \_regex_item_catcode_reverse:nT {catcode bitmap} {tests}
- \_regex_item_exact:nn {catcode} {char code}
- \_regex_item_reverse:n {tests}
- \_regex_item_exact_cs:n {csnames}, more precisely given as (csname) \scan_stop: (csname) \scan_stop: (csname) \scan_stop: (csname) and so on in a brace group.
- \_regex_item_cs:n {compiled regex}
46.3.1 Variables used when compiling

\l__regex_group_level_int

We make sure to open the same number of groups as we close.

\l__regex_mode_int
\l__regex_cs_in_class_mode_int
\l__regex_cs_mode_int
\l__regex_outer_mode_int
\l__regex_catcode_mode_int
\l__regex_class_mode_int
\l__regex_catcode_in_class_mode_int

While compiling, ten modes are recognized, labelled $-63, -23, -6, -2, 0, 2, 3, 6, 23, 63$. See section 46.3.3. We only define some of these as constants.

\l__regex_catcodes_int
\l__regex_default_catcodes_int
\l__regex_catcodes_bool

We wish to allow constructions such as \c[^BE](.. \cL[a-z]..), where the outer catcode test applies to the whole group, but is superseded by the inner catcode test. For this to work, we need to keep track of lists of allowed category codes: \l__regex_catcodes_int and \l__regex_default_catcodes_int are bitmaps, sums of $4^c$, for all allowed catcodes $c$. The latter is local to each capturing group, and we reset \l__regex_catcodes_int to that value after each character or class, changing it only when encountering a \c escape. The boolean records whether the list of categories of a catcode test has to be inverted: compare \c[^BE] and \c[BE].

\l__regex_internal_regex

The compilation step stores its result in this variable.
This sequence holds the prefix that makes up the line displayed to the user. The various items must be removed from the right, which is tricky with a token list, hence we use a sequence.

\seq_new:N \l__regex_show_prefix_seq
(End of definition for \l__regex_show_prefix_seq.)

A hack. To know whether a given class has a single item in it or not, we count the number of lines when showing the class.

\int_new:N \l__regex_show_lines_int
(End of definition for \l__regex_show_lines_int.)

46.3.2 Generic helpers used when compiling

\__regex_two_if_eq:NNNNTF
Used to compare pairs of things like \__regex_compile_special:N ? together. It’s often inconvenient to get the catcodes of the character to match so we just compare the character code. Besides, the expanding behaviour of \if:w is very useful as that means we can use \c_left_brace_str and the like.

\cs_new:Npn \__regex_two_if_eq:NNNNTF #1#2#3#4
{\if_meaning:w #1 #3 \if:w #2 #4 \exp_after:wN \exp_after:wN \exp_after:wN \use_ii:nn \fi:\fi:\use_ii:nn }

(End of definition for \__regex_two_if_eq:NNNNTF.)

\__regex_get_digits:NTFw \__regex_get_digits_loop:w
If followed by some raw digits, collect them one by one in the integer variable #1, and take the true branch. Otherwise, take the false branch.

\cs_new_protected:Npn \__regex_get_digits:NTFw #1#2#3#4#5
{ \__regex_if_raw_digit:NNTF #4#5 { #1 = #5 \__regex_get_digits_loop:nw {#2} } { #3 #4 #5 } }

\cs_new:Npn \__regex_get_digits_loop:nw #1#2#3
{ \__regex_if_raw_digit:NNTF #2#3 { #3 \__regex_get_digits_loop:nw {#1} } { \scan_stop: #1 #2 #3 } }

(End of definition for \__regex_get_digits:NTFw and \__regex_get_digits_loop:w.)

\__regex_if_raw_digit:NNTF
Test used when grabbing digits for the \{m,n\} quantifier. It only accepts non-escaped digits.

\cs_new:Npn \__regex_if_raw_digit:NNTF #1#2
{ \if_meaning:w \__regex_compile_raw:N #1 \if_int_compare:w \c_one_int < 1 #2 \exp_stop_f:
46.3.3 Mode

When compiling the NFA corresponding to a given regex string, we can be in ten distinct modes, which we label by some magic numbers:

-6 \[\text{\ldots}\] control sequence in a class,
-2 \text{\ldots} control sequence,
0 \ldots outer,
2 \text{\ldots} catcode test,
6 \[\text{\ldots}\] catcode test in a class,
-63 \[\text{\ldots}\] class inside mode −6,
-23 \text{\ldots} class inside mode −2,
3 \ldots class inside mode 0,
23 \text{\ldots} class inside mode 2,
63 \[\text{\ldots}\] class inside mode 6.

This list is exhaustive, because \text{\ldots} escape sequences cannot be nested, and character classes cannot be nested directly. The choice of numbers is such as to optimize the most useful tests, and make transitions from one mode to another as simple as possible.

- Even modes mean that we are not directly in a character class. In this case, a left bracket appends 3 to the mode. In a character class, a right bracket changes the mode as \( m \to (m - 15)/13 \), truncated.
- Grouping, assertion, and anchors are allowed in non-positive even modes (0, −2, −6), and do not change the mode. Otherwise, they trigger an error.
- A left bracket is special in even modes, appending 3 to the mode; in those modes, quantifiers and the dot are recognized, and the right bracket is normal. In odd modes (within classes), the left bracket is normal, but the right bracket ends the class, changing the mode from \( m \) to \( (m - 15)/13 \), truncated; also, ranges are recognized.
- In non-negative modes, left and right braces are normal. In negative modes, however, left braces trigger a warning; right braces end the control sequence, going from −2 to 0 or −6 to 3, with error recovery for odd modes.
- Properties (such as the \texttt{\d} character class) can appear in any mode.
\_\_regex\_if\_in\_class:TF  Test whether we are directly in a character class (at the innermost level of nesting). There, many escape sequences are not recognized, and special characters are normal. Also, for every raw character, we must look ahead for a possible raw dash.

```
\prg_new_conditional:Npnn \_\_regex\_if\_in\_class: { TF }
\{ 
  \if_int_odd:w \l\_regex\_mode\_int 
  \prg\_return\_true: 
  \else: 
  \prg\_return\_false: 
  \fi: 
\}
```

*(End of definition for \_\_regex\_if\_in\_class:TF.)*

\_\_regex\_if\_in\_cs:TF  Right braces are special only directly inside control sequences (at the inner-most level of nesting, not counting groups).

```
\cs_new:Npn \_\_regex\_if\_in\_cs:TF 
\{ 
  \if_int_odd:w \l\_regex\_mode\_int 
  \else: 
  \if_int_compare:w \l\_regex\_mode\_int < \c\_regex\_outer\_mode\_int 
  \exp\_after:w\N \exp\_after:w\N \exp\_after:w\N \use\_ii:nnn 
  \fi: 
  \fi: 
  \use\_ii:nn 
  \fi: 
\}
```

*(End of definition for \_\_regex\_if\_in\_cs:TF.)*

\_\_regex\_if\_in\_class\_or\_catcode:TF  Assertions are only allowed in modes 0, −2, and −6, i.e., even, non-positive modes.

```
\cs_new:Npn \_\_regex\_if\_in\_class\_or\_catcode:TF 
\{ 
  \if_int_odd:w \l\_regex\_mode\_int 
  \else: 
  \if_int_compare:w \l\_regex\_mode\_int > \c\_regex\_outer\_mode\_int 
  \else: 
  \exp\_after:w\N \exp\_after:w\N \exp\_after:w\N \use\_iii:nnn 
  \fi: 
  \fi: 
  \use\_i:nn 
  \fi: 
\}
```

*(End of definition for \_\_regex\_if\_in\_class\_or\_catcode:TF.)*

\_\_regex\_if\_within\_catcode:TF  This test takes the true branch if we are in a catcode test, either immediately following it (modes 2 and 6) or in a class on which it applies (modes 23 and 63). This is used to tweak how left brackets behave in modes 2 and 6.

```
\prg_new_conditional:Npnn \_\_regex\_if\_within\_catcode: { TF }
\{ 
  \if_int_compare:w \l\_regex\_mode\_int > \c\_regex\_outer\_mode\_int 
  \prg\_return\_true: 
  \else: 
  \prg\_return\_false: 
  \fi: 
\}
```

490
The \c escape sequence is only allowed in modes 0 and 3, i.e., not within any other \c escape sequence.

\__regex_mode_quit_c:  This function changes the mode as it is needed just after a catcode test.

46.3.4 Framework

\__regex_compile:w  Used when compiling a user regex or a regex for the \c{...} escape sequence within another regex. Start building a token list within a group (with e-expansion at the outset), and set a few variables (group level, catcodes), then start the first branch. At the end, make sure there are no dangling classes nor groups, close the last branch: we are done building \l__regex_internal_regex.
The compilation is done between \__regex_compile:w and \__regex_compile:end:, starting in mode 0. Then \__regex_escape_use:nnnn distinguishes special characters, escaped alphanumerics, and raw characters, interpreting \a, \x and other sequences. The 4 trailing \prg_do_nothing: are needed because some functions defined later look up to 4 tokens ahead. Before ending, make sure that any \c{...} is properly closed. No need to check that brackets are closed properly since \__regex Compile_end: does that. However, catch the case of a trailing \cL construction.
\_\_regex_compile_use:n Use a regex, regardless of whether it is given as a string (in which case we need to compile) or as a regex variable. This is used for \regex_match_case:nn and related functions to allow a mixture of explicit regex and regex variables.

\cs_new_protected:Npm \_\_regex_compile_use:n \#1
\{\tl_if_single_token:nT {#1}\{\exp_after:wN \_\_regex_compile_aux:w \token_to_meaning:N \#1 - \q__regex_nil \} \_\_regex_compile:n \#1 \l__regex_internal_regex \cs_new_protected:Npm \_\_regex_compile_use_aux:w \#1 - \#2 \q__regex_nil \{\str_if_eq:nnT { #1 - } { macro:->\_\_regex_branch:n } \{ \use_ii:nnn \} \}
(End of definition for \_\_regex_compile_use:n)

\_\_regex_compile_escaped:N \_\_regex_compile_special:N If the special character or escaped alphanumeric has a particular meaning in regexes, the corresponding function is used. Otherwise, it is interpreted as a raw character. We distinguish special characters from escaped alphanumeric characters because they behave differently when appearing as an end-point of a range.

\cs_new_protected:Npm \_\_regex_compile_special:N \#1
\{\cs_if_exist_use:cF { \_\_regex_compile_#1: } \{ \_\_regex_compile_raw:N \#1 \} \}
\cs_new_protected:Npm \_\_regex_compile_escaped:N \#1
\{\cs_if_exist_use:cF { \_\_regex_compile_/#1: } \{ \_\_regex_compile_raw:N \#1 \} \}
(End of definition for \_\_regex_compile_special:N)
This is used after finding one “test”, such as \d, or a raw character. If that followed a catcode test (e.g., \cL), then restore the mode. If we are not in a class, then the test is “standalone”, and we need to add \__regex_class:NnnnN and search for quantifiers. In any case, insert the test, possibly together with a catcode test if appropriate.

\__regex_compile_one:n
{\__regex_mode_quit_c:
  \__regex_if_in_class:TF { }
  {\tl_build_put_right:Nn \l__regex_build_tl
   {\__regex_class:NnnnN \c_true_bool \{ \if_false: \fi: \}
    \__regex_item_catcode:nT \int_use:N \l__regex_catcodes_int }
  {\exp_not:N \exp_not:n \{#1\}}
}
\tl_build_put_right:Ne \l__regex_build_tl
{\if_int_compare:w \l__regex_catcodes_int < \c__regex_all_catcodes_int
  \__regex_item_catcode:nT \c\regex_all_catcodes_int
  \{\int_use:N \l__regex_catcodes_int \exp_not:n \{#1\}
  \}
\else:
  \exp_not:N \exp_not:n \{#1\}
  \fi:
  }
\int_set_eq:NN \l__regex_catcodes_int \l__regex_default_catcodes_int
\__regex_if_in_class:TF { } { \__regex_compile_quantifier:w }

(End of definition for \__regex_compile_one:n.)

\__regex_compile_abort_tokens:n This function places the collected tokens back in the input stream, each as a raw character. Spaces are not preserved.

\__regex_compile_abort_tokens:e
{\cs_new_protected:Npn \__regex_compile_abort_tokens:n #1
  \use:e
  \exp_args:No \tl_map_function:nN { \tl_to_str:n \{#1\}}
  \_\regex_compile_raw:N
}
\cs_generate_variant:Nn \__regex_compile_abort_tokens:n { e }

(End of definition for \__regex_compile_abort_tokens:n.)

46.3.5 Quantifiers

This looks ahead and checks whether there are any quantifier (special character equal to either of \?\*\{). This is useful for the \u and \ur escape sequences.

\__regex_compile_if_quantifier:TFw}
{\cs_new_protected:Npn \__regex_compile_if_quantifier:TFw #1#2#3#4
  \token_if_eq_meaning:NNTF #3 \__regex_compile_special:N
  \{\cs_if_exist:cTF { \__regex_compile_quantifier:#4:w } \}
  \{ \use_ii:nn \}
  \{#1\} \{#2\} #3 #4

494
This looks ahead and finds any quantifier (special character equal to either of \+\*).\n\cs_new_protected:Npn \__regex_compile_quantifier:w #1#2\n\{\n  \token_if_eq_meaning:NNTF #1 \__regex_compile_special:N \n  \cs_if_exist_use:cF { \__regex_compile_quantifier_#2:w } \n  \{ \_\_regex_compile_quantifier_none: #1 #2 \} \n  \} \n(End of definition for \_\_regex_compile_quantifier:w.)\n\_\_regex_compile_quantifier_none: \_\_regex_compile_quantifier_abort:eNN\nThose functions are called whenever there is no quantifier, or a braced construction is invalid (equivalent to no quantifier, and whatever characters were grabbed are left raw).\n\cs_new_protected:Npn \__regex_compile_quantifier_none: \n\{ \tl_build_put_right:Nn \l__regex_build_tl \n  \{ \if_false: { \fi: } \{ 1 \} \{ 0 \} \c_false_bool \} \n\} \n\cs_new_protected:Npn \__regex_compile_quantifier_abort:eNN #1#2#3\n\{ \_\_regex_compile_quantifier_none: \n  \msg_warning:nnee { regex } { invalid-quantifier } {#1} {#3} \n  \_\_regex_compileAbort_tokens:e {#1} \n  \{ \_\_regex_compile_quantifier_abort:eNN \} \n}\n(End of definition for \_\_regex_compile_quantifier_none: and \_\_regex_compile_quantifier_abort:eNN.)\n\_\_regex_compile_quantifier_laziness:nnNN\nOnce the “main” quantifier (\?, \+, * or a braced construction) is found, we check whether it is lazy (followed by a question mark). We then add to the compiled regex a closing brace (ending \_\_regex_class:NnnnN and friends), the start-point of the range, its end-point, and a boolean, \texttt{true} for lazy and \texttt{false} for greedy operators.\n\cs_new_protected:Npn \_\_regex_compile_quantifier_laziness:nnNN #1#2#3#4\n\{ \_\_regex_two_if_eq:NNNNTF #3 #4 \_\_regex_compile_special:N ? \n  \{ \tl_build_put_right:Nn \l__regex_build_tl \n    \{ \if_false: { \fi: } \{ #1 \} \{ #2 \} \c_true_bool \} \n  \} \n  \{ \tl_build_put_right:Nn \l__regex_build_tl \n    \{ \if_false: { \fi: } \{ #1 \} \{ #2 \} \c_false_bool \} \n    \_\_regex_compile_quantifier_laziness:nnNN \} \n\}
(End of definition for \_\_regex_compile_quantifier_laziness:nnNN.)
For each “basic” quantifier, ?, *, +, feed the correct arguments to \_\_regex_compile_-
quantifier_laziness:nnNN, -1 means that there is no upper bound on the number of
repetitions.

\cs_new_protected:cpn { __regex_compile_quantifier_?:w }
\cs_new_protected:cpn { __regex_compile_quantifier_?:w } { \__regex_compile_quantifier_laziness:nnNN { 0 } { 1 } }  
\cs_new_protected:cpn { __regex_compile_quantifier_*:w }
\cs_new_protected:cpn { __regex_compile_quantifier_*:w } { \__regex_compile_quantifier_laziness:nnNN { 0 } { -1 } }  
\cs_new_protected:cpn { __regex_compile_quantifier_+:w }
\cs_new_protected:cpn { __regex_compile_quantifier_+:w } { \__regex_compile_quantifier_laziness:nnNN { 1 } { -1 } }  

(End of definition for \_\_regex_compile_quantifier_?:w, \_\_regex_compile_quantifier_*:w, and \_\_regex_compile_quantifier_+:w.)

Three possible syntaxes: {⟨int⟩}, {⟨int⟩}, or {⟨int⟩,⟨int⟩}. Any other syntax causes
us to abort and put whatever we collected back in the input stream, as raw
caracters, including the opening brace. Grab a number into \l\_\_regex_internal_a_int. If the
number is followed by a right brace, the range is \([a, a]\). If followed by a comma, grab one
more number, and call the \_ii or \_iii auxiliary. Those auxiliaries check for a closing
brace, leading to the range \([a, \infty]\) or \([a, b]\), encoded as \{a\}{−1} and \{a\}{b − a}.

\cs_new_protected:cpn { __regex_compile_quantifier_\c_left_brace_str :w }
\cs_new_protected:Npn \__regex_compile_quantifier_braced_auxi:w #1#2
\str_case_e:nnF { #1 #2 }
\__regex_get_digits:NTFw \l\_\_regex_internal_a_int
\{ \__regex_compile_quantifier_braced_auxii:w \}
\{ \__regex_compile_quantifier_abort:eNN { \\c_left_brace_str } \}
\cs_new_protected:Npn \\_\_regex_compile_quantifier_braced_auxii:w #1#2
\str_case_e:nnF { #1 #2 }
\__regex_get_digits:NTFw \l\_\_regex_internal_a_int
\{ \__regex_compile_quantifier_braced_auxii:w \}
\{ \__regex_compile_quantifier_abort:eNN { \\c_left_brace_str \\int_use:N \l\_\_regex_internal_a_int } \}
\cs_new_protected:Npn \\_\_regex_compile_quantifier_braced_auxii:w #1#2
\\_\_regex_two_if_eq:NNNNTF #1 #2 \\__regex_compile_special:N \\c_right_brace_str
\exp_args:No \\_\_regex_compile_quantifier_laziness:nnNN
\{ \\int_use:N \l\_\_regex_internal_a_int \} 0
\}  
\\_\_regex_get_digits:NTFw \l\_\_regex_internal_b_int
\{ \\_\_regex_compile_quantifier_braced_auxiiii:w \}
\{ \\_\_regex_compile_quantifier_braced_auxiiii:w \}
\\}  
\}  
\cs_new_protected:Npn \\_\_regex_compile_quantifier_braced_auxiiii:w #1#2
\str_case_e:nnF { #1 #2 }
\exp_args:No \\_\_regex_compile_special:N \\c_right_brace_str
\\int_use:N \l\_\_regex_internal_a_int \} { -1 }
5100 \{ \\
5101 \_\_regex_compile_quantifier_abort:eNN \\
5102 \{ \c_left_brace_str \int_use:N \l__regex_internal_a_int , \}
5103 #1 #2 \\
5104 \}
5105 \cs_new_protected:Npn \_\_regex_compile_quantifier_braced_auxiii:w #1#2
5106 \{ \\
5107 \_\_regex_two_if_eq:NNNNTF #1 #2 \_\_regex_compile_special:N \c_right_brace_str \\
5108 \{ \\
5109 \if_int_compare:w \l__regex_internal_a_int > \\
5110 \l__regex_internal_b_int \\
5111 \msg_error:nnee { regex } { backwards-quantifier } \\
5112 \{ \int_use:N \l__regex_internal_a_int \}
5113 \{ \int_use:N \l__regex_internal_b_int \}
5114 \int_zero:N \l__regex_internal_b_int \\
5115 \else: \\
5116 \int_sub:Nn \l__regex_internal_b_int \l__regex_internal_a_int \\
5117 \fi: \\
5118 \exp_args:Noo \_\_regex_compile_quantifier_laziness:mmNN \\
5119 \{ \int_use:N \l__regex_internal_a_int \}
5120 \{ \int_use:N \l__regex_internal_b_int \}
5121 \}
5122 \}
5123 \_\_regex_compile_quantifier_abort:eNN \\
5124 \{ \\
5125 \c_left_brace_str \\
5126 \int_use:N \l__regex_internal_a_int , \\
5127 \int_use:N \l__regex_internal_b_int \\
5128 \}
5129 #1 #2 \\
5130 \}
5131 \}
5132 \}

(End of definition for \_\_regex_compile_quantifier_{:w and others.})

\_\_regex_compile_raw_error:N

46.3.6 Raw characters

Within character classes, and following catcode tests, some escaped alphanumeric sequences such as \b do not have any meaning. They are replaced by a raw character, after spitting out an error.

\cs_new_protected:Npm \_\_regex_compile_raw_error:N \#1
5133 \msg_error:nnee { regex } { bad-escape } \#1
5134 \_\_regex_compile_raw:N \#1
5135 \}

(End of definition for \_\_regex_compile_raw_error:N.)

\_\_regex_compile_raw:N

If we are in a character class and the next character is an unescaped dash, this denotes a range. Otherwise, the current character \#1 matches itself.

\cs_new_protected:Npm \_\_regex_compile_raw:N \#1\#2\#3
5136 \_\_regex_if_in_class:TF
5137 
5138 497
We have just read a raw character followed by a dash; this should be followed by an end-point for the range. Valid end-points are: any raw character; any special character, except a right bracket. In particular, escaped characters are forbidden.
\msg_warning:nn { regex } { range-missing-end }
{{#1} \{ \c_backslash_str #3 \}}
\tl_build_put_right:Ne \l__regex_build_tl
{\__regex_item_equal:n { \int_value:w '#1 \exp_stop_f: }
 \__regex_item_equal:n { \int_value:w '- \exp_stop_f: }
}
#2#3

(End of definition for \__regex_compile_range:Nw and \__regex_if_end_range:NNTF.)

46.3.7 Character properties

\__regex_compile_.:  In a class, the dot has no special meaning. Outside, insert \__regex_prop_.:, which
matches any character or control sequence, and refuses −2 (end-marker).
The constants \__regex_prop_d:, etc. hold a list of tests which match the corresponding
character class, and jump to the \__regex_break_point:TF marker. As for a normal
character, we check for quantifiers.
\__regex_compile_/p:w \#1#2
{\cs_set_protected:Npn \__regex_tmp:w #1#2
 {\cs_new_protected:cpe { \__regex_compile_/p:w } \exp_not:cn { \__regex_prop_#1: }
 \cs_new_protected:cpe { \__regex_compile_/p:w } \exp_not:cn { \__regex_prop_#1: }
 \__regex_compile_one:n \exp_not:cn { \__regex_prop_#1: }
 \__regex_compile_one:n \exp_after:wN \__regex_break_true:w
 \fi:
}

(End of definition for \__regex_compile_.: and \__regex_prop_.:)

(End of definition for \__regex_compile_/d: and others.)
46.3.8 Anchoring and simple assertions

In modes where assertions are forbidden, anchors such as \A produce an error (\A is invalid in classes); otherwise they add an `\__regex_assertion:Nn` test as appropriate (the only negative assertion is \B). The test functions are defined later. The implementation for $ and ^ is only different from \A etc because these are valid in a class.

```latex
\cs_new_protected:Npn \__regex_compile_anchor_letter:NNN #1#2#3
\begin{verbatim}
\__regex_if_in_class_or_catcode:TF { \__regex_compile_raw_error:N #1 } \\
{ \tl_build_put_right:Nn \l__regex_build_tl \__regex_assertion:Nn \c_true_bool {#3} }
\end{verbatim}
```

```latex
\cs_new_protected:cpn { __regex_compile_/A: } { \__regex_compile_anchor_letter:NNN A \c_true_bool \__regex_A_test: }
\cs_new_protected:cpn { __regex_compile_/G: } { \__regex_compile_anchor_letter:NNN G \c_true_bool \__regex_G_test: }
\cs_new_protected:cpn { __regex_compile_/Z: } { \__regex_compile_anchor_letter:NNN Z \c_true_bool \__regex_Z_test: }
\cs_new_protected:cpn { __regex_compile_/z: } { \__regex_compile_anchor_letter:NNN z \c_true_bool \__regex_Z_test: }
\cs_new_protected:cpn { __regexCompile_/b: } { \__regex_compile_anchor_letter:NNN b \c_true_bool \__regex_b_test: }
\cs_new_protected:cpn { __regexCompile_/B: } { \__regex_compile_anchor_letter:NNN B \c_false_bool \__regex_b_test: }
\cs_set_protected:Npn \__regex_tmp:w #1#2
\begin{verbatim}
\cs_new_protected:cpn { __regexCompile_: } { \__regex_if_in_class:TF \{ \__regex_compile_raw:N #1 \} \\
{ \tl_build_put_right:Nn \l__regex_build_tl \__regex_assertion:Nn \c_true_bool {#2} }
\end{verbatim}
```

(End of definition for \__regex_compile_anchor_letter:NNN and others.)

46.3.9 Character classes

Outside a class, right brackets have no meaning. In a class, change the mode \( (m \rightarrow (m-15)/13, \text{truncated}) \) to reflect the fact that we are leaving the class. Look for quantifiers, unless we are still in a class after leaving one (the case of \([\ldots]\L[\ldots]\ldots\)). quantifiers.

```latex
\cs_new_protected:cpn { \__regexCompile_: } { \__regex_if_in_class:TF \\
{ \tl_build_put_right:Nn \l__regex_build_tl \__regex_assertion:Nn \c_true_bool {#2} }
\end{verbatim}
```

(End of definition for \__regex_compile_anchor_letter:NNN and others.)
In a class, left brackets might introduce a POSIX character class, or mean nothing. Immediately following \c{category}, we must insert the appropriate catcode test, then parse the class; we pre-expand the catcode as an optimization. Otherwise (modes 0, \texttt{--2} and \texttt{--6}) just parse the class. The mode is updated later.

\cs_new_protected:cpn { __regex_compile_[: ] }
\begin{verbatim}
\__regex_if_in_class:TF
  { \__regex_compile_class_posix_test:w }
\__regex_if_within_catcode:TF
  { \exp_after:wN \__regex_compile_class_catcode:w
    \int_use:N \l__regex_catcodes_int ;
  }
\__regex_compile_class_normal:w
\end{verbatim}

\cs_new_protected:Npn \__regex_compile_class_normal:w
\begin{verbatim}
\__regex_compile_class:TFNN
  { \__regex_class:NnnnN \c_true_bool }
  { \__regex_class:NnnnN \c_false_bool }
\end{verbatim}

\cs_new_protected:Npn \__regex_compile_class_catcode:w \l__regex_catcodes_int ;
\begin{verbatim}
\if_int_compare:w \l__regex_mode_int = \c__regex_catcode_mode_int
  \tl_build_put_right:Nn \l__regex_build_tl
  \__regex_compile_quantifier:w
\exp_after:wN \__regex_compile_raw:N \] }
\fi:
\tex_advance:D \l__regex_mode_int - 15 \exp_stop_f:
\tex_divide:D \l__regex_mode_int 13 \exp_stop_f:
\if_int_odd:w \l__regex_mode_int \else:
  \exp_after:wN \__regex_compile_quantifier:w
  \fi:
\}
{ \__regex_compile_raw:N \] }
\end{verbatim}

(End of definition for \__regex_compile_[: ])

\__regex_compile_class:_ In the “normal” case, we insert \__regex_class:NnnnN \langle boolean \rangle in the compiled code. The \langle boolean \rangle is true for positive classes, and false for negative classes, characterized by a leading \texttt{-}. The auxiliary \__regex_compile_class:TFNN also checks for a leading \texttt{]} which has a special meaning.

\cs_new_protected:Npm \__regex_compile_class_normal:w
\begin{verbatim}
\__regex_compile_class:TFNN
  { \__regex_class:NnnnN \c_true_bool }
  { \__regex_class:NnnnN \c_false_bool }
\end{verbatim}

(End of definition for \__regex_compile_class_normal:w)

\__regex_compile_class_catcode: In this function is called for a left bracket in modes 2 or 6 (catcode test, and catcode test within a class). In mode 2 the whole construction needs to be put in a class (like single character). Then determine if the class is positive or negative, inserting \__regex_item_catcode:nT or the reverse variant as appropriate, each with the current catcodes bitmap \#1 as an argument, and reset the catcodes.

\cs_new_protected:Npm \__regex_compile_class_catcode:w \#1;
\begin{verbatim}
\if_int_compare:w \l__regex_mode_int = \c__regex_catcode_mode_int
  \tl_build_put_right:Nn \l__regex_build_tl
\fi:
\end{verbatim}

501
If the first character is `^`, then the class is negative (use \#2), otherwise it is positive (use \#1). If the next character is a right bracket, then it should be changed to a raw one.

Here we check for a syntax such as `[:alpha:]`. We also detect `=` and `.` which have a meaning in POSIX regular expressions, but are not implemented in l3regex. In case we see `[:`, grab raw characters until hopefully reaching `]:`. If that’s missing, or the POSIX class is unknown, abort. If all is right, add the test to the current class, with an extra \__regex_item_reverse:n for negative classes (we make sure to wrap its argument in braces otherwise \regex_show:N would not recognize the regex as valid).

Here we check for a syntax such as `[:alpha:]`. We also detect `= and `. which have a meaning in POSIX regular expressions, but are not implemented in l3regex. In case we see `[:`, grab raw characters until hopefully reaching `]:`. If that’s missing, or the POSIX class is unknown, abort. If all is right, add the test to the current class, with an extra \__regex_item_reverse:n for negative classes (we make sure to wrap its argument in braces otherwise \regex_show:N would not recognize the regex as valid).
\__regex_compile_raw:N \[ #1 #2
\]
\cs_new_protected:Npn \__regex_compile_class_posix:NNNNw #1#2#3#4#5#6
{\__regex_two_if_eq:NNNNTF #5 #6 \__regex_compile_special:N ^
{\bool_set_false:N \l__regex_internal_bool
 \__kernel_tl_set:Nx \l__regex_internal_a_tl { \if_false: } \fi:
 \__regex_compile_class_posix_loop:w
}
{\bool_set_true:N \l__regex_internal_bool
 \__kernel_tl_set:Nx \l__regex_internal_a_tl { \if_false: } \fi:
 \__regex_compile_class_posix_loop:w #5 #6}
}
\cs_new:Npn \__regex_compile_class_posix_loop:w #1#2
{\token_if_eq_meaning:NNTF \__regex_compile_raw:N #1
{ #2 \__regex_compile_class_posix_loop:w }
{ \if_false: { \fi: } \__regex_compile_class_posix_end:w #1 #2 }
}
\cs_new_protected:Npn \__regex_compile_class_posix_end:w #1#2#3#4
{\__regex_two_if_eq:NNNNTF #1 #2 \__regex_compile_special:N :
{ \__regex_two_if_eq:NNNNTF #3 #4 \__regex_compile_special:N ]
}{ \use_ii:nn }
{\cs_if_exist:cTF { __regex_posix_ \l__regex_internal_a_tl : }
{ \__regex_compile_one:n
{ \bool_if:NTF \l__regex_internal_bool \use:n \__regex_item_reverse:n
{ \exp_not:c { __regex_posix_ \l__regex_internal_a_tl : } }
}
}
{\msg_warning:nne { regex } { posix-unknown }
{ \l__regex_internal_a_tl }
\__regex_compile_abort_tokens:e
{ [: \bool_if:NF \l__regex_internal_bool { ^ }
 \l__regex_internal_a_tl ]
}
}
{\msg_error:nnee { regex } { posix-missing-close }
{ [: \l__regex_internal_a_tl ] \{ #2 #4 }
\__regex_compile_abort_tokens:e { [: \l__regex_internal_a_tl }
#1 #2 #3 #4
46.3.10 Groups and alternations

The contents of a regex group are turned into compiled code in \l__regex_build_tl, which ends up with items of the form \l__regex_branch:n {⟨concatenation⟩}. This construction is done using \tl_build... functions within a \TeX group, which automatically makes sure that options (case-sensitivity and default catcode) are reset at the end of the group. The argument #1 is \l__regex_group:nnnN or a variant thereof. A small subtlety to support \cL(abc) as a shorthand for \cL(a\cLb\cLc): exit any pending catcode test, save the category code at the start of the group as the default catcode for that group, and make sure that the catcode is restored to the default outside the group.

\cs_new_protected:Npn \__regex_compile_group_begin:N #1
\tl_build_put_right:Nn \l__regex_build_tl { #1 { \if_false: } \fi: }
\__regex_mode_quit_c:
\group_begin:
\tl_build_begin:N \l__regex_build_tl
\int_set_eq:NN \l__regex_default_catcodes_int \l__regex_catcodes_int
\int_incr:N \l__regex_group_level_int
\tl_build_put_right:Nn \l__regex_build_tl
{ \l__regex_branch:n { \if_false: } \fi: }
\cs_new_protected:Npn \__regex_compile_group_end:
{ \if_int_compare:w \l__regex_group_level_int > \c_zero_int
\tl_build_put_right:Nn \l__regex_build_tl { \if_false: { \fi: } }
\tl_build_end:N \l__regex_build_tl
\exp_args:NNNe \group_end:
\tl_build_put_right:Nn \l__regex_build_tl { \l__regex_build_tl }
\int_set_eq:NN \l__regex_catcodes_int \l__regex_default_catcodes_int
\exp_after:wN \__regex_compile_quantifier:w
\else:
\msg_warning:nn { regex } { extra-rparen }
\exp_after:wN \__regex_compile_raw:N \exp_after:wN )
\fi: }

\cs_new_protected:Npn \__regex_compile_group_begin:N #1
\tl_build_put_right:Nn \l__regex_build_tl { #1 { \if_false: } \fi: }
\__regex_mode_quit_c:
\group_begin:
\tl_build_begin:N \l__regex_build_tl
\int_set_eq:NN \l__regex_default_catcodes_int \l__regex_catcodes_int
\int_incr:N \l__regex_group_level_int
\tl_build_put_right:Nn \l__regex_build_tl
{ \l__regex_branch:n { \if_false: } \fi: }
\cs_new_protected:Npn \__regex_compile_group_end:
{ \if_int_compare:w \l__regex_group_level_int > \c_zero_int
\tl_build_put_right:Nn \l__regex_build_tl { \if_false: { \fi: } }
\tl_build_end:N \l__regex_build_tl
\exp_args:NNNe \group_end:
\tl_build_put_right:Nn \l__regex_build_tl { \l__regex_build_tl }
\int_set_eq:NN \l__regex_catcodes_int \l__regex_default_catcodes_int
\exp_after:wN \__regex_compile_quantifier:w
\else:
\msg_warning:nn { regex } { extra-rparen }
\exp_after:wN \__regex_compile_raw:N \exp_after:wN )
\fi: }

\__regex_compile_(:

\cs_new_protected:cpn { __regex_compile_(:
{ \if_int_compare:w \l__regex_group_level_int > \c_zero_int
\tl_build_put_right:Nn \l__regex_build_tl { \if_false: { \fi: } }
\tl_build_end:N \l__regex_build_tl
\exp_after:wN \__regex_compile_quantifier:w
\else:
\msg_warning:nn { regex } { extra-rparen }
\exp_after:wN \__regex_compile_raw:N \exp_after:wN )
\fi: }

In a class, parentheses are not special. In a catcode test inside a class, a left parenthesis gives an error, to catch \a\cL(bcd)e. Otherwise check for a ?, denoting special groups, and run the code for the corresponding special group.

\cs_new_protected:cpn { __regex_compile_(:
{ \if_int_compare:w \l__regex_group_level_int > \c_zero_int
\tl_build_put_right:Nn \l__regex_build_tl { \if_false: { \fi: } }
\tl_build_end:N \l__regex_build_tl
\exp_after:wN \__regex_compile_quantifier:w
\else:
\msg_warning:nn { regex } { extra-rparen }
\exp_after:wN \__regex_compile_raw:N \exp_after:wN )
\fi: }
\c_regex_catcode_in_class_mode_int
\msg_error:nn { regex } { c-lparen-in-class }
\exp_after:wN \__regex_compile_raw:N \exp_after:wN ( \else:
\exp_after:wN \__regex_compile_lparen:w \fi:
\}
\cs_new_protected:Npn \__regex_compile_lparen:w #1#2#3#4
{ \__regex_two_if_eq:NNNNTF #1 #2 \__regex_compile_special:N ?
{ \cs_if_exist_use:cF
{ \__regex_compile_special_group\token_to_str:N #4 :w }
{ \msg_warning:nne { regex } { special-group-unknown }
{ (? #4 }
\__regex_compile_group_begin:N \__regex_group:nnnN
\__regex_compile_raw:N ? #3 #4
\}
\}
\__regex_compile_group_begin:N \__regex_group:nnnN
#1 #2 #3 #4
}
(End of definition for \__regex_compile_(:) )
\__regex_compile_|: In a class, the pipe is not special. Otherwise, end the current branch and open another one.
\cs_new_protected:cpn { \__regex_compile_|: } { \__regex_if_in_class:TF { \__regex_compile_raw:N | }
{ \tl_build_put_right:Nn \l__regex_build_tl \__regex_build_tl
{ \if_false: { \fi: } \__regex_branch:n { \if_false: } \fi: }
}
(End of definition for \__regex_compile_|: )
\__regex_compile_): Within a class, parentheses are not special. Outside, close a group.
\cs_new_protected:cpn { \__regex_compile_): } { \__regex_if_in_class:TF { \__regex_compile_raw:N ) }
{ \__regex_compile_group_end: }
}
(End of definition for \__regex_compile_): )
\__regex_compile_special_group::w \__regex_compile_special_group_|:w Non-capturing, and resetting groups are easy to take care of during compilation; for those groups, the harder parts come when building.
\cs_new_protected:cpn { \__regex_compile_special_group::w } { \__regex_compile_group_begin:N \__regex_group_no_capture:nnnN }
\cs_new_protected:cpn { \__regex_compile_special_group_|:w }
\texttt{\cs_new_protected:cpn \{ __regex_compile_special_group_|:w \}}

{ \__regex_compile_group_begin:N \__regex_group_resetting:nnnN }

\textit{(End of definition for \__regex_compile_special_group_::w and \__regex_compile_special_group_-:/w.)}

\texttt{\_regex_compile_special_group_i:w \_regex_compile_special_group_-:w}

The match can be made case-insensitive by setting the option with \texttt{(?!)}; the original behaviour is restored by \texttt{(?-i)}. This is the only supported option.

\texttt{\cs_new_protected:Npn \__regex_compile_special_group_i:w \#1\#2}

\begin{verbatim}
\{ \__regex_two_if_eq:NNNNTF \#1 \#2 \__regex_compile_special:N \}
\end{verbatim}

\texttt{\__regex_compile_/c: \__regex_compile_c_test:NN}

The \texttt{\c} escape sequence can be followed by a capital letter representing a character category, by a left bracket which starts a list of categories, or by a brace group holding a regular expression for a control sequence name. Otherwise, raise an error.

\texttt{\cs_new_protected:cpn \{ __regex_compile_/c: \}}

\textbf{46.3.11 Catcodes and csnames}

The \texttt{\c} escape sequence can be followed by a capital letter representing a character category, by a left bracket which starts a list of categories, or by a brace group holding a regular expression for a control sequence name. Otherwise, raise an error.
\_\_regex_compile_c_C:NN

If \texttt{cC} is not followed by \ or (...) then complain because that construction cannot
match anything, except in cases like \texttt{cC[...]}, where it has no effect.

\_\_regex_compile_c_[w

When encountering \texttt{c[}, the task is to collect uppercase letters representing character
categories. First check for ~ which negates the list of category codes.
The case of a left brace is easy, based on what we have done so far: in a group, compile the regular expression, after changing the mode to forbid nesting \c. Additionally, disable submatch tracking since groups don’t escape the scope of \c{...}. 

\cs_new_protected:cpn { __regex_compile_c_\c_left_brace_str :w }
\___regex_compile_w
\_\_regex_disable_submatches:
\l__regex_mode_int
  = \if_case:w \l__regex_mode_int
     \c\_\_regex_cs_mode_int
  \else:
     \c\_\_regex_cs_in_class_mode_int
  \fi:
\}

(End of definition for \_\_regex_compile_c_{:}
)

\_\_regex_compile_{:}
We forbid unescaped left braces inside a \c{...} escape because they otherwise lead to the confusing question of whether the first right brace in \c{{}x} should end \c or whether one should match braces.

\cs_new_protected:cpn { \_\_regex_compile_ \c_left_brace_str : }
\{
  \__regex_if_in_cs:TF
  { \msg_error:nnn { regex } { cu-lbrace } { c } }
  { \exp_after:wN \__regex_compile_raw:N \c_left_brace_str }
\}

(End of definition for \_\_regex_compile_{:}
)

\l__regex_cs_flag
\_\_regex_compile_end_cs:
\_\_regex_compile_cs_aux:Nn
\_\_regex_compile_cs_aux:Nn
Non-escaped right braces are only special if they appear when compiling the regular expression for a csname, but not within a class: \c{[{}]} matches the control sequences \{ and \}. So, end compiling the inner regex (this closes any dangling class or group). Then insert the corresponding test in the outer regex. As an optimization, if the control sequence test simply consists of several explicit possibilities (branches) then use \_\_regex_item_exact_cs:n with an argument consisting of all possibilities separated by \scan_stop:.

\cs_new_protected:cpm { \_\_regex_compile_ \c_right_brace_str : }
\{
  \__regex_if_in_cs:TF
  { \__regex_compile_end_cs: }
  { \exp_after:wN \__regex_compile_raw:N \c_right_brace_str }
\}
\cs_new_protected:Npm \_\_regex_compile_end_cs:
\{
  \__regex_compile_end:
  \flag_clear:N \l__regex_cs_flag
  \_\_kernel_tl_set:Nx \l__regex_internal_a_tl
    \exp_after:wN \__regex_compile_cs_aux:Nn \l__regex_internal_regex
      \q__regex_nil \q__regex_nil \q__regex_recursion_stop
  \exp_args:Ne \_\_regex_compile_one:n
    \flag_if_raised:NTF \l__regex_cs_flag
      \_\_regex_item_cs:n \exp_not:o \l__regex_internal_regex
    \_\_regex_item_exact_cs:n

509
46.3.12 Raw token lists with $u$

$\__regex_compile_/u$: The $u$ escape is invalid in classes and directly following a catcode test. Otherwise test for a following $r$ (for $\ur$), and call an auxiliary responsible for finding the variable name.
This enforces the presence of a left brace, then starts a loop to find the variable name.

\cs_new:Npn \__regex_compile_u_brace:NNN #1#2#3
\{\__regex_two_if_eq:NNNNTF #2 #3 \__regex_compile_special:N \c_left_brace_str
\{\tl_set:Nn \l__regex_internal_b_tl {#1}\__kernel_tl_set:Nx \l__regex_internal_a_tl { \if_false: } \fi:\__regex_compile_u_loop:NN
\}
\msg_error:nn { regex } { u-missing-lbrace }
\token_if_eq_meaning:NNTF #1 \__regex_compile_raw:N #2 #3
\{\__regex_compile_u_loop:NN\}
\}
\}
\}
\}
(End of definition for \__regex_compile_u_brace:NNN.)

\__regex_compile_u_loop:NN

We collect the characters for the argument of \textbackslash u within an \texttt{e}-expanding assignment. In principle we could just wait to encounter a right brace, but this is unsafe: if the right brace was missing, then we would reach the end-markers of the regex, and continue, leading to obscure fatal errors. Instead, we only allow raw and special characters, and stop when encountering a special right brace, any escaped character, or the end-marker.

\cs_new:Npn \__regex_compile_u_loop:NN #1#2
\{\token_if_eq_meaning:NNTF #1 \__regex_compile_raw:N #2 \__regex_compile_special:N
\{\token_if_eq_charcode:NNTF \c_right_brace_str #2 \__regex_compile_special:N \msg_expandable_error:nnn { regex } { cu-lbrace } \}
\else:
\__regex_compile_u_loop:NN
\}
\}
\}
\}
\}
(End of definition for \__regex_compile_u_loop:NN.)

511
For the \ur{...} construction, once we have extracted the variable's name, we replace all groups by non-capturing groups in the compiled regex (passed as the argument of \_regex_compile_ur:n). If that has a single branch (namely \tl_if_empty:oTF is false) and there is no quantifier, then simply insert the contents of this branch (obtained by use_ii:nn, which is expanded later). In all other cases, insert a non-capturing group and look for quantifiers to determine the number of repetition etc.

Once we have extracted the variable's name, we check for quantifiers, in which case we set up a non-capturing group with a single branch. Inside this branch (we omit it and the group if there is no quantifier), \_regex_compile_u_payload: puts the right tests corresponding to the contents of the variable, which we store in \_regex_internal_a_tl. The behaviour of \u then depends on whether we are within a \c{...} escape (in this case, the variable is turned to a string), or not.
\__regex_branch:n \{ \iffalse \fi: \\
    \__regex_compile_u_payload: \\
    \tl_build_put_right:Nn \l__regex_build_tl \{ \iffalse \{ \fi: \} \\
    \__regex_compile_quantifier:w \\
    \}
\} \{ \__regex_compile_u_payload: \}
\cs_new_protected:Npn \__regex_compile_u_payload: 
\{ \\
    \tl_set:Nv \l__regex_internal_a_tl \{ \l__regex_internal_a_tl \}
\if_int_compare:w \l__regex_mode_int = \c__regex_outer_mode_int \\
    \__regex_compile_u_not_cs: \\
    \else: \\
    \__regex_compile_u_in_cs: \\
\fi: \\
\}
\cs_new_protected:Npm \__regex_compile_u_end: \and \__regex_compile_u_payload:.
\__regex_compile_u_end: \(End of definition for \__regex_compile_u_end: and \__regex_compile_u_payload:.)
\__regex_compile_u_in_cs: \When \u appears within a control sequence, we convert the variable to a string with escaped spaces. Then for each character insert a class matching exactly that character, once.
\cs_new_protected:Npm \__regex_compile_u_in_cs: 
\{ \\
    \tl_analysis_map_inline:Nn \l__regex_internal_a_tl \\
    \{ \\
    \tl_build_put_right:Ne \l__regex_build_tl \\
    \{ \\
    \tl_map_function:NN \g__regex_internal_tl \__regex_compile_u_in_cs_aux:n \\
    \}
\}
\cs_new:Npm \__regex_compile_u_in_cs_aux:n \#1 
\{ \\
    \__regex_class:NnnnN \c_true_bool \{ \__regex_item_caseful_equal:n \{ \int_value:w '#1 \} \} \\
    \{ 0 \} \c_false_bool
\}
(End of definition for \__regex_compile_u_in_cs:.)
\__regex_compile_u_not_cs: \In mode 0, the \u escape adds one state to the NFA for each token in \l__regex_internal_a_tl. If a given \texttt{token} is a control sequence, then insert a string comparison test, otherwise, \__regex_item_exact:nn which compares catcode and character code.
\cs_new_protected:Npm \__regex_compile_u_not_cs: 
\{ \\
    \tl_analysis_map_inline:Nn \l__regex_internal_a_tl \\
    \{ \\
    \tl_build_put_right:Ne \l__regex_build_tl \\
    \}
\}
513
\_\_regex_class:NnnnN \c_true_bool
{
    \if_int_compare:w "##3 = \c_zero_int
    \_\_regex_item_exact_cs:n
    { \exp_after:wN \cs_to_str:N ##1 }
    \else:
    \_\_regex_item_exact:nn { \int_value:w "##3 } { ##2 }
    \fi:
}
{ 1 } { 0 } \c_false_bool
}
}
}
\end{definition}

\_\_regex_compile_/K:
The \K control sequence is currently the only “command”, which performs some action, rather than matching something. It is allowed in the same contexts as \b. At the compilation stage, we leave it as a single control sequence, defined later.

\begin{verbatim}
cs_new_protected:cpn { \_\_regex_compile_/K: }
{
    \int_compare:nNnTF \l__regex_mode_int = \c__regex_outer_mode_int
        { \tl_build_put_right:Nn \l__regex_build_tl { \__regex_command_K: } }
    { \__regex_compile_raw_error:N K }
}
\end{verbatim}

\begin{definition}
\_\_regex_clean_bool:n
\_\_regex_clean_int:n
\_\_regex_clean_int_aux:N
\_\_regex_clean_regex:n
\_\_regex_clean_regex_loop:w
\_\_regex_clean_branch:n
\_\_regex_clean_overוכח:n
\_\_regex_clean_class:n
\_\_regex_clean_class_loop:nnn
\_\_regex_clean_exact_cs:n
\_\_regex_clean_exact_cs:w
Before showing a regex we check that it is “clean” in the sense that it has the correct internal structure. We do this (in the implementation of \regex_show:N and \regex_-_log:N) by comparing it with a cleaned-up version of the same regex. Along the way we also need similar functions for other types: all \_\_regex_clean_{⟨type⟩}:n functions produce valid \langle type⟩ tokens (bool, explicit integer, etc.) from arbitrary input, and the output coincides with the input if that was valid.

\begin{verbatim}
cs_new:Npn \_\_regex_clean_bool:n #1
{
    \tl_if_single:nTF {#1}
        { \bool_if:NTF #1 \c_true_bool \c_false_bool }
    { \c_true_bool }
}
cs_new:Npn \_\_regex_clean_int:n #1
{
    \tl_if_head_eq_meaning:nNTF {#1} -
        { - \exp_args:No \_\_regex_clean_int:n { \use_none:n #1 } }
    { \int_eval:n { 0 \str_map_function:nN {#1} \_\_regex_clean_int_aux:N } }
}
cs_new:Npn \_\_regex_clean_int_aux:N #1
{
    \if_int_compare:w \c_one_int < 1 #1 ~
}\end{verbatim}

514
#1
\else:
    \str_map_break:n
\fi:

\cs_new:Npn \__regex_clean_regex:n #1
{\
    \__regex_clean_regex_loop:w #1
    \__regex_branch:n \{ q_recursion_tail \} \q_recursion_stop
}

\cs_new:Npn \__regex_clean_regex_loop:w #1 \__regex_branch:n #2
{\
    \quark_if_recursion_tail_stop:n {#2}
    \__regex_branch:n \{ \__regex_clean_branch:n {#2} \}
    \__regex_clean_regex_loop:w
}

\cs_new:Npn \__regex_clean_branch:n #1
{\
    \__regex_clean_branch_loop:n #1
    ? ? ? ? ? \prg_break_point:
}

\cs_new:Npn \__regex_clean_branch_loop:n #1
{\
    \tl_if_single:nF {#1} \prg_break:
      \token_case_meaning:NnF #1
      {\
        \__regex_command_K: { #1 \__regex_clean_branch_loop:n }
        \__regex_assertion:Nn { #1 \__regex_clean_assertion:Nn }
        \__regex_class:NnnnN { #1 \__regex_clean_class:NnnnN }
        \__regex_group:nnnN { #1 \__regex_clean_group:nnnN }
        \__regex_group_no_capture:nnnN { #1 \__regex_clean_group:nnnN }
        \__regex_group_resetting:nnnN { #1 \__regex_clean_group:nnnN }
      }
    \prg_break:
}

\cs_new:Npn \__regex_clean_assertion:Nn #1#2
{\
    \__regex_clean_bool:n {#1}
    \tl_if_single:nF {#2} { { \__regex_A_test: } \prg_break: }
    \token_case_meaning:NnTF #2
    {\
        \__regex_A_test: { }
        \__regex_G_test: { }
        \__regex_Z_test: { }
        \__regex_b_test: { }
    }
    { (#2) }
    { { \__regex_A_test: } \prg_break: }
    \__regex_clean_branch_loop:n
}

\cs_new:Npn \__regex_clean_class:NnnnN #1#2#3#4#5
{\
    \__regex_clean_bool:n {#1}
    \{ \__regex_clean_class:n {#2} \}
}

515
When cleaning a class there are many cases, including a dozen or so like `\_\_regex_prop_d:` or `\_\_regex_posix_alpha:`. To avoid listing all of them we allow any command that starts with the 13 characters `\_\_regex_prop_` or `\_\_regex_posix` (handily these have the same length, except for the trailing underscore).

When cleaning a class there are many cases, including a dozen or so like `\_\_regex_prop_d:` or `\_\_regex_posix_alpha:`. To avoid listing all of them we allow any command that starts with the 13 characters `\_\_regex_prop_` or `\_\_regex_posix` (handily these have the same length, except for the trailing underscore).
\exp_args:N \str_case:nnTF

\exp_args:N \str_range:nnn
  \cs_to_str:N \c_one_int { 13 }
\exp_args:N \str_range:nnn
  \{ \__regex_prop_ \} \{ \__regex_posix \} \{ \}
\exp_args:N \str_range:nnn
  \{ #1 \}
\__regex_clean_class_loop:nnn {#2} {#3}
\prg_break:
\endgroup
\endgroup
\cs_new:Npn \__regex_clean_exact_cs:n #1
  \exp_last_unbraced:Nf \use_none:n
  \{ \__regex_clean_exact_cs:w #1 \scan_stop: \q_recursion_tail \scan_stop:
  \q_recursion_stop
\}
\cs_new:Npn \__regex_clean_exact_cs:w #1 \scan_stop:
  \quark_if_recursion_tail_stop:n {#1}
  \scan_stop: \tl_to_str:n {#1}
  \__regex_clean_exact_cs:w
\endgroup
\endgroup
\cs_new:Npn \__regex_show:N #1
  \group_begin:
  \tl_build_begin:N \l__regex_build_tl
  \cs_set_protected:Npn \__regex_branch:n
    \seq_pop_right:NN \l__regex_show_prefix_seq
    \l__regex_internal_a_tl
    \__regex_show_one:n \{ +-branch \}
  \seq_put_right:No \l__regex_show_prefix_seq
  \l__regex_internal_a_tl
  \use:n
\endgroup
\endgroup
\cs_set_protected:Npn \__regex_group:nnnN
  \{ \__regex_show_group_aux:nnnnN \} \}
\endgroup
\endgroup
(End of definition for \__regex_clean_bool:n and others.)
\__regex_show:N

Within a group and within \tl_build_begin:N ... \tl_build_end:N we redefine all
the function that can appear in a compiled regex, then run the regex. The result stored
in \l__regex_internal_a_tl is then meant to be shown.
\cs_new_protected:Npn \__regex_show:N #1
  \group_begin:
  \tl_build_begin:N \l__regex_build_tl
  \cs_set_protected:Npn \__regex_branch:n
    \seq_pop_right:NN \l__regex_show_prefix_seq
    \l__regex_internal_a_tl
    \__regex_show_one:n \{ +-branch \}
  \seq_put_right:No \l__regex_show_prefix_seq
  \l__regex_internal_a_tl
  \use:n
\endgroup
\endgroup
\cs_set_protected:Npn \__regex_group:nnnN
  \{ \__regex_show_group_aux:nnnnN \} \}

517
Show a single character, together with its ascii representation if available. This could be
extended to beyond ascii. It is not ideal for parentheses themselves.

```
\cs_new:Npn \__regex_show_char:n #1
{ \int_eval:n {#1} \int_compare:nT { 32 <= #1 <= 126 } { \char_generate:nn {#1} {12} } }
(End of definition for \__regex_show_char:n)
\__regex_show_one:n
Every part of the final message go through this function, which adds one line to the output, with the appropriate prefix.

```
\cs_new_protected:Npn \__regex_show_one:n #1
{ \int_incr:N \l__regex_show_lines_int \tl_build_put_right:Ne \l__regex_build_tl { \exp_not:N \iow_newline: \seq_map_function:NN \l__regex_show_prefix_seq \use:n #1 \l__regex_internal_a_tl } }
(End of definition for \__regex_show_one:n)
\__regex_show_push:n
\__regex_show_pop:
\__regex_show_scope:nn
Enter and exit levels of nesting. The scope function prints its first argument as an “introduction”, then performs its second argument in a deeper level of nesting.

```
\cs_new_protected:Npn \__regex_show_push:n #1
{ \seq_put_right:Ne \l__regex_show_prefix_seq { #1 ~ } }
\cs_new_protected:Npn \__regex_show_pop: \cs_new_protected:Npn \__regex_show_scope:nn #1#2
{ \__regex_show_one:n {#1} \__regex_show_push:n { ~ } \use_ii:nn #2 \__regex_show_pop: \__regex_show_one:n \l__regex_examplenum } \__regex_show_pop:
(End of definition for \__regex_show_push:n, \__regex_show_pop:, and \__regex_show_scope:nn)
\__regex_show_group_aux:nnnnN
We display all groups in the same way, simply adding a message, (no capture) or (resetting), to special groups. The odd \use_ii:nn avoids printing a spurious +-branch for the first branch.

```
\cs_new_protected:Npn \__regex_show_group_aux:nnnnN #1#2#3#4#5
{ \__regex_show_one:n { , -group-begin #1 } \__regex_show_push:n { | } \use_ii:nn #2 \__regex_show_pop: \__regex_show_one:n { , -group-end \__regex_msg_repeated:nnN {#3} {#4} #5 } }
(End of definition for \__regex_show_group_aux:nnnnN)
```
I'm entirely unhappy about this function: I couldn't find a way to test if a class is a single
test. Instead, collect the representation of the tests in the class. If that had more than
one line, write Match or Don't match on its own line, with the repeating information if
any. Then the various tests on lines of their own, and finally a line. Otherwise, we need
to evaluate the representation of the tests again (since the prefix is incorrect). That's
chunky, but not too expensive, since it's only one test.

\begin{verbatim}
cs_new:Npn \__regex_show_class:NnnnN #1#2#3#4#5
  {
    \group_begin:
    \tl_build_begin:N \l__regex_build_tl
    \int_zero:N \l__regex_show_lines_int
    \__regex_show_push:n {~}
    #2
    \int_compare:nTF { \l__regex_show_lines_int = \c_zero_int }
      {
        \group_end:
        \__regex_show_one:n { \bool_if:NTF #1 { Fail } { Pass } }
      }
      {
        \bool_if:nTF \#1 && \int_compare:p:n { \l__regex_show_lines_int = \c_one_int } { #2 }
          {
            \group_end:
            \tl_build_put_right:Nn \l__regex_build_tl { \__regex_msg_repeated:nnN {#3} {#4} #5 }
          }
          {
            \tl_build_end:N \l__regex_build_tl
            \exp_args:NNNo
            \group_end:
            \tl_set:Nn \l__regex_internal_a_tl \l__regex_build_tl
            \__regex_show_one:n
              { \bool_if:NTF #1 { Match } { Don't-match } \__regex_msg_repeated:nnN {#3} {#4} #5 }
            \tl_build_put_right:Ne \l__regex_internal_a_tl
              { \exp_not:o \l__regex_internal_a_tl }
          }
        }
      }
    \tl_build_end:N \l__regex_build_tl
    \exp_args:NNNo
  }
\end{verbatim}

(End of definition for \__regex_show_class:NnnnN)

\__regex_show_item_catcode:NnT

Produce a sequence of categories which the catcode bitmap #2 contains, and show it,
indenting the tests on which this catcode constraint applies.

\begin{verbatim}
cs_new_protected:Npn \__regex_show_item_catcode:NnT #1#2
  {
    \seq_set_split:Nnn \l__regex_internal_seq { } { CBEMTPUDSLOA }
    \seq_set_filter:NNn \l__regex_internal_seq \l__regex_internal_seq
      { \int_if_odd_p:n { #2 / \int_use:c { c__regex_catcode_##1_int } } }
    \__regex_show_scope:nn
  }
\end{verbatim}

(End of definition for \__regex_show_item_catcode:NnT)
Building

46.4.1 Variables used while building

\_\_\_reg\_min\_state\_int \_\_\_reg\_max\_state\_int

The last state that was allocated is \_\_\_reg\_max\_state\_int \-\- 1, so that \_\_\_reg\_max\_state\_int always points to a free state. The \_\_\_reg\_min\_state variable is 1 to begin with, but gets shifted in nested calls to the matching code, namely in \c{...} constructions.

\_\_\_reg\_left\_state\_int \_\_\_reg\_right\_state\_int

Alternatives are implemented by branching from a \_\_\_reg\_left state into the various choices, then merging those into a \_\_\_reg\_right state. We store information about those states in two sequences. Those states are also used to implement group quantifiers. Most often, the left and right pointers only differ by 1.

\_\_\_reg\_left\_state\_seq \_\_\_reg\_right\_state\_seq

\_\_\_reg\_capturing\_group\_int

\_\_\_reg\_capturing\_group\_int is the next id number to be assigned to a capturing group. This starts at 0 for the group enclosing the full regular expression, and groups are counted in the order of their left parenthesis, except when encountering resetting groups.
46.4.2 Framework

This phase is about going from a compiled regex to an NFA. Each state of the NFA is stored in a \toks. The operations which can appear in the \toks are

- \_\_regex_action_start_wildcard:N \langle boolean \rangle inserted at the start of the regular expression, where a true \langle boolean \rangle makes it unanchored.
- \_\_regex_action_success: marks the exit state of the NFA.
- \_\_regex_action_cost:n \langle shift \rangle is a transition from the current \langle state \rangle to \langle state \rangle + \langle shift \rangle, which consumes the current character: the target state is saved and will be considered again when matching at the next position.
- \_\_regex_action_free:n \langle shift \rangle, and \_\_regex_action_free_group:n \langle shift \rangle are free transitions, which immediately perform the actions for the state \langle state \rangle + \langle shift \rangle of the NFA. They differ in how they detect and avoid infinite loops. For now, we just need to know that the group variant must be used for transitions back to the start of a group.
- \_\_regex_action_submatch:n \langle group \rangle \langle key \rangle where the \langle key \rangle is < or > for the beginning or end of group numbered \langle group \rangle. This causes the current position in the query to be stored as the \langle key \rangle submatch boundary.
- One of these actions, within a conditional.

We strive to preserve the following properties while building.

- The current capturing group is capturing_group − 1, and if a group opened now it would be labelled capturing_group.
- The last allocated state is max_state − 1, so max_state is a free state.
- The left_state points to a state to the left of the current group or of the last class.
- The right_state points to a newly created, empty state, with some transitions leading to it.
- The left/right sequences hold a list of the corresponding end-points of nested groups.

The n-type function first compiles its argument. Reset some variables. Allocate two states, and put a wildcard in state 0 (transitions to state 1 and 0 state). Then build the regex within a (capturing) group numbered 0 (current value of capturing_group). Finally, if the match reaches the last state, it is successful. A false boolean for argument #1 for the auxiliaries will suppress the wildcard and make the match anchored: used for \peek_regex:nTF and similar.

\begin{verbatim}
\cs_new_protected:Npn \_\_regex_build:n { \_\_regex_build_aux:Nn \c_true_bool }
\cs_new_protected:Npn \_\_regex_build:N { \_\_regex_build_aux:NN \c_true_bool }
\cs_new_protected:Npn \_\_regex_build_aux:Nn #1#2 { \_\_regex_compile:n {#2} }
\end{verbatim}
\_regex_build_aux:NN \#1 \l__regex_internal_regex

\cs_new_protected:Npn \_regex_build_aux:NN \#1\#2
\{ \_regex_standard_escapechar:
\int_zero:N \l__regex_capturing_group_int
\int_set_eq:NN \l__regex_max_state_int \l__regex_min_state_int
\_regex_build_new_state:
\_regex_build_new_state:
\_regex_toks_put_right:Nn \l__regex_left_state_int{ \_regex_action_start_wildcard:N \#1 }
\_regex_group:nnN \#2 \{ 0 \} \c_false_bool
\_regex_toks_put_right:Nn \l__regex_right_state_int{ \_regex_action_success: }
\}

(End of definition for \_regex_build:n and others.)

\g__regex_case_int Case number that was successfully matched in \regex_match_case:nn and related functions.
\int_new:N \g__regex_case_int

(End of definition for \g__regex_case_int.)

\l__regex_case_max_group_int The largest group number appearing in any of the (regex) in the argument of \regex_match_case:nn and related functions.
\int_new:N \l__regex_case_max_group_int

(End of definition for \l__regex_case_max_group_int.)

\_regex_case_build:n See \_regex_build:n, but with a loop.
\_regex_case_build:e
\_regex_case_build_aux:Nn
\_regex_case_build_loop:n
\_\_regex_case_build_loop:n \{#1\}

\int_set_eq:NN \l__regex_capturing_group_int \l__regex_case_max_group_int
\_\_regex_pop_lr_states:
}
\cs_new_protected:Npn \_\_regex_case_build_loop:n #1
{
\int_set_eq:NN \l__regex_capturing_group_int \c_one_int
\_\_regex_compile_use:n \{#1\}
\int_set:Nn \l__regex_case_max_group_int
{ \int_max:nn \l__regex_case_max_group_int \l__regex_capturing_group_int }
\seq_pop:NN \l__regex_right_state_seq \l__regex_internal_a_tl
\int_set:Nn \l__regex_right_state_int \l__regex_internal_a_tl
\_\_regex_toks_put_left:Ne \l__regex_right_state_int
\_\_regex_action_submatch:nN \c_zero_int >
\int_gset:Nn \g__regex_case_int
{ \int_use:N \g__regex_case_int }
\_\_regex_action_success:
\_\_regex_toks_clear:N \l__regex_max_state_int
\seq_pop:Nn \l__regex_right_state_seq
\int_set:Nn \l__regex_max_state_int
\_\_regex_push_lr_states:
#1
\_\_regex_pop_lr_states:
\_\_regex_toks_put_right:Nn \l__regex_right_state_int
{
\_if_int_compare:w -2 = \l__regex_curr_char_int

(End of definition for \_\_regex_case_build:n, \_\_regex_case_build_aux:Nn, and \_\_regex_case_build_loop:n.)

\_\_regex_build_for_cs:n
The matching code relies on some global intarray variables, but only uses a range of their entries. Specifically,

- \g__regex_state_active_intarray from \l__regex_min_state_int to \l__regex_max_state_int + 1;

Here, in this nested call to the matching code, we need the new versions of this range to involve completely new entries of the intarray variables, so we begin by setting (the new) \l__regex_min_state_int to (the old) \l__regex_max_state_int to use higher entries.

When using a regex to match a cs, we don’t insert a wildcard, we anchor at the end, and since we ignore submatches, there is no need to surround the expression with a group. However, for branches to work properly at the outer level, we need to put the appropriate left and right states in their sequence.
46.4.3 Helpers for building an nfa

When building the regular expression, we keep track of pointers to the left-end and right-end of each group without help from \TeX's grouping.

\begin{verbatim}
\cs_new_protected:Npn \__regex_push_lr_states: {
 \seq_push:No \l__regex_left_state_seq { \int_use:N \l__regex_left_state_int }
 \seq_push:No \l__regex_right_state_seq { \int_use:N \l__regex_right_state_int }
}
\cs_new_protected:Npn \__regex_pop_lr_states: {
 \seq_pop:NN \l__regex_left_state_seq \l__regex_internal_a_tl
 \int_set:Nn \l__regex_left_state_int \l__regex_internal_a_tl
 \seq_pop:NN \l__regex_right_state_seq \l__regex_internal_a_tl
 \int_set:Nn \l__regex_right_state_int \l__regex_internal_a_tl
}
\end{verbatim}

\begin{verbatim}
\cs_new_protected:Npn \__regex_build_transition_left:NNN \__regex_int_eval:w \__regex_int_eval:w #1 #2 #3 {
 \__regex_toks_put_left:Ne #2 { \__regex_int_eval:w \__regex_int_eval:w #1 { \tex_the:D \__regex_int_eval:w #3 - #2 } }
}
\cs_new_protected:Npn \__regex_build_transition_right:nNn #1 #2 #3 {
 \__regex_toks_put_right:Ne #2 { \__regex_int_eval:w \__regex_int_eval:w #1 { \tex_the:D \__regex_int_eval:w #3 - #2 } }
}
\end{verbatim}

\begin{verbatim}
\cs_new_protected:Npn \__regex_build_new_state: {
 \__regex_toks_clear:N \l__regex_max_state_int
 \int_set_eq:NN \l__regex_left_state_int \l__regex_right_state_int
 \int_set_eq:NN \l__regex_right_state_int \l__regex_max_state_int
 \int_incr:N \l__regex_max_state_int
}
\end{verbatim}

Add a new empty state to the NFA. Then update the left, right, and max states, so that the right state is the new empty state, and the left state points to the previously “current” state.
This function creates a new state, and puts two transitions starting from the old current state. The order of the transitions is controlled by \#1, true for lazy quantifiers, and false for greedy quantifiers.

\begin{verbatim}
\cs_new_protected:Npn \_\_regex_build_transitions_laziness:NNNNN #1#2#3#4#5
\begin{verbatim}
{ \_\_regex_build_new_state:
  \_\_regex_toks_put_right:Ne \l__regex_left_state_int
  \if_meaning:w \c_true_bool #1
    #2 { \tex_the:D \__regex_int_eval:w #3 - \l__regex_left_state_int }
    #4 { \tex_the:D \__regex_int_eval:w #5 - \l__regex_left_state_int }
  \else:
    #4 { \tex_the:D \__regex_int_eval:w #5 - \l__regex_left_state_int }
    #2 { \tex_the:D \__regex_int_eval:w #3 - \l__regex_left_state_int }
  \fi:
}
\end{verbatim}
\end{verbatim}
\end{verbatim}

(End of definition for \_\_regex_build_transitions_laziness:NNNNN.)

4.6.4.4 Building classes

The arguments are: \texttt{\{boolean\}} \texttt{\{(tests\}\}} \texttt{\{(min\}\}} \texttt{\{(more\}\}} \texttt{\{laziness\}}. First store the tests with a trailing \_\_regex_action_cost:n, in the true branch of \_\_regex_break_point:TF for positive classes, or the false branch for negative classes. The integer \texttt{(more)} is 0 for fixed repetitions, \texttt{-1} for unbounded repetitions, and \texttt{(max)} – \texttt{(min)} for a range of repetitions.

\begin{verbatim}
\cs_new_protected:Npn \_\_regex_class:NnnnN #1#2#3#4#5
\begin{verbatim}
{ \cs_set:Npe \_\_regex_tests_action_cost:n ##1
  \exp_not:n { \exp_not:n {#2} }
  \bool_if:NTF #1
  { \_\_regex_break_point:TF { \_\_regex_action_cost:n {##1} } { } }
  { \_\_regex_break_point:TF { } { \_\_regex_action_cost:n {##1} } }
  \fi:
}
\end{verbatim}
\end{verbatim}

(End of definition for \_\_regex_class:NnnnN and \_\_regex_tests_action_cost:n)

\_\_regex_class_repeat:n

This is used for a fixed number of repetitions. Build one state for each repetition, with a transition controlled by the tests that we have collected. That works just fine for \texttt{\#1 = 0} repetitions: nothing is built.

\begin{verbatim}
\cs_new_protected:Npn \_\_regex_class_repeat:n #1
\begin{verbatim}
{ \prg_replicate:nn {#1} }
\end{verbatim}
\end{verbatim}

526
This implements unbounded repetitions of a single class (e.g. the * and + quantifiers). If the minimum number of repetitions is 0, then build a transition from the current state to itself governed by the tests, and a free transition to a new state (hence skipping the tests). Otherwise, call \__regex_class_repeat:n for the code to match #1 repetitions, and add free transitions from the last state to the previous one, and to a new one. In both cases, the order of transitions is controlled by the laziness boolean #2.

\__regex_class_repeat:nnN
We want to build the code to match from #1 to #1 + #2 repetitions. Match #1 repetitions (can be 0). Compute the final state of the next construction as a. Build #2 > 0 states, each with a transition to the next state governed by the tests, and a transition to the final state a. The computation of a is safe because states are allocated in order, starting from max_state.
46.4.5 Building groups

Arguments: \( \langle \text{label} \rangle \) \{ \langle \text{contents} \rangle \} \{ \langle \text{min} \rangle \} \{ \langle \text{more} \rangle \} \langle \text{laziness} \rangle \). If \( \langle \text{min} \rangle \) is 0, we need to add a state before building the group, so that the thread which skips the group does not also set the start-point of the submatch. After adding one more state, the left state is the left end of the group, from which all branches stem, and the right state is the right end of the group, and all branches end their course in that state. We store those two integers to be queried for each branch, we build the NFA states for the contents of the group, and we forget about the two integers. Once this is done, perform the repetition: either exactly \#3 times, or \#3 or more times, or between \#3 and \#3 + \#4 times, with laziness \#5. The \( \langle \text{label} \rangle \) \#1 is used for submatch tracking. Each of the three auxiliaries expects left state and right state to be set properly.

\begin{verbatim}
\cs_new_protected:Npn \_regex_group_aux:nnnnN #1#2#3#4#5
{ \if_int_compare:w #3 = \c_zero_int
  \_regex_build_new_state:
  \_regex_build_transition_right:nNn \_regex_action_free_group:n
  \l__regex_left_state_int \l__regex_right_state_int
\fi:
  \_regex_build_new_state:
  \_regex_push_lr_states:
  #2
  \_regex_pop_lr_states:
  \if_case:w - #4 \exp_stop_f:
    \_regex_group_repeat:nn {#1} {#3}
  \or: \_regex_group_repeat:nnN {#1} {#3} \#5
  \else: \_regex_group_repeat:nnnN {#1} {#3} {#4} \#5
  \fi:
}
\end{verbatim}

(End of definition for \_regex_group_aux:nnnnN.)

\_regex_group:nnnN \_regex_group_no_capture:nnnN

Hand to \_regex_group_aux:nnnnN the label of that group (expanded), and the group itself, with some extra commands to perform.

\begin{verbatim}
\cs_new_protected:Npn \_regex_group:nnnN #1
{ \exp_args:No \_regex_group_aux:nnnnN
  { \int_use:N \l__regex_capturing_group_int }
  { \int_incr:N \l__regex_capturing_group_int \#1
  }
}
\end{verbatim}

(End of definition for \_regex_group:nnnN and \_regex_group_no_capture:nnnN.)

\_regex_group_resetting:nnnN \_regex_group_resetting_loop:nnNn

Again, hand the label \(-1\) to \_regex_group_aux:nnnnN, but this time we work a little bit harder to keep track of the maximum group label at the end of any branch, and to reset the group number at each branch. This relies on the fact that a compiled regex always is a sequence of items of the form \_regex_branch:n \{ \langle branch \rangle \}.

\begin{verbatim}
\cs_new_protected:Npn \_regex_group_resetting:nnnN #1
\end{verbatim}

528
\section*{\texttt{\_regex_branch:n}}

Add a free transition from the left state of the current group to a brand new state, starting point of this branch. Once the branch is built, add a transition from its last state to the right state of the group. The left and right states of the group are extracted from the relevant sequences.

\section*{\texttt{\_regex_group_repeat:nn}}

This function is called to repeat a group a fixed number of times \#2; if this is 0 we remove the group altogether (but don't reset the \texttt{\_regex\_capturing\_group} label). Otherwise, the auxiliary \texttt{\_regex\_group\_repeat\_aux:n} copies \#2 times the \texttt{\toks} for the group, and leaves \texttt{\_internal\_a} pointing to the left end of the last repetition. We only record the submatch information at the last repetition. Finally, add a state at the end (the transition to it has been taken care of by the replicating auxiliary).
\_\_regex_group_repeat:nnN

This function is called to repeat a group at least \( n \) times; the case \( n = 0 \) is very different from \( n > 0 \). Assume first that \( n = 0 \). Insert submatch tracking information at the start and end of the group, add a free transition from the right end to the “true” left state a
(remember: in this case we had added an extra state before the left state). This forms the loop, which we break away from by adding a free transition from a to a new state.

Now consider the case \( n > 0 \). Repeat the group \( n \) times, chaining various copies with a free transition. Add submatch tracking only to the last copy, then add a free transition from the right end back to the left end of the last copy, either before or after the transition to move on towards the rest of the NFA. This transition can end up before submatch tracking, but that is irrelevant since it only does so when going again through the group, recording new matches. Finally, add a state; we already have a transition pointing to it from \( \_\_\_\text{regex\_group\_repeat\_aux:nn} \).

\[
\text{\textbackslash cs\_new\_protected:Npn \textbackslash \_\_\_\text{regex\_group\_repeat:nnn} \ #1\#2\#3}
\]

\[
\{\text{\textbackslash if\_int\_compare:w \#2 = \c\_zero\_int}
\_\_\_\text{regex\_group\_submatches:nnn} \ #1\}
\_\_\_\text{regex\_left\_state\_int} \_\_\_\text{regex\_right\_state\_int}
\int\_set:Nn \_\_\_\text{regex\_internal\_a\_int}
\{ \_\_\_\text{regex\_left\_state\_int} - \c\_one\_int \}
\_\_\_\text{regex\_build\_transition\_right:nnn} \_\_\_\text{regex\_action\_free:n}
\_\_\_\text{regex\_right\_state\_int} \_\_\_\text{regex\_internal\_a\_int}
\_\_\_\text{regex\_build\_new\_state:}
\text{\textbackslash if\_meaning:w \c\_true\_bool \#3}
\_\_\_\text{regex\_build\_transition\_left:nnN} \_\_\_\text{regex\_action\_free:n}
\_\_\_\text{regex\_right\_state\_int}
\else:
\_\_\_\text{regex\_build\_transition\_right:nnn} \_\_\_\text{regex\_action\_free:n}
\_\_\_\text{regex\_internal\_a\_int} \_\_\_\text{regex\_right\_state\_int}
\fi:
\else:
\_\_\_\text{regex\_group\_repeat\_aux:n} \ #2\}
\_\_\_\text{regex\_group\_submatches:nnn} \ #1\}
\_\_\_\text{regex\_internal\_a\_int} \_\_\_\text{regex\_right\_state\_int}
\text{\textbackslash if\_meaning:w \c\_true\_bool \#3}
\_\_\_\text{regex\_build\_transition\_right:nnn} \_\_\_\text{regex\_action\_free\_group:n}
\_\_\_\text{regex\_right\_state\_int} \_\_\_\text{regex\_internal\_a\_int}
\else:
\_\_\_\text{regex\_build\_transition\_left:nnN} \_\_\_\text{regex\_action\_free\_group:n}
\_\_\_\text{regex\_right\_state\_int} \_\_\_\text{regex\_internal\_a\_int}
\fi:
\_\_\_\text{regex\_build\_new\_state:}
\fi:
\}

(End of definition for \( \_\_\_\_\text{regex\_group\_repeat:nnn} \))

We wish to repeat the group between \( \#2 \) and \( \#2 + \#3 \) times, with a laziness controlled by \( \#4 \). We insert submatch tracking up front: in principle, we could avoid recording submatches for the first \( \#2 \) copies of the group, but that forces us to treat specially the case \( \#2 = 0 \). Repeat that group with submatch tracking \( \#2 + \#3 \) times (the maximum number of repetitions). Then our goal is to add \( \#3 \) transitions from the end of the \( \#2 \)-th group, and each subsequent groups, to the end. For a lazy quantifier, we add those transitions to the left states, before submatch tracking. For the greedy case, we add the transitions to the right states, after submatch tracking and the transitions which go on with more repetitions. In the greedy case with \( \#2 = 0 \), the transition which skips over all
copies of the group must be added separately, because its starting state does not follow
the normal pattern: we had to add it “by hand” earlier.

46.4.6 Others

Usage: \_\_regex\_assertion:Nn \texttt{⟨boolean⟩ \{⟨test⟩⟩}, where the \texttt{⟨test⟩} is either of the
two other functions. Add a free transition to a new state, conditionally to the assertion
test. The \texttt{\_\_regex\_b\_test:} test is used by the \texttt{\b} and \texttt{\B} escape: check if the last
character was a word character or not, and do the same to the current character. The
boundary-markers of the string are non-word characters for this purpose.
\bool_if:NF #1 { { } }
{ \_regex_action_free:n
  \tex_the:D \_regex_int_eval:w
  \l__regex_right_state_int - \l__regex_left_state_int
}
\bool_if:NT #1 { { } }
}
\cs_new_protected:Npn \_regex_b_test:
{ \group_begin:
  \int_set_eq:NN \l__regex_curr_char_int \l__regex_last_char_int
  \_regex_prop_w:
  \_regex_break_point:TF
  { \group_end: \_regex_item_reverse:n \{ \_regex_prop_w: \} }
  { \group_end: \_regex_prop_w: \} }
\group_end:
\cs_new_protected:Npn \_regex_Z_test:
{ \if_int_compare:w -2 = \l__regex_curr_char_int \exp_after:wN \_regex_break_true:w \fi: }
\cs_new_protected:Npn \_regex_A_test:
{ \if_int_compare:w -2 = \l__regex_last_char_int \exp_after:wN \_regex_break_true:w \fi: }
\cs_new_protected:Npn \_regex_G_test:
{ \if_int_compare:w \l__regex_curr_pos_int = \l__regex_start_pos_int \exp_after:wN \_regex_break_true:w \fi: }

(End of definition for \_regex_assertion:Nn and others.)

\_regex_command_K: \cs_new_protected:Npn \_regex_command_K:
{ \_regex_build_new_state:
  \_regex_toks_put_right:Nn \l__regex_left_state_int
  \_regex_action_submatch:nN \c_zero_int <
  \bool_set_true:N \l__regex_fresh_thread_bool
  \_regex_action_free:n
  \tex_the:D \_regex_int_eval:w
  \l__regex_right_state_int - \l__regex_left_state_int

533
46.5 Matching

We search for matches by running all the execution threads through the NFA in parallel, reading one token of the query at each step. The NFA contains “free” transitions to other states, and transitions which “consume” the current token. For free transitions, the instruction at the new state of the NFA is performed immediately. When a transition consumes a character, the new state is appended to a list of “active states”, stored in \g__regex_thread_info_intarray (together with submatch information): this thread is made active again when the next token is read from the query. At every step (for each token in the query), we unpack that list of active states and the corresponding submatch props, and empty those.

If two paths through the NFA “collide” in the sense that they reach the same state after reading a given token, then they only differ in how they previously matched, and any future execution would be identical for both. (Note that this would be wrong in the presence of back-references.) Hence, we only need to keep one of the two threads: the thread with the highest priority. Our NFA is built in such a way that higher priority actions always come before lower priority actions, which makes things work.

The explanation in the previous paragraph may make us think that we simply need to keep track of which states were visited at a given step: after all, the loop generated when matching (a?)∗ against a is broken, isn’t it? No. The group first matches a, as it should, then repeats; it attempts to match a again but fails; it skips a, and finds out that this state has already been seen at this position in the query: the match stops. The capturing group is (wrongly) a. What went wrong is that a thread collided with itself, and the later version, which has gone through the group one more times with an empty match, should have a higher priority than not going through the group.

We solve this by distinguishing “normal” free transitions \__regex_action_free:n from transitions \__regex_action_free_group:n which go back to the start of the group. The former keeps threads unless they have been visited by a “completed” thread, while the latter kind of transition also prevents going back to a state visited by the current thread.

46.5.1 Variables used when matching

The tokens in the query are indexed from min_pos for the first to max_pos – 1 for the last, and their information is stored in several arrays and \toks registers with those numbers. We match without backtracking, keeping all threads in lockstep at the curr_pos in the query. The starting point of the current match attempt is start_pos, and success_pos, updated whenever a thread succeeds, is used as the next starting position.
The character and category codes of the token at the current position and a token list expanding to that token; the character code of the token at the previous position; the character code of the token just before a successful match; and the character code of the result of changing the case of the current token (A-Z ↔ a-z). This last integer is only computed when necessary, and is otherwise \c_max_int. The curr_char variable is also used in various other phases to hold a character code.

6548 \int_new:N \l__regex_curr_char_int
6549 \int_new:N \l__regex_curr_catcode_int
6550 \tl_new:N \l__regex_curr_token_tl
6551 \int_new:N \l__regex_last_char_int
6552 \int_new:N \l__regex_last_char_success_int
6553 \int_new:N \l__regex_case_changed_char_int

(End of definition for \l__regex_curr_char_int and others.)

\l__regex_curr_state_int

For every character in the token list, each of the active states is considered in turn. The variable \l__regex_curr_state_int holds the state of the nfa which is currently considered: transitions are then given as shifts relative to the current state.

6554 \int_new:N \l__regex_curr_state_int

(End of definition for \l__regex_curr_state_int.)

\l__regex_curr_submatches_tl \l__regex_success_submatches_tl

The submatches for the thread which is currently active are stored in the curr_submatches list, which is almost a comma list, but ends with a comma. This list is stored by \__regex_store_state:n into an intarray variable, to be retrieved when matching at the next position. When a thread succeeds, this list is copied to \l__regex_success_submatches_tl: only the last successful thread remains there.

6555 \tl_new:N \l__regex_curr_submatches_tl
6556 \tl_new:N \l__regex_success_submatches_tl

(End of definition for \l__regex_curr_submatches_tl and \l__regex_success_submatches_tl.)

\l__regex_step_int

This integer, always even, is increased every time a character in the query is read, and not reset when doing multiple matches. We store in \g__regex_state_active_intarray the last step in which each ⟨state⟩ in the NFA was encountered. This lets us break infinite loops by not visiting the same state twice in the same step. In fact, the step we store is equal to step when we have started performing the operations of \toks⟨state⟩, but not finished yet. However, once we finish, we store step + 1 in \g__regex_state_active_intarray. This is needed to track submatches properly (see building phase). The step is also used to attach each set of submatch information to a given iteration (and automatically discard it when it corresponds to a past step).

6557 \int_new:N \l__regex_step_int

(End of definition for \l__regex_step_int.)

\l__regex_min_thread_int \l__regex_max_thread_int

All the currently active threads are kept in order of precedence in \g__regex_thread_info_intarray together with the corresponding submatch information. Data in this intarray is organized as blocks from min_thread (included) to max_thread (excluded). At the start of every step, the whole array is unpacked, so that the space can immediately be reused, and max_thread is reset to min_thread, effectively clearing the array.

6558 \int_new:N \l__regex_min_thread_int
6559 \int_new:N \l__regex_max_thread_int

535
\g__regex_state_active_intarray
\g__regex_thread_info_intarray

\g__regex_state_active_intarray stores the last (step) in which each (state) was active. \g__regex_thread_info_intarray stores threads to be considered in the next step, more precisely the states in which these threads are.

\intarray_new:Nn \g__regex_state_active_intarray { 65536 }
\intarray_new:Nn \g__regex_thread_info_intarray { 65536 }

The list \l__regex_matched_analysis_tl consists of a brace group containing three brace groups corresponding to the current token, with the same syntax as \tl_analysis_map_inline:nn. The list \l__regex_matched_analysis_tl (constructed under the \tl_build machinery) has one item for each token that has already been treated so far in a given match attempt: each item consists of three brace groups with the same syntax as \tl_analysis_map_inline:nn.

\tl_new:N \l__regex_matched_analysis_tl
\tl_new:N \l__regex_curr_analysis_tl

Every time a match is found, this token list is used. For single matching, the token list is empty. For multiple matching, the token list is set to repeat the matching, after performing some operation which depends on the user function. See \__regex_single_match: and \__regex_multi_match:n.

\tl_new:N \l__regex_every_match_tl

When doing multiple matches, we need to avoid infinite loops where each iteration matches the same empty token list. When an empty token list is matched, the next successful match of the same empty token list is suppressed. We detect empty matches by setting \l__regex_fresh_thread_bool to true for threads which directly come from the start of the regex or from the \K command, and testing that boolean whenever a thread succeeds. The function \__regex_if_two_empty_matches:F is redefined at every match attempt, depending on whether the previous match was empty or not: if it was, then the function must cancel a purported success if it is empty and at the same spot as the previous match; otherwise, we definitely don’t have two identical empty matches, so the function is \use:n.

\bool_new:N \l__regex_fresh_thread_bool
\bool_new:N \l__regex_empty_success_bool
\bool_new:N \l__regex_match_success_bool
\cs_new_eq:NN \__regex_if_two_empty_matches:F \use:n

The boolean \l__regex_match_success_bool is true if the current match attempt was successful, and \g__regex_success_bool is true if there was at least one successful match. This is the only global variable in this whole module, but we would need it to be local when matching a control sequence with \c{...}. This is done by saving the global variable into \l__regex_saved_success_bool, which is local, hence not affected by the changes due to inner regex functions.

\bool_new:N \g__regex_success_bool
\bool_new:N \l__regex_saved_success_bool
\bool_new:N \l__regex_match_success_bool

536
46.5.2 Matching: framework

Initialize the variables that should be set once for each user function (even for multiple matches). Namely, the overall matching is not yet successful; none of the states should be marked as visited (\_\_regex_state_active_intarray), and we start at step 0; we pretend that there was a previous match ending at the start of the query, which was not empty (to avoid smothering an empty match at the start). Once all this is set up, we are ready for the ride. Find the first match.

\begin{verbatim}
\cs_new_protected:Npn \__regex_match:n #1
\__regex_match_init:
\__regex_match_once_init:
\tl_analysis_map_inline:nn {#1}
\__regex_match_one_token:nnN {##1} {##2} ##3
\__regex_match_one_token:nnN { } { -2 } F
\prg_break_point:Nn \__regex_maplike_break: { }
\end{verbatim}

\begin{verbatim}
\cs_new_protected:Npn \__regex_match_cs:n #1
\int_set_eq:NN \l__regex_min_thread_int \l__regex_max_thread_int
\__regex_match_init:
\__regex_match_once_init:
\str_map_inline:nn {#1}
\tl_if_blank:nTF {##1}
\{ \__regex_match_one_token:nnN {##1} {'##1} A \}
\{ \__regex_match_one_token:nnN {##1} {'##1} C \}
\__regex_match_one_token:nnN { } { -2 } F
\prg_break_point:Nn \__regex_maplike_break: { }
\end{verbatim}

\begin{verbatim}
\cs_new_protected:Npn \__regex_match_init:
\{ \bool_gset_false:N \g__regex_success_bool
\int_step_inline:nnn \l__regex_min_state_int { \l__regex_max_state_int - \c_one_int }
\{ \__kernel_intarray_gset:Nnn
\g__regex_state_active_intarray (#1) \c_one_int
\int_zero:N \l__regex_step_int
\int_set:Nn \l__regex_min_pos_int { 2 }
\int_set_eq:NN \l__regex_min_pos_int \l__regex_max_pos_int
\int_set:Nn \l__regex_last_char_success_int { -2 }
\tl_build_begin:N \l__regex_matched_analysis_tl
\tl_clear:N \l__regex_curr_analysis_tl
\int_set_eq:NN \l__regex_min_submatch_int \c_one_int
\int_set_eq:NN \l__regex_submatch_int \l__regex_min_submatch_int
\bool_set_false:N \l__regex_empty_success_bool
\end{verbatim}

537
\__regex_match_once_init:  
This function resets various variables used when finding one match. It is called before the loop through characters, and every time we find a match, before searching for another match (this is controlled by the every_match token list).

First initialize some variables: set the conditional which detects identical empty matches; this match attempt starts at the previous success_pos, is not yet successful, and has no submatches yet; clear the array of active threads, and put the starting state 0 in it. We are then almost ready to read our first token in the query, but we actually start one position earlier than the start because \_\_regex_match_one_token:nnN increments \l__regex_curr_pos_int and saves \l__regex_curr_char_int as the last_char so that word boundaries can be correctly identified.

\cs_new_protected:Npn \__regex_match_once_init: 
{ 
\if_meaning:w \c_true_bool \l__regex_empty_success_bool 
\cs_set:Npn \__regex_if_two_empty_matches:F 
\else: 
\cs_set_eq:NN \__regex_if_two_empty_matches:F \use:n 
\fi:
\int_set_eq:NN \l__regex_start_pos_int \l__regex_success_pos_int
\bool_set_false:N \l__regex_match_success_bool
\tl_set:Ne \l__regex_curr_submatches_tl
\exp_hyphenate:nn { 2 * \l__regex_capturing_group_int } { 0 , }
\tl_set:Nn \l__regex_max_thread_int { \l__regex_min_thread_int }
\tl_set:Nn \l__regex_curr_pos_int { \l__regex_start_pos_int - \c_one_int }
\tl_set:Nn \l__regex_start_pos_int { \l__regex_start_pos_int - \c_one_int }
\tl_set:Nn \l__regex_last_char_success_int { \l__regex_start_pos_int - \c_one_int }
\tl_set:Nn \l__regex_mached_analysis_tl { \l__regex_mached_analysis_tl \l__regex_last_char_success_int }
\exp_after:wN \l__regex_maplike_break: { }
\cs_new_protected:Npn \__regex_match_once_init_aux:
{ 
\tl_map_inline:nn { \exp_after:wN \l__regex_internal_a_tl \l__regex_curr_analysis_tl }
\prg_break_point:Nn \__regex_maplike_break: { }
}
\cs_new_protected:Npn \__regex_match_once_init:
{ 
\tl_set:Nn \l__regex_every_match_tl
\}

\__regex_single_match:  
For a single match, the overall success is determined by whether the only match attempt is a success. When doing multiple matches, the overall matching is successful as soon as any match succeeds. Perform the action #1, then find the next match.

\cs_new_protected:Npn \__regex_single_match:
{ 
\tl_set:Nn \l__regex_every_match_tl
\}

\__regex_multi_match:n  
(End of definition for \__regex_match:n, \__regex_match_cs:n, and \__regex_match_init::)
\bool_gset_eq:NN
\g__regex_success_bool
\l__regex_match_success_bool
\__regex_maplike_break:
}
\cs_new_protected:Npn \__regex_multi_match:n #1
{
\tl_set:Nn \l__regex_every_match_tl
{
\if_meaning:w \c_false_bool \l__regex_match_success_bool
\exp_after:wN \__regex_maplike_break:
\fi:
\bool_gset_true:N \g__regex_success_bool
#1
}
\__regex_match_once_init:
}
(End of definition for \__regex_single_match: and \__regex_multi_match:n.)
\__regex_match_one_token:nnN\__regex_match_one_active:n
At each new position, set some variables and get the new character and category from
the query. Then unpack the array of active threads, and clear it by resetting its length
(max_thread). This results in a sequence of \__regex_use_state_and_submatches:w
\langle state \rangle, \langle submatch-clist \rangle; and we consider those states one by one in order. As soon
as a thread succeeds, exit the step, and, if there are threads to consider at the next
position, and we have not reached the end of the string, repeat the loop. Otherwise, the
last thread that succeeded is the match. We explain the fresh_thread business when
describing \__regex_action_wildcard:.
\cs_new_protected:Npm \__regex_match_one_token:nnN #1#2#3
{
\int_add:Nn \l__regex_step_int { 2 }
\int_incr:N \l__regex_curr_pos_int
\int_set_eq:NN \l__regex_last_char_int \l__regex_curr_char_int
\cs_set_eq:NN \__regex_maybe_compute_ccc: \__regex_compute_case_changed_char:
\tl_set:Nn \l__regex_curr_token_tl {#1}
\int_set:Nn \l__regex_curr_char_int {#2}
\int_set:Nn \l__regex_curr_catcode_int { #3 }
\tl_build_put_right:Ne \l__regex_matched_analysis_tl
{ \exp_not:o \l__regex_curr_analysis_tl }
\tl_set:Nn \l__regex_curr_analysis_tl { { {#1} {#2} #3 } }
\use:e
{

\int_set_eq:NN \l__regex_max_thread_int \l__regex_curr_thread_tl {#1}
\int_set:Nn \l__regex_last_char_int \l__regex_curr_char_int
\cs_set_eq:NN \__regex_match_one_token:nnN \__regex_match_one_active:n
\int_set_eq:NN \l__regex_min_thread_int \l__regex_max_thread_int
\int_step_function:nnN
\l__regex_min_thread_int
{ \l__regex_max_thread_int - \c_one_int }
\__regex_match_one_active:n

\prg_break_point:
\bool_set_false:N \l__regex_fresh_thread_bool
\if_int_compare:w -2 < \l__regex_curr_char_int
\exp_after:wN \use_i:nn

539
46.5.3 Using states of the NFA

\_\_regex\_use\_state:

Use the current NFA instruction. The state is initially marked as belonging to the current step: this allows normal free transition to repeat, but group-repeating transitions won’t. Once we are done exploring all the branches it spawned, the state is marked as step + 1: any thread hitting it at that point will be terminated.

\_\_regex\_use\_state\_and\_submatches:w

This function is called as one item in the array of active threads after that array has been unpacked for a new step. Update the curr_state and curr_submatches and use the state if it has not yet been encountered at this step.

(End of definition for \_\_regex\_match\_one\_token:mmN and \_\_regex\_match\_one\_active:mmN.)
46.5.4 Actions when matching

For an unanchored match, state 0 has a free transition to the next and a costly one to itself, to repeat at the next position. To catch repeated identical empty matches, we need to know if a successful thread corresponds to an empty match. The instruction resetting \l__regex_fresh_thread_bool may be skipped by a successful thread, hence we had to add it to \l__regex_match_one_token too.

\cs_new_protected:Npn \__regex_action_start_wildcard:N #1
\begin{verbatim}
\bool_set_true:N \l__regex_fresh_thread_bool
\__regex_action_free:n {1}
\bool_set_false:N \l__regex_fresh_thread_bool
\bool_if:NT #1 { \__regex_action_cost:n {0} }
\end{verbatim}

(End of definition for \__regex_action_start_wildcard:N.)

These functions copy a thread after checking that the NFA state has not already been used at this position. If not, store submatches in the new state, and insert the instructions for that state in the input stream. Then restore the old value of \l__regex_curr_state_int and of the current submatches. The two types of free transitions differ by how they test that the state has not been encountered yet: the group version is stricter, and will not use a state if it was used earlier in the current thread, hence forcefully breaking the loop, while the “normal” version will revisit a state even within the thread itself.

\cs_new_protected:Npn \__regex_action_free:n
\begin{verbatim}
\cs_new_protected:Npn \__regex_action_free_group:n
\cs_new_protected:Npn \__regex_action_free_aux:nn #1#2
\use:e
\begin{verbatim}
\int_add:Nn \l__regex_curr_state_int {#2}
\exp_not:n
\begin{verbatim}
\if_int_compare:w
\__kernel_intarray_item:Nn \g__regex_state_active_intarray \l__regex_curr_state_int
\int_set:Nn \l__regex_curr_state_int { \int_use:N \l__regex_curr_state_int }
\tl_set:Nn \exp_not:N \l__regex_curr_submatches_tl
\exp_after:wN \__regex_use_state:
\fi:
\end{verbatim}
\begin{verbatim}
\int_set:Nn \l__regex_curr_state_int
{ \int_use:N \l__regex_curr_state_int }
\tl_set:Nn \exp_not:N \l__regex_curr_submatches_tl
{ \exp_not:o \l__regex_curr_submatches_tl }
\end{verbatim}
\end{verbatim}
\end{verbatim}

(End of definition for \__regex_action_free:n, \__regex_action_free_group:n, and \__regex_action_free_aux:nn.)

A transition which consumes the current character and shifts the state by \#1. The resulting state is stored in the appropriate array for use at the next position, and we also store the current submatches.
\cs_new_protected:Npn \__regex_action_cost:n #1 \\
{ \exp_args:No \__regex_store_state:n \\
{ \tex_the:D \__regex_int_eval:w \l__regex_curr_state_int + #1 } }

(End of definition for \__regex_action_cost:n.)

\__regex_store_state:n \\
\__regex_store_submatches: \\
Put the given state and current submatch information in \g__regex_thread_info_intarray, and increment the length of the array.
\cs_new_protected:Npn \__regex_store_state:n #1 \\
{ \exp_args:No \__regex_store_submatches:nn \\
\l__regex_curr_submatches_tl {#1} \\
\int_incr:N \l__regex_max_thread_int }
\cs_new_protected:Npn \__regex_store_submatches:nn #1#2 \\
{ \__kernel_intarray_gset_range_from_clist:Nnn \\
\g__regex_thread_info_intarray \\
\__regex_int_eval:w \c_one_int + \l__regex_max_thread_int * \\
(\l__regex_capturing_group_int * 2 + \c_one_int) }
{ #2 , #1 }

(End of definition for \__regex_store_state:n and \__regex_store_submatches:)

\__regex_disable_submatches: \\
Some user functions don’t require tracking submatches. We get a performance improvement by simply defining the relevant functions to remove their argument and do nothing with it.
\cs_new_protected:Npn \__regex_disable_submatches: \\
{ \cs_set_protected:Npn \__regex_store_submatches:n ##1 { } \\
\cs_set_protected:Npn \__regex_action_submatch:nN ##1##2 { } }

(End of definition for \__regex_disable_submatches:)

\__regex_action_submatch:nN \\
\__regex_action_submatch_aux:w \\
\__regex_action_submatch_auxii:w \\
\__regex_action_submatch_auxiii:w \\
\__regex_action_submatch_auxiv:w \\
Update the current submatches with the information from the current position. Maybe a bottleneck.
\cs_new_protected:Npn \__regex_action_submatch:nN #1#2 \\
{ \exp_after:wN \__regex_action_submatch_aux:w \\
\l__regex_curr_submatches_tl \{#1\} #2 }
\cs_new_protected:Npn \__regex_action_submatch_aux:w #1#2 \\
{ \tl_set:Ne \l__regex_curr_submatches_tl \\
\prg_replicate:nn \\
{ #2 \if_meaning:w > #3 + \l__regex_capturing_group_int \fi: } }

542
\_\_regex\_action\_success:  There is a successful match when an execution path reaches the last state in the NFA, unless this marks a second identical empty match. Then mark that there was a successful match; it is empty if it is “fresh”; and we store the current position and submatches. The current step is then interrupted with \verb+\prg\_break+, and only paths with higher precedence are pursued further. The values stored here may be overwritten by a later success of a path with higher precedence.

\_\_regex\_action\_success:  \{  \_\_regex\_if\_two\_empty\_matches:F  \{  \bool\_set\_true:N \_\_regex\_match\_success\_bool  \bool\_set\_eq:NN \_\_regex\_empty\_success\_bool  \_\_regex\_fresh\_thread\_bool  \int\_set\_eq:NN \_\_regex\_success\_pos\_int \_\_regex\_curr\_pos\_int  \int\_set\_eq:NN \_\_regex\_last\_char\_success\_int \_\_regex\_last\_char\_int  \tl\_build\_begin:N \_\_regex\_matched\_analysis\_tl  \tl\_set\_eq:NN \_\_regex\_success\_submatches\_tl  \prg\_break:  \}  \}  (End of definition for \_\_regex\_action\_success:)

46.6  Replacement

46.6.1  Variables and helpers used in replacement

\_\_regex\_replacement\_csnames\_int  The behaviour of closing braces inside a replacement text depends on whether a sequences \verb+\c+ or \verb+\u+ has been encountered. The number of “open” such sequences that should be closed by } is stored in \_\_regex\_replacement\_csnames\_int, and decreased by 1 by each }.

\int\_new:N \_\_regex\_replacement\_csnames\_int  (End of definition for \_\_regex\_replacement\_csnames\_int.)

\_\_regex\_replacement\_category\_tl \_\_regex\_replacement\_category\_seq  This sequence of letters is used to correctly restore categories in nested constructions such as \verb+\cL(abc\cD(_)d)+.

\tl\_new:N \_\_regex\_replacement\_category\_tl  \seq\_new:N \_\_regex\_replacement\_category\_seq
\g__regex_balance_tl

This token list holds the replacement text for \__regex_replacement_balance_one_match:n while it is being built incrementally.

6823 \cs_new:Npn \__regex_replacement_balance_one_match:n #1
\{
- \__regex_submatch_balance:n {#1}
\}

(End of definition for \__regex_replacement_balance_one_match:n.)

\__regex_replacement_do_one_match:n

The input is the same as \__regex_replacement_balance_one_match:n. This function is redefined to expand to the part of the token list from the end of the previous match to a given match, followed by the replacement text. Hence concatenating the result of this function with all possible arguments (one call for each match), as well as the range from the end of the last match to the end of the string, produces the fully replaced token list. The initialization does not matter, but (as an example) we set it as for an empty replacement.

6832 \cs_new:Npn \__regex_replacement_exp_not:N #1 { \exp_not:n {#1} }

(End of definition for \__regex_replacement_exp_not:N.)

\__regex_replacement_exp_not:V

This function lets us navigate around the fact that the primitive \exp_not:n requires a braced argument. As far as I can tell, it is only needed if the user tries to include in the replacement text a control sequence set equal to a macro parameter character, such as \c_parameter_token. Indeed, within an e/x-expanding assignment, \exp_not:N # behaves as a single #, whereas \exp_not:n {#} behaves as a doubled ##.

6831 \cs_new:Npn \__regex_replacement_exp_not:V \# { \exp_not:n {\#} }

(End of definition for \__regex_replacement_exp_not:V.)
46.6.2 Query and brace balance

When it is time to extract submatches from the token list, the various tokens are stored in \toks registers numbered from \l__regex_min_pos_int inclusive to \l__regex_max_pos_int exclusive. The function \_regex_query_range:nn \{\langle min\rangle\} \{\langle max\rangle\} unpacks registers from the position (\langle min\rangle) to the position (\langle max\rangle)−1 included. Once this is expanded, a second e-expansion results in the actual tokens from the query. That second expansion is only done by user functions at the very end of their operation, after checking (and correcting) the brace balance first.

\begin{verbatim}
\cs_new:Npn \__regex_query_range:nn #1#2 { \exp_after:wN \__regex_query_range_loop:ww \int_value:w \__regex_int_eval:w #1 \exp_after:wN ; \int_value:w \__regex_int_eval:w #2 ; \prg_break_point: \cs_new:Npn \__regex_query_range_loop:ww #1 ; #2 ; { \if_int_compare:w #1 < #2 \exp_stop_f: \else: \prg_break:n \fi: \__regex_toks_use:w #1 \exp_stop_f: \exp_after:wN \__regex_query_range_loop:ww \int_value:w \__regex_int_eval:w #1 + \c_one_int ; #2 ; } \end{verbatim}

(End of definition for \_regex_query_range:nn and \_regex_query_range_loop:ww.)

\_regex_query_submatch:n

Find the start and end positions for a given submatch (of a given match).

\begin{verbatim}
\cs_new:Npn \__regex_query_submatch:n #1 \__regex_query_range:nn \{ \__kernel_intarray_item:Nn \g__regex_submatch_begin_intarray {#1} \} \{ \__kernel_intarray_item:Nn \g__regex_submatch_end_intarray {#1} \}
\end{verbatim}

(End of definition for \_regex_query_submatch:n.)

\_regex_submatch_balance:n

Every user function must result in a balanced token list (unbalanced token lists cannot be stored by TeX). When we unpacked the query, we kept track of the brace balance, hence the contribution from a given range is the difference between the brace balances at the \langle max pos \rangle and \langle min pos \rangle. These two positions are found in the corresponding "submatch" arrays.

\begin{verbatim}
\cs_new_protected:Npn \__regex_submatch_balance:n #1 \__regex_query_range:nn { \tex_the:D \__regex_int_eval:w \__regex_intarray_item:NnF \g__regex_balance_intarray \{ \__kernel_intarray_item:Nn \g__regex_submatch_begin_intarray {#1} \} \{ \__kernel_intarray_item:Nn \g__regex_submatch_end_intarray {#1} \} \c_zero_int
\end{verbatim}
(End of definition for \_regex_submatch_balance:n.)

46.6.3 Framework

The replacement text is built incrementally. We keep track in \_regex_balance_int of the balance of explicit begin- and end-group tokens and we store in \_regex_balance_tl some code to compute the brace balance from submatches (see its description). Detect unescaped right braces, and escaped characters, with trailing \prg_do_nothing: because some of the later function look-ahead. Once the whole replacement text has been parsed, make sure that there is no open csname. Finally, define the balance_one_match and do_one_match functions.

\cs_new_protected:Npn \_regex_replace:n { \_regex_replace_apply:Nn \_regex_replace_set:n }
\msg_error:nne \{ \text{regex} \} \{ \text{replacement-missing-rparen} \}
\{ \text{seq_count:N} \ \_\text{regex}_\text{replacement}_\text{category_seq} \}
\text{seq_clear:N} \ \_\text{regex}_\text{replacement}_\text{category_seq}
\}
\text{tl_gput_right:N} \ \_\text{regex}_\text{balance_tl}
\{ \ + \ \text{int_use:N} \ \_\text{regex}_\text{balance_int} \}
\text{tl_build_end:N} \ \_\text{regex}_\text{build_tl}
\text{exp_args:NNo}
\text{group_end:}
\#1 \ \_\text{regex}_\text{build_tl}
\}
\text{cs_generate_variant:Nn} \ \_\text{regex}_\text{replacement:n} \ \{ \ \text{e} \}
\text{cs_new_protected:NNp} \ \_\text{regex}_\text{replacement_set:n} \ #1
\{ \}
\text{cs_set:NNp} \ \_\text{regex}_\text{replacement_do_one_match:n} \ #1
\{
\_\text{regex}_\text{query_range:nn}
\{
\_\text{kernel_intarray_item:NN}
\text{g\_regex_submatch_prev_intarray} \ (#1)
}\}
\{
\_\text{kernel_intarray_item:NN}
\text{g\_regex_submatch_begin_intarray} \ (#1)
}\#1
\}
\text{exp_args:NNo} \ \text{use:n}
\{ \text{cs_gset:NNp} \ \_\text{regex}_\text{replacement_balance_one_match:n} \ #1 \}
\{
\text{g\_regex_balance_tl}
- \ \_\text{regex}_\text{submatch_balance:n} \ (#1)
\}
\)
(End of definition for \_\text{regex}_\text{replacement:n}, \_\text{regex}_\text{replacement_apply:Nn}, and \_\text{regex}_
\text{replacement_set:n})

\_\text{regex}_\text{case replacement:n}
\_\text{regex}_\text{case replacement:e}
\exp_args:No \__regex_replacement_set:n
{ \g__regex_case_replacement_tl \fi: }
\cs_generate_variant:Nn \__regex_case_replacement:n { e }
\cs_new_protected:Npn \__regex_case_replacement_aux:n #1
{ \tl_gput_right:Nn \g__regex_case_replacement_tl { \or: #1 }
\tl_gput_right:No \g__regex_case_balance_tl
{ \exp_after:wN \or: \g__regex_balance_tl }
}

(End of definition for \__regex_case_replacement:n.)

\__regex_replacement_put:n This gets redefined for \peek_regex_replace_once:nnTF.
\cs_new_protected:Npn \__regex_replacement_put:n
{ \tl_build_put_right:Nn \l__regex_build_tl }

(End of definition for \__regex_replacement_put:n.)

\__regex_replacement_normal:n \__regex_replacement_normal_aux:N Most characters are simply sent to the output by \tl_build_put_right:NN, unless a particular category code has been requested: then \__regex_replacement_c_A:w or a similar auxiliary is called. One exception is right parentheses, which restore the category code in place before the group started. Note that the sequence is non-empty there: it contains an empty entry corresponding to the initial value of \l__regex_replacement_category_tl. The argument #1 is a single character (including the case of a catcode-other space). In case no specific catcode is requested, we take into account the current catcode regime (at the time the replacement is performed) as much as reasonable, with all impossible catcodes (escape, newline, etc.) being mapped to “other”.
\cs_new_protected:Npn \__regex_replacement_normal:n #1
{ \int_compare:nNnTF \l__regex_replacement_csnames_int > \c_zero_int
{ \exp_args:No \__regex_replacement_put:n { \token_to_str:N #1 } }
{ \tl_if_empty:NTF \l__regex_replacement_category_tl
{ \__regex_replacement_normal_aux:N #1 }
{ \use:c { \__regex_replacement_c_ \l__regex_replacement_category_tl :w } #1 }
}
\seq_pop:NN \l__regex_replacement_category_seq
\l__regex_replacement_category_tl
}
\cs_new_protected:Npn \__regex_replacement_normal_aux:N #1
{ \token_if_eq_charcode:NNTF #1 \c_space_token
{ \__regex_replacement_c_S:w }
{ \exp_after:wN \exp_after:wN }
As in parsing a regular expression, we use an auxiliary built from \texttt{#1} if defined. Otherwise, check for escaped digits (standing from submatches from 0 to 9): anything else is a raw character.

Later on, \texttt{##1} will be replaced by a pointer to the 0-th submatch for a given match.
\cs_new_protected:Npn \__regex_replacement_put_submatch_aux:n #1
  \tl_build_put_right:Nn \l__regex_build_tl
  \{ \__regex_query_submatch:n \{ \__regex_int_eval:w \#1 + \#1 \scan_stop: \} \}
\if_int_compare:w \l__regex_replacement_csnames_int = \c_zero_int
  \tl_gput_right:Nn \g__regex_balance_tl \{ + \__regex_submatch_balance:n \{ \__regex_int_eval:w \#1 + \#1 \scan_stop: \} \}
\fi:
\}

(End of definition for \__regex_replacement_put_submatch:n and \__regex_replacement_put_submatch_-
aux:n.)

\__regex_replacement_g:w
\__regex_replacement_g_digits:NN
Grab digits for the \g escape sequence in a primitive assignment to the integer \l__-
regex_internal_a_int. At the end of the run of digits, check that it ends with a right
brace.
\cs_new_protected:Npn \__regex_replacement_g:w #1#2
  \token_if_eq_meaning:NNTF #1 \__regex_replacement_lbrace:N
  { \l__regex_internal_a_int = \__regex_replacement_g_digits:NN \}
  { \__regex_replacement_error:NNN g #1 #2 \}
\}
\cs_new:Npn \__regex_replacement_g_digits:NN #1#2
  \token_if_eq_meaning:NNTF #1 \__regex_replacement_normal:n
  { \if_int_compare:w \c_one_int < 1#2 \exp_stop_f: \#2
    \exp_after:wN \use_i:nnn
    \exp_after:wN \__regex_replacement_g_digits:NN \else:
    \exp_stop_f:
    \exp_after:wN \__regex_replacement_error:NNN \exp_after:wN g \fi:
  \}
\)
\cs_new_protected:Npn \__regex_replacement_g:NN \#1 \#2
  \token_if_eq_meaning:NNTF \#1 \__regex_replacement_rbrace:N \#1
  \exp_args:No \__regex_replacement_put_submatch:n \{ \int_use:N \l__regex_internal_a_int \}
  \exp_after:wN \use_none:nn \else:
  \exp_after:wN \__regex_replacement_error:NNN \exp_after:wN g \fi:
\)
#1 #2
\}

(End of definition for \__regex_replacement_g:w and \__regex_replacement_g_digits:NN.)
46.6.5 Csnames in replacement

\csnewprotected\m__regexreplacementc\aw
\c may only be followed by an unescaped character. If followed by a left brace, start a
control sequence by calling an auxiliary common with \u. Otherwise test whether the
category is known; if it is not, complain.

\begin{verbatim}
s__newprotected\Np m __regexreplacementc\aw #1#2
\{
 \tokenifeqmeaning\NNTF #1 __regexreplacementnormal\an
 \{
 \csifeexist\cTF { __regexreplacementc#2\aw }
 \{ __regexreplacementcat\NN #2 \}
 \{ __regexreplacementerror\NNN c #1#2 \}
 \}
 \{ \tokenifeqmeaning\NNTF #1 __regexreplacementlbrace\N
 \{ __regexreplacementcuaux\Nw __regexreplacementexpnot\N \}
 \{ __regexreplacementerror\NNN c #1#2 \}
 \}
\}
\end{verbatim}

(End of definition for __regexreplacementc\aw.)

\__regexreplacementcuaux\Nw
Start a control sequence with \cs\aw, protected from expansion by \#1 (either __regexreplacementexpnot\N or \expnot\V), or turned to a string by \tl_to_str\V if inside
another c-name construction \c or \u. We use \tl_to_str\V rather than \tl_to_str\N
to deal with integers and other registers.

\begin{verbatim}
s__newprotected\Np m __regexreplacementcuaux\Nw #1
\{
 \ifcase\w \l__regexreplacementcsnames_int
 \tlbuildputright\NN \l__regexbuildtl
 \{ \expnot\N \expafter\WN \#1 \cs\aw \}
 \else:
 \tlbuildputright\NN \l__regexbuildtl
 \{ \expnot\N \expafter\WN \tltostr\V \cs\aw \}
 \fi:
 \intincr\N \l__regexreplacementcsnames_int
\}
\end{verbatim}

(End of definition for __regexreplacementcuaux\Nw.)

\__regexreplacementu\aw
Check that \u is followed by a left brace. If so, start a control sequence with \cs\aw,
which is then unpacked either with \expnot\V or \tl_to_str\V depending on the
current context.

\begin{verbatim}
s__newprotected\Np m __regexreplacementu\aw #1#2
\{
 \tokenifeqmeaning\NNTF #1 __regexreplacementlbrace\N
 \{ __regexreplacementcuaux\Nw __regexreplacementexpnot\V \}
 \{ __regexreplacementerror\NNN u #1#2 \}
 \}
\end{verbatim}

(End of definition for __regexreplacementu\aw.)
Within a \(\texttt{c}\{\ldots\} \) or \(\texttt{u}\{\ldots\} \) construction, end the control sequence, and decrease the brace count. Otherwise, this is a raw right brace.

\[\text{cs_new_protected:Np} \__regex_replacement_rbrace:N \#1\]
\[\text{if_int_compare:w} \l__regex_replacement_csnames_int > \texttt{c_zero_int}\]
\[\tl_build_put_right:Nn \l__regex_build_tl \{ \texttt{cs_end: } \}
\[\int_decr:N \l__regex_replacement_csnames_int\]
\[\text{else:}\]
\[\__regex_replacement_normal:n \{#1\}\]
\[\text{fi:}\]

(End of definition for \(\__regex_replacement_rbrace:N\).)

Within a \(\texttt{c}\{\ldots\} \) or \(\texttt{u}\{\ldots\} \) construction, this is forbidden. Otherwise, this is a raw left brace.

\[\text{cs_new_protected:Np} \__regex_replacement_lbrace:N \#1\]
\[\text{if_int_compare:w} \l__regex_replacement_csnames_int > \texttt{c_zero_int}\]
\[\text{msg_error:nnn} \{ \texttt{regex} \} \{ \texttt{cu-lbrace } \} \{ \texttt{u } \}
\[\text{else:}\]
\[\__regex_replacement_normal:n \{#1\}\]
\[\text{fi:}\]

(End of definition for \(\__regex_replacement_lbrace:N\).)

46.6.6 Characters in replacement

Here, \#1 is a letter among BEMTPUDSLOA and \#2#3 denote the next character. Complain if we reach the end of the replacement or if the construction appears inside \(\texttt{c}\{\ldots\} \) or \(\texttt{u}\{\ldots\} \), and detect the case of a parenthesis. In that case, store the current category in a sequence and switch to a new one.

\[\text{cs_new_protected:Np} \__regex_replacement_cat:NNN \#1#2#3\]
\[\text{token_if_eq_meaning:NNTF} \text{prg_do_nothing: } \#3\]
\[\text{msg_error:nn} \{ \texttt{regex} \} \{ \texttt{replacement-catcode-end } \} \}
\[\text{int_compare:nNnTF} \l__regex_replacement_csnames_int > \texttt{c_zero_int}\]
\[\text{msg_error:nnnnn}\]
\[\text{repl-catcode-in-cs } \{#1\} \{#3\} \#2 \#3\]
\[\text{token_if_eq_meaning:NNTF} \#2 \#3 \__regex_replacement_normal:n \}
\[\text{seq_push:NV} \l__regex_replacement_category_seq \]
\[\l__regex_replacement_category_tl\]
\[\tl_set:Nn \l__regex_replacement_category_tl \{#1\}\]
\[\text{token_if_eq_meaning:NNTF} \#2 \__regex_replacement_escaped:N\]
We now need to change the category code of the null character many times, hence work in a group. The catcode-specific macros below are defined in alphabetical order; if you are trying to understand the code, start from the end of the alphabet as those categories are simpler than active or begin-group.

\group_begin:
\__regex_replacement_char:nNN

The only way to produce an arbitrary character–catcode pair is to use the \lowercase or \uppercase primitives. This is a wrapper for our purposes. The first argument is the null character with various catcodes. The second and third arguments are grabbed from the input stream: \#3 is the character whose character code to reproduce. We could use \char_generate:nn but only for some catcodes (active characters and spaces are not supported).

\cs_new_protected:Npn \__regex_replacement_char:nNN \#1\#2\#3
\end_of_definition_for \__regex_replacement_char:nNN

\__regex_replacement_c_A:w

For an active character, expansion must be avoided, twice because we later do two \exp_after:wN expansions, to unpack \toks for the query, and to expand their contents to tokens of the query.

\char_set_catcode_active:N \^^@\cs_new_protected:Npn \__regex_replacement_c_A:w \__regex_replacement_char:nNN \#1 \exp_not:n \exp_not:N \^^@\end_of_definition_for \__regex_replacement_c_A:w

\__regex_replacement_c_B:w

An explicit begin-group token increases the balance, unless within a \c {...} or \u {...} construction. Add the desired begin-group character, using the standard \if_false: trick. We eventually \exp_after:wN expands twice. The first time must yield a balanced token list, and the second one gives the bare begin-group token. The \exp_after:wN is not strictly needed, but is more consistent with \tl_analysis.

\char_set_catcode_group_begin:N \^^@\cs_new_protected:Npn \__regex_replacement_c_B:w \__regex_replacement_char:nNN \exp_not:n \exp_not:N \^^@\end_of_definition_for \__regex_replacement_c_B:w
This is not quite catcode-related: when the user requests a character with category “control sequence”, the one-character control symbol is returned. As for the active character, we prepare for two e-expansions.

Subscripts fit the mould: \texttt{\texttt{lowercase}} the null byte with the correct category.

Similar to the begin-group case, the second e-expansion produces the bare end-group token.

Simply \texttt{\texttt{lowercase}} a letter null byte to produce an arbitrary letter.

No surprise here, we lowercase the null math toggle.
\_\_regex_replacement_c\_O:w \ Lowercase an other null byte.
\char_set_catcode_other:N \ ^^@ 
\cs_new_protected:Npn \_\_regex_replacement_c\_O:w 
\{ \_\_regex_replacement_char:nNN \{ ^\{ @ \} \} \}

\End{definition for \_\_regex_replacement_c\_O:w}

\_\_regex_replacement_c\_P:w \ For macro parameters, expansion is a tricky issue. \ We need to prepare for two e-
expansions and passing through various macro definitions. \ Note that we cannot replace \texttt{one \exp_not:n} by doubling the macro parameter characters because this would mis-
beweif a mischievous user asks for \texttt{\c\{\cP\#}} , since that macro parameter character
would be doubled.
\char_set_catcode_parameter:N \ ^^@ 
\cs_new_protected:Npn \_\_regex_replacement_c\_P:w 
\{ \_\_regex_replacement_char:nNN 
\exp_not:n \{ \exp_not:n \{ ^\{ @ ^\{ @ ^\{ @ ^\{ @ \} \} \} \}
\}

\End{definition for \_\_regex_replacement_c\_P:w}

\_\_regex_replacement_c\_S:w \ Spaces are normalized on input by \TeX to have character code 32. \ It is in fact impossible
to get a token with character code 0 and category code 10. \ Hence we use 32 instead of 0 as our base character.
\cs_new_protected:Npn \_\_regex_replacement_c\_S:w #1#2 
\if_int_compare:w '#2 = \c_zero_int \msg_error:nn \{ regex \} \{ replacement-null-space \} 
\fi:
\tex_lccode:D '\ = '#2 \scan_stop:
\tex_lowercase:D \{ \_\_regex_replacement_put:n \{-\} \}

\End{definition for \_\_regex_replacement_c\_S:w}

\_\_regex_replacement_c\_T:w \ No surprise for alignment tabs here. \ Those are surrounded by the appropriate braces
whenever necessary, hence they don’t cause trouble in alignment settings.
\char_set_catcode_alignment:N \ ^^@ 
\cs_new_protected:Npn \_\_regex_replacement_c\_T:w 
\{ \_\_regex_replacement_char:nNN \{ ^\{ @ \} \} \}

\End{definition for \_\_regex_replacement_c\_T:w}

\_\_regex_replacement_c\_U:w \ Simple call to \_\_regex_replacement_char:nNN which lowercases the math superscript
\texttt{\^\{ @ \}.  
\char_set_catcode_math_superscript:N \ ^^@ 
\cs_new_protected:Npn \_\_regex_replacement_c\_U:w 
\{ \_\_regex_replacement_char:nNN \{ ^\{ @ \} \} \}

\End{definition for \_\_regex_replacement_c\_U:w}

\ Restore the catcode of the null byte.
\group_end:
46.6.7 An error

Simple error reporting by calling one of the messages `replacement-c`, `replacement-g`, or `replacement-u`.

```latex
\cs_new_protected:Npn \_\regex_replacement_error:NNN \#1#2#3
\{\msg_error:nne { regex } { replacement-#1 } {#3} \#2 \#3
\}
```

*(End of definition for `\_\regex_replacement_error:NNN`)*

46.7 User functions

Before being assigned a sensible value, a regex variable matches nothing.

```latex
\cs_new_protected:Npn \regex_new:N \#1\{ \cs_new_eq:NN \#1 \c__regex_no_match_regex \}
```

*(End of definition for `\regex_new:N`. This function is documented on page 55.)*

The usual scratch space.

<table>
<thead>
<tr>
<th>\l_tma_regex</th>
<th>\l_tmbr_regex</th>
<th>\gtmpa_regex</th>
<th>\g_tmb_regex</th>
</tr>
</thead>
<tbody>
<tr>
<td><code>\regex_new:N \l_tma_regex</code></td>
<td><code>\regex_new:N \l_tmbr_regex</code></td>
<td><code>\regex_new:N \gtmpa_regex</code></td>
<td><code>\regex_new:N \g_tmb_regex</code></td>
</tr>
</tbody>
</table>

*(End of definition for `\l_tma_regex` and others. These variables are documented on page 60.)*

Compile, then store the result in the user variable with the appropriate assignment function.

```latex
\cs_new_protected:Npn \regex_set:Nn \#1#2\{ \__regex_compile:n {#2} \tl_set_eq:NN \#1 \l__regex_internal_regex \}
\cs_new_protected:Npn \regex_gset:Nn \#1#2\{ \__regex_compile:n {#2} \tl_gset_eq:NN \#1 \l__regex_internal_regex \}
\cs_new_protected:Npn \regex_const:Nn \#1#2\{ \__regex_compile:n {#2} \tl_const:Ne \#1 { \exp_not:o \l__regex_internal_regex } \}
```

*(End of definition for `\regex_set:Nn`, `\regex_gset:Nn`, and `\regex_const:Nn`. These functions are documented on page 55.)*

User functions: the `n` variant requires compilation first. Then show the variable with some appropriate text. The auxiliary `\_\regex_show:N` is defined in a different section.

```latex
\cs_new_protected:Npn \regex_show:n \#1\{ \__regex_show:Nn \msg_show:nneeee \}
\cs_new_protected:Npn \regex_log:n \{ \__regex_show:Nn \msg_log:nneeee \}
\cs_new_protected:Npn \__regex_show:NN \#1\{ \exp_not:o \l__regex_internal_regex \}
```

*(End of definition for `\regex_show:n`, `\regex_log:n`, `\_\regex_show:N`, `\regex_show:N`, `\regex_log:N`, `\_\regex_show:NN`)*
\regex_match:nn\regex_match:nV\regex_match:NN\regex_match:NV

Those conditionals are based on a common auxiliary defined later. Its first argument builds the NFA corresponding to the regex, and the second argument is the query token list. Once we have performed the match, convert the resulting boolean to \prg_return_-true: or false.

\prg_new_protected_conditional:Npnn \regex_match:nn \regex_match:nV \regex_match:NN \regex_match:NV

(End of definition for \regex_match:nnTF and \regex_match:NnTF. These functions are documented on page 56.)

\regex_count:nnN\regex_count:nVN\regex_count:NnN\regex_count:NVN

Again, use an auxiliary whose first argument builds the NFA.

\cs_new_protected:Npn \regex_count:nnN \regex_count:nVN \regex_count:NnN \regex_count:NVN

(End of definition for \regex_count:nnN and \regex_count:nVN. These functions are documented on page 56.)
The auxiliary errors if #1 has an odd number of items, and otherwise it sets \g__regex_case_int according to which case was found (zero if not found). The true branch leaves the corresponding code in the input stream.

\cs_set_protected:Npn \__regex_tmp:w #1#2#3
{ \cs_new_protected:Npn #2 ##1 { #1 { \__regex_build:n {##1} } } \cs_new_protected:Npn #3 ##1 { #1 { \__regex_build:N ##1 } } \prg_new_protected_conditional:Npnn #2 ##1##2##3 { T , F , TF } { #1 { \__regex_build:n {##1} } {##2} ##3 \__regex_return: } \prg_new_protected_conditional:Npnn #3 ##1##2##3 { T , F , TF } { #1 { \__regex_build:N ##1 } {##2} ##3 \__regex_return: } \cs_generate_variant:Nn #2 { nV } \prg_generate_conditional_variant:Nnn #2 { nV } { T , F , TF } \cs_generate_variant:Nn #3 { NV } \prg_generate_conditional_variant:Nnn #3 { NV } { T , F , TF } }

\__regex_tmp:w \__regex_extract_once:nnN \regex_extract_once:nnN \regex_extract_once:NnN
\__regex_tmp:w \__regex_extract_all:nnN \regex_extract_all:nnN \regex_extract_all:NnN
\__regex_tmp:w \__regex_replace_once:nnN \regex_replace_once:nnN \regex_replace_once:NnN
\__regex_tmp:w \__regex_split:nnN \regex_split:nnN \regex_split:NnN

(End of definition for \regex_match_case:nnTF. This function is documented on page 56.)

\regex_replace_case_once:nN
\regex_replace_case_once:nN
\regex_replace_case_once:nN
\regex_replace_case_once:nN
\regex_replace_case_once:nN
\regex_replace_case_once:nN

If the input is bad (odd number of items) then take the false branch. Otherwise, use the same auxiliary as \regex_replace_once:nnN, but with more complicated code to
build the automaton, and to find what replacement text to use. The `\tl_item:nn` is only expanded once we know the value of `\g__regex_case_int`, namely which case matched.

```latex
\cs_new_protected:Npn \regex_replace_case_once:nNTF #1#2
\int_if_odd:nTF { \tl_count:n (#1) }
\msg_error:nneee { regex } { case-odd }
\token_to_str:N \regex_replace_case_once:nN(TF) { code }
\tl_count:n (#1) \tl_to_str:n (#1) \use_ii:nn

\__regex_replace_case_once_aux:nnN
\__regex_case_build:e { \__regex_tl_odd_items:n (#1) }
\__regex_case_replacement:e { \tl_item:nn {#1} { 2 * \g__regex_case_int } }
\bool_if:NTF \g__regex_success_bool
\cs_new_protected:Npn \regex_replace_case_once:nN #1#2
{ \regex_replace_case_once:nNTF {#1} {#2} { } { } }
\cs_new_protected:Npn \regex_replace_case_once:nNT #1#2#3
{ \regex_replace_case_once:nNTF {#1} {#2} {#3} { } }
\cs_new_protected:Npn \regex_replace_case_once:nNF #1#2
{ \regex_replace_case_once:nNTF {#1} {#2} { } }

(End of definition for `\regex_replace_case_once:nNTF`. This function is documented on page 59.)
```

If the input is bad (odd number of items) then take the false branch. Otherwise, use the same auxiliary as `\regex_replace_all:nnN`, but with more complicated code to build the automaton, and to find what replacement text to use.

```latex
\cs_new_protected:Npn \regex_replace_case_all:nNTF #1#2
\int_if_odd:nTF { \tl_count:n (#1) }
\msg_error:nneee { regex } { case-odd }
\token_to_str:N \regex_replace_case_all:nN(TF) { code }
\tl_count:n (#1) \tl_to_str:n (#1) \use_ii:nn

\__regex_replace_all_aux:nnN
\__regex_case_build:e { \__regex_tl_odd_items:n (#1) }
\__regex_case_replacement:e { \__regex_tl_even_items:n (#1) }
\bool_if:NTF \g__regex_success_bool
\cs_new_protected:Npn \regex_replace_case_all:nN #1#2
{ \regex_replace_case_all:nNTF {#1} {#2} { } { } }
\cs_new_protected:Npn \regex_replace_case_all:nNT #1#2#3
{ \regex_replace_case_all:nNTF {#1} {#2} {#3} { } }
\cs_new_protected:Npn \regex_replace_case_all:nNF #1#2
{ \regex_replace_case_all:nNTF {#1} {#2} { } }

(End of definition for `\regex_replace_case_all:nNTF`. This function is documented on page 59.)
```
46.7.1 Variables and helpers for user functions

The number of matches found so far is stored in \l__regex_match_count_int. This is only used in the \regex_count functions.

\l__regex_match_count_int

Those flags are raised to indicate begin-group or end-group tokens that had to be added when extracting submatches.

\l__regex_begin_flag \l__regex_end_flag

The end-points of each submatch are stored in two arrays whose index \langle submatch \rangle ranges from \l__regex_min_submatch_int (inclusive) to \l__regex_submatch_int (exclusive). Each successful match comes with a 0-th submatch (the full match), and one match for each capturing group: submatches corresponding to the last successful match are labelled starting at \l__regex_zeroth_submatch_int. The entry \l__regex_zeroth_submatch_int in \g__regex_submatch_prev_intarray holds the position at which that match attempt started: this is used for splitting and replacements.

\l__regex_min_submatch_int \l__regex_submatch_int \l__regex_zeroth_submatch_int

Hold the place where the match attempt begun, the end-points of each submatch, and which regex case the match corresponds to, respectively.

\g__regex_submatch_prev_intarray \g__regex_submatch_begin_intarray \g__regex_submatch_end_intarray \g__regex_submatch_case_intarray

The first thing we do when matching is to store the balance of begin-group/end-group characters into \g__regex_balance_intarray.

\g__regex_balance_intarray

Keep track of the number of left/right braces to add when performing a regex operation such as a replacement.

\l__regex_added_begin_int \l__regex_added_end_int
\_\_regex_return: This function triggers either \texttt{prg\_return\_false}: or \texttt{prg\_return\_true}: as appropriate to whether a match was found or not. It is used by all user conditionals.

\begin{verbatim}
\cs_new_protected:Npn \_\_regex_return:
\{ \if_meaning:w \c_true_bool \g__regex_success_bool \prg\_return\_true:\ \else:\ \prg\_return\_false:\ \fi:\ }
\end{verbatim}

(End of definition for \_\_regex_return:.)

\_\_regex_query_set:n \_\_regex_query_set_aux:nN
To easily extract subsets of the input once we found the positions at which to cut, store the input tokens one by one into successive \texttt{toks} registers. Also store the brace balance (used to check for overall brace balance) in an array.

\begin{verbatim}
\cs_new_protected:Npn \_\_regex_query_set:n #1
\{ \int_zero:N \l__regex_balance_int \int_zero:N \l__regex_curr_pos_int \_\_regex_query_set_aux:nN { } F \tl_analysis_map_inline:nn {#1} \{ \_\_regex_query_set_aux:nN {##1} ##3 \} \_\_regex_query_set_aux:nN { } F \int_set_eq:NN \l__regex_max_pos_int \l__regex_curr_pos_int \}
\cs_new_protected:Npn \_\_regex_query_set_aux:nN #1#2
\{ \int_incr:N \l__regex_curr_pos_int \_\_regex_toks_set:Nn \l__regex_curr_pos_int {#1} \_\kernel_intarray_gset:Nnn \g__regex_balance_intarray \l__regex_curr_pos_int \l__regex_balance_int \if_case:w "#2 \exp_stop_f:\or: \int_incr:N \l__regex_balance_int \or: \int_decr:N \l__regex_balance_int \fi:\ }
\end{verbatim}

(End of definition for \_\_regex_query_set:n and \_\_regex_query_set_aux:nN.)

46.7.2 Matching

\_\_regex_if_match:nn
We don’t track submatches, and stop after a single match. Build the NFA with \#1, and perform the match on the query \#2.

\begin{verbatim}
\cs_new_protected:Npn \_\_regex_if_match:nn #1#2
\{ \group_begin:\ \_\_regex_disable_submatches:\ \_\_regex_single_match:\ #1 \_\_regex_match:n \{#2\} \group_end:\ }
\end{verbatim}
The code would get badly messed up if the number of items in \#1 were not even, so we catch this case, then follow the same code as __regex_match:nnTF but using __regex_case_build:n and without returning a result.

7423 \cs_new_protected:Npn \__regex_match_case:nnTF #1#2
7424 {  
7425 \int_if_odd:nTF { \tl_count:n {#1} } 
7426 {  
7427 \msg_error:nneee { regex } { case-odd } 
7428 { \token_to_str:N \regex_match_case:nn(TF) } { code } 
7429 { \tl_count:n {#1} } { \tl_to_str:n {#1} } 
7430 \use_ii:nn 
7431 }  
7432 {  
7433 \__regex_if_match:nn 
7434 { \__regex_case_build:e { \__regex_tl_odd_items:n {#1} } } 
7435 {#2}  
7436 \bool_if:NTF \g__regex_success_bool 
7437 }  
7438 }  
7439 \cs_new:Npn \__regex_match_case_aux:nn #1#2 { \exp_not:n { {#1} } }

(End of definition for __regex_match_case:nnTF and __regex_match_case_aux:nn.)

Again, we don’t care about submatches. Instead of aborting after the first “longest match” is found, we search for multiple matches, incrementing \l__regex_match_count_int every time to record the number of matches. Build the NFA and match. At the end, store the result in the user’s variable.

7440 \cs_new_protected:Npn \__regex_count:nnN #1#2#3
7441 {  
7442 \group_begin:  
7443 \__regex_disable_submatches:  
7444 \int_zero:N \l__regex_match_count_int  
7445 \__regex_multi_match:n { \int_incr:N \l__regex_match_count_int } 
7446 {#1}  
7447 \__regex_match:n {#2}  
7448 \exp_args:NNNo  
7449 \group_end:  
7450 \int_set:Nn #3 { \int_use:N \l__regex_match_count_int }  
7451 }

(End of definition for __regex_count:nnN.)

46.7.3 Extracting submatches

Match once or multiple times. After each match (or after the only match), extract the submatches using __regex_extract:. At the end, store the sequence containing all the submatches into the user variable \#3 after closing the group.

7452 \cs_new_protected:Npn \__regex_extract_once:nnN \__regex_extract_all:nnN #1#2#3
7453 {  
7454 \group_begin:  
7455 \__regex_single_match:
\_\_regex_split:nnN Splitting at submatches is a bit more tricky. For each match, extract all submatches, and replace the zeroth submatch by the part of the query between the start of the match attempt and the start of the zeroth submatch. This is inhibited if the delimiter matched an empty token list at the start of this match attempt. After the last match, store the last part of the token list, which ranges from the start of the match attempt to the end of the query. This step is inhibited if the last match was empty and at the very end: decrement \l__regex_submatch_int, which controls which matches will be used.
The end-points of submatches are stored as entries of two arrays from \texttt{l\_regex\_min\_submatch\_int} to \texttt{l\_regex\_submatch\_int} (exclusive). Extract the relevant ranges into \texttt{g\_regex\_internal\_tl}, separated by \texttt{\_regex\_tmp:w \{\}. We keep track in the two flags \texttt{\_regex\_begin} and \texttt{\_regex\_end} of the number of begin-group or end-group tokens added to make each of these items overall balanced. At this step, \texttt{\_regex\_tmp:w} is counted as being balanced (same number of begin-group and end-group tokens). This problem is caught by \texttt{\_regex\_extract\_check:w}, explained later. After complaining about any begin-group or end-group tokens we had to add, we are ready to construct the user's sequence outside the group.

\begin{verbatim}
\cs_new_protected:Npn \_regex\_group\_end\_extract\_seq:N #1
\{ \_regex\_group\_end\_extract\_seq:N #2 \}
\cs_gset_protected:Npn \_regex\_extract\_seq:N #1
\{ \_regex\_extract\_seq_loop:Nw \}
\end{verbatim}
\_\regex_extract_seq_aux:n
\_\regex_extract_seq_aux:ww

The \_\regex_extract_seq_aux:n auxiliary builds one item of the sequence of submatches. First compute the brace balance of the submatch, then extract the submatch from the query, adding the appropriate braces and raising a flag if the submatch is not balanced.

\cs_new:Npn \_\regex_extract_seq_aux:n #1
\{ \_\regex_tmp:w { } \exp_after:wN \_\regex_extract_seq_aux:ww \int_value:w \_\regex_submatch_balance:n {#1} ; #1; \}
\cs_new:Npn \_\regex_extract_seq_aux:ww #1; #2;
\{ \if_int_compare:w #1 < \c_zero_int \prg_replicate:nn {-#1}
\{ \flag_raise:N \l__regex_begin_flag \exp_not:n { { \if_false: } \fi: } \}
\} \if_int_compare:w #1 > \c_zero_int \prg_replicate:nn {#1}
\{ \flag_raise:N \l__regex_end_flag \exp_not:n { \if_false: \{ \fi: } \}
\}
\}

(End of definition for \_\regex_group_end_extract_seq:N and others.)

\_\regex_extract_check:w
\_\regex_extract_check:n
\_\regex_extract_check_loop:w
\_\regex_extract_check_end:w

In \_\regex_group_end_extract_seq:N we had to expand \g__regex_internal_tl to turn \if_false: constructions into actual begin-group and end-group tokens. This is done with a \_\kernel_tl_gset:Nx assignment, and \_\regex_extract_check:w is run immediately after this assignment ends, thanks to the \afterassignment primitive. If all of the items were properly balanced (enough begin-group tokens before end-group tokens, so } is not) then \_\regex_extract_check:w is called just before the closing brace of the \_\kernel_tl_gset:Nx (thanks to our sneaky \if_false: \{ \fi: \} construction), and finds that there is nothing left to expand. If any of the items is unbalanced, the
assignment gets ended early by an extra end-group token, and our check finds more
tokens needing to be expanded in a new \_kernel_tl_gset:Nx assignment. We need
to add a begin-group and an end-group tokens to the unbalanced item, namely to the
last item found so far, which we reach through a loop.

\cs_new_protected:Nnpn \_regex_extract_check:w
\exp_after:wN \_regex_extract_check:n
\exp_after:wN { \if_false: } \fi:
\cs_new_protected:Nnpn \_regex_extract_check:n #1
\tl_if_empty:nF {#1}
\int_incr:N \l__regex_added_begin_int
\int_incr:N \l__regex_added_end_int
\tex_afterassignment:D \_regex_extract_check:w
\_kernel_tl_gset:Nx \g__regex_internal_tl
\exp_after:wN \_regex_extract_check_loop:w
\g__regex_internal_tl
\_regex_tmp:w \_regex_extract_check_end:w
#1
}

\cs_new:Npn \_regex_extract_check_loop:w #1 \_regex_tmp:w #2
\exp_not:o #1\_regex_extract_check_loop:w \_regex_tmp:w \prg_do_nothing:

Arguments of \_regex_extract_check_end:w are: #1 is the part of the item before
the extra end-group token; #2 is junk; #3 is \prg_do_nothing: followed by the not-yetexpanded part of the item after the extra end-group token. In the replacement text,
the first brace and the \if_false: \fi: construction are the added begin-group and
end-group tokens (the latter being not-yet expanded, just like #3), while the closing brace
after \exp_not:o #1 replaces the extra end-group token that had ended the assignment
early. In particular this means that the character code of that end-group token is lost.

\cs_new:Npn \_regex_extract_check:end:w
\exp_not:o \_regex_extract_check_loop:w \_regex_tmp:w \prg_do_nothing:

(End of definition for \_regex_extract_check:w and others.)

\_regex_extract: \_regex_extract_aux:w

Our task here is to store the list of end-points of submatches, and store them in appro-
priate array entries, from \_regex_zeroth_submatch_int upwards. First, we store in
\g__regex_submatch_prev_intarray the position at which the match attempt started.
We extract the rest from the comma list \l__regex_success_submatches_tl, which starts with entries to be stored in \g__regex_submatch_begin_intarray and continues with entries for \g__regex_submatch_end_intarray.

```latex
\cs_new_protected:Npn \__regex_extract:
{ \if_meaning:w \c_true_bool \g__regex_success_bool
  \int_set_eq:NN \l__regex_zeroth_submatch_int \l__regex_submatch_int
  \prg_replicate:nn \l__regex_capturing_group_int
  \{ \__kernel_intarray_gset:Nnn \g__regex_submatch_prev_intarray
  \l__regex_zeroth_submatch_int \c_zero_int
  \__kernel_intarray_gset:Nnn \g__regex_submatch_case_intarray
  \l__regex_zeroth_submatch_int \c_zero_int
  \int_incr:N \l__regex_submatch_int
  \__kernel_intarray_gset:Nnn \g__regex_submatch_prev_intarray
  \l__regex_zeroth_int \l__regex_start_pos_int
  \__kernel_intarray_gset:Nnn \g__regex_submatch_case_intarray
  \l__regex_zeroth_submatch_int \c_zero_int
  \int_zero:N \l__regex_internal_a_int
  \exp_after:wN \__regex_extract_aux:w \l__regex_success_submatches_tl
  \prg_break_point: \__regex_use_none_delimit_by_q_recursion_stop:w,
  \q__regex_recursion_stop
  \fi:
}
\cs_new_protected:Npn \__regex_extract_aux:w #1 ,
{ \prg_break: #1 \prg_break_point:
  \if_int_compare:w \l__regex_internal_a_int < \l__regex_capturing_group_int
  \__kernel_intarray_gset:Nnn \g__regex_submatch_begin_intarray
  \l__regex_zeroth_submatch_int \l__regex_start_pos_int
  \{ \__regex_int_eval:w \l__regex_zeroth_submatch_int + \l__regex_internal_a_int \}
  \else:
    \__kernel_intarray_gset:Nnn \g__regex_submatch_end_intarray
    \{ \__regex_int_eval:w
      \l__regex_zeroth_submatch_int + \l__regex_internal_a_int
      - \l__regex_capturing_group_int
    \}
  \fi:
  \int_incr:N \l__regex_internal_a_int
  \__regex_extract_aux:w
}(End of definition for \__regex_extract: and \__regex_extract_aux:w.)
```

46.7.4 Replacement

Build the NFA and the replacement functions, then find a single match. If the match failed, simply exit the group. Otherwise, we do the replacement. Extract submatches. Compute the brace balance corresponding to replacing this match by the replacement (this depends on submatches). Prepare the replaced token list: the replacement function produces the tokens from the start of the query to the start of the match and the replacement text for
this match; we need to add the tokens from the end of the match to the end of the query.
Finally, store the result in the user’s variable after closing the group: this step involves
an additional e-expansion, and checks that braces are balanced in the final result.

\cs_new_protected:Npn \__regex_replace_once:nnN #1#2
\{ \__regex_replace_once_aux:nnN (#1) \{ \__regex_replacement:n (#2) \} \}
\cs_new_protected:Npn \__regex_replace_once_aux:nnN #1#2#3
\{ \group_begin:
\__regex_single_match:
#1
\exp_args:No \__regex_match:n (#3)
\bool_if:NTF \g__regex_success_bool
{ \__regex_extract:
\exp_args:No \__regex_query_set:n (#3)
#2
\int_set:Nn \l__regex_balance_int
\{ \__regex_replacement_balance_one_match:n \l__regex_zeroth_submatch_int 
\__kernel_tl_set:Nx \l__regex_internal_a_tl
{ \__regex_replacement_do_one_match:n \l__regex_zeroth_submatch_int
\__regex_query_range:nn
{ \__kernel_intarray_item:Nn \g__regex_submatch_end_intarray
\l__regex_zeroth_submatch_int
\l__regex_max_pos_int
}
\__regex_group_end_replace:N #3
}
\} \group_end: \}
\)

\__regex_replace_all:nnN Match multiple times, and for every match, extract submatches and additionally store
the position at which the match attempt started. The entries from \l__regex_min_submatch_int
to \l__regex_submatch_int hold information about submatches of every
match in order; each match corresponds to \l__regex_capturing_group_int consecutive entries. Compute the brace balance corresponding to doing all the replacements:
this is the sum of brace balances for replacing each match. Join together the replacement
texts for each match (including the part of the query before the match), and the end of
the query.

\cs_new_protected:Npn \__regex_replace_all:nnN #1#2
\{ \__regex_replace_all_aux:nnN (#1) \{ \__regex_replacement:n (#2) \} \}
\cs_new_protected:Npn \__regex_replace_all_aux:nnN #1#2#3
\{ \group_begin:
\__regex_multi_match:n \{ \__regex_extract: \}
#1
\exp_args:No \__regex_match:n (#3)
\exp_args:No \__regex_query_set:n (#3)
#2
#3
\} \group_end: \}
\int_set:Nn \l__regex_balance_int
{
\c_zero_int
\int_step_function:nnnN
\l__regex_min_submatch_int
\l__regex_capturing_group_int
{ \l__regex_submatch_int \c_one_int }
\l__regex_replacement_balance_one_match:n
}
\__kernel_tl_set:Nx \l__regex_internal_a_tl
{
\int_step_function:nnnN
\l__regex_min_submatch_int
\l__regex_capturing_group_int
{ \l__regex_submatch_int \c_one_int }
\__regex_replacement_do_one_match:n
\__regex_query_range:nn
\l__regex_start_pos_int \l__regex_max_pos_int
}
\__regex_group_end_replace:N #3
}

(End of definition for \__regex_replace_all:nnN.)

At this stage \l__regex_internal_a_tl (e-expands to the desired result). Guess from \l__regex_balance_int the number of braces to add before or after the result then try expanding. The simplest case is when \l__regex_internal_a_tl together with the braces we insert via \prg_replicate:nn give a balanced result, and the assignment ends at the \if_false: { \fi: } construction: then \__regex_group_end_replace_check:w sees that there is no material left and we successfully found the result. The harder case is that expanding \l__regex_internal_a_tl may produce extra closing braces and end the assignment early. Then we grab the remaining code using; importantly, what follows has not yet been expanded so that \__regex_group_end_replace_check:n grabs everything until the last brace in \__regex_group_end_replace_try:, letting us try again with an extra surrounding pair of braces.

\cs_new_protected:Npn \__regex_group_end_replace:N #1
{
\int_set:Nn \l__regex_added_begin_int
{ \int_max:nn { - \l__regex_balance_int } \c_zero_int }
\int_set:Nn \l__regex_added_end_int
{ \int_max:nn \l__regex_balance_int \c_zero_int }
\__regex_group_end_replace_try:
\int_compare:nNnT { \l__regex_added_begin_int + \l__regex_added_end_int }
> \c_zero_int
{
\msg_error:nnee { regex } { result-unbalanced }
{ replacing } \{ \int_use:N \l__regex_added_begin_int }
{ \int_use:N \l__regex_added_end_int }
}
\group_end:
\tl_set_eq:NN #1 \g__regex_internal_tl
}
\cs_new_protected:Npn \__regex_group_end_replace_try:
46.7.5 Peeking ahead

\_\_regex\_peek\_true\_tl
\_\_regex\_peek\_false\_tl

True/false code arguments of \peek\_regex\_nTF or similar.

\_\_regex\_replacement\_tl

When peeking in \peek\_regex\_replace\_once\_nnTF we need to store the replacement text.

\_\_regex\_input\_tl
\_\_regex\_input\_item\_n

Stores each token found as \_\_regex\_input\_item\_n \{\langle tokens\rangle\}, where the \langle tokens\rangle o-expand to the token found, as for \tl\_analysis\_map\_inline\_nn.

The T and F functions just call the corresponding TF function. The four TF functions differ along two axes: whether to remove the token or not, distinguished by using \_\_regex\_\_peek\_end: or \_\_regex\_peek\_remove\_end\_n (the latter case needs an argument, as we will see), and whether the regex has to be compiled or is already in an N-type variable, distinguished by calling \_\_regex\_build\_aux\_NN or \_\_regex\_build\_aux\_NN. The first argument of these functions is \c\_false\_bool to indicate that there should be no implicit
insertion of a wildcard at the start of the pattern: otherwise the code would keep looking further into the input stream until matching the regex.

\begin{verbatim}
\cs_new_protected:Npn \peek_regex:nTF #1
\{ \__regex_peek:nnTF
\{ \__regex_build_aux:Nn \c_false_bool {#1} \}
\{ \__regex_peek_end: \}
\}
\cs_new_protected:Npn \peek_regex:nT #1#2
\{ \peek_regex:nTF {#1} {#2} { } \}
\cs_new_protected:Npn \peek_regex:nF #1 { \peek_regex:nTF {#1} { } }
\cs_new_protected:Npn \peek_regex:NTF #1
\{ \__regex_peek:nnTF
\{ \__regex_build_aux:NN \c_false_bool #1 \}
\{ \__regex_peek_end: \}
\}
\cs_new_protected:Npn \peek_regex:NT #1#2
\{ \peek_regex:NTF #1 {#2} { } \}
\cs_new_protected:Npn \peek_regex:NF #1 { \peek_regex:NTF {#1} { } }
\cs_new_protected:Npn \peek_regex_remove_once:nTF #1
\{ \__regex_peek:nnTF
\{ \__regex_build_aux:Nn \c_false_bool {#1} \}
\{ \__regex_peek_remove_end:n {##1} \}
\}
\cs_new_protected:Npn \peek_regex_remove_once:nT #1#2
\{ \peek_regex_remove_once:nTF {#1} {#2} { } \}
\cs_new_protected:Npn \peek_regex_remove_once:nF #1
\{ \peek_regex_remove_once:nTF {#1} { } \}
\cs_new_protected:Npn \peek_regex_remove_once:NTF #1
\{ \__regex_peek:nnTF
\{ \__regex_build_aux:NN \c_false_bool #1 \}
\{ \__regex_peek_remove_end:n {##1} \}
\}
\cs_new_protected:Npn \peek_regex_remove_once:NT #1#2
\{ \peek_regex_remove_once:NTF #1 {#2} { } \}
\cs_new_protected:Npn \peek_regex_remove_once:NF #1
\{ \peek_regex_remove_once:NTF #1 { } \}
\end{verbatim}

(End of definition for \peek_regex:nTF and others. These functions are documented on page 209.)

\_\_regex_peek:nnTF \_\_regex_peek_aux:nnTF

Store the user’s true/false codes (plus \texttt{\textbackslash group_end:}) into two token lists. Then build the automaton with \texttt{#1}, without submatch tracking, and aiming for a single match. Then start matching by setting up a few variables like for any regex matching like \texttt{\regex_match:nnTF}, with the addition of \texttt{\_\_regex_input_tl} that keeps track of the tokens seen, to reinsert them at the end. Instead of \texttt{\tl_analysis_map_inline:nn} on the input, we call \texttt{\peek_analysis_map_inline:n} to go through tokens in the input stream. Since \texttt{\_\_regex_match_one_token:nnN} calls \texttt{\_\_regex_maplike_break}: we need to catch that and break the \texttt{\peek_analysis_map_inline:n} loop instead.
% __regex_peek_aux:nnTF
{
    __regex_disable_submatches:
    \}
}\cs_new_protected:Npn __regex_peek_aux:nnTF #1#2#3#4
{
    \group_begin:
    \tl_set:Nn \l__regex_peek_true_tl { \group_end: #3 }
    \tl_set:Nn \l__regex_peek_false_tl { \group_end: #4 }
    __regex_single_match:
    #1
    __regex_match_init:
    \tl_build_begin:N \l__regex_input_tl
    \__regex_match_once_init:
    \peek_analysis_map_inline:n
    \group_begin:
    \tl_build_put_right:Nn \l__regex_input_tl
    \{ \__regex_input_item:n {##1} \}
    \__regex_match_one_token:nnN {##1} {##2} ##3
    \use_none:nn
    \prg_break_point:Nn \__regex_maplike_break:
    \{ \peek_analysis_map_break:n {#2} \}
    \}
    \}
}(End of definition for __regex_peek:nnTF and __regex_peek_aux:nnTF.)

\__regex_peek_end:
\__regex_remove_end:n
Once the regex matches (or permanently fails to match) we call __regex_peek_end:, or \__regex_remove_end:n with argument the last token seen. For \peek_regex:nTF we reinsert tokens seen by calling __regex_remove_end:n regardless of the result of the match. For \peek_regex_remove_once:nTF we reinsert the tokens seen only if the match failed; otherwise we just reinsert the tokens #1, with one expansion. To be more precise, #1 consists of tokens that \o-expand and \e-expand to the last token seen, for example it is \exp_not:N \langle cs \rangle for a control sequence. This means that just doing \exp_after:wN \l__regex_peek_true_tl #1 would be unsafe because the expansion of \langle cs \rangle would be suppressed.
\__regex_peek_end:
\__regex_remove_end:n
{(End of definition for __regex_peek_end: and __regex_remove_end:n.)}
Insert the true/false code \#1, followed by the tokens found, which were stored in \_\_\_\_regex_input_tl. For this, loop through that token list using \_\_\_\_regex_reinsert_item:n, which expands \#1 once to get a single token, and jumps over it to expand what follows, with suitable \exp:w and \exp_end:. We cannot just use \use:e on the whole token list because the result may be unbalanced, which would stop the primitive prematurely, or let it continue beyond where we would like.

\cs_new_protected:Npn \_\_\_\_regex_peek_reinsert:N #1
\tl_build_end:N \l__\_\_\_\_regex_input_tl
\cs_set_eq:NN \_\_\_\_regex_input_item:n \_\_\_\_regex_reinsert_item:n
\exp_after:wN #1 \exp:w \l__\_\_\_\_regex_input_tl \exp_end:
\)

(End of definition for \_\_\_\_regex_peek_reinsert:N and \_\_\_\_regex_reinsert_item:n.)

Similar to \peek_regex_replace_once:nnTF above.

\cs_new_protected:Npn \peek_regex_replace_once:nnTF #1 #2 #3
\peek_regex_replace_once:nnTF #1 { #2 } { } { } \)

(End of definition for \peek_regex_replace_once:nnTF and \peek_regex_replace_once:nnTF. These functions are documented on page 210.)

Same as \_\_\_\_regex_peeck:nnTF (used for \peek_regex_replace_once:nnTF above), but without disabling submatches, and with a different end. The replacement text \#2 is stored, to be analyzed later.

\cs_new_protected:Npn \_\_\_\_regex_peeck_replace:nnTF #1 #2
\tl_set:Nn \l__\_\_\_\_regex_replacement_tl {#2}
\_\_\_\_regex_peek_aux:nnTF {#1} \{ \_\_\_\_regex_peeck_replace_end: \}
\)

(End of definition for \_\_\_\_regex_peeck_replace:nnTF.)
If the match failed \__regex_peek_reinsert:N reinserts the tokens found. Otherwise, finish storing the submatch information using \__regex_extract:, and store the input into \toks. Redefine a few auxiliaries to change slightly their expansion behaviour as explained below. Analyse the replacement text with \__regex_replacement:n, which as usual defines \__regex_replacement_do_one_match:n to insert the tokens from the start of the match attempt to the beginning of the match, followed by the replacement text. The \use:e expands for instance the trailing \__regex_query_range:nn down to a sequence of \__regex_insert_item:n {⟨tokens⟩} where ⟨tokens⟩ o-expand to a single token that we want to insert. After e-expansion, \use:e does \use:n, so we have \exp_after:wN \l__regex.Peek_true_tl \exp:w ... \exp_end:. This is set up such as to obtain \l__regex.Peek_true_tl followed by the replaced tokens (possibly unbalanced) in the input stream.

\cs_new_protected:Npn \__regex_peek_replace_end:
\bool_if:NTF \g__regex_success_bool
{\__regex_extract:
  \__regex_query_set_from_input_tl:
  \cs_set_eq:NN \__regex_replacement_put:n \__regex_peek_replacement_put:n
  \cs_set_eq:NN \__regex_replacement_put_submatch_aux:n
  \__regex_peek_replacement_put_submatch_aux:n
  \cs_set_eq:NN \__regex_input_item:n \__regex_reinsert_item:n
  \cs_set_eq:NN \__regex_replacement_exp_not:N \__regex_peek_replacement_token:n
  \cs_set_eq:NN \__regex_replacement_exp_not:V \__regex_peek_replacement_var:N
  \exp_args:No \__regex_replacement:n { \l__regex_replacement_tl }
  \use:e
  {\exp_not:n { \exp_after:wN \l__regex.Peek_true_tl \exp:w }
    \__regex_replacement_do_one_match:n \l__regex_zeroth_submatch_int
    \__regex_query_range:nn
    { \kernel_intarray_item:Nn \g__regex_submatch_end_intarray
      \l__regex_zeroth_submatch_int
    } \l__regex_max_pos_int
    \exp_end:
  }
}{\__regex_peek_reinsert:N \l__regex.Peek_false_tl }

(End of definition for \__regex_peek_replace_end:)
While building the replacement function \_\_regex_replacement_do_one_match:n, we often want to put simple material, given as \#1, whose e-expansion o-expands to a single token. Normally we can just add the token to \l__regex_build_tl, but for \peek__regex_replace_once:nnTF we eventually want to do some strange expansion that is basically using \exp_after:wN to jump through numerous tokens (we cannot use e-expansion like for \regex_replace_once:nnNTF because it is ok for the result to be unbalanced since we insert it in the input stream rather than storing it. When within a csnam we don’t do any such shenanigan because \cs:w ... \cs_end: does all the expansion we need.

\cs_new_protected:Npn \__regex_peek_replacement_put:n #1
\if_case:w \l__regex_replacement_csnames_int
\tl_build_put_right:Nn \l__regex_build_tl
{ \exp_not:N \__regex_reinsert_item:n {#1} }
\else:
\tl_build_put_right:Nn \l__regex_build_tl {#1}
\fi:

(End of definition for \_\_regex_query_set_from_input_tl: and \_\_regex_query_set_item:n.)

\_\_regex_replacement_token:n
When hit with \exp:w, \_\_regex_replacement_token:n \{(token)\} stops \exp_end: and does \exp_after:wN \{token\} \exp:w to continue expansion after it.

\cs_new_protected:Npn \_\_regex_replacement_token:n \#1
{ \exp_after:wN \exp_end: \exp_after:wN #1 \exp:w }

(End of definition for \_\_regex_replacement_token:n.)

\_\_regex_replacement_put_submatch_aux:n
While analyzing the replacement we also have to insert submatches found in the query. Since query items \_\_regex_input_item:n \{(tokens)\} expand correctly only when surrounded by \exp:w ... \exp_end:, and since these expansion controls are not there within csnames (because \cs:w ... \cs_end: make them unnecessary in most cases), we have to put \exp:w and \exp_end: by hand here.

\cs_new_protected:Npn \_\_regex_replacement_put_submatch_aux:n #1
{ \if_case:w \l__regex_replacement_csnames_int
\tl_build_put_right:Nn \l__regex_build_tl
{ \_\_regex_query_submatch:n \{ \_\_regex_int_eval:w #1 + ##1 \scan_stop: \} }
\else:
\tl_build_put_right:Nn \l__regex_build_tl {#1}
\exp:w
\_\_regex_query_submatch:n \{ \_\_regex_int_eval:w #1 + ##1 \scan_stop: \} }
This is used for \u outside csnames. It makes sure to continue expansion with \exp:w before expanding the variable #1 and stopping the \exp:w that precedes.

(End of definition for \_regex_peek_replacement_var:N)

46.8 Messages

Messages for the preparser phase.

Invalid quantifier.

Messages for missing or extra closing brackets and parentheses, with some fancy singular/plural handling for the case of parentheses.
\msg_new:nnnn { regex } { missing-rparen }
{
    Missing-right-
    \int_compare:nTF { #1 = 1 } { parenthesis } { parentheses } -
    inserted-in-regular-expression.
}
{
    LaTeX-was-given-a-regular-expression-with-\int_eval:n {#1} -
    more-left-parentheses-than-right-parentheses.
}
\msg_new:nnnn { regex } { extra-rparen }
{
    Extra-right-parenthesis-ignored-in-regular-expression. }
{
    LaTeX-came-across-a-closing-parenthesis-when-no-submatch-group-
    was-open.-The-parenthesis-will-be-ignored.
}

Some escaped alphanumerics are not allowed everywhere.
\msg_new:nnnn { regex } { bad-escape }
{
    Invalid-escape-'\iow_char:N\#1'--
    \_regex_if_in_cs:TF { within-a-control-sequence. }
    \_regex_if_in_class:TF
    { in-a-character-class. }
    { following-a-category-test. }
}
{
    The-escape-sequence-'\iow_char:N\#1'-may-not-appear-
    \_regex_if_in_cs:TF
    { within-a-control-sequence-test-introduced-by-
      '\iow_char:N\c\iow_char:N\c\{'. }
    { \_regex_if_in_class:TF
      { within-a-character-class- }
      { following-a-category-test-such-as-'\iow_char:N\cL'- } }
    because-it-does-not-match-exactly-one-character.
}

Range errors.
\msg_new:nnnn { regex } { range-missing-end }
{
    Invalid-end-point-for-range-'#1-#2'-in-character-class. }
{
    The-end-point-'#2'-of-the-range-'#1-#2'-may-not-serve-as-an-
    end-point-for-a-range:-alphanumeric-characters-should-not-be-
    escaped,-and-non-alphanumeric-characters-should-be-escaped.
}
\msg_new:nnnn { regex } { range-backwards }
{
    Range-['#1-#2']'-out-of-order-in-character-class. }

In-ranges-of-characters-’[x-y]’-appearing-in-character-classes,-
the-first-character-code-must-not-be-larger-than-the-second.-
Here,-’#1’-has-character-code-\int_eval:n {’#1},-while-
’#2’-has-character-code-\int_eval:n {’#2}.
}

Errors related to \c and \u.
\msg_new:nnnn { regex } { c-bad-mode }
{ Invalid-nested-’\iow_char:N\c’-escape-in-regular-expression. }
{ The-’\iow_char:N\c’-escape-cannot-be-used-within-
a-control-sequence-test-’\iow_char:N\c(...)'-
or-another-category-test.-
To-combine-several-category-tests,-use-’\iow_char:N\c[...]’. }
\msg_new:nnnn { regex } { c-C-invalid }
{ ’\iow_char:N\cC’-should-be-followed-by-’.’-or-’(’,-not-’#1’. }
{ The-’\iow_char:N\cC’-construction-restricts-the-next-item-to-be-a-
control-sequence-or-the-next-group-to-be-made-of-control-sequences.-
It-only-makes-sense-to-follow-it-by-’.’-or-by-a-group. }
\msg_new:nnnn { regex } { cu-lbrace }
{ Left-braces-must-be-escaped-in-’\iow_char:N\#{...}’. }
{ Constructions-such-as-’\iow_char:N\#{...\iow_char:N\{...}’-are-
not-allowed-and-should-be-replaced-by-
’\iow_char:N\#{...\token_to_str:N\{...}’. }
\msg_new:nnnn { regex } { c-lparen-in-class }
{ Catcode-test-cannot-apply-to-group-in-character-class }
{ Construction-such-as-’\iow_char:N\cL(abc)’-are-not-allowed-inside-a-
class-’[...]’-because-classes-do-not-match-multiple-characters-at-once. }
\msg_new:nnnn { regex } { c-missing-rbrace }
{ Missing-right-brace-inserted-for-’\iow_char:N\c’-escape. }
{ LaTeX-was-given-a-regular-expression-where-a-
’\iow_char:N\c\{...-constr-struction-was-not-ended-
with-a-closing-brace-’\iow_char:N\}’. }
\msg_new:nnnn { regex } { c-missing-rbrack }
{ Missing-right-bracket-inserted-for-’\iow_char:N\c’-escape. }
{ A-construction-’\iow_char:N\c[...’-appears-in-a-
regular-expression,-but-the-closing-’]’-is-not-present. }
\msg_new:nnnn { regex } { c-missing-category }
{ Invalid-character-’#1’-following-’\iow_char:N\c’-escape. }
{ In-regular-expressions,-the-’\iow_char:N\c’-escape-sequence-
may-only-be-followed-by-a-left-brace,-a-left-bracket,-or-a-
capital-letter-representing-a-character-category,-namely-
one-of-’ABCDELMOPSTU’. }

578
\msg_new:nnnn { regex } { c-trailing }
  { Trailing-category-code-escape-\iow_char:N\c'... }
  { A-regular-expression-ends-with-\iow_char:N\c'-followed-by-a-letter.-It-will-be-ignored. }
\msg_new:nnnn { regex } { u-missing-lbrace }
  { Missing-left-brace-following-\iow_char:N\u'escape. }
  { The-\iow_char:N\u'-escape-sequence-must-be-followed-by-a-brace-group-with-the-name-of-the-variable-to-use. }
\msg_new:nnnn { regex } { u-missing-rbrace }
  { Missing-right-brace-inserted-for-\iow_char:N\u'-escape. }
  { LaTeX-\str_if_eq:eeTF { } {#2} { reached-the-end-of-the-string- }
  { encountered-an-escaped-alphanumeric-character-\iow_char:N\#2' }
  { when-parsing-the-argument-of-an-\iow_char:N\u\iow_char:N\{...\}'-escape. }
}\msg_new:nnnn { regex } { posix-unsupported }
  { POSIX-collating-element-\#1 - \#1'-not-supported. }
  { The-\[.foo.\]-and-\=[bar]='-syntaxes-have-a-special-meaning-in-POSIX-regular-expressions.-This-is-not-supported-by-LaTeX.-Maybe-you-forget-to-escape-a-left-bracket-in-a-character-class? }
\msg_new:nnnn { regex } { posix-unknown }
  { POSIX-class-\#1'unknown. }
  { \#1'-is-not-among-the-known-POSIX-classes-\[:alnum:\]',-\[:alpha:\]',-\[:ascii:\]',-\[:blank:\]',-\[:cntrl:\]',-\[:digit:\]',-\[:graph:\]',-\[:lower:\]',-\[:print:\]',-\[:punct:\]',-\[:space:\]',-\[:upper:\]',-\[:word:\]',-\[:xdigit:\]]. }
\msg_new:nnnn { regex } { posix-missing-close }
  { Missing-closing-\#1'-for-POSIX-class. }
  { The-POSIX-syntax-\#1'-must-be-followed-by-\',-not-\#2'. }

In various cases, the result of a \l3regex operation can leave us with an unbalanced token list, which we must re-balance by adding begin-group or end-group character tokens.
\msg_new:nnnn { regex } { result-unbalanced }
  { Missing-brace-inserted-when-\#1. }
  { LaTeX-was-asked-to-do-some-regular-expression-operation,-and-the-resulting-token-list-would-not-have-the-same-number-of-begin-group-and-end-group-tokens.-Braces-were-inserted:-\#2-left,-\#3-right. }
Error message for unknown options.
\msg_new:nnnn { regex } { unknown-option }
{ Unknown-option-`\1' for-regular-expressions. }
{ The-only-available-option-is-`case-insensitive',-toggled-by-'
\(?i\)` and -`\(?-i\)`.
}\msg_new:nnnn { regex } { special-group-unknown }
{ Unknown-special-group-`\1...' in-regular-expression. }
{ The-only-valid-constructions-starting-with-`\(?\)'-are-'
\(?::\).\)`-`\(?|-\)\)`-`\(?i\)` and -`\(?-i\)`.
}
Errors in the replacement text.
\msg_new:nnnn { regex } { replacement-c }
{ Misused-`\iow_char:N\c' in-a-replacement-text. }
{ In-a-replacement-text,-the-`\iow_char:N\c' escape-sequence- can-be-followed-by-one-of-the-letters-`ABCDLMOPSTU'- or-a-brace-group,-not-by-`\1'.
}\msg_new:nnnn { regex } { replacement-u }
{ Misused-`\iow_char:N\u' in-a-replacement-text. }
{ In-a-replacement-text,-the-`\iow_char:N\u' escape-sequence- must-be-followed-by-a-brace-group-holding-the-name-of-the- variable-to-use.
}\msg_new:nnnn { regex } { replacement-g }
{ Missing-brace-for-the-`\iow_char:N\g' construction-in-a-replacement-text.
}\msg_new:nnnn { regex } { replacement-catcode-end }
{ Missing-character-for-the-`\iow_char:N\c<category><character>' construction-in-a-replacement-text.
}\msg_new:nnnn { regex } { replacement-catcode-escaped }
{ Escaped-letter-or-digit-after-category-code-in-replacement-text.
In a replacement text, the `\bow_trans:nnn{\c}\c\{...\}` escape sequence can be followed by one of the letters ‘ABCDELMOPSTU’ representing the character category. Then, a character must follow, not `\bow_trans:nnn{\c}\c\#2`.

\msg_new:nnnn { regex } { replacement-catcode-in-cs }
{ Category-code `\bow_trans:nnn{\c}\c\#1#3`-ignored-inside `\bow_trans:nnn{\c}\c\{...\}`-in-a-replacement-text. }

\msg_new:nnnn { regex } { replacement-null-space }
{ TeX cannot build a space token with character code 0. }
{ You asked for a character token with category space, and character code 0, for instance through `\bow_trans:nnn{\c}\c\{\char\x00}`. This specific case is impossible and will be replaced by a normal space. }

\msg_new:nnnn { regex } { replacement-missing-rbrace }
{ There `\int_compare:nTF { #1 = 1 } { was } { were } - #1- missing-right-`\int_compare:nTF { #1 = 1 } { brace } { braces }. }

\msg_new:nnnn { regex } { replacement-missing-rparen }
{ There `\int_compare:nTF { #1 = 1 } { was } { were } - #1- missing-right- `\int_compare:nTF { #1 = 1 } { parenthesis } { parentheses }. }

\msg_new:nnnn { regex } { submatch-too-big }
{ Submatch #1 used but regex only has #2 group(s) }

Some escaped alphanumerics are not allowed everywhere.

\msg_new:nnnn { regex } { backwards-quantifier }
{ Quantifier "#1,#2" is backwards. }
{ The values given in a quantifier must be in order. }

Used in user commands, and when showing a regex.

\msg_new:nnnn { regex } { case-odd }
{ #1 with odd number of items }
{ There must be a #2 part for each regex: - found odd number of items (#3) in `\bow_trans:nnn{\c}\c\#4`}
\msg_new:nnn { regex } { show }
{ \-Compiled-regex-
  \tl_if_empty:nTF (#1) { variable- #2 } { (#1) } :
  #3
}
\prop_gput:Nnn \g_msg_module_name_prop { regex } { LaTeX }
\prop_gput:Nnn \g_msg_module_type_prop { regex } { }
\__regex_msg_repeated:nnN
This is not technically a message, but seems related enough to go there. The arguments are: #1 is the minimum number of repetitions; #2 is the number of allowed extra repetitions (−1 for infinite number), and #3 tells us about laziness.
\cs_new:Npn \__regex_msg_repeated:nnN #1#2#3
{ \str_if_eq:eeF { #1 #2 } { 1 0 }
  { , ~ repeated ~
    \int_case:nnF {#2}
    { -1 } { #1-or-more-times,-\bool_if:NTF #3 { lazy } { greedy } }
    { 0 } { #1-times }
  }
  { between-#1-and-\int_eval:n {#1+#2}-times,-
    \bool_if:NTF #3 { lazy } { greedy }
  }
}
(End of definition for \__regex_msg_repeated:nnN.)

\__regex_trace_push:nnN \__regex_trace_pop:nnN \__regex_trace:nne
Here #1 is the module name (regex) and #2 is typically 1. If the module’s current tracing level is less than #2 show nothing, otherwise write #3 to the terminal.
\cs_new_protected:Npn \__regex_trace_push:nnN #1#2#3
{ \__regex_trace:nne {#1} {#2} { entering~ \token_to_str:N #3 } }
\cs_new_protected:Npn \__regex_trace_pop:nnN #1#2#3
{ \__regex_trace:nne {#1} {#2} { leaving~ \token_to_str:N #3 } }
\cs_new_protected:Npn \__regex_trace:nne #1#2#3
{ \int_compare:nNnF { \int_use:c { g__regex_trace_#1_int } } < {#2}
  { \iow_term:e { Trace:-#3 } }
}
(End of definition for \__regex_trace_push:nnN, \__regex_trace_pop:nnN, and \__regex_trace:nne.)
\g__regex_trace_regex_int
No tracing when that is zero.
\int_new:N \g__regex_trace_regex_int

46.9 Code for tracing

There is a more extensive implementation of tracing in the l3trial package \texttt{l3trace}. Function names are a bit different but could be merged.
This function lists the contents of all states of the NFA, stored in \toks from 0 to \_\_\_\- \regex_max_state_int (excluded).

\cs_new_protected:Npn \_\_\_\_\_\regex_trace_states:n #1
\int_step_inline:nnn \_\_\_\_\_\regex_min_state_int { \_\_\_\_\_\regex_max_state_int - \c_one_int }
{ \_\_\_\_\_\regex_trace:nne { regex } { #1 } }
\_\_\_\_\_\regex_trace:nne { \toks ##1 = { \_\_\_\_\_\regex_toks_use:w ##1 } }
\_\_\_\_\_\regex_trace_states:n
Chapter 47

\texttt{l3prg} implementation

The following test files are used for this code: \texttt{m3prg001.lvt, m3prg002.lvt, m3prg003.lvt}.

\begin{verbatim}
\texttt{package}
n
\section{Primitive conditionals}

Those two primitive \TeX conditionals are synonyms. \texttt{if\_bool:N} is defined in \texttt{l3basics}, as it's needed earlier to define quark test functions.

\begin{verbatim}
\cs_new_eq:NN \if\_predicate:w \tex_ifodd:D

(End of definition for \texttt{if\_bool:N} and \texttt{if\_predicate:w}. These functions are documented on page 72.)
\end{verbatim}

\section{Defining a set of conditional functions}

These are all defined in \texttt{l3basics}, as they are needed “early”. This is just a reminder!

(End of definition for \texttt{prg\_set\_conditional:Nppn} and others. These functions are documented on page 63.)

\section{The boolean data type}

Boolean variables have to be initiated when they are created. Other than that there is not much to say here.

\begin{verbatim}
\cs_new_protected:Npn \bool\_const:Nn \tl\_const:Nn \bool\_set:Nn \bool\_clear:Nn \bool\_clear\_all

(End of definition for \texttt{bool\_new:N}. This function is documented on page 66.)
\end{verbatim}

\begin{verbatim}
\cs_new_protected:Npm \bool\_new:N \bool\_new:c \bool\_new\_c

A merger between \texttt{tl\_const:Nn} and \texttt{bool\_set:Nn}.
\end{verbatim}
Setting is already pretty easy. When check-declarations is active, the definitions are patched to make sure the boolean exists. This is needed because booleans are not based on token lists nor on \TeX registers.

\begin{verbatim}
\cs_new_protected:Npn \bool_set_true:N #1 { \cs_set_eq:NN #1 \c_true_bool }
\cs_new_protected:Npn \bool_set_false:N #1 { \cs_set_eq:NN #1 \c_false_bool }
\cs_new_protected:Npn \bool_gset_true:N #1 { \cs_gset_eq:NN #1 \c_true_bool }
\cs_new_protected:Npn \bool_gset_false:N #1 { \cs_gset_eq:NN #1 \c_false_bool }
\cs_generate_variant:Nn \bool_set_true:N { c }
\cs_generate_variant:Nn \bool_set_false:N { c }
\cs_generate_variant:Nn \bool_gset_true:N { c }
\cs_generate_variant:Nn \bool_gset_false:N { c }
\end{verbatim}

The usual copy code. While it would be cleaner semantically to copy the \cs_set_eq:NN family of functions, we copy \tl_set_eq:NN because that has the correct checking code.

\begin{verbatim}
\cs_new_eq:NN \bool_set_eq:NN \tl_set_eq:NN
\cs_new_eq:NN \bool_gset_eq:NN \tl_gset_eq:NN
\cs_generate_variant:Nn \bool_set_eq:NN { Nc, cN, cc }
\cs_generate_variant:Nn \bool_gset_eq:NN { Nc, cN, cc }
\end{verbatim}

This function evaluates a boolean expression and assigns the first argument the meaning \c_true_bool or \c_false_bool. Again, we include some checking code. It is important to evaluate the expression before applying the \chardef primitive, because that primitive sets the left-hand side to \scan_stop: before looking for the right-hand side.

\begin{verbatim}
\exp_last_unbraced:NNNf \tex_chardef:D #1 = { \bool_if_p:n {#2} }
\exp_last_unbraced:NNNNf \tex_global:D \tex_chardef:D #1 = { \bool_if_p:n {#2} }
\cs_generate_variant:Nn \bool_set:Nn { c }
\cs_generate_variant:Nn \bool_gset:Nn { c }
\end{verbatim}

Set to false or true locally or globally.

\begin{verbatim}
\exp_last_unbraced:NNNf \bool_if:Nf {#1} { \bool_set_inverse:N #1 }
\exp_last_unbraced:NNNf \bool_if:Nf {#1} { \bool_gset_inverse:N #1 }
\exp_last_unbraced:NNNf \bool_if:Nf {#1} { \bool_gset_inverse:N #1 }
\end{verbatim}
47.4 Internal auxiliaries

\q__bool_recursion_tail\n\q__bool_recursion_stop

Internal recursion quarks.

\quark_new:N \q__bool_recursion_tail
\quark_new:N \q__bool_recursion_stop

Functions to gobble up to a quark.

\__kernel_quark_new_test:N \__bool_if_recursion_tail_stop_do:nn

Functions to query recursion quarks.

\bool_if_p:N \bool_if_p:c \bool_if:N \bool_if:cTF
Straight forward here. We could optimize here if we wanted to as the boolean can just
be input directly.

\prg_new_conditional:Npnn \bool_if:N #1 { p , T , F , TF }
\prg_generate_conditional_variant:Nnn \bool_if:N { c } { p , T , F , TF }

\bool_to_str:N \bool_to_str:c \bool_to_str:n
Expands to string literal true or false.

\cs_new:Npe \bool_to_str:N #1
\exp_not:N \bool_if:NTF #1 \tl_to_str:n { true } \tl_to_str:n { false }
\cs_generate_variant:Nn \bool_to_str:N { c }
\cs_new:Npe \bool_to_str:n #1
\exp_not:N \bool_if:nTF { #1 } \tl_to_str:n { true } \tl_to_str:n { false }

(End of definition for \bool_set_inverse:N and \bool_gset_inverse:N. These functions are documented on page 66.)
\bool_show:n \bool_log:n
\bool_show:N \bool_log:N \bool_show:c \bool_log:c
\__bool_show:NN

Show the truth value of the boolean.
\cs_new_protected:Npn \bool_show:n \bool_log:n
\cs_new_protected:Npn \bool_to_str:n
\__kernel_msg_show_eval:Nn \bool_to_str:n
\__kernel_msg_log_eval:Nn \bool_to_str:n

(End of definition for \bool_show:n and \bool_log:n. These functions are documented on page 67.)

\bool_show:N \bool_log:N \bool_show:c \bool_log:c \__bool_show:NN

Show the truth value of the boolean, as true or false.
\cs_new_protected:Npn \bool_show:N \cs_generate_variant:Nn \bool_show:N { c }
\cs_new_protected:Npn \bool_log:N \cs_generate_variant:Nn \bool_log:N { c }
\cs_new_protected:Npn \__bool_show:NN #1#2
\__kernel_chk_defined:NT #2
\token_case_meaning:NnF #2
\c_true_bool { \exp_args:Ne #1 { \token_to_str:N #2 = true } }
\c_false_bool { \exp_args:Ne #1 { \token_to_str:N #2 = false } }
\msg_error:nneee { kernel } { bad-type }
\token_to_str:N #2 \token_to_meaning:N #2 \bool

(End of definition for \bool_show:N, \bool_log:N, and \__bool_show:NN. These functions are documented on page 67.)

\l_tmpa_bool \l_tmpb_bool \g_tmpa_bool \g_tmpb_bool

A few booleans just if you need them.
\bool_new:N \l_tmpa_bool \bool_new:N \l_tmpb_bool
\bool_new:N \g_tmpa_bool \bool_new:N \g_tmpb_bool

(End of definition for \l_tmpa_bool and others. These variables are documented on page 67.)

\bool_if_exist_p:N \bool_if_exist_p:c \bool_if_exist:NTF \bool_if_exist:cTF
\prg_new_eq_conditional:NNn \bool_if_exist:N \cs_if_exist:N
\prg_new_eq_conditional:NNn \bool_if_exist:c \cs_if_exist:c

(End of definition for \bool_if_exist:NTF. This function is documented on page 67.)

587
47.5 Boolean expressions

\begin{verbatim}
\bool_if:n \bool_if:nTF \__bool_if_p:n \__bool_if_p_aux:w
\end{verbatim}

Evaluating the truth value of a list of predicates is done using an input syntax somewhat similar to the one found in other programming languages with ( and ) for grouping, \texttt{!} for logical “Not”, \texttt{&} for logical “And” and \texttt{||} for logical “Or”. However, they perform eager evaluation. We shall use the terms \texttt{Not}, \texttt{And}, \texttt{Or}, \texttt{Open} and \texttt{Close} for these operations.

Any expression is terminated by a \texttt{Close} operation. Evaluation happens from left to right in the following manner using a \texttt{GetNext} function:

- If an \texttt{Open} is seen, start evaluating a new expression using the \texttt{Eval} function and call \texttt{GetNext} again.
- If a \texttt{Not} is seen, remove the \texttt{!} and call a \texttt{GetNext} function with the logic reversed.
- If none of the above, reinsert the token found (this is supposed to be a predicate function) in front of an \texttt{Eval} function, which evaluates it to the boolean value \texttt{⟨true⟩} or \texttt{⟨false⟩}.

The \texttt{Eval} function then contains a post-processing operation which grabs the instruction following the predicate. This is either \texttt{And}, \texttt{Or} or \texttt{Close}. In each case the truth value is used to determine where to go next. The following situations can arise:

\texttt{⟨true⟩And} Current truth value is true, logical And seen, continue with \texttt{GetNext} to examine truth value of next boolean (sub-)expression.

\texttt{⟨false⟩And} Current truth value is false, logical And seen, stop using the values of predicates within this sub-expression until the next \texttt{Close}. Then return \texttt{⟨false⟩}.

\texttt{⟨true⟩Or} Current truth value is true, logical Or seen, stop using the values of predicates within this sub-expression until the nearest \texttt{Close}. Then return \texttt{⟨true⟩}.

\texttt{⟨false⟩Or} Current truth value is false, logical Or seen, continue with \texttt{GetNext} to examine truth value of next boolean (sub-)expression.

\texttt{⟨true⟩Close} Current truth value is true, \texttt{Close} seen, return \texttt{⟨true⟩}.

\texttt{⟨false⟩Close} Current truth value is false, \texttt{Close} seen, return \texttt{⟨false⟩}.

\begin{verbatim}
\prg_new_conditional:Npnn \bool_if:n #1 { T , F , TF } {
\if_predicate:w \bool_if_p:n {#1} \prg_return_true: \else: \prg_return_false: \fi: }
\end{verbatim}

\texttt{\prg_new_conditional:Npnn \bool_if:nTF #1 { T , F , TF } {}}

\texttt{} (End of definition for \texttt{\bool_if:nTF}. This function is documented on page 69.)

\texttt{\bool_if_p:n \__bool_if_p:n \__bool_if_p_aux:w} To speed up the case of a single predicate, \texttt{f}-expand and check whether the result is one token (possibly surrounded by spaces), which must be \texttt{\c_true_bool} or \texttt{\c_false_bool}. We use a version of \texttt{\tl_if_single:nTF} optimized for speed since we know that an empty \#1 is an error. The auxiliary \texttt{\__bool_if_p_aux:w} removes the trailing parenthesis and gets rid of any space, then returns \texttt{\c_true_bool} or \texttt{\c_false_bool} as appropriate. This extra work around is because in a \texttt{\bool_set:Nn}, the underlying \texttt{\chardef} turns
the bool being set temporarily equal to \relax, thus assigning a boolean to itself would fail (gh/1055). For the general case, first issue a \group_align_safe_begin: as we are using && as syntax shorthand for the And operation and we need to hide it for \TeX. This group is closed after \__bool_get_next:NN returns \c_true_bool or \c_false_bool. That function requires the trailing parenthesis to know where the expression ends.

\cs_new:Npn \bool_if_p:n { \exp_args:Nf \__bool_if_p:n }
\cs_new:Npn \__bool_if_p:n #1
\tl_if_empty:oT { \use_none:nn #1 . } { \__bool_if_p_aux:w }
\group_align_safe_begin:
\exp_after:wN \group_align_safe_end:
\exp:w \exp_end_continue_f:w % ( \__bool_get_next:NN \use_i:nnnn #1 )
\cs_new:Npn \__bool_if_p_aux:w #1 \use_i:nnnn #2#3
\bool_if:NTF #2 \c_true_bool \c_false_bool (End of definition for \bool_if_p:n, \__bool_if_p:n, and \__bool_if_p_aux:w. This function is documented on page 69.)

\__bool_get_next:NN The GetNext operation. Its first argument is \use_i:nnnn, \use_ii:nnnn, \use_iii:nnnn, or \use_iv:nnnn (we call these “states”). In the first state, this function eventually expand to the truth value \c_true_bool or \c_false_bool of the expression which follows until the next unmatched closing parenthesis. For instance “\__bool_get_next:NN \use_i:nnnn \c_true_bool && \c_true_bool )” (including the closing parenthesis) expands to \c_true_bool. In the second state (after a !) the logic is reversed. We call these two states “normal” and the next two “skipping”. In the third state (after \c_true_bool||) it always returns \c_true_bool. In the fourth state (after \c_false_bool&&) it always returns \c_false_bool and also stops when encountering ||, not only parentheses. This code itself is a switch: if what follows is neither ! nor [, we assume it is a predicate.

\cs_new:Npn \__bool_get_next:NN #1#2
\use:c \__bool_!Nw \if_meaning:w !#2 \else: \if_meaning:w #2 ( \else: p \fi: \fi:
\use_i:nnnn \use_ii:nnnn \use_iii:nnnn \use_iv:nnnn #1 #2
\)

(End of definition for \__bool_get_next:NN.)

\__bool_!Nw The Not operation reverses the logic: it discards the ! token and calls the GetNext operation with the appropriate first argument. Namely the first and second states are interchanged, but after \c_true_bool|| or \c_false_bool&& the ! is ignored.

\cs_new:cpn { \__bool_!Nw } #1#2
\exp_after:wN \__bool_get_next:NN \use_i:nnnn \use_ii:nnnn \use_iii:nnnn \use_iv:nnnn #1 #2

589
\__bool_(:Nw
The Open operation starts a sub-expression after discarding the open parenthesis. This
is done by calling GetNext (which eventually discards the corresponding closing paren-
thesis), with a post-processing step which looks for And, Or or Close after the group.
\cs_new:cpn { __bool_(:Nw } #1#2
{ \exp_after:wN \__bool_choose:NNN \exp_after:wN #1
\int_value:w \__bool_get_next:NN \use_i:nnnn
}
(End of definition for \__bool_(:Nw.)
\__bool_p:Nw
If what follows GetNext is neither \texttt{!} nor \texttt{()}, evaluate the predicate using the primitive
\int_value:w. The canonical \texttt{true} and \texttt{false} values have numerical values 1 and 0
respectively. Look for And, Or or Close afterwards.
\cs_new:cpn { __bool_p:Nw } #1
{ \exp_after:wN \__bool_choose:NNN \exp_after:wN #1 \int_value:w }
(End of definition for \__bool_p:Nw.)
\__bool_choose:NNN
The arguments are #1: a function such as \texttt{\use_i:nnnn}, #2: 0 or 1 encoding the current
truth value, #3: the next operation, And, Or or Close. We distinguish three cases
\__bool_():0:
according to a combination of #1 and #2. Case 2 is when #1 is \texttt{\use_iii:nnnn} (state 3),
\__bool_():1:
namely after \texttt{c\_true\_bool ||}. Case 1 is when #1 is \texttt{\use_i:nnnn} and #2 is \texttt{true}
or
\__bool_&:0:
when #1 is \texttt{\use_i:nnnn} and #2 is \texttt{false}, for instance for \texttt{!c\_false\_bool}. Case 0
\__bool_&:1:
includes the same with \texttt{true/false} interchanged and the case where #1 is \texttt{\use_iv:nnnn}
\__bool_&:2:
namely after \texttt{c\_false\_bool &&}. In case 1, namely when the argument is \texttt{true}
and we are in a normal state continue in
\__bool_():0:
the normal state 1. In case 2, namely when skipping alternatives in an Or, continue in
\__bool_():1:
same state. When seeing \texttt{&} in case 0 go into state 4, equivalent to having seen \texttt{c\_false\_bool &&}.
\__bool_():2:
In case 1, namely when the argument is \texttt{true} and we are in a normal state continue in
the normal state 1. In case 2, namely when skipping alternatives in an Or, continue in
the same state. When seeing \texttt{\|} in case 0, continue in a normal state; in particular stop
\__bool_():0:
skipping for \texttt{c\_false\_bool &&} because that binds more tightly than \texttt{||}. In the other
two cases start skipping for \texttt{c\_true\_bool ||}.
\cs_new:Npn \__bool_choose:NNN #1#2#3
{ \use:c
  \__bool_\token_to_str:N #3 _
  #1 #2 { \if_meaning:w 0 #2 1 \else: 0 \fi: } 2 0 : 
}
\cs_new:cpn { __bool_():0: } { \c_false_bool }
\cs_new:cpn { __bool_():1: } { \c_true_bool }
\cs_new:cpn { __bool_():2: } { \c_true_bool }
\cs_new:cpn { __bool_&:0: } \& \{ \__bool_get_next:NN \use_iv:nnnn
\cs_new:cpn { __bool_&:1: } \& \{ \__bool_get_next:NN \use_i:nnnn
\cs_new:cpn { __bool_&:2: } \& \{ \__bool_get_next:NN \use_iii:nnnn
\cs_new:cpn { __bool_():0: } \& \{ \__bool_get_next:NN \use_i:nnnn
\cs_new:cpn { __bool_():1: } \& \{ \__bool_get_next:NN \use_iii:nnnn
\cs_new:cpn { __bool_():2: } \& \{ \__bool_get_next:NN \use_iii:nnnn

590
Go through the list of expressions, stopping whenever an expression is false. If the end is reached without finding any false expression, then the result is true.

\cs_new:Npn \bool_lazy_all_p:n #1
\prg_new_conditional:Npnn \bool_lazy_all:n #1 { T , F , TF }
\if_predicate:w \bool_lazy_all_p:n {#1}
\prg_return_true:
\else:
\prg_return_false:
\fi:
\cs_new:Npn \__bool_lazy_all:n #1
\__bool_if_recursion_tail_stop_do:nn {#1} { \c_true_bool }
\bool_if:nF {#1}{ \__bool_use_i_delimit_by_q_recursion_stop:nw { \c_false_bool } }
\__bool_lazy_all:n
\end_of_definition_for \bool_lazy_all:nTF and \__bool_lazy_all:n. This function is documented on page 69.

Only evaluate the second expression if the first is true. Note that #2 must be removed as an argument, not just by skipping to the \else: branch of the conditional since #2 may contain unbalanced \TeX conditionals.

\prg_new_conditional:Npnn \bool_lazy_and:nn #1#2 { p , T , F , TF }
\if_predicate:w \bool_lazy_and:nn #1\{ \bool_if_p:n {#2} \} { \c_false_bool }
\prg_return_true:
\else:
\prg_return_false:
\fi:
\end_of_definition_for \bool_lazy_and:nnTF. This function is documented on page 69.

Go through the list of expressions, stopping whenever an expression is true. If the end is reached without finding any true expression, then the result is false.

\cs_new:Npn \bool_lazy_any_p:n #1
\prg_new_conditional:Npnn \bool_lazy_any:n #1 { T , F , TF }
\if_predicate:w \bool_lazy_any_p:n {#1}
\prg_return_true:
\else:
\prg_return_false:
\fi:
\cs_new:Npn \__bool_lazy_any:n #1
\end_of_definition_for \bool_lazy_any:nTF and \__bool_lazy_any:n. This function is documented on page 69.
\bool_lazy_or_p:nn\bool_lazy_or:nnTF
\bool_not_p:n\bool_xor_p:nn\bool_xor:nnTF

\bool_while_do:Nn\bool_until_do:Nn

\bool_lazy_or_p:nn
\bool_lazy_or:nnTF

Only evaluate the second expression if the first is false.

\bool_not_p:n
The Not variant just reverses the outcome of \bool_if_p:n. Can be optimized but this is nice and simple and according to the implementation plan. Not even particularly useful to have it when the infix notation is easier to use.

\bool_xor_p:nn\bool_xor:nnTF
Exclusive or. If the boolean expressions have same truth value, return false, otherwise return true.

47.6 Logical loops

\bool_while_do:Nn\bool_until_do:Nn
A while loop where the boolean is tested before executing the statement. The “while” version executes the code as long as the boolean is true; the “until” version executes the code as long as the boolean is false.
A do-while loop where the body is performed at least once and the boolean is tested after executing the body. Otherwise identical to the above functions.

\begin{verbatim}
\cs_new:Npn \bool_do_while:Nn #1#2
{ #2 \bool_if:NT #1 { \bool_do_while:Nn #1 {#2} } }
\cs_new:Npn \bool_do_until:Nn #1#2
{ #2 \bool_if:NF #1 { \bool_do_until:Nn #1 {#2} } }
\cs_generate_variant:Nn \bool_do_while:Nn { c }
\cs_generate_variant:Nn \bool_do_until:Nn { c }
\end{verbatim}

Loop functions with the test either before or after the first body expansion.

\begin{verbatim}
\cs_new:Npn \bool_while_do:nn #1#2
{ \bool_if:nT {#1} { #2 \bool_while_do:nn {#1} {#2} } }
\cs_new:Npn \bool_do_while:nn #1#2
{ #2 \bool_if:nT {#1} { \bool_do_while:nn {#1} {#2} } }
\cs_new:Npn \bool_until_do:nn #1#2
{ \bool_if:nF {#1} { #2 \bool_until_do:nn {#1} {#2} } }
\cs_new:Npn \bool_do_until:nn #1#2
{ #2 \bool_if:nF {#1} { \bool_do_until:nn {#1} {#2} } }
\end{verbatim}

For boolean cases the overall idea is the same as for \texttt{\str_case:nnTF} as described in \texttt{l3str}.

\begin{verbatim}
\cs_new:Npn \bool_case:nTF
{ \exp:w \__bool_case:nTF }
\cs_new:Npn \bool_case:n
{ \__bool_case:NnTF }
\__bool_case_end:nw
\end{verbatim}
47.7 Producing multiple copies

This function uses a cascading csname technique by David Kastrup (who else :-)

The idea is to make the input 25 result in first adding five, and then 20 copies of the code to be replicated. The technique uses cascading csnames which means that we start building several csnames so we end up with a list of functions to be called in reverse order. This is important here (and other places) because it means that we can for instance make the function that inserts five copies of something to hand down to the next function in line. This is exactly what happens here: in the example with 25 then the next function is the one that inserts two copies but it sees the ten copies handed down by the previous function. In order to avoid the last function to insert say, 100 copies of the original argument just to gobble them again we define separate functions to be inserted first. These functions also close the expansion of \exp:w, which ensures that \prg_replicate:nn only requires two steps of expansion.

This function has one flaw though: Since it constantly passes down ten copies of its previous argument it severely affects the main memory once you start demanding hundreds of thousands of copies. Now I don’t think this is a real limitation for any ordinary use, and if necessary, it is possible to write \prg_replicate:nn \{1000\} \{\prg_replicate:nn \{1000\} \{\langle\text{code}\rangle\}\}. An alternative approach is to create a string of m’s with \exp:w which can be done with just four macros but that method has its own problems since it can exhaust the string pool. Also, it is considerably slower than what we use here so the few extra csnames are well spent I would say.
Then comes all the functions that do the hard work of inserting all the copies. The first function takes :n as a parameter.

Users shouldn’t ask for something to be replicated once or even not at all but...

(End of definition for \prg_replicate:nn and others. This function is documented on page 71.)

### 47.8 Detecting \TeX’s mode

\mode_if_vertical_p: For testing vertical mode. Strikes me here on the bus with David, that as long as we are just talking about returning true and false states, we can just use the primitive
conditionals for this and gobbling the $\exp_end$: in the input stream. However this requires knowledge of the implementation so we keep things nice and clean and use the return statements.

\begin{verbatim}
\prg_new_conditional:Npnn \mode_if_vertical: { p , T , F , TF } 
{ \if_mode_vertical: \prg_return_true: \else: \prg_return_false: \fi: }
\end{verbatim}

(End of definition for $\mode_if_vertical$:TF. This function is documented on page 72.)

\begin{verbatim}
\prg_new_conditional:Npnn \mode_if_horizontal: { p , T , F , TF } 
{ \if_mode_horizontal: \prg_return_true: \else: \prg_return_false: \fi: }
\end{verbatim}

(End of definition for $\mode_if_horizontal$:TF. This function is documented on page 71.)

\begin{verbatim}
\prg_new_conditional:Npnn \mode_if_inner: { p , T , F , TF } 
{ \if_mode_inner: \prg_return_true: \else: \prg_return_false: \fi: }
\end{verbatim}

(End of definition for $\mode_if_inner$:TF. This function is documented on page 72.)

\begin{verbatim}
\prg_new_conditional:Npnn \mode_if_math: { p , T , F , TF } 
{ \if_mode_math: \prg_return_true: \else: \prg_return_false: \fi: }
\end{verbatim}

(End of definition for $\mode_if_math$:TF. This function is documented on page 72.)

47.9 Internal programming functions

TEX’s alignment structures present many problems. As Knuth says himself in *TEX: The Program*: “It’s sort of a miracle whenever $\halign$ or $\valign$ work, [...]” One problem relates to commands that internally issue a $\cr$ but also peek ahead for the next character for use in, say, an optional argument. If the next token happens to be a $&$ with category code 4 we get some sort of weird error message because the underlying $\futurelet$ stores the token at the end of the alignment template. This could be a $&_4$ giving a message like ! Misplaced $\cr$. or even worse: it could be the $\endtemplate$ token causing even more trouble! To solve this we have to open a special group so that TEX still thinks it’s on safe ground but at the same time we don’t want to introduce any brace group that may find its way to the output. The following functions help with this by using behaviour documented only in Appendix D of *The TEXbook*... In short evaluating ‘{ and ‘} as numbers will not change the counter TEX uses to keep track of its state in an alignment, whereas gobbling a brace using $\if_false$: will affect TEX’s state without producing any real group. We place the $\if_false$: \fi: part at that place so that the successive expansions of $\group_align_safe_begin/end$: are always brace balanced.

\begin{verbatim}
\tex_catcode:D '\''' = 2 \exp_stop_f:
\cs_new:Npn \group_align_safe_begin:
{ \exp:w \if_false: \{ \fi: \exp_stop_f: }
\tex_catcode:D '\''' = 1 \exp_stop_f:
\cs_new:Npn \group_align_safe_end:
{ \exp:w \if_false: \} \fi: \exp_stop_f: }
\end{verbatim}

596
\g__kernel_prg_map_int

A nesting counter for mapping.

\int_new:N \_kernel_prg_map_int

(End of definition for \g__kernel_prg_map_int.)

\prg_break_point:Nn \prg_map_break:Nn

These are defined in l3basics, as they are needed “early”. This is just a reminder that is the case!

(End of definition for \prg_break_point:Nn and \prg_map_break:Nn. These functions are documented on page 72.)

\prg_break_point: \prg_break: \prg_break:n

Also done in l3basics.

(End of definition for \prg_break_point:, \prg_break:, and \prg_break:n. These functions are documented on page 73.)

\endinput
Chapter 48

\textbf{\texttt{l3sys} implementation}

\section{Kernel code}

\subsection{Detecting the engine}

Set the \texttt{T}, \texttt{F}, \texttt{TF}, \texttt{p} forms of \texttt{#1} to be constants equal to the result of evaluating the boolean expression \texttt{#2}.

\begin{verbatim}
\cs_new_protected:Npn \_\_sys_const:nn #1#2
{
\bool_if:nTF {#2}
{\cs_new_eq:cN { #1 :T } \use:n \cs_new_eq:cN { #1 :F } \use_none:n \cs_new_eq:cN { #1 :TF } \use_i:nn \cs_new_eq:cN { #1 _p: } \c_true_bool}
{\cs_new_eq:cN { #1 :T } \use_none:n \cs_new_eq:cN { #1 :F } \use:n \cs_new_eq:cN { #1 :TF } \use_ii:nn \cs_new_eq:cN { #1 _p: } \c_false_bool}
}
\end{verbatim}

(End of definition for \_\_sys_const:nn.)

Set up the engine tests on the basis exactly one test should be true. Mainly a case of looking for the appropriate marker primitive.

\begin{verbatim}
\sys_if_engine_luatex_p:\sys_if_engine_luatex:TF \sys_if_engine_pdftex_p:\sys_if_engine_pdftex:TF \sys_if_engine_ptex_p:\sys_if_engine_ptex:TF \sys_if_engine_uptex_p:\sys_if_engine_uptex:TF \sys_if_engine_xetex_p:\sys_if_engine_xetex:TF \c_sys_engine_str
\end{verbatim}

598
{ \cs_if_exist:NT \tex_luatexversion:D \ { \luatex } \\
{ \cs_if_exist:NT \tex_pdfxversion:D \ { \pdfx } } \\
{ \cs_if_exist:NT \tex_kanjiskip:D } \\
{ { \cs_if_exist:NTF \tex_enablecjktoken:D \ { \upx } \ { \ptx } } } \\
{ \cs_if_exist:NT \tex_XeTeXversion:D \ { \xetex } } \\
\tl_map_inline:nn { { \luatex } { \pdfx } { \ptx } { \upx } } \{ \uptx \ \{ \xetex \} } \\
\{ \str_if_eq_p:Vn \ c_sys_engine_str \ { \#1 \} \} \\
\} (End of definition for \sys_if_engine_luatex:TF and others. These functions are documented on page 75.)

c_sys_engine_exec_str
c_sys_engine_format_str

Take the functions defined above, and set up the engine and format names. \c_sys_engine_exec_str differs from \c_sys_engine_str as it is the actual engine name, not a “filtered” version. It differs for \ptx and \upx, which have a leading e, and for \luatex, because \LaTeX uses the LuaHBTeX engine.

\c_sys_engine_format_str is quite similar to \c_sys_engine_str, except that it differentiates \pdflatex from \latex (which is pdfTEX in DVI mode). This differentiation, however, is reliable only if the user doesn’t change \tex_pdfoutput:D before loading this code.

\group_begin:
\cs_set_eq:NN \lua_now:e \tex_directlua:D
\str_const:Ne \c_sys_engine_exec_str
{ \sys_if_engine_pdfx:T \ { \pdf } } \sys_if_engine_xetex:T \ { \xe } \sys_if_engine_ptx:T \ { \ep } \sys_if_engine_uptex:T \ { \eup } \sys_if_engine_luatex:T
{ \lua \lua_now:e } \{ if \ (pcall(require, \ 'luaharfbuzz')) then - \tex.print("hb") - \end \}
{ \str_if_eq_p:Vn \ c_sys_engine_str \ { \#1 } \} \\
{ \group_end:
\str_const:Ne \c_sys_engine_format_str
{ \cs_if_exist:NTF \fmtname } \\
{ \bool_lazy_or:nTF \ \{ \str_if_eq_p:Vn \ fmtname \ { \plain } \} }
599
{ \str_if_eq_p:Vn \fmtname { LaTeX2e } }
{
{ \sys_if_engine_pdftex:T
  { \int_compare:nNnT { \tex_pdfoutput:D } = { 1 } { pdf } }
  \sys_if_engine_xetex:T { xe }
  \sys_if_engine_ptex:T { p }
  \sys_if_engine_uptex:T { up }
}
{ \sys_if_engine_luatex:T
  \int_compare:nNnT { \tex_pdfoutput:D } = { 0 } { dvi }
  lua
}
{ \str_if_eq:VnTF \fmtname { LaTeX2e } }
{ latex }
{ bool_lazy_and:nnT
  { \sys_if_engine_exec_str: }
  { \int_compare_p:nN { \tex_pdfoutput:D } = { 0 } }
  e
}
}
{ unknown }

(End of definition for \c_sys_engine_exec_str and \c_sys_engine_format_str. These variables are documented on page 75.)

\c_sys_engine_version_str
Various different engines, various different ways to extract the data!
\str_const:Ne \c_sys_engine_version_str
\str_case:on \c_sys_engine_str
{ pdftex
  \int_div_truncate:nn { \tex_pdfxversion:D } { 100 }
  .
  \int_mod:nn { \tex_pdfxversion:D } { 100 }
  .
  \tex_pdfxrevision:D
}
{ ptex
  \cs_if_exist:NT \tex_ptxversion:D
  \int_use:N \tex_ptxversion:D
  .
  \int_use:N \tex_ptxminorversion:D
  \tex_ptxrevision:D
  -
  \int_use:N \tex_epTeXversion:D

600
48.1.2 Platform

Setting these up requires the file module (file lookup), so is actually implemented there.

(End of definition for `$\sys_if_platform_unix_p$, `$\sys_if_platform_windows_p$`, and `$\c_sys_platform_str$`. These functions are documented on page 76.)

48.1.3 Configurations

Loading the backend code is pretty simply: check that the backend is valid, then load it up.

```latex
\cs_new_protected:Npn \sys_load_backend:n #1
\__sys_load_backend_check:N
\c_sys_backend_str
\c_sys_engine_version_str

(End of definition for `$\c_sys_engine_version_str$`. This variable is documented on page 75.)
\msg_error:nn \sys \backend-set
}
{
\tl_if_blank:nF \#1
{
\tl_gset:Nn \g__sys_backend_tl \#1
\__sys_load_backend_check:N \g__sys_backend_tl
\str_const:N \c_sys_backend_str \g__sys_backend_tl
\__kernel_sys_configuration_load:n
{
l3backend- \c_sys_backend_str
}
}\cs_new_protected:Npn \__sys_load_backend_check:N \#1
{
\sys_if_engine_xetex:TF
{
\str_case:V \#1
{
\dvisvgm \}
\xdvipdfmx \{
\tl_gset:Nn \#1 \xetex
\}
\xetex
\}
\msg_error:nnee \sys \wrong-backend \#1 \xetex
\tl_gset:Nn \#1 \xetex
\}
\sys_if_engine_luatex:TF
{
\bool_lazy_or:nnF
{
\str_if_eq_p:V \#1 \luatex
\str_if_eq_p:V \#1 \pdftex
{\msg_error:nnee \sys \wrong-backend \#1 \xetex
\tl_gset:Nn \#1 \xetex
\sys_if_engine_luatex:TF
{
\tl_gset:Nn \#1 \luatex
\}
\tl_gset:Nn \#1 \pdftex
}
\}
\str_case:V \#1
{
\dvisvgm \}
\xdvipdfmx \{
\tl_gset:Nn \#1 \xetex
\}
\xetex
\}
\str_case:V \#1
{
\dvips \}
\dvipdfmx \{
\str_if_eq:V \#1 \dvips
\}}
}
\sys_load_backend:n \__sys_load_backend_check:N, and \c_sys_backend_str. These functions are documented on page 79.

\sys_ensure_backend: A simple wrapper.
\cs_new_protected:Npn \sys_ensure_backend:
\str_if_exist:NF \c_sys_backend_str
{ \sys_load_backend:n { } }
(End of definition for \sys_ensure_backend:. This function is documented on page 79.)

\g__sys_debug_bool
\bool_new:N \g__sys_debug_bool
(End of definition for \g__sys_debug_bool.)

\sys_load_debug: The most complicated thing here is that we can only use \__kernel_sys_configuration_load:n in the preamble in \LaTeX.
\cs_new_protected:Npn \sys_load_debug:
\bool_if:NF \g__sys_debug_bool
{ \__kernel_sys_configuration_load:n { l3debug } }
\bool_gset_true:N \g__sys_debug_bool
\cs_if_exist:NT \@expl@finalise@setup@@
{ \tl_gput_right:Nn \@expl@finalise@setup@@
  { \tl_gput_right:Nn \@kernel@after@begindocument
    { \cs_gset_protected:Npn \sys_load_debug:
      { \msg_error:nn { sys } { load-debug-in-preamble } } }
  }
}
(End of definition for \sys_load_debug:. This function is documented on page 79.)
48.1.4 Access to the shell

\_\_sys\_internal\_tl

\tl\_new:N \_\_sys\_internal\_tl

(End of definition for \_\_sys\_internal\_tl.)

\c\_\_sys\_marker\_tl

The same idea as the marker for rescanning token lists.

\tl\_const:N \c\_\_sys\_marker\_tl \{ \token\_to\_str:N \}

(End of definition for \c\_\_sys\_marker\_tl.)

\sys\_get\_shell:nnNF \sys\_get\_shell:nnN \_\_sys\_get\_do:Nw

Setting using a shell is at this level just a slightly specialised file operation, with an
additional check for quotes, as these are not supported.
\sys_shell_now:n \sys_shell_now:e \sys_shell_now:x \_sys_shell_now:e
Execute commands through shell escape at shipout.
For Lua\TeX{}, we use the same helper as above but delayed using a late\_lua whatsit.
Creating a late\_lua whatsit works a bit different if we are running under Con\TeX{}.

\sys_shell_shipout:n \sys_shell_shipout:e \sys_shell_shipout:x \_sys_shell_shipout:e
Execute commands through shell escape at shipout.
For Lua\TeX{}, we use the same helper as above but delayed using a late\_lua whatsit.
Creating a late\_lua whatsit works a bit different if we are running under Con\TeX{}.
return function(f)
  local n = node.new(whatsit_id, latelua_sub)
  setfield(n, 'data', f)
  return n
end

local node_write = node.direct.write

lua(cmd("_sys_shell_shipout:e", function()
  local cmd = scan_string()
  node_write(new_latelua(function() shellescape(cmd) end))
  end, "global", "protected")
end

luacmd("__sys_shell_shipout:e", function()

(End of definition for \sys_shell_shipout:n and \_sys_shell_shipout:e. This function is documented on page 78.)

48.2 Dynamic (every job) code

\_kernel_sys_everyjob:
\_sys_everyjob:n
\g_sys_everyjob_tl
\cs_new_protected:Npn \_kernel_sys_everyjob:
{ \tl_use:N \g_sys_everyjob_tl
  \tl_gclear:N \g_sys_everyjob_tl
}
\cs_new_protected:Npn \_sys_everyjob:n #1
{ \__sys_shell_shipout:e \exp_not:n \#1 }
\cs_new_protected:Npn \_sys_shell_shipout:n #1
{ \iow_shipout:Nn \c__sys_shell_stream_int \#1 }
\cs_generate_variant:Nn \sys_shell_shipout:n { e , x }

(End of definition for \_kernel_sys_everyjob:, \_sys_everyjob:n, and \g_sys_everyjob_tl.)

48.2.1 The name of the job
\c_sys_jobname_str

Inherited from the \LaTeX3 name for the primitive. This has to be the primitive as it’s set in \everyjob. If the user does

\pdf\latex \input some-file-name

then \everyjob is inserted before \jobname is changed form texput, and thus we would have the wrong result.

\__sys_everyjob:n
{ \cs_new_eq:NN \c_sys_jobname_str \tex_jobname:D }

(End of definition for \_kernel_sys_everyjob:, \_sys_everyjob:n, and \g_sys_everyjob_tl.)
48.2.2 Time and date

Copies of the information provided by \TeX. There is a lot of defensive code in package mode: someone may have moved the primitives, and they can only be recovered if we have \texttt{primitive} and it is working correctly. For Ini\TeX of course that is all redundant but does no harm.

\begin{verbatim}
\__sys_everyjob:n
  { group_begin:
    \cs_set:Npn \__sys_tmp:w #1
      { \str_if_eq:eeTF \cs_meaning:N \token_to_str:N \#1 \{ \#1 \}
        { \cs_if_exist:NTF \tex_primitive:D
            { \bool_lazy_and:nnTF
                \{ \sys_if_engine_xetex_p: \}
                \int_compare_p:nNn
                  \exp_after:wN \use_none:n \tex_XeTeXrevision:D
                  \#1
                { \#1 }{ 0 }
            }{ 0 }
          \}
        { \boots.lazy_and:nnTF
            \{ \sys_if_engine_xetex_p: \}
            \if_int_compare_p:nNn
              \exp_after:wN \use_none:n \tex_XeTeXrevision:D
              \#1
            { \#1 }{ 0 }
          \}
        { \tex_primitive:D \#1 }
      }{ 0 }

    \int_const:Nn \c_sys_minute_int
      { \int_mod:nn \__sys_tmp:w \time \{ \#1 \} }
    \int_const:Nn \c_sys_hour_int
      { \int_div_truncate:nn \__sys_tmp:w \time \{ \#1 \} }
    \int_const:Nn \c_sys_day_int
      { \__sys_tmp:w \day }
    \int_const:Nn \c_sys_month_int
      { \__sys_tmp:w \month }
    \int_const:Nn \c_sys_year_int
      { \__sys_tmp:w \year }
  } group_end:
\end{verbatim}

\begin{verbatim}
\__sys_everyjob:n
  { group_begin:
    \cs_set:Npn \__sys_tmp:w #1
      { \str_if_eq:eeTF \cs_meaning:N \token_to_str:N \#1 \{ \#1 \}
        { \cs_if_exist:NTF \tex_primitive:D
            { \bool_lazy_and:nnTF
                \{ \sys_if_engine_xetex_p: \}
                \int_compare_p:nNn
                  \exp_after:wN \use_none:n \tex_XeTeXrevision:D
                  \#1
                { \#1 }{ 0 }
            }{ 0 }
          \}
        { \boots.lazy_and:nnTF
            \{ \sys_if_engine_xetex_p: \}
            \if_int_compare_p:nNn
              \exp_after:wN \use_none:n \tex_XeTeXrevision:D
              \#1
            { \#1 }{ 0 }
          \}
        { \tex_primitive:D \#1 }
      }{ 0 }

    \int_const:Nn \c_sys_minute_int
      { \int_mod:nn \__sys_tmp:w \time \{ \#1 \} }
    \int_const:Nn \c_sys_hour_int
      { \int_div_truncate:nn \__sys_tmp:w \time \{ \#1 \} }
    \int_const:Nn \c_sys_day_int
      { \__sys_tmp:w \day }
    \int_const:Nn \c_sys_month_int
      { \__sys_tmp:w \month }
    \int_const:Nn \c_sys_year_int
      { \__sys_tmp:w \year }
  } group_end:
\end{verbatim}

(End of definition for \c_sys_minute_int and others. These variables are documented on page 74.)

\c_sys_timestamp_str

A simple expansion: Lua\TeX chokes if we use \texttt{pdf\texttt{feedback}} here, hence the direct use of Lua. Notice that the function there is in the \texttt{pdf} library but is not actually tied to PDF.

\begin{verbatim}
\__sys_everyjob:n
  { \str_const:Nn \c_sys_timestamp_str
      { \cs_if_exist:NTF \tex_directlua:D
          { \tex_directlua:D \{ tex.print(pdf.getcreationdate()) \} }
        { \tex_creationdate:D }
      }
  }
\end{verbatim}

(End of definition for \c_sys_jobname_str. This variable is documented on page 74.)
48.2.3 Random numbers

\sys_rand_seed:
Unpack the primitive.

\sys_gset_rand_seed:n
The primitive always assigns the seed globally.

\sys_timer:
In \LaTeX, create a pseudo-primitive, otherwise try to locate the real primitive. The elapsed time will be available if this succeeds.
48.2.4 Access to the shell

Expose the engine’s shell escape status to the user.

\c_sys_shell_escape_int

\sys_if_shell_p
\sys_if_shell:TF
\sys_if_shell_unrestricted:TF
\sys_if_shell_restricted:TF

Performs a check for whether shell escape is enabled. The first set of functions returns true if either of restricted or unrestricted shell escape is enabled, while the other two sets of functions return true in only one of these two cases.

\sys_get_query:nN
\sys_get_query:nnN
\sys_get_query:nnnN
\__sys_get_query_auxi:nnnN
\__sys_get_query_auxi:neeN
\__sys_get_query_auxii:nnnN
\__sys_get_query_auxii:neeN

Calling the query system is quite straight-forward: most of the effort is in making the read-back catcode-safe. We also want to trim off the trailing \texttt{\textasciitilde\textbackslash m} from the last line.
A wrapper for convenience.
\cs_new_protected:Npn \sys_split_query:nN #1\#2
\{ \sys_split_query:nnN {#1} \} \} \#2\}
\cs_new_protected:Npn \sys_split_query:nnN #1#2#3
\{ \sys_split_query:nnnN {#1} {#2} \} \#3\}
\group_begin:
\tex_lccode:D '\* = 13 \scan_stop:
\tex_lowercase:D
{\group_end:
\cs_new_protected:Npn \sys_split_query:nnnN #1#2#3#4
\{ \seq_clear:N #4
\sys_get_query:nnnN {#1} {#2} {#3} \l__sys_tmp_tl
\tl_if_empty:NF \l__sys_tmp_tl
\{ \seq_set_split:NnV #4 * \l__sys_tmp_tl \}
\}
\group_end:
\cs_new_protected:Npn \sys_split_query:nN \sys_split_query:nnN \sys_split_query:nnnN
\group_begin:
\tex_lccode:D '\* = 13 \scan_stop:
\tex_lowercase:D
{\group_end:
\cs_gset_eq:NN \g_file_curr_name_str \tex_jobname:D
\group_begin:
\cs_new_protected:Npn \__sys_everyjob:n
\{ \cs_gset_eq:NN \g_file_curr_name_str \tex_jobname:D \}
\group_end:
(End of definition for \sys_split_query:nN, \sys_split_query:nnN, and \sys_split_query:nnnN. These functions are documented on page 79.)
\g_file_curr_name_str

48.3.1 Held over from l3file
\c_sys_jobname_str
See comments about \c_sys_jobname_str: here, as soon as there is file input/output, things get “tided up”,
\__sys_everyjob:n
\{ \cs_gset_eq:NN \g_file_curr_name_str \tex_jobname:D \}
(End of definition for \g_file_curr_name_str. This variable is documented on page 99.)

48.4 Last-minute code
\sys_finalise:
\\c_sys_finalise:
\\c_sys_finalise:
\tl_new:N \g__sys_finalise_tl
\cs_new_protected:Npn \__sys_finalise:n #1
\{ \tl_gput_right:Nn \g__sys_finalise_tl {#1} \}
\cs_new_protected:Npn \__sys_finalise:n #1
\{ \tl_gput_right:Nn \g__sys_finalise_tl {#1} \}
\cs_new_protected:Npn \__sys_finalise:n #1
\{ \tl_gput_right:Nn \g__sys_finalise_tl {#1} \}
(End of definition for \sys_finalise:, \__sys_finalise:n, and \g__sys_finalise_tl. This function is documented on page 79.)

48.4.1 Detecting the output
\sys_if_output_dvi_p:
\sys_if_output_dvi:TF
\sys_if_output_pdf_p:
\sys_if_output_pdf:TF
\c_sys_output_str

This is a simple enough concept: the two views here are complementary.
\{\int_compare:nNnTF { \cs_if_exist_use:NF \tex_pdfoutput:D \{ 0 \} } > \{ 0 \} \\
\{ \texttt{pdf} } \\
\{ \texttt{dvi} \\
\}
\}
\__sys_const:nn { \sys_output_dvi:TF } \\
\{ \str_if_eq_p:Vn \c_sys_output_str { \texttt{dvi} } \}
\__sys_const:nn { \sys_output_pdf:TF } \\
\{ \str_if_eq_p:Vn \c_sys_output_str { \texttt{pdf} } \}
\}

(End of definition for \sys_output_dvi:TF, \sys_output_pdf:TF, and \c_sys_output_str. These functions are documented on page 76.)

### 48.4.2 Configurations

\g__sys_backend_tl

As the backend has to be checked and possibly adjusted, the approach here is to create a variable and use that in a one-shot to set a constant.

\tl_new:N \g__sys_backend_tl \\
\__sys_finalise:n \\
\__kernel_tl_gset:Nx \g__sys_backend_tl \\
\{ \sys_if_engine_xetex:TF \\
\{ \sys_if_output_pdf:TF \\
\{ \sys_if_engine_pdfTeX:TF \\
\{ \texttt{pdftex} \\
\{ \texttt{luatex} \\
\} \\
\{ \texttt{dvips} \\
\} \\
\}
\}
\} \\
\}

If there is a class option set, and recognised, we pick it up: these will over-ride anything set automatically but will themselves be over-written if there is a package option.

\__sys_finalise:n \\
\{ \cs_if_exist:NT \@classoptionslist \\
\{ \cs_if_eq:NNF \@classoptionslist \scan_stop: \\
\{ \clist_map_inline:Nn \@classoptionslist \\
\{ \str_case:nnT \{ \#1 \\
\{ \dvipdfmx \\
\{ \tl_gset:Nn \g__sys_backend_tl { \dvipdfmx } \} \\
\{ \dvips \\
\{ \tl_gset:Nn \g__sys_backend_tl { \dvips } \} \\
\}

612
\{ dvisvgm \}
\{ \tl_gset:Nn \g__sys_backend_tl { dvisvgm } \}
\{ pdftex \}
\{ \tl_gset:Nn \g__sys_backend_tl { pdfmode } \}
\{ xetex \}
\{ \tl_gset:Nn \g__sys_backend_tl { xdvipdfmx } \}
\{ \clist_remove_all:Nn \@unusedoptionlist {#1} \}
\{ \clist_remove_all:Nn \@unusedoptionlist {#1} \}
\begin{verbatim}
(End of definition for \g__sys_backend_tl.)
\end{verbatim}
\endinput
Chapter 49

l3msg implementation

\l__msg_internal_tl A general scratch for the module.
\tl_new:N \l__msg_internal_tl
(End of definition for \l__msg_internal_tl.)

\l__msg_name_str Used to save module info when creating messages.
\str_new:N \l__msg_name_str
\str_new:N \l__msg_text_str
(End of definition for \l__msg_name_str and \l__msg_text_str.)

49.1 Internal auxiliaries

\s__msg_mark Internal scan marks.
\s__msg_stop
\scan_new:N \s__msg_mark
\scan_new:N \s__msg_stop
(End of definition for \s__msg_mark and \s__msg_stop.)

\msg_use_none_delimit_by_s_stop:w Functions to gobble up to a scan mark.
\cs_new:Npn \__msg_use_none_delimit_by_s_stop:w #1 \s__msg_stop { }
(End of definition for \msg_use_none_delimit_by_s_stop:w.)

49.2 Creating messages

Messages are created and used separately, so there two parts to the code here. First, a mechanism for creating message text. This is pretty simple, as there is not actually a lot to do.

\c__msg_text_prefix_tl Locations for the text of messages.
\c__msg_more_text_prefix_tl
\tl_const:Nn \c__msg_text_prefix_tl { msg~text~>~ }
\tl_const:Nn \c__msg_more_text_prefix_tl { msg~extra~text~>~ }

614
\texttt{\msg_if_exist_p:nn} \texttt{\msg_if_exist:nnTF} \par

Test whether the control sequence containing the message text exists or not.

\begin{verbatim}
\prg_new_conditional:Npnn \msg_if_exist:nn { p , T , F , TF } { \cs_if_exist:cTF { \c__msg_text_prefix_tl #1 / #2 } { \prg_return_true: } { \prg_return_false: } }
\end{verbatim}

(End of definition for \texttt{\msg_if_exist:nnTF}. This function is documented on page 81.)

\texttt{\__msg_chk_if_free:nn} \par

This auxiliary is similar to \texttt{\__kernel_chk_if_free_cs:N}, and is used when defining messages with \texttt{\msg_new:nnnn}.

\begin{verbatim}
\cs_new_protected:Npn \__msg_chk_free:nn { \msg_if_exist:nnT {#1} {#2} { \msg_error:nnnn { msg } { already-defined } {#1} {#2} } { } }
\end{verbatim}

(End of definition for \texttt{\__msg_chk_if_free:nn}.)

\texttt{\msg_new:nnnn} \texttt{\msg_new:nnee} \texttt{\msg_new:nnxx} \texttt{\msg_new:nnn} \texttt{\msg_new:nne} \texttt{\msg_new:nnx} \texttt{\msg_set:nnnn} \texttt{\msg_set:nnn} \texttt{\msg_set:nne} \texttt{\msg_set:nnx} \texttt{\msg_set:nnnn} \texttt{\msg_set:nnee} \texttt{\msg_set:nnxx} \texttt{\msg_set:nnn} \texttt{\msg_set:nne} \texttt{\msg_set:nnx} \texttt{\msg_set:nnnn} \par

Setting a message simply means saving the appropriate text into two functions. A sanity check first.

\begin{verbatim}
\cs_new_protected:Npn \msg_new:nnnn #1#2#3#4 { \__msgchk_if_free:nn {#1} {#2} { \cs_gset:cpn { \c__msg_text_prefix_tl #1 / #2 } {#3} } {#4} }
\end{verbatim}

(End of definition for \texttt{\msg_new:nnnn} and others. These functions are documented on page 81.)
49.3  Messages: support functions and text

Simple pieces of text for messages.

\c__msg_coding_error_text_tl
\c__msg_continue_text_tl
\c__msg_critical_text_tl
\c__msg_fatal_text_tl
\c__msg_help_text_tl
\c__msg_no_info_text_tl
\c__msg_on_line_text_tl
\c__msg_return_text_tl
\c__msg_trouble_text_tl

\tl_const:Nn \c__msg_coding_error_text_tl
{ This-is-a-coding-error.
  \ \ \}
\tl_const:Nn \c__msg_continue_text_tl
{ Type<return>-to-continue }
\tl_const:Nn \c__msg_critical_text_tl
{ Reading-the-current-file-'\g_file_curr_name_str'-will-stop. }
\tl_const:Nn \c__msg_fatal_text_tl
{ This-is-a-fatal-error:-LaTeX-will-abort. }
\tl_const:Nn \c__msg_help_text_tl
{ For-immediate-help-type-H<return> }
\tl_const:Nn \c__msg_no_info_text_tl
{ LaTeX-does-not-know-anything-more-about-this-error,-sorry. }
\tl_const:Nn \c__msg_return_text_tl
{ on-line }
\tl_const:Nn \c__msg_trouble_text_tl
{ Try-typing<return>-to-proceed.
  \ \ \If-that-doesn’t-work,-type-X<return>-to-quit. }
\tl_const:Nn \c__msg_on_line_text_tl
{ More-errors-will-almost-certainly-follow: \ \ }
\tl_const:Nn \c__msg_return_text_tl
{ the-LaTeX-run-should-be-aborted. }

(End of definition for \c__msg_coding_error_text_tl and others.)

\msg_line_number:  For writing the line number nicely.  \msg_line_context: was set up earlier, so this is not new.
\cs_new:Npn \msg_line_number: { \int_use:N \tex_inputlineno:D }
\cs_gset:Npn \msg_line_context:
{ \c__msg_on_line_text_tl \c_space_tl \msg_line_number: } (End of definition for \msg_line_number: and \msg_line_context: These functions are documented on page 82.)

616
49.4 Showing messages: low level mechanism

The low-level interruption macro is rather opaque, unfortunately. Depending on the availability of more information there is a choice of how to set up the further help. We feed the extra help text and the message itself to a wrapping auxiliary, in this order because we must first setup \TeX's \texttt{\textbackslash errhelp} register before issuing an \texttt{\textbackslash errmessage}. To deal with the various cases of critical or fatal errors with and without help text, there is a bit of argument-passing to do.

\begin{verbatim}
\cs_new_protected:Npn \__msg_interrupt:NnnnN #1#2#3#4#5
\begin{Verbatim}
\str_set:Ne \l__msg_text_str { #1 {#2} }
\str_set:Ne \l__msg_name_str { \msg_module_name:n {#2} }
\cs_if_eq:cNTF
\begin{Verbatim}
{ \c__msg_more_text_prefix_tl #2 / #3 }
\__msg_no_more_text:nnnn
\end{Verbatim}
\begin{Verbatim}
{ \use:c { \c__msg_text_prefix_tl #2 / #3 } #4 }
\__msg_interrupt_wrap:nnn
\begin{Verbatim}
\__msg_interrupt_text:n
\__msg_interrupt_more_text:n
\end{Verbatim}
\begin{Verbatim}
\c__msg_no_info_text_tl
\tl_if_empty:NF #5
\begin{Verbatim}
{ \ \ \ #5 }
\end{Verbatim}
\end{Verbatim}
\end{Verbatim}
\begin{Verbatim}
\c__msg_no_more_text:nnnn
\end{Verbatim}
\begin{Verbatim}
\cs_new:Npn \__msg_no_more_text:nnnn #1#2#3#4 { }
\end{Verbatim}
\end{Verbatim}
\end{verbatim}

\begin{verbatim}
\cs_new_protected:Npn \__msg_interrupt_wrap:nnn #1#2#3
\begin{Verbatim}
\iow_wrap:nnnN { \ #3 } { } { } \__msg_interrupt_more_text:n
\group_begin:
\begin{Verbatim}
\int_sub:Nn \l_iow_line_count_int { 2 }
\iow_wrap:nenN { \l__msg_text_str : ~ #1 }
\end{Verbatim}
\end{verbatim}

\textit{End of definition for \__msg_interrupt:Nnnn and \__msg_no_more_text:nnnn.}
\end{verbatim}

First setup \TeX's \texttt{\textbackslash errhelp} register with the extra help \#1, then build a nice-looking error message with \#2. Everything is done using e-type expansion as the new line markers are different for the two type of text and need to be correctly set up. The auxiliary \texttt{\__msg_interrupt_more_text:n} receives its argument as a line-wrapped string, which is thus unaffected by expansion. We ave to split the main text into two parts as only the “message” itself is wrapped with a leader: the generic help is wrapped at full width. We also have to allow for the two characters used by \texttt{\textbackslash errmessage} itself.

\begin{verbatim}
\cs_new_protected:Npn \__msg_interrupt_wrap:nnn #1#2#3
\begin{Verbatim}
\iof_wrap:nnnn { \ \ #3 } { } { } \__msg_interrupt_more_text:n
\group_begin:
\begin{Verbatim}
\int_sub:Nn \l_iow_line_count_int { 2 }
\iof_wrap:nenN { \l__msg_text_str : ~ #1 }
\end{Verbatim}
\end{verbatim}

\textit{End of definition for \__msg_interrupt_wrap:nnn and \__msg_interrupt_more_text:n.}
\end{verbatim}
The business end of the process starts by producing some visual separation of the message from the main part of the log. The error message needs to be printed with everything made “invisible”: \texttt{TEX}'s own information involves the macro in which \texttt{errmessage} is called, and the end of the argument of the \texttt{errmessage}, including the closing brace. We use an active \texttt{!} to call the \texttt{errmessage} primitive, and end its argument with \texttt{use-\texttt{none}:n \{ \langle \text{spaces} \rangle \}} which fills the output with spaces. Two trailing closing braces are turned into spaces to hide them as well. The group in which we alter the definition of the active \texttt{!} is closed before producing the message: this ensures that tokens inserted by typing \texttt{I} in the command-line are inserted after the message is entirely cleaned up.

The \texttt{\_kernel_iow\_with:Nnn} auxiliary, defined in \texttt{l3file}, expects an \langle \text{integer variable} \rangle, an integer \langle \text{value} \rangle, and some \langle \text{code} \rangle. It runs the \langle \text{code} \rangle after ensuring that the \langle \text{integer variable} \rangle takes the given \langle \text{value} \rangle, then restores the former value of the \langle \text{integer variable} \rangle if needed. We use it to ensure that the \texttt{newlinechar} is 10, as needed for \texttt{\_kernel\_io\_newline:} to work, and that \texttt{\_errorcontextlines} is \texttt{-1}, to avoid showing irrelevant context. Note that restoring the former value of these integers requires inserting tokens after the \texttt{\_kernel\_err\_message}, which go in the way of tokens which could be inserted by the user. This is unavoidable.
49.5 Displaying messages

\textit{\texttt{\LaTeX}} is handling error messages and so the \TeX{} ones are disabled.

A function for issuing messages: both the text and order could in principle vary. The module name may be empty for kernel messages, hence the slightly contorted code path for a space.

```latex
\cs_new:Npn \msg_fatal_text:n #1
\msg_error_text:n {#1}
\msg_critical_text:n {Critical ~ \msg_error_text:n {#1}}
\msg_info_text:n {Info}}
\__msg_text:nn
```

(End of definition for \texttt{\_msg_interrupt:n})
For storing public module information: the kernel data is set up in advance.

\prop_new:N \g_msg_module_name_prop
\prop_new:N \g_msg_module_type_prop
\prop_gput:Nnn \g_msg_module_type_prop { LaTeX } { }

(End of definition for \g_msg_module_name_prop and \g_msg_module_type_prop. These variables are documented on page 81.)

\msg_module_type:n
Contextual footer information, with the potential to give modules an alternative name.

\cs_new:Npn \msg_module_type:n #1
{\prop_if_in:NnTF \g_msg_module_type_prop {#1} { \prop_item:Nn \g_msg_module_type_prop {#1} } { Package } }

(End of definition for \msg_module_type:n. This function is documented on page 81.)

\msg_module_name:n
\msg_see_documentation_text:n
Contextual footer information, with the potential to give modules an alternative name.

\cs_new:Npn \msg_module_name:n #1
{\prop_if_in:NnTF \g_msg_module_name_prop {#1} { \prop_item:Nn \g_msg_module_name_prop {#1} } {#1} }
\cs_new:Npn \msg_see_documentation_text:n #1
{See-the- \msg_module_name:n {#1}~-documentation-for-further-information.}

(End of definition for \msg_module_name:n and \msg_see_documentation_text:n. These functions are documented on page 81.)

\__msg_class_new:nn
\group_begin:
\cs_set_protected:Npn \__msg_class_new:nn #1#2
{\prop_new:c { l__msg_redirect_ #1 _prop }
\cs_new_protected:cpn { __msg_ #1 _code:nnnnnn } ##1##2##3##4##5##6 {#2}
\cs_new_protected:cpn { msg_ #1 :nnnnnn } ##1##2##3##4##5##6
{\use:e}
\group_end:
For fatal errors, after the error message \TeX{} bails out. We force a bail out rather than using \texttt{\end} as this means it does not matter if we are in a context where normally the run cannot end.

Not quite so bad: just end the current file.
For an error, the interrupt routine is called. We check if there is a “more text” by comparing that control sequence with a permanently empty text. We have to undefine the bootstrap versions here.

\cs_undefine:N \msg_error:nnee
\cs_undefine:N \msg_error:nne
\cs_undefine:N \msg_error:nn
\__msg_class_new:nn { error }
\begin{verbatim}
\__msg_interrupt:NnnnN \msg_error_text:n {#1} {#2}
\{ {#3} {#4} {#5} {#6} \}
c_empty_tl
\end{verbatim}

(End of definition for \msg_error:nnnnnn and others. These functions are documented on page 84.)

Warnings and information messages have no decoration. Warnings are printed to the terminal while information can either go to the log or both log and terminal.

\cs_new_protected:Npn \__msg_info_aux:NNnnnnnn #1#2#3#4#5#6#7#8
\begin{verbatim}
\str_set:Ne \l__msg_text_str { #2 {#3} }
\str_set:Ne \l__msg_name_str { \msg_module_name:n {#3} }
#1 { }
\iow_wrap:nenN
\begin{verbatim}
{ \l__msg_text_str : ~
\use:c { \c__msg_text_prefix_tl #3 / #4 } {#5} {#6} {#7} {#8}}
{ ( \l__msg_name_str )
\prg_replicate:nn
{ \str_count:N \l__msg_text_str
\- \str_count:N \l__msg_name_str }
\} #1
\end{verbatim}
\end{verbatim}
\__msg_class_new:nn { warning }
\__msg_info_aux:NNnnnnnn \iow_term:n \msg_warning_text:n
\{#1} {#2} {#3} {#4} {#5} {#6}
\__msg_class_new:nn { note }
\__msg_info_aux:NNnnnnnn \iow_log:n \msg_info_text:n
{#1} {#2} {#3} {#4} {#5} {#6}
\__msg_class_new:nn { info }
\__msg_info_aux:NNnnnnnn \iow_log:n \msg_info_text:n
{#1} {#2} {#3} {#4} {#5} {#6}
“Log” data is very similar to information, but with no extras added. “Term” is used for communicating with the user through the terminal, like diagnostic messages, and debugging. This is similar to “log” messages, but uses the terminal output.

The `none` message type is needed so that input can be gobbled.

The `show` message type is used for \seq_show:N and similar complicated data structures. Wrap the given text with a trailing dot (important later) then pass it to `\__msg_show:n`. If there is `\>~` (or if the whole thing starts with `>~`) we split there, print the first part and show the second part using `\showtokens` (the `\exp_after:wN` ensure a nice display).

Note that this primitive adds a leading `>~` and trailing dot. That is why we included a trailing dot before wrapping and removed it afterwards. If there is no `\>~` do the same but with an empty second part which adds a spurious but inevitable `>~`.

```latex
\__msg_class_new:nn { log } \{ \protect \iow_wrap:nnnN \protect \{ \use:c { \c__msg_text_prefix_tl #1 / #2 } {#3} {#4} {#5} {#6} \} \protect \} \iow_log:n \}
\__msg_class_new:nn { term } \{ \protect \iow_wrap:nnnN \protect \{ \use:c { \c__msg_text_prefix_tl #1 / #2 } {#3} {#4} {#5} {#6} \} \protect \} \iow_term:n \}
\__msg_class_new:nn { none } \{ \}
```

```
\__msg_class_new:nn { show } \{ \protect \iow_wrap:nnnN \protect \{ \use:c { \c__msg_text_prefix_tl #1 / #2 } {#3} {#4} {#5} {#6} \} \protect \} \__msg_show:n \}
```
\__msg_show:nn {#1} {#2}
\cs_new_protected:Npn \__msg_show:nn #1#2
{
\tl_if_empty:nF {#1}
{ \exp_args:No \iow_term:n { \use_none:n #1 } }
\tl_set:Nn \l__msg_internal_tl {#2}
\__kernel_iow_with:Nnn \tex_newlinechar:D { 10 }
{
\__kernel_iow_with:Nnn \tex_errorcontextlines:D { -1 }
{
\tex_showtokens:D \exp_after:wN \exp_after:wN \exp_after:wN
{ \exp_after:wN \l__msg_internal_tl }
}
}
}

(End of definition for \msg_show:nnnnnn and others. These functions are documented on page 87.)
End the group to eliminate \__msg_class_new:nn.
\group_end:
\msg_show_item:n \msg_show_item_unbraced:n \msg_show_item:nn \msg_show_item_unbraced:nn
Each item in the variable is formatted using one of the following functions. We cannot use \ and so on because these short-hands cannot be used inside the arguments of messages, only when defining the messages. We need to use ^\^J here directly as l3file is not yet loaded.
\cs_new:Npe \msg_show_item:n #1
\cs_new:Npe \msg_show_item_unbraced:n #1
\cs_new:Npe \msg_show_item:nn #1#2
\cs_new:Npe \msg_show_item_unbraced:nn #1#2

(End of definition for \msg_show_item:n and others. These functions are documented on page 87.)
\__msg_class_chk_exist:nT
Checking that a message class exists. We build this from \cs_if_free:cTF rather than \cs_if_exist:cTF because that avoids reading the second argument earlier than necessary.
\cs_new:Npn \__msg_class_chk_exist:nT #1
{\cs_if_free:cTF { \_msg__ #1 _code:nnnnn }
{ \msg_error:nnn { msg } { class_unknown } {#1} }
}

624
\_msg_class_tl  
Support variables needed for the redirection system.
9624 \tl_new:N \_msg_class_tl 
9625 \tl_new:N \_msg_current_class_tl
(End of definition for \_msg_class_tl and \_msg_current_class_tl.)
\_msg_redirect_prop For redirection of individually-named messages
9626 \prop_new:N \_msg_redirect_prop
(End of definition for \_msg_redirect_prop.)
\_msg_hierarchy_seq During redirection, split the message name into a sequence: {/module/submodule}, {/module}, and {}/.
9627 \seq_new:N \_msg_hierarchy_seq
(End of definition for \_msg_hierarchy_seq.)
\_msg_class_loop_seq Classes encountered when following redirections to check for loops.
9628 \seq_new:N \_msg_class_loop_seq
(End of definition for \_msg_class_loop_seq.)
\_msg_class_chk_exist:nT Actually using a message is a multi-step process. First, some safety checks on the message and class requested. The code and arguments are then stored to avoid passing them around. The assignment to \_msg_use_code: is similar to \tl_set:Nn. The message is eventually produced with whatever \_msg_class_tl is when \_msg_use_code: is called. Here is also a good place to suppress tracing output if the trace package is loaded since all (non-expandable) messages go through this auxiliary.
9629 \cs_new_protected:Npn \_msg_use:nnnnnnn #1#2#3#4#5#6#7
9630 \cs_if_exist_use:N \conditionally@traceoff
9631 \msg_if_exist:nnTF {#2} {#3}
9632 { \_msg_class_chk_exist:nT {#1}
9633 { \tl_set:Nn \_msg_current_class_tl {#1}
9634 \cs_set_protected:Npe \_msg_use_code:
9635 { \exp_not:n
9636 \use:c { \_msg_ \_msg_class_tl_code:nnnnnn } \{#2\} \{#3\} \{#4\} \{#5\} \{#6\} \{#7\}
9637 }
9638 } \_msg_use_redirect_name:n \{ \_msg_use:nnnnnn \{ \_msg_class_tl \{ \_msg_use_code: \}
9639 \cs_if_exist_use:N \conditionally@traceon
9640 \}
9641 \cs_new_protected:Npm \_msg_use:nnnnnnn #1#2#3#4#5#6#7
9642 \cs_if_exist_use:N \conditionally@traceoff
9643 \msg_if_exist:nnTF {#2} {#3}
9644 { \_msg_class_chk_exist:nT {#1}
9645 { \tl_set:Nn \_msg_current_class_tl {#1}
9646 \cs_set_protected:Npe \_msg_use_code:
9647 { \exp_not:n
9648 \use:c { \_msg_ \_msg_class_tl_code:nnnnnn } \{#2\} \{#3\} \{#4\} \{#5\} \{#6\} \{#7\}
9649 }
9650 } \_msg_use_redirect_name:n \{ \_msg_use:nnnnnn \{ \_msg_class_tl \{ \_msg_use_code: \}
9651 \cs_if_exist_use:N \conditionally@traceon
9652 \}
9653 \cs_new_protected:Npm \_msg_use_code: \{ \}

625
The first check is for an individual message redirection. If this applies then no further redirection is attempted. Otherwise, split the message name into ⟨module⟩, ⟨submodule⟩ and ⟨message⟩ (with an arbitrary number of slashes), and store {/module/submodule}, {/module} and {} into \_\_msg\_{hierarchy\_seq}. We then map through this sequence, applying the most specific redirection.

At this point, the items of \_\_msg\_{hierarchy\_seq} are the various levels at which we should look for a redirection. Redirections which are less specific than the argument of \_\_msg\_{use\_redirect\_module\_n} are not attempted. This argument is empty for a class redirection, /module for a module redirection, etc. Loop through the sequence to find the most specific redirection, with module ##1. The loop is interrupted after testing for a redirection for #1 equal to the argument #1 (least specific redirection allowed). When a redirection is found, break the mapping, then if the redirection targets the same class, output the code with that class, and otherwise set the target as the new current class, and search for further redirections. Those redirections should be at least as specific as ##1.
\msg_redirect_name:nn

Named message always use the given class even if that class is redirected further. An empty target class cancels any existing redirection for that message.

\cs_new_protected:Npn \msg_redirect_name:nnn #1#2#3
{ \tl_if_empty:nTF {#3}
  \prop_remove:Nn \l__msg_redirect_prop { / #1 / #2 }
  \prop_put:Nnn \l__msg_redirect_prop { / #1 / #2 } {#3}
{ \tl_set:Nn \l__msg_current_class_tl {#2}
  \seq_clear:N \l__msg_class_loop_seq
  \__msg_redirect_loop_chk:nnn {#2} {#3} {#1}
}
}

(End of definition for \msg_redirect_name:nnn and others.)

\msg_redirect_class:nn
\msg_redirect_module:nn
\__msg_redirect:n
\__msg_redirect_loop_chk:nnn
\__msg_redirect_loop_list:n

If the target class is empty, eliminate the corresponding redirection. Otherwise, add the redirection. We must then check for a loop: as an initialization, we start by storing the initial class in \l__msg_current_class_tl.

\cs_new_protected:Npn \msg_redirect_class:nn
{ \__msg_redirect:nnn { } }
\cs_new_protected:Npn \msg_redirect_module:nnn #1
{ \__msg_redirect:nnn { / #1 } }
\cs_new_protected:Npn \__msg_redirect:nnn #1#2#3
{ \__msg_class_chk_exist:nT {#2}
  \tl_if_empty:nTF {#3}
    \prop_remove:cn { l__msg_redirect_ #2 _prop } {#1}
  { \__msg_class_chk_exist:nT {#3}
    \prop_put:cnn { l__msg_redirect_ #2 _prop } {#1} {#3}
    \tl_set:Nn \l__msg_current_class_tl {#2}
    \seq_clear:N \l__msg_class_loop_seq
    \__msg_redirect_loop_chk:nnn {#2} {#3} {#1}
  }
}

(End of definition for \msg_redirect_name:nnn. This function is documented on page 89.)

Since multiple redirections can only happen with increasing specificity, a loop requires that all steps are of the same specificity. The new redirection can thus only create a loop with other redirections for the exact same module, #1, and not submodules. After some initialization above, follow redirections with \l__msg_class_tl, and keep track in \l__msg_class_loop_seq of the various classes encountered. A redirection from a class to
itself, or the absence of redirection both mean that there is no loop. A redirection to the
initial class marks a loop. To break it, we must decide which redirection to cancel. The
user most likely wants the newly added redirection to hold with no further redirection.
We thus remove the redirection starting from #2, target of the new redirection. Note
that no message is emitted by any of the underlying functions: otherwise we may get an
infinite loop because of a message from the message system itself.

\begin{verbatim}
\cs_new_protected:Npn \__msg_redirect_loop_chk:nnn #1#2#3
\seq_put_right:Nn \l__msg_class_loop_seq {#1}
\prop_get:cnNT { l__msg_redirect_ #1 _prop } {#3} \l__msg_class_tl
{ \str_if_eq:VnF \l__msg_class_tl \l__msg_class_tl \l__msg_current_class_tl
\{ \prop_put:cnn \l__msg_redirect_ #2 _prop \l__msg_class_tl \l__msg_class_tl \l__msg_current_class_tl
\msg_warning:nneee{ msg }{ redirect-loop \seq_item:Nn \l__msg_class_loop_seq {1} \seq_item:Nn \l__msg_class_loop_seq {2} \l__msg_class_tl \l__msg_current_class_tl
{ \__msg_redirect_loop_list:n \seq_map_function:NN \l__msg_class_loop_seq \__msg_redirect_loop_list:n
{ \seq_item:Nn \l__msg_class_loop_seq {1} }
\__msg_redirect_loop_chk:onn \l__msg_class_tl \l__msg_class_tl \l__msg_current_class_tl \l__msg_class_tl \l__msg_current_class_tl
\} \prop_get:cnNT { l__msg_redirect_ #2 _prop } \l__msg_class_tl \tl_show:n
\} \cs_generate_variant:Nn \__msg_redirect_loop_chk:nnn { o }
\cs_new:Npn \__msg_redirect_loop_list:n #1 { \{#1} ~ => ~ }
\end{verbatim}

(End of definition for \msg_redirect_class:nn and others. These functions are documented on page
89.)

### 49.6 Kernel-specific functions

A short-hand used for \int_show:n and similar functions that passes to \tl_show:n the
result of applying #1 (a function such as \int_eval:n) to the expression #2. The use
of f-expansion ensures that #1 is expanded in the scope in which the show command is
called, rather than in the group created by \iow_wrap:nnn. This is only important for
expressions involving the \currentgrouplevel or \currentgrouptype. On the other
hand we want the expression to be converted to a string with the usual escape character,

\begin{verbatim}
\cs_new_protected:Npn \__kernel_msg_show_eval:nnN
( \exp_args:Nf \__msg_show_eval:nnN { \tl_eval:n {#1 = \tl_show:n {#2}} })
\cs_new_protected:Npn \__kernel_msg_show_eval:nnN
{ \tl_log:n \tl_show:n {\tl_eval:n {#1 = \tl_show:n {#2}}}}
\cs_new_protected:Npn \__msg_show_eval:nnN
\end{verbatim}

628
49.7 Internal messages

Error messages needed to actually implement the message system itself.

\begin{verbatim}
\msg_new:nnnn { msg } { already-defined } { \_msg_coding_error_text_tl }
\msg_new:nnnn { msg } { unknown } { \_msg_coding_error_text_tl }
\end{verbatim}
LaTeX was asked to display a message called ‘#2’
by the module ‘#1’: this message does not exist.
\c__msg_return_text_tll
}
\msg_new:nnnn { msg } { class-unknown }
{ Unknown-message-class=’#1’. }
{ LaTeX has been asked to redirect messages to a class=’#1’:
this was never defined.
\c__msg_return_text_tll
}
\msg_new:nnnn { msg } { redirect-loop }
{ Message redirection loop caused by {#1} => {#2}
\tl_if_empty:nF {#3} { -for-module=’ \use_none:n #3 ’ } .
}
{ Adding the message redirection {#1} => {#2}
\tl_if_empty:nF {#3} { -for-the-module=’ \use_none:n #3 ’ } -
created an infinite loop\\
\low-indent:n { #4 \\\
}
}

Messages for earlier kernel modules plus a few for l3keys which cover coding errors.
\msg_new:nnnn { kernel } { bad-number-of-arguments }
{ Function ‘#1’ cannot be defined with #2 arguments. }
{ \c__msg_coding_error_text_tl
LaTeX has been asked to define a function ‘#1’ with
#2 arguments. -
TeX allows between 0 and 9 arguments for a single function.
}
\msg_new:nnnn { kernel } { command-already-defined }
{ Control sequence ‘#1’ already defined. }
{ \c__msg_coding_error_text_tl
LaTeX has been asked to create a new control sequence ‘#1’-
but this name has already been used elsewhere. \ \ \The current meaning is:\
\ \ #2
}
\msg_new:nnnn { kernel } { command-not-defined }
{ Control sequence ‘#1’ undefined. }
{ \c__msg_coding_error_text_tl
LaTeX has been asked to use a control sequence ‘#1’:
this has not been defined yet.
}
\msg_new:nnnn { kernel } { empty-search-pattern }
{ Empty search pattern. }
{ \c__msg_coding_error_text_tl
LaTeX has been asked to replace an empty pattern by ‘#1’: - that would lead to an infinite loop!
}
\cs_if_exist:NF \tex_elapsedtime:D
  \msg_new:nnnn { kernel } { no-elapsed-time }
    { No-clock-detected-for-\texttt{#1}. }
    { The-current-engine-provides-no-way-to-access-the-system-time. }
  \}
\msg_new:nnnn { kernel } { non-base-function }
  \{ Function-'\texttt{#1}'-is-not-a-base-function \}
  \{
  \c__msg_coding_error_text_tl
    Functions-defined-through-\iow_char:N\cs_new:Nn\must-have-
    a-signature-consisting-of-only-normal-arguments-'N'-and-'n'.-
    The-signature-'\texttt{#2}'-of-'\texttt{#1}'-contains-other-arguments-'#3'.-
    To-define-variants-use-\iow_char:N\cs_generate_variant:Nn\and-to-define-other-functions-use-\iow_char:N\cs_new:Npm.
  \}
\msg_new:nnnn { kernel } { missing-colon }
  \{ Function-'\texttt{#1}'-contains-no-':'. \}
  \{
  \c__msg_coding_error_text_tl
    Code-level-functions-must-contain-':'-to-separate-the-
    argument-specification-from-the-function-name.-This-is-
    needed-when-defining-conditionals-or-variants,-or-when-building-a-
    parameter-text-from-the-number-of-arguments-of-the-function.
  \}
\msg_new:nnnn { kernel } { overflow }
  \{ Integers-larger-than-2^{30}-1-cannot-be-stored-in-arrays. \}
  \{
    \tl_if_empty:nF {#2} { at-position-#2- } in-the-array-'\texttt{#1}'.-
    The-largest-allowed-value-#4-will-be-used-instead.
  \}
\msg_new:nnnn { kernel } { out-of-bounds }
  \{ Access-to-an-entry-beyond-an-array's-bounds. \}
  \{
    An-attempt-was-made-to-access-or-store-data-at-position-#2-of-the-
    array-'\texttt{#1}',-but-this-array-has-entries-at-positions-from-1-to-#3.
  \}
\msg_new:nnnn { kernel } { protected-predicate }
  \{ Predicate-'\texttt{#1}'-must-be-expandable. \}
  \{
  \c__msg_coding_error_text_tl
    LaTeX-has-been-asked-to-define-'\texttt{#1}'-as-a-protected-predicate.-
    Only-expandable-tests-can-have-a-predicate-version.
  \}
\msg_new:nnnn { kernel } { randint-backward-range }
  \{ Wrong-order-of-bounds-in-\iow_char:N\int_rand:nn{#1}{#2}. \}
\msg_new:nnnn { kernel } { conditional-form-unknown }
  \{ Conditional-form-'\texttt{#1}'-for-function-'\texttt{#2}'-unknown. \}
  \{
  \c__msg_coding_error_text_tl
    LaTeX-has-been-asked-to-define-the-conditional-form-'\texttt{#1}'-of-
    the-function-'\texttt{#2}',-but-only-'TF','T','F','and-'p'-forms-exist.
  \}

631
9090 \msg_new:nnnn { kernel } { variant-too-long }
9091 { Variant-form-‘\#1’-longer-than-base-signature-of-‘\#2’. }
9092 { \c__msg_coding_error_text_tl
9093 \LaTeX{} has been asked to create a variant of the function-‘\#2’-
9094 with a signature starting with-‘\#1’, but that is longer than-
9095 the signature (part after the colon) of-‘\#2’. }
9097 \msg_new:nnnn { kernel } { invalid-variant }
9098 { Variant-form-‘\#1’-invalid-for-base-form-‘\#2’. }
9099 { \c__msg_coding_error_text_tl
9100 \LaTeX{} has been asked to create a variant of the function-‘\#2’-
9101 with a signature starting with-‘\#1’, but cannot change an argument-
9102 from type-‘\#3’ to type-‘\#4’. }
9111 \msg_new:nnnn { kernel } { invalid-exp-args }
9112 { Invalid-variant-specifier-‘\#1’-in-‘\#2’. }
9113 { \c__msg_coding_error_text_tl
9114 \LaTeX{} has been asked to create an -\{\iow_char:N\}exp_args:N...-
9115 function with signature-‘\#2’-but-‘\#1’ is not a valid argument-
9116 specifier. }
9121 \msg_new:nnn { kernel } { deprecated-variant }
9122 { Variant-form-‘\#1’-deprecated-for-base-form-‘\#2’.
9123 One should not change an argument from type-‘\#3’ to type-‘\#4’
9124 \str_case:nnF {\#3}
9125 { \{ n \} { :-use-a-\{\token_if_eq_charcode:NNTF \#4 c v V’}-variant? }
9126 \{ N \} { :-base-form-only-accepts-a-single-token-argument. }
9127 \{\#4\} { :-base-form-is-already-a-variant. }
9128 } { . } }
9134 \msg_new:nnn { char } { active }
9135 { Cannot-generate-active-chars. }
9136 \msg_new:nnn { char } { invalid-catcode }
9137 { Invalid-catcode-for-char-generation. }
9138 \msg_new:nnn { char } { null-space }
9139 { Cannot-generate-null-char-as-a-space. }
9141 \msg_new:nnn { char } { out-of-range }
9142 \msg_new:nnn { dim } { zero-of-range }
9143 { Cannot-generate-null-char-as-a-space. }
9144 \msg_new:nnnn { kernel } { quote-in-shell }
9145 { Quotes-in-shell-command-‘\#1’. }
9146 \msg_new:nnnn { keys } { no-property }
9148 { No-property-given-in-definition-of-key-‘\#1’.
9149 } { \c__msg_coding_error_text_tl
9150 Inside-\{\keys_define:nn\} each-key-name-
9151 needs-a-property: \\ \}
LaTeX did not find a '.' to indicate the start of a property.

The property '#1' accepts boolean values only.

LaTeX was asked to set property '#1' for key '#2'. No value was given for the property, and one is required.

LaTeX has been asked to create quark test function '#1' but that name with signature '#2', but that signature is not valid.

LaTeX has been asked to create a new scan mark '#1' but this name has already been used for a scan mark.

An attempt was made to push or pop the item at position #3 of '#1', but this position does not exist.

Sequence only has #2 item \int_compare:nF { #2 = 1 } {s}.}

Sequence '#1' does not have an item #3

The sequence '#1' is too long to be shuffled by TeX.
\TeX\ has \intevaln { \cmax_register_int + 1 } ~
toks-registers:-this-only-allows-to-shuffle-up-to-
\intuseN \cmax_register_int \ items.-
The-list-will-not-be-shuffled.
}
}  \msgnew:nnnn { kernel } { variable-not-defined }
{ Variable-#1-undefined. }
{ \c__msg_coding_error_text_tl
LaTeX\-has\-been\-asked\-to\-show\-a\-variable-#1,\-but\-this\-has\-not-
been\-defined\-yet.
}
}  \msgnew:nnnn { kernel } { bad-type }
{ Variable-`#1'-is-not-a-valid-#3. }
{ \c__msg_coding_error_text_tl
The-variable-`#1'-with-\tlifemptyTF {#4} {meaning} {value}\\\
\iowindentn {#2}\\\
should-be-a-#3-variable,-but-
\tlifemptyTF {#4}
{ it\-is\-not \strifeqnnF {#3} { bool } \{-a\-short\-macro \}. }
{ it\-does\-not\-have\-the\-correct-
\strifeqnnTF {#2} {#4}
{ category\-codes. }
{ internal\-structure:\\\\iowindentn {#4} }
}
}
}  \msgnew:nnnn { prop } { bad-link }
{ Variable-`#1'-is-not-a-valid-(linked)-prop. }
{ \c__msg_coding_error_text_tl
The-variable-`#1'-has-an-incorrect-internal-structure.-
Its-internal-entry-`#2'-points-to-`#3',-whose-name-is-not-of-the-
form-`#4-<key>'.
}
}  \msgnew:nnnn { clist } { non-clist }
{ Variable-`#1'-is-not-a-valid-clist. }
{ \c__msg_coding_error_text_tl
The-variable-`#1'-with-value\\\
\iowindentn {#2}\\\
should-be-a-clist-variable,-but-it-includes-empty-or-blank-items-
without-braces.
}
}  \msgnew:nnnn { prop } { misused }
{ A\-property\-list\-was\-misused. }
{ \c__msg_coding_error_text_tl
A\-property\-list\-variable\-was\-used\-without\-an\-accessor\-function.-
It-
\tlifemptyTF {#1}
{ is\-empty. }
{ contains\-the\-key\-value\-pairs \usenone:n \#1 . }
Some errors only appear in expandable settings, hence don’t need a “more-text” argument.

Messages used by the "show" functions.

The-comma-list: \tl_if_empty:nF \#1 \{ \#1 \} \tl_if_empty:nTF \#2 \{ is-empty \} \#2 .

The-integer-array-#1-contains-#2-items: \#3 .

The-sequence-#1-\tl_if_empty:nTF \#2 \{ is-empty \} \#2 .

The-following-
\_msg\_expandable\_error:nn

In expansion only context, we cannot use the normal means of reporting errors. Instead, we rely on a low-level \TeX\ error caused by expanding a macro `\???` with parameter text “?” (this could be any token) which we used followed by something else (here, a space). This shows the context, which thanks to the odd-looking \use:n is

<argument> \???

! mypkg Error: The error message.

In other words, \TeX\ is processing the argument of `\use:n`, which is \??? ⟨space⟩ ! ⟨error type⟩ : (error message).

\cs_set_protected:Npn \_msg\_tmp:w #1
{\cs_new:Npn #1 ? { }}
\cs_new:Npn \_msg\_expandable\_error:nn #1#2
{\exp_after:wN \exp_after:wN \_msg\_use\_none\_delimit\_by\_s\_stop:w \use:n { #1 ! #2 : #1 } \s__msg\_stop}
\exp_args:Nc \_msg\_tmp:w { ??? }

(End of definition for \_msg\_expandable\_error:nn)
The command built from the csname \c__msg_text_prefix_tl #1 / #2 takes four arguments and builds the error text, which is fed to \msg_expandable_error:n with appropriate expansion: just as for usual messages the arguments are first turned to strings, then the message is fully expanded. The module name also has to be determined.

\begin{verbatim}
\exp_args_generate:n { oooo }
\cs_new:Npn \msg_expandable_error:nnnnnn #1#2#3#4#5#6
{ \exp_args:Nee \__msg_expandable_error:nn
  \exp_args:Nc \exp_args:Noooo
  { \c__msg_text_prefix_tl #1 / #2 }
  { \tl_to_str:n {#3} }
  { \tl_to_str:n {#4} }
  { \tl_to_str:n {#5} }
  { \tl_to_str:n {#6} }
}
\msg_error_text:n {#1} }
\cs_new:Npn \msg_expandable_error:nnnnn #1#2#3#4#5
{ \msg_expandable_error:nnnnnn {#1} {#2} {#3} {#4} {#5} { } }
\cs_new:Npn \msg_expandable_error:nnnnf #1#2#3#4
{ \msg_expandable_error:nnnnnn {#1} {#2} {#3} {#4} { } { } }
\cs_new:Npn \msg_expandable_error:nnnff #1#2#3
{ \msg_expandable_error:nnnnnn {#1} {#2} {#3} { } { } { } }
\cs_new:Npn \msg_expandable_error:nnn #1#2#3#4
{ \msg_expandable_error:nnnnnn {#1} {#2} {#3} {#4} { } { } { } }
\cs_generate_variant:Nn \msg_expandable_error:nnnnnn { nnffff }
\cs_generate_variant:Nn \msg_expandable_error:nnnnn { nnfff }
\cs_generate_variant:Nn \msg_expandable_error:nnnn { nnff }
\cs_generate_variant:Nn \msg_expandable_error:nnn { nnf }\end{verbatim}

(End of definition for \msg_expandable_error:nnnnnn and others. These functions are documented on page 88.)

49.9 Message formatting

\begin{verbatim}
\prop_gput:Nnn \g_msg_module_name_prop { kernel } { LaTeX }
\prop_gput:Nnn \g_msg_module_type_prop { kernel } { }
\clist_map_inline:nn
{ char , clist , coffin , debug , deprecation , dim, msg ,
  quark , prg , prop , scanmark , seq , sys }
\prop_gput:Nnn \g_msg_module_name_prop {#1} { LaTeX }
\prop_gput:Nnn \g_msg_module_type_prop {#1} { }
\end{verbatim}

637
Chapter 50

\textbf{l3file implementation}

The following test files are used for this code: \texttt{m3file001}.

\section{Input operations}

\subsection{Variables and constants}

\begin{itemize}
  \item \texttt{\_ior\_internal\_tl} Used as a short-term scratch variable.
  \item \texttt{\_ior\_term\_ior} Reading from the terminal (with a prompt) is done using a positive but non-existent stream number. Unlike writing, there is no concept of reading from the log.
  \item \texttt{\_ior\_stream\_tl} Used to recover the raw stream number from the stack.
  \item \texttt{\_ior\_streams\_prop} The name of the file attached to each stream is tracked in a property list. To get the correct number of reserved streams in package mode the underlying mechanism needs to be queried. For \LaTeX\ and plain \TeX\ this data is stored in \texttt{\count16}: with the \texttt{etex} package loaded we need to subtract 1 as the register holds the number of the next stream to use. In Con\TeXt, we need to look at \texttt{\count38} but there is no subtraction: like the original plain \TeX/\LaTeX\ mechanism it holds the value of the last stream allocated.
\end{itemize}
50.1.2 Stream management

Reserving a new stream is done by defining the name as equal to using the terminal.
\begin{verbatim}
\ior_new:N #1 { \cs_new_eq:NN #1 \c__ior_term_ior }
\end{verbatim}

The usual scratch space.
\begin{verbatim}
\ior_new:N \g_tmpa_ior
\ior_new:N \g_tmpb_ior
\end{verbatim}

An auxiliary searches for the file in the \TeX, \LaTeX\ 2? and \LaTeX\ 3 paths. Then pass the file found to the lower-level function which deals with streams. The \texttt{full_name} is empty when the file is not found.
\begin{verbatim}
\ior_open:NnTF \ior_open:cnTF
\end{verbatim}
Streams are reserved using \texttt{\newread} before they can be managed by \texttt{ior}. To prevent \texttt{ior} from being affected by redefinitions of \texttt{\newread} (such as done by the third-party package \texttt{morewrites}), this macro is saved here under a private name. The complicated code ensures that \texttt{\__ior_new:N} is not \texttt{\outer} despite plain \TeX’s \texttt{\newread} being \texttt{\outer}. For Con\TeX xt, we have to deal with the fact that \texttt{\newread} works like our own: it actually checks before altering definition.

\begin{verbatim}
\exp_args:NNf \cs_new_protected:Npn \__ior_new:N
{ \exp_args:NNc \exp_after:wN \exp_stop_f: { newread } }
\cs_if_exist:NT \contextversion
{ \cs_new_eq:NN \__ior_new_aux:N \__ior_new:N
\cs_gset_protected:Npn \__ior_new:N #1
{ \cs_undefine:N #1 \__ior_new_aux:N #1 }
}
\cs_generate_variant:Nn \__kernel_ior_open:Nn { No }
\cs_new_protected:Npe \__kernel_ior_open:Nn { No }
\cs_defend:NN \__kernel_ior_open:Nn { No }
\exp_after:wN \__kernel_ior_open:Nn { No }
\exp_after:wN \__ior_open_stream:Nn
{ \__ior_open_stream:Nn { No }
\ior_close:N #1
\seq_gpop:NNTF \g__ior_streams_seq \l__ior_stream_tl
{ \__ior_open_stream:Nn #1 {#2} }
\__ior_new:N #1
\seq_gpop:NNTF \g__ior_streams_seq \l__ior_stream_tl
{ \__ior_new:N #1 \l__ior_stream_tl { \int_eval:n {#1} }
\__ior_open_stream:Nn #1 {#2} }
\__kernel_ior_open:Nn { No }
\cs_gset_protected:Npn \__ior_open_stream:Nn #1#2
{ \ior_openin:N #1 \sys_if_engine_luatex:TF
{ {#2} }
\prop_gput:NVn \exp_not:N \g__ior_streams_prop #1 {#2}
\exp_not:N \__kernel_file_name_quote:n {#2} \scan_stop: }
\exp_after:wN \__ior_open_stream:Nn
\exp_after:wN \__ior_open_stream:Nn
\end{verbatim}

The stream allocation itself uses the fact that there is a list of all of those available. Life gets more complex as it’s important to keep things in sync. That is done using a two-part approach: any streams that have already been taken up by \texttt{ior} but are now free are tracked, so we first try those. If that fails, ask plain \TeX or \LaTeXe for a new stream and use that number (after a bit of conversion).

Here, we act defensively in case Lua\TeX is in use with an extensionless file name.

\begin{verbatim}
\exp_after:wN \__ior_open_stream:Nn
\cs_new_protected:Npe \__kernel_ior_open:Nn { No }
\cs_gset_protected:Npn \__kernel_ior_open:Nn { No }
\cs_gset_protected:Npn \__ior_open_stream:Nn
{ \ior_openin:N #1 \sys_if_engine_luatex:TF
{ {#2} }
\prop_gput:NVn \exp_not:N \g__ior_streams_prop #1 {#2}
\exp_not:N \__kernel_file_name_quote:n {#2} \scan_stop: }
\exp_after:wN \__ior_open_stream:Nn
\exp_after:wN \__ior_open_stream:Nn
\end{verbatim}

\begin{verbatim}
\exp_after:wN \__ior_open_stream:Nn
\cs_new_protected:Npe \__kernel_ior_open:Nn { No }
\cs_gset_protected:Npn \__kernel_ior_open:Nn { No }
\cs_gset_protected:Npn \__ior_open_stream:Nn
{ \ior_openin:N #1 \sys_if_engine_luatex:TF
{ {#2} }
\prop_gput:NVn \exp_not:N \g__ior_streams_prop #1 {#2}
\exp_not:N \__kernel_file_name_quote:n {#2} \scan_stop: }
\exp_after:wN \__ior_open_stream:Nn
\exp_after:wN \__ior_open_stream:Nn
\end{verbatim}
Actually much easier than either the standard open or input versions! When calling \_kernel_ior_open:Nn the file the pipe is added to signal a shell command, but the quotes are not added yet—they are added later by \_kernel_file_name_quote:n.

\cs_new_protected:Npn \ior_shell_open:Nn #1#2
\sys_if_shell:TF
\__ior_shell_open:oN { \tl_to_str:n {#2} } #1 }
\msg_error:nn { kernel } { pipe-failed }
\cs_new_protected:Npm \ior_shell_open:nN #1#2
\tl_if_in:nnTF {#1} { " }
\msg_error:nne
\msg_error:nnne { kernel } { quote-in-shell } {#1}
\__kernel_ior_open:Nn #2 { |#1 } }
\cs_generate_variant:Nn \__ior_shell_open:nN { o }
\msg_new:nnnn { kernel } { pipe-failed }
\msg_new:nnnn { kernel } { Cannot-run-piped-system-commands. }
\LaTeX-tried-to-call-a-system-process-but-this-was-not-possible.\ LaT\m-ex-tried-to-call-a-system-process-but-this-was-not-possible.\ Try-the-\texttt{\--shell-escape\texttt{(or-\texttt{--enable-pipes\texttt{)}}\--option.}

(End of definition for \ior_shell_open:Nn and \__ior_shell_open:nN. This function is documented on page 91.)

\cs_new_protected:Npn \ior_close:N #1
\int_compare:nT { -1 < #1 < \c__ior_term_ior }
\tex_closein:D #1
\prop_gremove:NV \g__ior_streams_prop #1
\seq_if_in:NVF \g__ior_streams_seq #1
\seq_gpush:NV \g__ior_streams_seq #1
\cs_gset_eq:NN #1 \c__ior_term_ior
\cs_gset_eq:NN \c__ior_term_ior #1
\cs_generate_variant:Nn \ior_close:N { c }

(End of definition for \ior_close:N. This function is documented on page 92.)

\cs_new_protected:Npm \ior_show:N { \__ior_show:NN { \tl_show:n } }
\cs_generate_variant:Nn \ior_show:N { c }
\cs_new_protected:Npm \ior_log:N { \__ior_log:NN { \tl_log:n } }
\cs_generate_variant:Nn \ior_log:N { c }

(End of definition for \ior_log:N. This function is documented on page 92.)
\ior_show:NN \ior_log:N, and \ior_show:N. These functions are documented on page 92.

Show the property lists, but with some “pretty printing”. See the l3msg module. The first argument of the message is ior (as opposed to iow) and the second is empty if no read stream is open and non-empty (the list of streams formatted using \msg_show_item_unbraced:nn) otherwise. The code of the message show-streams takes care of translating ior/iow to English.

\ior_show_list: \ior_log_list: \ior_list:N

Show the property lists, but with some “pretty printing”. See the l3msg module. The first argument of the message is ior (as opposed to iow) and the second is empty if no read stream is open and non-empty (the list of streams formatted using \msg_show_item_unbraced:nn) otherwise. The code of the message show-streams takes care of translating ior/iow to English.

50.1.3 Reading input

\if_eof:w The primitive conditional

\ior_if_eof_p:N \ior_if_eof:NTF

To test if some particular input stream is exhausted the following conditional is provided. The primitive test can only deal with numbers in the range \[0, 15\] so we catch outliers (they are exhausted).
\begin{verbatim}
\if\endlinechar=\-
\@text:line
{ \text_read:D @ to @ }
\else
{ \tl_set:Nn @ { \q_no_value } }
\fi:
\else:
\prg_return_true:
\fi:
\else:
\prg_return_true:
\fi:
\}
End of definition for \ior_if_eof:NTF. This function is documented on page 95.)
\end{verbatim}

And here we read from files.
\begin{verbatim}
\cs_new_protected:Npn \ior_get:NN #1#2
{ \ior_get:NNF #1 #2 { \tl_set:Nn #2 { \q_no_value } } }
\cs_new_protected:Npn \__ior_get:NN #1#2
{ \tex_read:D #1 to #2 }
\prg_new_protected_conditional:Npnn \ior_get:NN #1#2 { T , F , TF }
{ \ior_if_eof:NTF #1
{ \prg_return_false: }
{ \__ior_get:NN #1#2
\prg_return_true: }
}
(End of definition for \ior_get:NN, \__ior_get:NN, and \ior_get:NTTF. These functions are documented on page 93.)
\end{verbatim}

Reading as strings is a more complicated wrapper, as we wish to remove the endline character and restore it afterwards.
\begin{verbatim}
\cs_new_protected:Npn \ior_str_get:NN #1#2
{ \ior_str_get:NNF #1 #2 { \tl_set:Nn #2 { \q_no_value } } }
\cs_new_protected:Npn \__ior_str_get:NN #1#2
{ \exp_args:Nno \use:n
{ \int_set:Nn \tex_endlinechar:D { -1 }
\tex_readline:D @ to @ }
\int_set:Nn \tex_endlinechar:D
}
\prg_new_protected_conditional:Npnn \ior_str_get:NN #1#2 { T , F , TF }
{ \ior_if_eof:NTF #1
{ \prg_return_false: }
{ \__ior_str_get:NN #1#2
\prg_return_true: }
}
(End of definition for \ior_str_get:NN, \__ior_str_get:NN, and \ior_str_get:NTTF. These functions are documented on page 93.)
\end{verbatim}
For reading without a prompt.

Getting from the terminal is better with pretty-printing.

Usual map breaking functions.

Mapping over an input stream can be done on either a token or a string basis, hence the set up. Within that, there is a check to avoid reading past the end of a file, hence the two applications of \ior_if_eof:N and its lower-level analogue \if_eof:w. This mapping cannot be nested with twice the same stream, as the stream has only one “current line”.

644
Since the \TeX{} primitive (\texttt{\read} or \texttt{\readline}) assigns the tokens read in the same way as a token list assignment, we simply call the appropriate primitive. The end-of-loop is checked using the primitive conditional for speed.

\begin{verbatim}
\cs_new_protected:Npn \ior_map_variable:NNn { \__ior_map_variable:NNNn \ior_get:NN }
\cs_new_protected:Npn \ior_str_map_variable:NNn { \__ior_map_variable:NNNn \ior_str_get:NN }
\cs_new_protected:Npn \__ior_map_variable:NNNn #1#2#3#4 {
    \ior_if_eof:NF #2 { \__ior_map_variable_loop:NNNn #1#2#3 {#4} }
    \prg_break_point:Nn \ior_map_break: { }
}
\cs_new_protected:Npn \__ior_map_variable_loop:NNNn #1#2#3#4 {
    #1 #2 #3
    \if_eof:w #2
    \exp_after:wN \ior_map_break:
    \fi:
    #4
    \__ior_map_variable_loop:NNNn #1#2#3 {#4}
}
\end{verbatim}

(End of definition for \texttt{\ior_map_variable:NNn} and others. These functions are documented on page 94.)

50.2 Output operations

There is a lot of similarity here to the input operations, at least for many of the basics. Thus quite a bit is copied from the earlier material with minor alterations.

50.2.1 Variables and constants

\texttt{\_\_iow\_internal\_tl}

Used as a short-term scratch variable.

\begin{verbatim}
\tl_new:N \_\_iow\_internal\_tl
\end{verbatim}

(End of definition for \texttt{\_\_iow\_internal\_tl}.)

\texttt{\c\_log\_iow} \texttt{\c\_term\_iow}

Here we allocate two output streams for writing to the transcript file only (\texttt{\c\_log\_iow}) and to both the terminal and transcript file (\texttt{\c\_term\_iow}). Recent \LaTeX{} provides 128...
write streams; we also use \texttt{\_c_term_iow} as the first non-allowed write stream so its value depends on the engine.

\begin{verbatim}
10438 \int_const:Nn \c_log_iow { -1 }
10439 \int_const:Nn \c_term_iow
10440 { 10441 \bool_lazy_and:nnTF 10442 { \sys_if_engine_luatex_p: } 10443 { \int_compare_p:nNn \tex_luatexversion:D > { 80 } } 10444 { 128 } 10445 { 16 } } 10446 
\end{verbatim}

\textit{(End of definition for \texttt{\_c_log_iow} and \texttt{\_c_term_iow}. These variables are documented on page 99.)}

\texttt{\_g_iow_streams_seq} A list of the currently-available output streams to be used as a stack.

\begin{verbatim}
10447 \seq_new:N \g__iow_streams_seq
\end{verbatim}

\textit{(End of definition for \texttt{\_g_iow_streams_seq}.)}

\texttt{\_l_iow_stream_tl} Used to recover the raw stream number from the stack.

\begin{verbatim}
10448 \tl_new:N \l__iow_stream_tl
\end{verbatim}

\textit{(End of definition for \texttt{\_l_iow_stream_tl}.)}

\texttt{\_g_iow_streams_prop} As for reads with the appropriate adjustment of the register numbers to check on.

\begin{verbatim}
10449 \prop_new:N \g__iow_streams_prop
10450 \int_step_inline:nnn 10451 { 0 } 10452 { 10453 \cs_if_exist:NTF \contextversion 10454 { \tex_count:D 39 ~ } 10455 { 10456 \tex_count:D 17 ~ 10457 \cs_if_exist:NT \loccount { - 1 } 10458 } 10459 } 10460 { 10461 \prop_gput:Nnn \g__iow_streams_prop {#1} { Reserved-by-format } 10462 }
\end{verbatim}

\textit{(End of definition for \texttt{\_g_iow_streams_prop}.)}

\texttt{\_s_iow_mark} \texttt{\_s_iow_stop} Internal scan marks.

\begin{verbatim}
10453 \scan_new:N \s__iow_mark
10454 \scan_new:N \s__iow_stop
\end{verbatim}

\textit{(End of definition for \texttt{\_s_iow_mark} and \texttt{\_s_iow_stop}.)}

\texttt{\_iow_use_i_delimit_by_s_stop:nw} Functions to gobble up to a scan mark.

\begin{verbatim}
10461 \cs_new:Npn \_iow_use_i_delimit_by_s_stop:nw #1 #2 \s__iow_stop {#1}
\end{verbatim}

\textit{(End of definition for \texttt{\_iow_use_i_delimit_by_s_stop:nw}.)}
50.3 Stream management

Reserving a new stream is done by defining the name as equal to writing to the terminal:
odd but at least consistent.

\cs_new_protected:N \iow_new:N \iow_new:c
\cs_generate_variant:Nn \iow_new:N { c }

(End of definition for \iow_new:N. This function is documented on page 91.)

\g_tmap_a_iow \g_tmap_b_iow
\tl_new:N \l__iow_file_name_tl

(End of definition for \g_tmap_a_iow and \g_tmap_b_iow. These variables are documented on page 99.)

\__iow_new:N
As for read streams, copy \newwrite, making sure that it is not \outer. For Con\TeXt, we have to deal with the fact that \newwrite works like our own: it actually checks before altering definition.

\exp_args:NNf \cs_new_protected:Npn \__iow_new:N #1
\cs_if_exist:NT \contextversion
\cs_new_eq:NN \__iow_new_aux:N \__iow_new:N
\cs_gset_protected:Npn \__iow_new:N #1
\cs_undefine:N #1
\__iow_new_aux:N #1

(End of definition for \__iow_new:N.)

\l__iow_file_name_tl
\tl_new:N \l__iow_file_name_tl
(End of definition for \l__iow_file_name_tl.)

\iow_open:Nn \iow_open:NV \iow_open:cn \iow_open:cV
\__kernel_iow_open:Nn \__kernel_iow_open:No \__iow_open_stream:Nn \__iow_open_stream:NV
\cs_new_protected:Npn \iow_open:Nn \iow_open:NV \iow_open:cn \iow_open:cV
\__kernel_tl_set:Nx \l__iow_file_name_tl
\__kernel_file_name_sanitiz.n (#2)
\__kernel_iow_open:No \l__iow_file_name_tl
\cs_generate_variant:Nn \iow_open:Nn \iow_open:NV \iow_open:cn \iow_open:cV
\cs_new_protected:Npn \__kernel_iow_open:Nn #1 \iow_open:NV \iow_open:cn \iow_open:cV

(End of definition for \iow_open:Nn. The same idea as for reading, but without the path and without the need to allow for a conditional version.)
\texttt{\iow_close:N} #1
\texttt{\seq_gpop:NNTF \g__iow_streams_seq \l__iow_stream_tl}
\{ \__iow_open_stream:Nn #1 {#2} \}
\{
  \__iow_new:N #1
  \__kernel_tl_set:Nx \l__iow_stream_tl { \int_eval:n {#1} }
  \__iow_open_stream:Nn #1 {#2}
\}
\cs_generate_variant:Nn \__kernel_iow_open:Nn { No }
\cs_new_protected:Npn \__iow_open_stream:Nn #1#2
  \{\tex_global:D \tex_chardef:D #1 = \l__iow_stream_tl \scan_stop:
  \prop_gput:NVn \g__iow_streams_prop #1 {#2} \scan_stop:
  \tex_immediate:D \tex_openout:D #1 \__kernel_file_name_quote:n {#2} \scan_stop:
\}
\cs_generate_variant:Nn \__iow_open_stream:Nn { NV }

(End of definition for \iow_open:Nn, \__kernel_iow_open:Nn, and \__iow_open_stream:Nn. This function is documented on page 91.)

\texttt{\iow_shell_open:Nn} \texttt{\__iow_shell_open:nN} \texttt{\__iow_shell_open:oN}

Very similar to the ior version
\texttt{\cs_generate_variant:Nn \__kernel_iow_open:Nn { No }}
\texttt{\cs_new_protected:Npn \__iow_shell_open:nN #1#2}
\} \sys_if_shell:TF
\{ \__iow_shell_open:oN { \tl_to_str:n {#2} } #1 \}
\{ \msg_error:nn { kernel } { pipe-failed } \}
\}
\cs_new_protected:Npm \__iow_shell_open:nN \#1#2
\{ \tl_if_in:nnTF {#1} { " } \{ \msg_error:nne { kernel } { quote-in-shell } {#1} \}
\} \__kernel_iow_open:Nn \#2 { "#1"
\}
\cs_generate_variant:Nn \__kernel_iow_open:Nn { o }

(End of definition for \iow_shell_open:Nn and \__iow_shell_open:nN. This function is documented on page 91.)

\texttt{\iow_close:N} \texttt{\iow_close:c}

Closing a stream is not quite the reverse of opening one. First, the close operation is easier than the open one, and second as the stream is actually a number we can use it directly to show that the slot has been freed up.
\texttt{\cs_new_protected:Npn \iow_close:N #1}
\{ \int_compare:nT { \c_log_iow < #1 < \c_term_iow } \{
  \tex_immediate:D \tex_closeout:D #1
  \prop_gremovex:Nv \g__iow_streams_prop \#1
  \seq_if_in:Nx \g__iow_streams_seq \#1
  \{ \seq_gpushx:Nv \g__iow_streams_seq \#1 \}
  \cs_gset_eq:NN #1 \c_term_iow
\}

(End of definition for \iow_close:N and \iow_close:n. This function is documented on page 91.)

648
Seek the stream in the \texttt{\g__iow Streams_prop} list, then show the stream as open or closed accordingly.

\begin{verbatim}
\cs_new_protected:Npn \iow_show:N { \__iow_show:NN \tl_show:n }
\cs_generate_variant:Nn \iow_show:N { c }
\cs_new_protected:Npn \iow_log:N { \__iow_show:NN \tl_log:n }
\cs_generate_variant:Nn \iow_log:N { c }
\cs_new_protected:Npn \__iow_show:NN #1#2
{
  \__kernel_chk_defined:NT #2
  {
    \prop_get:NVNTF \g__iow_streams_prop #2 \l__iow_internal_tl
    {
      \exp_args:Ne #1
      {
        \token_to_str:N #2 \ closed \ l__iow_internal_tl }
    }
    {
      \exp_args:Ne #1 \token_to_str:N #2 \ open: \ l__iow_internal_tl }
  }
}
\end{verbatim}

Done as for input, but with a copy of the auxiliary so the name is correct.

\begin{verbatim}
\cs_new_protected:Npn \iow_show_list: { \__iow_list:N \msg_show:nneeee }
\cs_new_protected:Npn \iow_log_list: { \__iow_list:N \msg_log:nneeee }
\cs_new_protected:Npn \__iow_list:N #1
{
  \prop_map_function:NN \g__iow_streams_prop \msg_show_item_unbraced:nn
  \msg_show_item_unbraced:nn
}
\end{verbatim}

First the easy part, this is the primitive, which expects its argument to be braced.

\begin{verbatim}
\cs_new_protected:Npn \iow_shipout_e:Nn \iow_shipout_e:Ne \iow_shipout_e:cn \iow_shipout_e:ce
\cs_new_protected:Npn \iow_shipout_e:Nn { \__iow_shipout_e:NN #1#2 }
\cs_new_protected:Npn \iow_shipout_e:Ne { \tex_write:D #1 {#2} }
\cs_new_protected:Npn \iow_shipout_e:cn { \cs_generate_variant:Nn \iow_shipout_e:Nn { c, ce } }
\end{verbatim}

\section{Deferred writing}

\subsection*{Deferred writing}

(End of definition for \texttt{\iow_show:N}, \texttt{\iow_log:N}, and \texttt{\__iow_show:NN}. These functions are documented on page 92.)

(End of definition for \texttt{\iow_show_list:}, \texttt{\iow_log_list:}, and \texttt{\__iow_list:N}. These functions are documented on page 92.)

(End of definition for \texttt{\iow_shipout_e:Nn}. This function is documented on page 97.)
With \textsf{\texttt{\textbackslash iow\_shipout:Nn}} available deferred writing without expansion is easy.

\begin{verbatim}
(End of definition for \texttt{\textbackslash iow\_shipout:Nn}. This function is documented on page 96.)
\end{verbatim}

\subsection*{50.3.2 Immediate writing}

If the integer \#1 is equal to \#2, just leave \#3 in the input stream. Otherwise, pass the old value to an auxiliary, which sets the integer to the new value, runs the code, and restores the integer.

\begin{verbatim}
(End of definition for \texttt{\textbackslash \_kernel\_iow\_with:Nnn} and \texttt{\_iow\_with:nNnn}.)
\end{verbatim}

\texttt{\textbackslash iow\_now:Nn} This routine writes the second argument onto the output stream without expansion. If this stream isn’t open, the output goes to the terminal instead. If the first argument is no output stream at all, we get an internal error. We don’t use the expansion done by \texttt{\textbackslash write} to get the \texttt{\textbackslash Nx} variant, because it differs in subtle ways from \textit{x}-expansion, namely, macro parameter characters would not need to be doubled. We set the \texttt{\_newlinechar} to 10 using \texttt{\_kernel\_iow\_with:Nnn} to support formats such as plain \textsf{\texttt{\textbackslash \textbackslash ^{}J}} which \textsf{\texttt{\textbackslash \textbackslash ^{}J}} looks at the value of the \texttt{\_newlinechar} at shipout time in those cases.

\begin{verbatim}
(End of definition for \texttt{\textbackslash iow\_now:Nn}. This function is documented on page 96.)
\end{verbatim}

Writing to the log and the terminal directly are relatively easy; as we need the two \texttt{\textbackslash e}-type variants for bootstrapping, they are redefinitions here.

\begin{verbatim}
(650)
\end{verbatim}
50.3.3 Special characters for writing

\texttt{\iow_newline}: Global variable holding the character that forces a new line when something is written to an output stream.

\texttt{\iow_char:N} Function to write any escaped char to an output stream.

50.3.4 Hard-wrapping lines to a character count

This is the “raw” number of characters in a line which can be written to the terminal. The standard value is the line length typically used by \TeX Live and MiK\TeX.

\texttt{\l__iow_newline_tl} The token list inserted to produce a new line, with the \texttt{\langle run-on text \rangle}.

\texttt{\l__iow_line_target_int} This stores the target line count: the full number of characters in a line, minus any part for a leader at the start of each line.

\texttt{\__iow_set_indent:n} \texttt{\__iow_unindent:w} \texttt{\l__iow_one_indent_tl} \texttt{\l__iow_one_indent_int} The one indent variables hold one indentation marker and its length. The \texttt{\__iow_unindent:w} auxiliary removes one indentation. The function \texttt{\__iow_set_indent:n} (that could possibly be public) sets the indentation in a consistent way. We set it to four spaces by default.
\l_\_iow_indent_tl \l_\_iow_indent_int
The current indentation (some copies of \l_\_iow_one_indent_tl) and its number of characters.

\l_\_iow_line_tl \l_\_iow_line_part_tl
These hold the current line of text and a partial line to be added to it, respectively.

\l_\_iow_line_break_bool
Indicates whether the line was broken precisely at a chunk boundary.

\l_\_iow_wrap_tl
Used for the expansion step before detokenizing, and for the output from wrapping text:
fully expanded and with lines which are not overly long.

\c__iow_wrap_marker_tl
\c__iow_wrap_end_marker_tl
\c__iow_wrap_newline_marker_tl
\c__iow_wrap_allow_break_marker_tl
\c__iow_wrap_indent_marker_tl
\c__iow_wrap_unindent_marker_tl
Every special action of the wrapping code is starts with the same recognizable string,
\c__iow_wrap_marker_tl. Upon seeing that “word”, the wrapping code reads one space-delimited argument to know what operation to perform. The setting of \escapechar here is not very important, but makes \c__iow_wrap_marker_tl look marginally nicer.
\iow_wrap_allow_break: We set \iow_wrap_allow_break:n to produce an error when outside messages. Within wrapped message, it is set to \_\iow_wrap_allow_break: when valid and otherwise to \_\iow_wrap_allow_break_error:. The second produces an error expandably.

\iow_indent:n We set \iow_indent:n to produce an error when outside messages. Within wrapped message, it is set to \_\iow_indent:n when valid and otherwise to \_\iow_indent_error:n. The first places the instruction for increasing the indentation before its argument, and the instruction for unindenting afterwards. The second produces an error expandably. Note that there are no forced line-break, so the indentation only changes when the next line is started.

\iow_wrap:nnnN The main wrapping function works as follows. First give \, \_ and other formatting commands the correct definition for messages and perform the given setup #3. The definition of \_ uses an “other” space rather than a normal space, because the latter might be absorbed by \TeX to end a number or other \_\_ expansion. Use \conditionally@traceoff if defined; it is introduced by the trace package and suppresses uninteresting tracing of the wrapping code.
Then fully-expand the input: in package mode, the expansion uses \LaTeX{}\textsc{2e}'s \protect\typeout mechanism in the same way as \typeout. In generic mode this setting is useless but harmless. As soon as the expansion is done, reset \iow-indent:n to its error definition: it only works in the first argument of \iow-wrap:nnnN.

Afterwards, set the newline marker (two assignments to fully expand, then convert to a string) and initialize the target count for lines (the first line has target count \l_iow-line_count_int instead).

Sanity check.

There is then a loop over the input, which stores the wrapped result in \l__iow-wrap-tl. After the loop, the resulting text is passed on to the function which has been given as a post-processor. The f-expansion removes a leading space from \l__iow-wrap-tl.

(End of definition for \iow-wrap:nnnN. This function is documented on page 98.)
f-expansion of the main wrapping function and \use_none:n removes a newline marker inserted by later code. The main loop consists of repeatedly calling the chunk auxiliary to wrap chunks delimited by (newline or indentation) markers.

```latex
\cs_new_protected:Npn \__iow_wrap_do:
\begin{verbatim}
\__kernel_tl_set:Nx \l__iow_wrap_tl
\{\exp_args:No \__kernel_str_to_other_fast:n \l__iow_wrap_tl
\c__iow_wrap_end_marker_tl
\}
\__kernel_tl_set:Nx \l__iow_wrap_tl
\{\exp_after:wN \__iow_wrap_fix_newline:w \l__iow_wrap_tl
\^^J \q__iow_nil \^^J \s__iow_stop
\}
\exp_after:wN \__iow_wrap_start:w \l__iow_wrap_tl
\end{verbatim}
\cs_new:Npn \__iow_wrap_fix_newline:w #1 \^^J #2 \^^J
\begin{verbatim}
\__iow_wrap_chunk:nw { \l_iow_line_count_int }
\end{verbatim}
\end{verbatim}

(End of definition for \__iow_wrap_do:, \__iow_wrap_fix_newline:w, and \__iow_wrap_start:w.)

\__iow_wrap_chunk:nw The chunk and next auxiliaries are defined indirectly to obtain the expansions of \c_catcode_other_space_tl and \c__iow_wrap_marker_tl in their definition. The next auxiliary calls a function corresponding to the type of marker (its \#2), which can be newline or indent or unindent or end. The first argument of the chunk auxiliary is a target number of characters and the second is some string to wrap. If the chunk is empty simply call next. Otherwise, set up a call to \__iow_wrap_line:nw, including the indentation if the current line is empty, and including a trailing space (#1) before the \__iow_wrap_end_chunk:w auxiliary.

\cs_set_protected:Npn \__iow_tmp:w #1#2
\begin{verbatim}
\__iow_wrap_chunk:nw { \l_iow_line_count_int }
\end{verbatim}
\end{verbatim}

\begin{verbatim}
\end{verbatim}
The code snippet describes a function that processes text to break it into lines. The function, \_iow_wrap_next:nw, is used to wrap text into lines. It checks if the current line is empty and, if not, it processes the line using the \_iow_wrap_line:nw function, which adds indentation and stores the line part. If the line is too long, it breaks it into seven parts,Calling \_iow_wrap_line_seven:nnnnnnn function, and processes the remaining text using \_iow_wrap_end_chunk:w function. This function works by grabbing 8 characters at a time and storing them in the line part. When there are less than 8 characters left, it calls \_iow_wrap_line_end:NnnnnnnnN function to clean up the remaining characters. The weird \use_none:nnnnn line is used to ensure the required data is in the right place.
Functions here are defined indirectly: \_\_iow_wrap_line:nw is eventually called with an “other” space as its argument. The goal is to remove from \_\_iow_line_part_tl the part after the last space. In most cases this is done by repeatedly calling the break_loop auxiliary, which leaves “words” (delimited by spaces) until it hits the trailing space: then its argument ##3 is \_\_iow_wrap_line:seven instead of a single token, and that break_end auxiliary leaves in the assignment the line until the last space, then calls \_\_iow_wrap_line_end:nw to finish up the line and move on to the next. If there is no space in \_\_iow_line_part_tl then the break_first auxiliary calls the break_none auxiliary. In that case, if the current line is empty, the complete word (including
##4, characters beyond what we had grabbed) is added to the line, making it over-long. Otherwise, the word is used for the following line (and the last space of the line so far is removed because it was inserted due to the presence of a marker).

\cs_set_protected:Npn \__iow_tmp:w #1
{\tex_edef:D \l__iow_line_part_tl
  {\if_false: } \fi:
  \exp_after:wN \__iow_wrap_break_first:w
  \l__iow_line_part_tl
  #1
  { ? \__iow_wrap_break_end:w }
  \s__iow_mark
}

\cs_new:Npn \__iow_wrap_break_first:w ##1 #1 ##2
{\use_none:nn ##2 \__iow_wrap_break_none:w
  \__iow_wrap_break_loop:w ##1 #1 ##2}

\cs_new:Npn \__iow_wrap_break_none:w ##1##2 #1 ##3 \s__iow_mark ##4 #1
{\tl_if_empty:NTF \l__iow_line_tl
  {##2 ##4 \__iow_wrap_line_end:nw {}}
  {\__iow_wrap_line_end:nw {\__iow_wrap_trim:N} ##2 ##4 #1}}

\cs_new:Npn \__iow_wrap_break_loop:w ##1 #1 ##2 #1 ##3
{\use_none:n ##3
  ##1 #1
  \__iow_wrap_break_loop:w ##2 #1 ##3}

\cs_new:Npn \__iow_wrap_break_end:w ##1 #1 ##2 ##3 #1 ##4 \s__iow_mark
{##1 \__iow_wrap_line_end:nw {}}

\exp_args:NV \__iow_tmp:w \c_catcode_other_space_tl
(End of definition for \__iow_wrap_break:w and others.)

\__iow_wrap_next_line:w The special case where the end of a line coincides with the end of a chunk is detected here, to avoid a spurious empty line. Otherwise, call \__iow_wrap_line:nw to find characters for the next line (remembering to account for the indentation).

\cs_new_protected:Npn \__iow_wrap_next_line:w #1#2 \s__iow_stop
{\tl_clear:N \l__iow_line_tl
  \token_if_eq_meaning:NNTF #1 \__iow_wrap_end_chunk:w
  {\tl_clear:N \l__iow_line_part_tl
    \bool_set_true:N \l__iow_line_break_bool
    \__iow_wrap_next:nw {\l__iow_line_target_int}
  }
  \__iow_wrap_line:nw
  {\l__iow_indent_tl}
}
\_\_iow\_wrap\_allow\_break:n

This is called after a chunk has been wrapped. The \_\_iow\_line\_part\_tl typically ends with a space (except at the beginning of a line?), which we remove since the allow-break marker should not insert a space. Then move on with the next chunk, making sure to adjust the target number of characters for the line in case we did remove a space.

\cs_new_protected:Npn \_\_iow\_wrap\_allow\_break:n \#1
\{\tl\set:Nx \l__iow\_line\_tl { \l__iow\_line\_part\_tl \__iow\_wrap\_trim:N \l__iow\_line\_part\_tl }
\bool_set_false:N \l__iow\_line\_break\_bool
\tl_if_empty:NTF \l__iow\_line\_part\_tl
\{ \__iow\_wrap\_chunk:nw {\#1} \}
\{ \exp_args:Nf \__iow\_wrap\_chunk:nw { \int_eval:n { \#1 + 1 } } \}
\}

(End of definition for \_\_iow\_wrap\_allow\_break:n.)

\_\_iow\_wrap\_indent:n \_\_iow\_wrap\_unindent:n

These functions are called after a chunk has been wrapped, when encountering indent/unindent markers. Add the line part (last line part of the previous chunk) to the line so far and reset a boolean denoting the presence of a line-break. Most importantly, add or remove one indent from the current indent (both the integer and the token list). Finally, continue wrapping.

\cs_new_protected:Npn \_\_iow\_wrap\_indent:n \#1
\{\tl\put_right:Ne \l__iow\_line\_tl { \l__iow\_line\_part\_tl }
\bool_set_false:N \l__iow\_line\_break\_bool
\int_add:Nn \l__iow\_indent\_int { \l__iow\_one\_indent\_int }
\tl\put_right:No \l__iow\_indent\_tl { \l__iow\_one\_indent\_tl }
\__iow\_wrap\_chunk:nw {\#1}
\}

\cs_new_protected:Npn \_\_iow\_wrap\_unindent:n \#1
\{\tl\put_right:Ne \l__iow\_line\_tl { \l__iow\_line\_part\_tl }
\bool_set_false:N \l__iow\_line\_break\_bool
\int_sub:Nn \l__iow\_indent\_int { \l__iow\_one\_indent\_int }
\__kernel_tl\set:Nx \l__iow\_indent\_tl { \exp_after:wN \__iow\_unindent:w \l__iow\_indent\_tl }
\__iow\_wrap\_chunk:nw {\#1}
\}

(End of definition for \_\_iow\_wrap\_indent:n and \_\_iow\_wrap\_unindent:n.)

\_\_iow\_wrap\_newline:n \_\_iow\_wrap\_end:n

These functions are called after a chunk has been line-wrapped, when encountering a newline/end marker. Unless we just took a line-break, store the line part and the line so far into the whole \l__iow\_wrap\_tl, trimming a trailing space. In the newline case look for a new line (of length \l__iow\_line\_target\_int) in a new chunk.

\cs_new_protected:Npn \_\_iow\_wrap\_newline:n \#1
\{
First add the last line part to the line, then append it to \l\_iow\_wrap\_tl with the appropriate new line (with “run-on” text), possibly with its last space removed (#1 is empty or \l\_iow\_wrap\_trim:N).

\cs_set_protected:Npn \__iow\_tmp:w #1
\cs_new:Npn \__iow\_wrap\_trim:N ##1
\exp_after:wN \__iow\_wrap\_trim:w ##1 \s\__iow\_mark #1 \s\__iow\_mark \s\__iow\_stop
\cs_new:Npn \__iow\_wrap\_trim:w ##1 #1 \s\__iow\_mark
\cs_new:Npn \__iow\_wrap\_trim\_aux:w ##1 \s\__iow\_mark ##2 \s\__iow\_stop {##1}
\exp_args:NV \__iow\_tmp:w \c\catcode\_other_space_tl

(End of definition for \_iow\_wrap\_newline:n and \_iow\_wrap\_end:n.)

\_iow\_wrap\_store\_do:n

\l\_file\_internal\_tl

Used as a short-term scratch variable.

(End of definition for \_file\_internal\_tl.)

50.4 File operations
The name of the current file should be available at all times: the name itself is set dynamically.
\begin{verbatim}
\str_new:N \g_file_curr_dir_str
\str_new:N \g_file_curr_ext_str
\str_new:N \g_file_curr_name_str
\end{verbatim}

The input list of files is stored as a sequence stack. In package mode we can recover the information from the details held by \LaTeX{}\TeX\,\epsilon{} (we must be in the preamble and loaded using \usepackage or \RequirePackage). As \LaTeX{}\TeX\,\epsilon{} doesn't store directory and name separately, we stick to the same convention here. In pre-loading, \@currnamestack is empty so is skipped.
\begin{verbatim}
\seq_new:N \g__file_stack_seq
\group_begin:
\cs_set_protected:Npn \__file_tmp:w #1#2#3
{\tl_if_blank:nTF {#1}
{ \cs_set:Npn \__file_tmp:w ##1 " ##2 " ##3 \s__file_stop
{ { } {##2} { } }
\seq_gput_right:Ne \g__file_stack_seq
{ \exp_after:wN \__file_tmp:w \tex_jobname:D
" \tex_jobname:D " \s__file_stop
}

} \seq_gput_right:Nn \g__file_stack_seq \{ { } {#1} {#2} \}
\__file_tmp:w
\seq_if_exist:NT \@currnamestack
{ \tl_if_empty:NF \@currnamestack
{ \exp_after:wN \__file_tmp:w \@currnamestack \}
\seq_gput_right:Ne \g__file_stack_seq \{ { } {#1} {#2} \}
\__file_tmp:w
\cs_if_exist:NT \@currnamestack
{ \tl_if_empty:NF \@currnamestack
{ \exp_after:wN \__file_tmp:w \@currnamestack \}
\group_end:
\end{verbatim}

The total list of files used is recorded separately from the current file stack, as nothing is ever popped from this list. The current file name should be included in the file list! We will eventually copy the contents of \@filelist.
\begin{verbatim}
\seq_new:N \g__file_record_seq
\end{verbatim}

For storing the basename and full path whilst passing data internally.
\begin{verbatim}
\tl_new:N \l__file_base_name_tl
\tl_new:N \l__file_full_name_tl
\end{verbatim}
\l__file_dir_str  Used in parsing a path into parts: in contrast to the above, these are never used outside
of the current module.
\l__file_ext_str
\l__file_name_str
\str_new:N \l__file_dir_str
\str_new:N \l__file_ext_str
\str_new:N \l__file_name_str

(End of definition for \l__file_dir_str, \l__file_ext_str, and \l__file_name_str.)
\l_file_search_path_seq  The current search path.
\seq_new:N \l_file_search_path_seq

(End of definition for \l_file_search_path_seq. This variable is documented on page 100.)
\l__file_tmp_seq  Scratch space for comma list conversion.
\seq_new:N \l__file_tmp_seq

(End of definition for \l__file_tmp_seq.)

50.4.1 Internal auxiliaries
\s__file_stop  Internal scan marks.
\scan_new:N \s__file_stop

(End of definition for \s__file_stop.)
\q__file_nil  Internal quarks.
\quark_new:N \q__file_nil

(End of definition for \q__file_nil.)
\__file_quark_if_nil_p:n  Branching quark conditional.
\__file_quark_if_nil:nTF
\__kernel_quark_new_conditional:Nn \__file_quark_if_nil:n { TF }

(End of definition for \__file_quark_if_nil:nTF.)
\q__file_recursion_tail \q__file_recursion_stop  Internal recursion quarks.
\quark_new:N \q__file_recursion_tail
\quark_new:N \q__file_recursion_stop

(End of definition for \q__file_recursion_tail and \q__file_recursion_stop.)
\__file_if_recursion_tail_break:NN \__file_if_recursion_tail_stop_do:Nn
\__kernel_quark_new_test:N \__file_if_recursion_tail_break:NN \__file_if_recursion_tail_stop:NN
\__kernel_quark_new_test:N \__file_if_recursion_tail_stop_do:nn

(End of definition for \__file_if_recursion_tail_break:NN and \__file_if_recursion_tail_stop_-
do:Nn.)
Expanding the file name uses a `\csname`-based approach, and relies on active characters (for example from UTF-8 characters) being properly set up to expand to an expansion-safe version using `\ifsfn`. This is less conservative than the token-by-token approach used before, but it is much faster.

```
\cs_new:Npn \__kernel_file_name_sanitize:n #1
\{ \exp_args:Ne \__file_name_trim_spaces:n { \exp_args:Ne \__file_name_strip_quotes:n { \__file_name_expand:n {#1} } } \}
```

We’ll use `\cs:w` to start expanding the file name, and to avoid creating csnames equal to `\relax` with “common” names, there’s a prefix `__file_name=` to the csname. There’s also a guard token at the end so we can check if there was an error during the process and (try to) clean up gracefully.

```
\cs_new:Npn \__file_name_expand:n #1
\{ \exp_after:wN \__file_name_expand_cleanup:Nw \cs:w __file_name = #1 \cs_end: \__file_name_expand_end: \}
```

With the csname built, we grab it, and grab the remaining tokens delimited by `\__file_name_expand_end:`. If there are any remaining tokens, something bad happened, so we’ll call the error procedure `\__file_name_expand_error:Nw`. If everything went according to plan, then use `\token_to_str:N` on the csname built, and call `\__file_name_expand_cleanup:w` to remove the prefix we added a while back. `\__file_name_expand_cleanup:w` takes a leading argument so we don’t have to bother about the value of `\tex_escapechar:D`.

```
\cs_new:Npn \__file_name_expand_cleanup:Nw #1 #2 \__file_name_expand_end:
\{ \tl_if_empty:nF {#2} \{ \__file_name_expand_error:Nw #2 \__file_name_expand_end: \} \exp_after:wN \__file_name_expand_cleanup:w \token_to_str:N #1 \}
\exp_last_unbraced:NNNNo
```

In non-error cases `\__file_name_expand_end:` should not expand. It will only do so in case there is a `\csname` too much in the file name, so it will throw an error (while expanding), then insert the missing `\cs_end:` and yet another `\__file_name_expand_end:` that will be used as a delimiter by `\__file_name_expand_cleanup:Nw` (or that will expand again if yet another `\endcsname` is missing).

```
\cs_new:Npn \__file_name_expand_end:
\{ \msg_expandable_error:nn { kernel } { filename-missing-endsname } \cs_end: \__file_name_expand_end:
\}
```
Now to the error case. \__file_name\_expand\_error: \Nw adds an extra \cs\_end: so that in case there was an extra \cs\_name in the file name, then \__file_name\_expand\_error\_aux: \Nw throws the error.

```
1099 \cs\_new:Npn \__file_name\_expand\_error: \Nw \#1 \#2 \__file_name\_expand\_end: 
1100 \{ \__file_name\_expand\_error\_aux: \Nw \#1 \#2 \cs\_end: \__file_name\_expand\_end: \}
1101 \cs\_new:Npn \__file_name\_expand\_error\_aux: \Nw \#1 \#2 \cs\_end: \#3
1102 \__file_name\_expand\_end: 
1103 \{ 
1104 \msg\_expandable\_error:nnff 
1105 \{ kernel \} \{ filename\_chars\_lost \} 
1106 \{ \token\_to\_str:N \#1 \} \{ \exp\_stop\_f: \#2 \}
1107 \}
```

Quoting file name uses basically the same approach as for \luaquotejobname: count the " tokens and remove them.

```
1108 \cs\_new:Npn \__file\_name\_strip\_quotes:n \#1 
1109 \{ 
1110 \__file\_name\_strip\_quotes:nw \{ 0 \} \#1 " \q\_file\_recursion\_tail " \q\_file\_recursion\_stop \{\#1\}
1111 \}
1112 \cs\_new:Npn \__file\_name\_strip\_quotes:nw \#1\#2 " 
1113 \{ 
1114 \if\_meaning:w \q\_file\_recursion\_tail \#2 
1115 \__file\_name\_strip\_quotes\_end:wnwn 
1116 \fi: \#2 
1117 \__file\_name\_strip\_quotes:nw \{ \#1 + 1 \}
1118 \}
1119 \cs\_new:Npn \__file\_name\_strip\_quotes\_end:wnwn \fi: \#1 
1120 \__file\_name\_strip\_quotes:nw \#2 \q\_file\_recursion\_stop \#3
1121 \{ 
1122 \fi: 
1123 \int\_if\_odd:nT \#2 
1124 \{ 
1125 \msg\_expandable\_error:nnn 
1126 \{ kernel \} \{ unbalanced\_quote\_in\_filename \} \{\#3\} 
1127 \}
1128 \}
```

Spaces need to be trimmed from the start of the name and from the end of any extension. However, the name we are passed might not have an extension: that means we have to look for one. If there is no extension, we still use the standard trimming function but deliberately prevent any spaces being removed at the end.

```
1131 \cs\_new:Npn \__file\_name\_trim\_spaces:n \#1 
1132 \{ \__file\_name\_trim\_spaces:nw \{\#1\} \#1 . \q\_file\_nil . \s\_file\_stop \}
1133 \cs\_new:Npn \__file\_name\_trim\_spaces:nw \#1\#2 \#3 \#4 \s\_file\_stop 
1134 \{ 
1135 \__file\_quark\_if\_nil:nTF \#3 
1136 \{ 
1137 \tl\_trim\_spaces\_apply:nW \{ \#1 \s\_file\_stop \}
1138 \__file\_name\_trim\_spaces\_aux:n 
1139 \} 
1140 \{ \tl\_trim\_spaces:n \{\#1\} \}
1141 \}
```
\cs_new:Npn \__file_name_trim_spaces_aux:n #1
\cs_new:Npn \__file_name_trim_spaces_aux:w #1 \s__file_stop {#1}

(End of definition for \__kernel_file_name_sanitize:n and others.)

\cs_new:Npn \__kernel_file_name_quote:n #1
\cs_new:Npn \__file_name_quote:nw
\__kernel_file_name_quote:n #1
\__file_name_quote:nw {#1} #1 ~ \q__file_nil \s__file_stop }

(End of definition for \__kernel_file_name_quote:n and \__file_name_quote:nw.)

\c__file_marker_tl
The same idea as the marker for rescanning token lists: this pair of tokens cannot appear
in a file that is being input.
\tl_const:Ne \c__file_marker_tl { : \token_to_str:N : }

(End of definition for \c__file_marker_tl.)

\file_get:nnN \file_get:VnN
\file_get:nnN \file_get_aux:nnN \file_get_do:Nw
The approach here is similar to that for \tl_set_rescan:Nnn. The file contents are
grabbed as an argument delimited by \c__file_marker_tl. A few subtleties: braces in
\if_false: ... \fi: to deal with possible alignment tabs, \tracingnesting to avoid
a warning about a group being closed inside the \scantokens, and \prg_return_true:
is placed after the end-of-file marker.
\cs_new_protected:Npn \file_get:nnN #1#2#3
\cs_generate_variant:Nn \file_get:nnN { V }
\prg_new_protected_conditional:Nnn \file_get:nnN #1#2#3 { T , F , TF }
\prg_generate_conditional_variant:Nnn \file_get:nnN { V } { T , F , TF }
\cs_new_protected:Npe \__file_get_aux:nnN \__file_get_do:Nw
\prg_generate_conditional_variant:Nnn \file_get:nnN \file_get:nnN { V } { T , F , TF }
\exp_not:N \if_false: { \exp_not:N \fi:
\group_begin:
\\exp_not:N \if_false: { \exp_not:N \fi:
665
\begin{verbatim}
\__file_size:n
\__file_full_name:n
\__file_full_name_aux:n
\__file_full_name_auxi:n
\__file_full_name_auxii:n
\__file_full_name_auxiiw:n
\__file_full_name_auxi:j
\__file_full_name_auxii:j
\__file_name_cleanup:w
\__file_name_ext_check:nn
\__file_name_ext_check:nnw
\__file_name_ext_check:nnnw
\__file_name_ext_check:nnn
\__file_name_ext_check:nnnn

File searching can be carried out if the \texttt{pdffilesize} primitive or an equivalent is available. That of course means we need to arrange for everything else to here to be done by expansion too. We start off by sanitizing the name and quoting if required: we may need to remove those quotes, so the raw name is passed too.

\cs_new:Npn \file_full_name:n #1
  { \exp_args:Ne \__file_full_name:n { \__kernel_file_name_sanitize:n {#1} } }

First, we check of the file is just here: no mapping so we do not need the break part of the broader auxiliary. We are using the fact that the primitive here returns nothing if the file is entirely absent. To avoid unnecessary filesystem lookups, the result of \texttt{pdffilesize} is kept available as an argument. For package mode, \texttt{\input@path} is a token list not a sequence.

\cs_new:Npn \__file_full_name_aux:n #1
  { \tl_if_blank:nF {#1} { \exp_args:Nne \__file_full_name_auxii:nn {#1} { \__file_full_name_aux:n {#1} } } }

To avoid repeated reading of files we need to cache the loading: this is important as the code here is used by \texttt{all} file checks. The same marker is used in the \texttt{kpX2e} kernel, meaning that we get a double-saving with for example \texttt{IfFileExists}. As this is all about performance, we use the low-level approach for the conditionals. For a file already seen, the size is reported as \texttt{-1} so it’s distinct from any non-cached ones.

\cs_new:Npn \__file_full_name_aux:n #1
  { \tl_if_blank:nF {#1} { \exp_args:Nne \__file_full_name_auxii:nn {#1} { \__file_full_name_aux:n {#1} } } }

\end{verbatim}

(End of definition for \texttt{\file_get:nnNTF} and others. These functions are documented on page 103.)
We will need the size of files later, and we have to avoid the \scan_stop: causing issues if we are raising the flag. Thus there is a slightly odd gobble here.

Two pars to the auxiliary here so we can avoid doing quoting twice in the event we find the right file.
As TeX automatically adds .tex if there is no extension, there is a little clean up to do here. First, make sure we are not in the directory part, saving that. Then check for an extension.

```latex
\cs_new:Npn \__file_name_cleanup:w #1 \__file_name_end: { }
\cs_new:Npn \__file_name_end: { }
\exp_args:No \__file_ext_check:nnnw
{ \use_none:n #2 } #1 #3 . \q__file_nil . \s__file_stop
\cs_new:Npe \__file_ext_check:nnnw #1#2#3#4 . #5 . #6 \s__file_stop
\exp_not:N \__file_quark_if_nil:nTF {#5}
{ #1 #3 \tl_to_str:n { .tex } } #1 #3 } {#2}
\cs_new:Npn \__file_ext_check:nnn #1
{ \exp_args:Nne \__file_ext_check:nnnn {#1} { 
\__file_full_name_aux:n {#1} } }
\cs_new:Npn \__file_ext_check:nnnn #1#2#3#4
{ \tl_if_blank:nTF {#2}
#3
{ \bool_lazy_or:nnTF
{ \int_compare_p:nNn {#4} = {#2} }
{ \int_compare_p:nNn {#2} = {-1} }
{#1}
{#3}
}
\endinput
(End of definition for \file_full_name:n and others. This function is documented on page 102.)
```

These functions pre-date using \texttt{\tex_filesize:D} for file searching, so are get functions with protection. To avoid having different search set ups, they are simply wrappers.
around the code above.

```latex
\cs_new_protected:Npn \file_get_full_name:nN #1#2
\{
\file_get_full_name:nNF {#1} #2
\{ \tl_set:Nn #2 { \q_no_value } \}
\}
\cs_generate_variant:Nn \file_get_full_name:nN { V }
\prg_new_protected_conditional:Nppnn \file_get_full_name:nN #1#2 { T , F , TF }
\{
\__kernel_tl_set:Nx #2
\{ \file_full_name:n {#1} \}
\tl_if_empty:NTF #2
\{ \prg_return_false: \}
\{ \prg_return_true: \}
\}
\prg_generate_conditional_variant:Nnn \file_get_full_name:nN { V } { T , F , TF }
```

(End of definition for \file_get_full_name:nN, \file_get_full_name:nNTF, and \__file_get_full_name_search:nN. These functions are documented on page 102.)

\g__file_internal_ior A reserved stream to test for opening a shell.

\ior_new:N \g__file_internal_ior

(End of definition for \g__file_internal_ior.)

\file_mdfive_hash:n Getting file details by expansion is relatively easy if a bit repetitive. As the MD5 function has a slightly different syntax from the other commands, there is a little cleaning up to do.

\cs_new:Npn \file_size:n #1
\{ \__file_details:nn {#1} { size } \}
\cs_generate_variant:Nn \file_size:n { V }
\cs_new:Npn \file_timestamp:n #1
\{ \__file_details:nn {#1} { moddate } \}
\cs_generate_variant:Nn \file_timestamp:n { V }
\cs_new:Npn \__file_details:nn #1#2
\{ \exp_args:Ne \__file_details_aux:nn
\{ \file_full_name:n {#1} \} {#2} \}
\}
\cs_new:Npn \__file_details_aux:nn #1#2
\{ \tl_if_blank:nF {#1}
\{ \use:c { tex_file #2 :D } {#1} \}
\}
\cs_new:Npn \__file_mdfive_hash:n #1
\{ \exp_args:Ne \__file_mdfive_hash:n
\{ \file_full_name:n {#1} \} {#2} \}
\}
\cs_new:Npn \__file_mdfive_hash:n #1
\{ \exp_args:Ne \__file_mdfive_hash:n
\{ \file_full_name:n {#1} \} {#2} \}
\cs_generate_variant:Nn \file_mdfive_hash:n { V }
\cs_new:Npn \__file_mdfive_hash:n #1
\{ \tex_mdfivesum:D file {#1} \}
```

(End of definition for \file_mdfive_hash:n and others. These functions are documented on page 101.)

669
These are separate as they need multiple arguments or the file size. For LuaTeX, the emulation does not need the file size so we save a little on expansion.

\file_hex_dump:nnn
\file_hex_dump:Vnn
\__file_hex_dump_auxi:nnn
\__file_hex_dump_auxii:nnnn
\__file_hex_dump_auxiii:nnnn
\__file_hex_dump_auxiv:nnnn
\file_hex_dump:n
\file_hex_dump:V
\__file_hex_dump:n
\cs_new:Npn \file_hex_dump:nnn #1#2#3
\exp_args:Neee \__file_hex_dump_auxi:nnn
\file_full_name:n {#1}
\int_eval:n {#2}
\int_eval:n {#3}
\cs_generate_variant:Nn \file_hex_dump:nnn { V }
\cs_new:Npn \__file_hex_dump_auxi:nnn #1#2#3
\bool_lazy_any:nF
{ \tl_if_blank_p:n {#1} }
{ \int_compare_p:NnNn {#2} = 0 }
{ \int_compare_p:NnNn {#3} = 0 }
\exp_args:Ne \__file_hex_dump_auxii:nnnn
\__file_details_aux:nn {#1} { size }
{#1} {#2} {#3}
\cs_new:Npn \__file_hex_dump_auxii:nnnn #1#2#3#4
\int_compare:nNnTF {#3} > 0
\__file_hex_dump_auxiii:nnnn {#3}
\exp_args:Ne \__file_hex_dump_auxiii:nnnn
\int_eval:n {#1 + #3}
{#1} {#2} {#4}
\cs_new:Npn \__file_hex_dump_auxiii:nnnn #1#2#3#4
\int_compare:nNnTF {#4} > 0
\__file_hex_dump_auxiv:nnn {#4}
\exp_args:Ne \__file_hex_dump_auxiv:nnn
\int_eval:n {#2 + #4}
{#1} {#3}
\cs_new:Npn \__file_hex_dump_auxiv:nnn #1#2#3
\tex_filedump:D
\int_eval:n {#2 - 1} -
\int_eval:n {#1 - #2 + 1}
{#3}
\cs_new:Npn \file_hex_dump:n #1
\exp_args:Ne \__file_hex_dump:n { \file_full_name:n {#1} }
\cs_generate_variant:Nn \file_hex_dump:n { V }
Non-expandable wrappers around the above in the case where appropriate primitive support exists.

\cs_new_protected:Npn \file_get_hex_dump:nN #1#2
\{ \file_get_hex_dump:nNF {#1} #2 { \tl_set:Nn #2 { \q_no_value } } \}
\cs_generate_variant:Nn \file_get_hex_dump:nN { V }
\cs_new_protected:Npn \file_get_mdfive_hash:nN #1#2
\{ \file_get_mdfive_hash:nNF {#1} #2 { \tl_set:Nn #2 { \q_no_value } } \}
\cs_generate_variant:Nn \file_get_mdfive_hash:nN { V }
\cs_new_protected:Npn \file_get_size:nN #1#2
\{ \file_get_size:nNF {#1} #2 { \tl_set:Nn #2 { \q_no_value } } \}
\cs_generate_variant:Nn \file_get_size:nN { V }
\cs_new_protected:Npn \file_get_timestamp:nN #1#2
\{ \file_get_timestamp:nNF {#1} #2 { \tl_set:Nn #2 { \q_no_value } } \}
\cs_generate_variant:Nn \file_get_timestamp:nN { V }
\prg_new_protected_conditional:Npnn \file_get_hex_dump:nN #1#2 { T , F , TF }
\{ \__file_get_details:nnN {#1} { hex_dump } #2 \}
\prg_generate_conditional_variant:Nnn \file_get_hex_dump:nN
{ V } { T , F , TF }
\prg_new_protected_conditional:Npnn \file_get_mdfive_hash:nN #1#2 { T , F , TF }
\{ \__file_get_details:nnN {#1} { mdfive_hash } #2 \}
\prg_generate_conditional_variant:Nnn \file_get_mdfive_hash:nN
{ V } { T , F , TF }
\prg_new_protected_conditional:Npnn \file_get_size:nN #1#2 { T , F , TF }
\{ \__file_get_details:nnN {#1} { size } #2 \}
\prg_generate_conditional_variant:Nnn \file_get_size:nN
{ V } { T , F , TF }
\prg_new_protected_conditional:Npnn \file_get_timestamp:nN #1#2 { T , F , TF }
\{ \__file_get_details:nnN {#1} { timestamp } #2 \}
\prg_generate_conditional_variant:Nnn \file_get_timestamp:nN
{ V } { T , F , TF }
\cs_new_protected:Npn \__file_get_details:nnN \#1#2
\{ \__kernel_tl_set:Nx \#2
\{ \use:c { file_ #1 :n } \{#1\} \}
\tl_if_empty:NTF \#3
\{ \prg_return_false: \} \}
Custom code due to the additional arguments.

file_get_hex_dump:nnnN
file_get_hex_dump:VnnN
file_get_hex_dump:nnnNTF
file_get_hex_dump:VnnNTF

As we are doing a fixed-length “big” integer comparison, it is easiest to use the low-level behavior of string comparisons.

__file_str_cmp:nn

Comparison of file date can be done by using the low-level nature of the string comparison functions.

__file_timestamp:n

(End of definition for file_get_hex_dump:nnnNTF. This function is documented on page 100.)
\file_if_exist_p:n \file_if_exist_p:V
\file_if_exist:nTF \file_if_exist:VF
\file_if_exist_input:n \file_if_exist_input:V
\file_if_exist_input:nF \file_if_exist_input:VF

The test for the existence of a file is a wrapper around the function to add a path to a file. If the file was found, the path contains something, whereas if the file was not located then the return value is empty.

\prg_new_conditional:Npnn \file_if_exist:n #1 { p , T , F , TF }
\prg_generate_conditional_variant:Nnn \file_if_exist:n { V } { p , T , F , TF }

(End of definition for \file_compare_timestamp:nNnTF, \__file_compare_timestamp:nnN, and \__file_timestamp:n. This function is documented on page 102.)

\file_if_exist_input:n \file_if_exist_input:V
\file_if_exist_input:nF \file_if_exist_input:VF

Input of a file with a test for existence. We do not define the T or TF variants because the most useful place to place the \texttt{true code} would be inconsistent with other conditionals.

\cs_new_protected:Npm \file_if_exist_input:n #1
\cs_generate_variant:Nn \file_if_exist_input:n { V }
\cs_new_protected:Npm \file_if_exist_input:nF #1#2
\cs_generate_variant:Nn \file_if_exist_input:nF { V }

(End of definition for \file_if_exist:nTF. This function is documented on page 100.)
\file_input_stop:
A simple rename.
\begin{verbatim}
\cs_new_protected:Npn \file_input_stop: { \tex_endinput:D }
\end{verbatim}

__kernel_file_missing:n
An error message for a missing file, also used in \ior_open:Nn.
\begin{verbatim}
\cs_new_protected:Npn \__kernel_file_missing:n #1
\{ \msg_error:nne { kernel } { file-not-found } \{ \__kernel_file_name_sanitze:n {#1} \} \}
\end{verbatim}

\file_input:n
\file_input:V
\__file_input:n
\__file_input:V
\__file_input_push:n
\__kernel_file_input_push:n
\__file_input_pop:
\__kernel_file_input_pop:
\__file_input_pop:nnn

\begin{verbatim}
\cs_new_protected:Npn \file_input:n #1
\{ \file_get_full_name:nNTF {#1} \l__file_full_name_tl
\{ \__file_input:V \l__file_full_name_tl \}
\{ \__kernel_file_missing:n {#1} \}
\}
\end{verbatim}

\begin{verbatim}
\cs_new_protected:Npn \__file_input_push:n #1
\{ \seq_gpush:Ne \g__file_stack_seq
\{ \g_file_curr_dir_str \}
\{ \g_file_curr_name_str \}
\{ \g_file_curr_ext_str \}
\file_parse_full_name:nNNN {#1} \l__file_dir_str \l__file_name_str \l__file_ext_str
\str_gset_eq:NN \g_file_curr_dir_str \l__file_dir_str
\str_gset_eq:NN \g_file_curr_name_str \l__file_name_str
\str_gset_eq:NN \g_file_curr_ext_str \l__file_ext_str
\}
\end{verbatim}

\begin{verbatim}
\cs_new_protected:Npn \__file_input_pop:
\}
\end{verbatim}

\begin{verbatim}
\cs_new_protected:Npn \__file_input:n #1
\{ \exp_not:N \clist_if_exist:NTF \exp_not:N \@filelist
\{ \exp_not:N \@addtofilelist {#1} \}
\{ \seq_gput_right:Nn \exp_not:N \g__file_record_seq {#1} \}
\exp_not:N \__file_input_push:n {#1}
\exp_not:N \sys_if_engine_luatex:TF
\{ {#1} \}
\exp_not:N \__kernel_file_name_quote:n {#1} \scan_stop: \}
\exp_not:N \__file_input_pop:
\}
\end{verbatim}

\begin{verbatim}
\cs_new_protected:Npn \__file_input_pop:nnn
\}
\end{verbatim}

\begin{verbatim}
\cs_new_protected:Npn \__file_input_push:n #1
\{ \exp_not:N \clist_if_exist:NTF \exp_not:N \@filelist
\{ \exp_not:N \@addtofilelist {#1} \}
\{ \exp_not:N \__file_input_push:n {#1} \}
\exp_not:N \sys_if_engine_luatex:TF
\{ {#1} \}
\exp_not:N \__kernel_file_name_quote:n {#1} \scan_stop: \}
\exp_not:N \__file_input_pop:
\}
\end{verbatim}

\begin{verbatim}
\cs_new_protected:Npn \__file_input_pop:
\}
\end{verbatim}

\begin{verbatim}
\cs_new_protected:Npn \__file_input_pop:nnn
\}
\end{verbatim}

Loading a file is done in a safe way, checking first that the file exists and loading only if it does. Push the file name on the \g__file_stack_seq, and add it to the file list, either \g__file_record_seq, or \@filelist in package mode.
\begin{verbatim}
\cs_new_protected:Npn \__file_input:n #1
\{ \exp_not:N \clist_if_exist:NTF \exp_not:N \@filelist
\{ \exp_not:N \@addtofilelist {#1} \}
\{ \exp_not:N \__file_input_push:n {#1} \}
\exp_not:N \sys_if_engine_luatex:TF
\{ {#1} \}
\exp_not:N \__kernel_file_name_quote:n {#1} \scan_stop: \}
\exp_not:N \__file_input_pop:
\}
\end{verbatim}

Keeping a track of the file data is easy enough: we store the separated parts so we do not need to parse them twice.
\file_input_raw:n
\file_input_raw:V
\__file_input_raw:nn
\file_parse_full_name:n
\file_parse_full_name:V
\file_parse_full_name_apply:n
\file_parse_full_name_apply:V
\file_parse_full_name

No error checking, no tracking.

(End of definition for \file_input:n and others. This function is documented on page 103.)

The main parsing macro \file_parse_full_name_apply:n passes the file name #1 through \_kernel_file_name_sanitze:n so that we have a single normalised way to treat files internally. \file_parse_full_name:n uses the former, with \prg_do_nothing: to leave each part of the name within a pair of braces.

(End of definition for \file_input_raw:n and \__file_input_raw:nn. This function is documented on page 103.)
\_\__file\_parse\_full\_name\_area:nw splits the file name into chunks separated by /, until the last one is reached. The last chunk is the file name plus the extension, and everything before that is the path. When \_\__file\_parse\_full\_name\_area:nw is done, it leaves the path within braces after the scan mark \s\_\_file\_stop and proceeds parsing the actual file name.

\_\__file\_parse\_full\_name\_base:nw does roughly the same as above, but it separates the chunks at each period. However here there’s some extra complications: In case #1 is empty, it is assumed that the extension is actually empty, and the file name is #2. Besides, an extra . has to be added to #2 because it is later removed in \_\__file\_parse\_full\_name\_tidy:nnnN. In any case, if there’s an extension, it is returned with a leading ..

Now we just need to tidy some bits left loose before. The loop used in the two macros above start with a leading / and . in the file path an name, so here we need to remove them, except in the path, if it is a single /, in which case it’s left as is. After all’s done, pass to #4.
(End of definition for \texttt{\file_parse_full_name:n} and others. These functions are documented on page 103.)

\texttt{\file_parse_full_name:nNNN}
\texttt{\file_parse_full_name:VNNN}

\begin{verbatim}
cs_new_protected:Npn \file_parse_full_name:nNNN #1 #2 #3 #4}
  { \file_parse_full_name_apply:nN (#1)
  \__file_full_name_assign:nnnNNN #2 #3 #4 }
\cs_new_protected:Npn \__file_full_name_assign:nnnNNN #1 #2 #3 #4 #5 #6
  { \str_set:Nn #4 {#1}
  \str_set:Nn #5 {#2}
  \str_set:Nn #6 {#3}
  }
\cs_generate_variant:Nn \file_parse_full_name:nNNN { V }
\end{verbatim}

(End of definition for \texttt{\file_parse_full_name:nNNN}. This function is documented on page 102.)

A function to list all files used to the log, without duplicates. In package mode, if \texttt{@filelist} is still defined, we need to take this list of file names into account (we capture it \texttt{\AtBeginDocument} into \texttt{\g__file_record_seq}), turning it to a string (this does not affect the commas of this comma list).

\begin{verbatim}
cs_new_protected:Npn \file_show_list: { \__file_list:N \msg_show:nneeee }
\cs_new_protected:Npn \file_log_list: { \__file_list:N \msg_log:nneeee }
\cs_new_protected:Npn \__file_list:N #1
  { \seq_clear:N \l__file_tmp_seq
  \clist_if_exist:NT \@filelist
  { \exp_args:NNe \seq_set_from_clist:Nn \l__file_tmp_seq { \tl_to_str:N \@filelist }
  }
  \seq_concat:NNN \l__file_tmp_seq \l__file_tmp_seq \g__file_record_seq
  \seq_remove_duplicates:N \l__file_tmp_seq
  \seq_map_function:NN \l__file_tmp_seq \__file_list_aux:n
  { } { } { }
  \}
\cs_new:Npn \__file_list_aux:n #1 { \iow_newline: #1 }
\end{verbatim}

(End of definition for \texttt{\file_show_list:} and others. These functions are documented on page 104.)

When used as a package, there is a need to hold onto the standard file list as well as the new one here. File names recorded in \texttt{@filelist} must be turned to strings before being added to \texttt{\g__file_record_seq}.

\begin{verbatim}
cs_if_exist:NT \@filelist
  { \AtBeginDocument
    { }
  }
\end{verbatim}

677
50.5 GetIdInfo

As documented in expl3.dtx this function extracts file name etc from an SVN Id line. This used to be how we got version number and so on in all modules, so it had to be defined in l3bootstrap. Now it’s more convenient to define it after we have set up quite a lot of tools, and l3file seems the least unreasonable place for it.

The idea here is to extract out the information needed from a standard SVN Id line, but to avoid a line that would get changed when the file is checked in. Hence the fact that none of the lines here include both a dollar sign and the Id keyword!

A first check for a completely empty SVN field. If that is not the case, there is a second case when a file created using svn cp but has not been checked in. That leaves a special marker -1 version, which has no further data. Dealing correctly with that is the reason for the space in the line to use __file_id_info_auxii:w.

Here, #1 is Id, #2 is the file name, #3 is the extension, #4 is the version, #5 is the check in date and #6 is the check in time and user, plus some trailing spaces. If #4 is the marker -1 value then #5 and #6 are empty.
Convert an SVN-style date into a \LaTeX\-style one.

\begin{verbatim}
\cs_new_protected:Npn \__file_id_info_auxiii:w #1 - #2 - #3 - #4 \s__file_stop
    { \tl_set:Nn \ExplFileDate { #1/#2/#3 } }
\end{verbatim}

(End of definition for \GetIdInfo and others. This function is documented on page 10.)

## 50.6 Checking the version of kernel dependencies

This function is responsible for checking if dependencies of the \LaTeX\3 kernel match the version preloaded in the \LaTeX\2\epsilon kernel. If versions don’t match, the function attempts to tell why by searching for a possible stray format file.

The function starts by checking that the kernel date is defined, and if not zero is used to force the error route. The kernel date is then compared with the argument requested date (usually the packaging date of the dependency). If the kernel date is less than the required date, it’s an error and the loading should abort.

\begin{verbatim}
\cs_new_protected:Npn \__kernel_dependency_version_check:Nn #1
    { \exp_args:NV \__kernel_dependency_version_check:nn #1 }
\cs_new_protected:Npn \__kernel_dependency_version_check:nn #1
    { \cs_if_exist:NTF \c__kernel_expl_date_tl
        { \exp_args:NV \__file_kernel_dependency_compare:nnn \c__kernel_expl_date_tl {#1} }
        { \__file_kernel_dependency_compare:nnn { 0000-00-00 } {#1} } }
\cs_new_protected:Npn \__file_kernel_dependency_compare:nnn #1 #2 #3
    { \int_compare:nNnT \__file_parse_version:w #1 \s__file_stop < \__file_parse_version:w #2 \s__file_stop
        { \__file_mismatched_dependency_error:nn {#2} {#3} } }
\cs_new_protected:Npn \__file_mismatched_dependency_error:nn #1 #2
    { \exp_args:NNe \ior_shell_open:Nn \g__file_internal_ior
        { \__file_parse_version:w #1 \s__file_stop } <
        \__file_parse_version:w #2 \s__file_stop
        \__file_mismatcheddependency_error:nn {#2} {#3} }
\end{verbatim}

If the versions differ, then we try to give the user some guidance. This function starts by taking the engine name \c_sys_engine_str and replacing \texttt{tex} by \texttt{latex}, then building a command of the form: \texttt{kpswitch --all --engine=\{engine\} \{format\}[-dev].fmt} to query the format files available. A shell is opened and each line is read into a sequence.

\begin{verbatim}
\cs_new_protected:Npn \__file_mismatcheddependency_error:nn #1 #2
    { \exp_args:NNe \ior_shell_open:Nn \g__file_internal_ior
        \__file_parse_version:w #1 - #2 - #3 \s__file_stop {\#1\#2\#3} }
\end{verbatim}

If the versions differ, then we try to give the user some guidance. This function starts by taking the engine name \c_sys_engine_str and replacing \texttt{tex} by \texttt{latex}, then building a command of the form: \texttt{kpswitch --all --engine=\{engine\} \{format\}[-dev].fmt} to query the format files available. A shell is opened and each line is read into a sequence.

\begin{verbatim}
\cs_new_protected:Npn \__file_mismatcheddependency_error:nn #1 #2
    { \exp_args:NNe \ior_shell_open:Nn \g__file_internal_ior
        \__file_parse_version:w #1 - #2 - #3 \s__file_stop {\#1\#2\#3} }
\end{verbatim}

679
kpsewhich --all
--engine = \c_sys_engine_exec_str
\c_space_tl \c_sys_engine_format_str
\bool_lazy_and:nnT
{ \tl_if_exist_p:N \development@branch@name }
{ ! \tl_if_empty_p:N \development@branch@name }
{ -dev } .fmt
\seq_clear:N \l__file_tmp_seq
\ior_map_inline:Nn \g__file_internal_ior
{ \seq_put_right:Nn \l__file_tmp_seq {##1} }
\ior_close:N \g__file_internal_ior
\msg_error:nnnn { kernel } { mismatched-support-file }
{#1} {#2}
And finish by ending the current file.
\tex_endinput:D
)

Now define the actual error message:
\msg_new:nnnn { kernel } { mismatched-support-file }
{ Mismatched-LaTeX-support-files-detected. \ }
{ Loading-'#2'-aborted! }
\c__kernel_expl_date_tl may not exist, due to an older format, so only print the dates when the sentinel token list exists:
\tl_if_exist:NT \c__kernel_expl_date_tl
{ \ \ \ \ The-L3-programming-layer-in-the-LaTeX-format \ 
  is-dated-\c__kernel_expl_date_tl,-but-in-your-TeX-
  tree-the-files-require \ at-least-#1. }
}
{ The sequence containing the format files should have exactly one item: the format file currently being run. If that’s the case, the cause of the error is not that, so print a generic help with some possible causes. If more than one format file was found, then print the list to the user, with appropriate indications of what’s in the system and what’s in the user tree.
\int_compare:nNnTF { \seq_count:N \l__file_tmp_seq } > 1
{ The-cause-seems-to-be-an-old-format-file-in-the-user-tree. \ 
  LaTeX-found-these-files: \ 
  \seq_map_tokens:Nn \l__file_tmp_seq { \ \---\use:n } \ 
  Try-deleting-the-file-in-the-user-tree-then-run-LaTeX-again. }
{ The-most-likely-causes-are: \ 
  \---A-recent-format-generation-failed; \ 
  \---A-stray-format-file-in-the-user-tree-which-needs-
  to-be-removed-or-rebuilt; \ 
  \---You-are-running-a-manually-installed-version-of-#2 \ 
}

680
which is incompatible with the version in \LaTeX. \}

\LaTeX will abort loading the incompatible support files but this may lead to later errors. Please ensure that your \LaTeX format is correctly regenerated.

(End of definition for \_kernel_dependency_version_check:Nn and others.)

## 50.7 Messages

\msg_new:nnnn { kernel } { file-not-found }
\{ File-`#1'-not-found. \}
\{
  The-requested-file-could-not-be-found-in-the-current-directory,-
in-the-\TeX-search-path-or-in-the-LaTeX-search-path.
\}
\msg_new:nnn { kernel } { file-list }
\{
  >-File-List-<
  #1 \}

.............
\}
\msg_new:nnnn { kernel } { filename-chars-lost }
\{ #1-invalid-in-file-name.-Lost:-#2. \}
\{
  There was an invalid token in the file name that caused the characters following it to be lost.
\}
\msg_new:nnnn { kernel } { filename-missing-endcsname }
\{ Missing `\iow_char:N\endcsname' inserted in filename. \}
\{
  The file name had more `\iow_char:N\csname commands than `\iow_char:N\endcsname ones. \LaTeX will add the missing `\iow_char:N\endcsname and try to continue as best as it can.
\}
\msg_new:nnnn { kernel } { unbalanced-quote-in-filename }
\{ Unbalanced-quotes-in-file-name-'#1'. \}
\{
  File names must contain balanced numbers of quotes (`').
\}
\msg_new:nnnn { kernel } { iow-indent }
\{ Only-#1 allows-#2 \}
\{
  The command-#2 can only be used in messages which will be wrapped using-#1.
  \tl_if_empty:nF {#3} { - It was called with argument-`#3'. }
\}

## 50.8 Functions delayed from earlier modules

<@@@=sys>
Detecting the platform on LuaTeX is easy: for other engines, we use the fact that the two common cases have special null files. It is possible to probe further (see package platform), but that requires shell escape and seems unlikely to be useful. This is set up here as it requires file searching.

\begin{verbatim}
\sys_if_engine_luatex:TF
{ \str_const:Nn \c_sys_platform_str { \tex_directlua:D { tex.print(os.type) } } }
{ \file_if_exist:nTF { nul: } }
{ \file_if_exist:nF { /dev/null } }
{ \str_const:Nn \c_sys_platform_str { windows } }
{ \file_if_exist:nT { /dev/null } }
{ \str_const:Nn \c_sys_platform_str { unix } }
}\cs_if_exist:NF \c_sys_platform_str
{ \str_const:Nn \c_sys_platform_str { unknown } }
\end{verbatim}

(End of definition for \c_sys_platform_str. This variable is documented on page 76.)

We can now set up the tests.

\begin{verbatim}
\clist_map_inline:nn { unix , windows }
{ \__file_const:nn { sys_if_platform_ #1 } }
{ \str_if_eq_p:Vn \c_sys_platform_str { #1 } }
\end{verbatim}

(End of definition for \sys_if_platform_unix:TF and \sys_if_platform_windows:TF. These functions are documented on page 76.)
Chapter 51

l3luatex implementation

51.1 Breaking out to Lua

Copies of primitives.
\cs_new_eq:NN \__lua_escape:n \tex_luaescapestring:D
\cs_new_eq:NN \__lua_now:n \tex_directlua:D
\cs_new_eq:NN \__lua_shipout:n \tex_latelua:D
(End of definition for \__lua_escape:n, \__lua_now:n, and \__lua_shipout:n.)
These functions are set up in l3str for bootstrapping: we want to replace them with a “proper” version at this stage, so clean up.
\cs_undefine:N \lua_escape:e
\cs_undefine:N \lua_now:e
\lua_now:n
\lua_now:e
\lua_shipout_e:n
\lua_shipout:n
\lua_escape:n
\lua_escape:e
(End of definition for \lua_now:n and others. These functions are documented on page 105.)
\lua_load_module:n
Wrapper around require’(module)’.
\str_new:N \l__lua_err_msg_str
\cs_new_protected:Npn \lua_load_module:n #1 { \__lua_load_module_p:n { #1 } }
\cs_new_protected:Npn \lua_load_module:e #1 { \exp_not:n { #1 } }
\cs_new_protected:Npn \lua_load_module:n #1 { \__lua_load_module_p:n { #1 } }
\cs_new_protected:Npn \lua_load_module:e #1 { \exp_not:n { #1 } }
\cs_new:Npn \lua_load_module:n #1 { \lua_load_module:e { \exp_not:n { #1 } } }
\cs_new:Npn \lua_load_module:e #1 { \lua_load_module:e { \exp_not:n { #1 } } }
(End of definition for \lua_load_module:n and others. These functions are documented on page 105.)
As with engines other than LuaTeX these have to be macros, we give them the same status in all cases. When LuaTeX is not in use, simply give an error message:

```latex
\sys_if_engine_luatex:F
{\clist_map_inline:nn
{\lua_escape:n, \lua_escape:e,
\lua_now:n, \lua_now:e}
}\clist_map_inline:nn{\cs_gset:Npn \lua_shipout_e:n, \lua_shipout:n, \lua_load_module:n}
\clist_map_inline:nn{msg_expandable_error:nnn lualatex lualatex-required \#1}
\clist_map_inline:nn{\cs_gset_protected:Npn \lua_shipout_e:n, \lua_shipout:n, \lua_load_module:n}
\clist_map_inline:nn{msg_error:nnn lualatex lualatex-required \#1}
}
```

51.2 Messages

```latex
\msg_new:nnnn lualatex lualatex-required
{LuaTeX-engine-not-in-use!-Ignoring-\#1.}
\msg_new:nnnn lualatex module-not-found
{The-file-`\#1.lua'-could-not-be-found.-Please-ensure-
that-the-file-was-properly-installed-and-that-the-
filename-database-is-current. \& \&
The-Lua-loader-provided-this-additional-information:
}`
\prop_gput:Nnn \g_msg_module_name_prop lualatex LaTeX
\prop_gput:Nnn \g_msg_module_type_prop lualatex }
```

51.3 Lua functions for internal use
Most of the emulation of pdfTeX here is based heavily on Heiko Oberdiek’s `pdftexcmds` package.

**ltx.utils** Create a table for the kernel’s own use.

```
ltx = ltx or {utils={}}
ltx.utils = ltx.utils or { }
local ltxutils = ltx.utils
```

(End of definition for ltx.utils. This function is documented on page 106.)

Local copies of global tables.

```
local io = io
local kpse = kpse
local lfs = lfs
local math = math
local md5 = md5
local os = os
local string = string
local tex = tex
local texio = texio
```

Local copies of standard functions.

```
local abs = math.abs
local byte = string.byte
local floor = math.floor
local format = string.format
local gsub = string.gsub
local lfs_attr = lfs.attributes
local open = io.open
local os_date = os.date
local setcatcode = tex.setcatcode
local sprint = tex.sprint
local cprint = tex.cprint
local write = tex.write
local write_nl = texio.write_nl
local utf8_char = utf8.char
local package_loaded = package.loaded
local package_searchers = package.searchers
local table_concat = table.concat
```

```
local scan_int = token.scan_int or token.scan_integer
local scan_string = token.scan_string
local scan_keyword = token.scan_keyword
local put_next = token.put_next
local token_create = token.create
local token_new = token.new
local set_macro = token.set_macro
```

Since `token.create` only returns useful values after the tokens has been added to TeX’s hash table, we define a variant which defines it first if necessary.

```
local token_create_safe
local is_defined = token.is_defined
local set_char = token.set_char
```
local runtoks = tex.runtoks
local let_token = token_create('let')

function token_create_safe(s)
  local orig_token = token_create(s)
  if is_defined(s, true) then
    return orig_token
  end
  set_char(s, 0)
  local new_token = token_create(s)
  runtoks(function()
    put_next(let_token, new_token, orig_token)
  end)
  return new_token
end

token.command_id

local true_tok = token_create_safe('prg_return_true:,')
local false_tok = token_create_safe('prg_return_false:,')

In ConTeXt lmtx token.command_id does not exist, but it can easily be emulated with
ConTeXt’s tokens.commands.

local command_id = token.command_id
if not command_id and tokens and tokens.commands then
  local id_map = tokens.commands
  function command_id(name)
    return id_map[name]
  end
end

Deal with ConTeXt: doesn’t use kpse library.
local kpse_find = (resolvers and resolvers.findfile) or kpse.find_file

escapehex An internal auxiliary to convert a string to the matching hex escape. This works on a byte
basis: extension to handled UTF-8 input is covered in pdftexcmds but is not currently
required here.

local function escapehex(str)
  return (gsub(str, ".", function (ch) return format("%02X", byte(ch)) end))
end
(End of definition for escapehex.)

ltx.utils.filedump Similar comments here to the next function: read the file in binary mode to avoid any
line-end weirdness.

local function filedump(name,offset,length)
  local file = kpse_find(name,"tex",true)
  if not file then return end
  local f = open(file,"rb")
  if not f then return end
  if offset and offset > 0 then
    f:seek("set", offset)
  end
  local data = f:read(length or 'a')
  f:close()

return escapehex(data)
end
ltxutils.filedump = filedump

(End of definition for ltx.utils.filedump. This function is documented on page 106.)

md5.HEX
Hash a string and return the hash in uppercase hexadecimal format. In some engines, this is built-in. For traditional Lua\TeX, the conversion to hexadecimal has to be done by us.

local md5_HEX = md5.HEX
if not md5_HEX then
  local md5_sum = md5.sum
  function md5_HEX(data)
    return escapehex(md5_sum(data))
  end
  md5.HEX = md5_HEX
end

(End of definition for md5.HEX.)

ltx.utils.filemd5sum
Read an entire file and hash it: the hash function itself is a built-in. As Lua is byte-based there is no work needed here in terms of UTF-8 (see \texttt{pdftexcmds} and how it handles strings that have passed through Lua\TeX). The file is read in binary mode so that no line ending normalisation occurs.

local function filemd5sum(name)
  local file = kpse_find(name, "tex", true) if not file then return end
  local f = open(file, "rb") if not f then return end
  local data = f:read("*a")
  f:close()
  return md5_HEX(data)
end
ltxutils.filemd5sum = filemd5sum

(End of definition for ltx.utils.filemd5sum. This function is documented on page 106.)

ltx.utils.filemoddate
There are two cases: If the C standard library is C99 compliant, we can use \texttt{%z} to get the timezone in almost the right format. We only have to add primes and replace a zero or missing offset with \Z.

Of course this would be boring, so Windows does things differently. There we have to manually calculate the offset. See procedure \texttt{makepdftime} in \texttt{utils.c} of pdftex\).

local filemoddate
if os_date(’%z’:match’^[+-]%d%d%d%d$’ then
  local pattern = lpeg.Cs(16 *
  * -1)
  function filemoddate(name)
    local file = kpse_find(name, "tex", true)
    if not file then return end
    local date = lfs_attr(file, "modification")
    if not date then return end
    return pattern:match(os_date("D:\%Y%m%d%H%M%S\z", date))
local function filemoddate(name)
    local file = kpse_find(name, "tex", true)
    if not file then return end
    local date = lfs_attr(file, "modification")
    if not date then return end
    local d = os_date("*t", date)
    local u = os_date("!*t", date)
    local off = 60 * (d.hour - u.hour) + d.min - u.min
    if d.year ~= u.year then
        if d.year > u.year then
            off = off + 1440
        else
            off = off - 1440
        end
    elseif d.yday ~= u.yday then
        if d.yday > u.yday then
            off = off + 1440
        else
            off = off - 1440
        end
    end
    local timezone
    if off == 0 then
        timezone = "Z"
    elseif off < 0 then
        timezone = "-"
        off = -off
    else
        timezone = "+
    end
    timezone = format("%s%02d:%02d", timezone, hours // 60, hours % 60)
    return format("D:%04d%02d%02d%02d%02d%02d%s", d.year, d.month, d.day, d.hour, d.min, d.sec, timezone)
end

ltxutils.filemoddate = filemoddate

(End of definition for ltx.utils.filemoddate. This function is documented on page 106.)

local function filesize(name)
    local file = kpse_find(name, "tex", true)
    if file then
        local size = lfs_attr(file, "size")
        if size then
            return size
        end
    end
end
ltxutils.filesize = filesize

ltx.utils.filesize
A simple disk lookup.
luadef  An internal function for defining control sequences form Lua which behave like primitives. This acts as a wrapper around \texttt{token.setlua} which accepts a function instead of an index into the functions table.

```lua
local luacmd do
    local set_lua = token.set_lua
    local undefined_cs = command_id'undefined_cs'

    if not context and not luatexbase then require'ltluatex' end
    if luatexbase then
        local new_luafunction = luatexbase.new_luafunction
        local functions = lua.get_functions_table()
        function luacmd(name, func, ...)
            local id
            local tok = token_create(name)
            if tok.command == undefined_cs then
                id = new_luafunction(name)
                set_lua(name, id, ...)
            else
                id = tok.index or tok.mode
            end
            functions[id] = func
        end
    elseif context then
        local register = context.functions.register
        local functions = context.functions.known
        function luacmd(name, func, ...)
            local tok = token_create(name)
            if tok.command == undefined_cs then
                token.set_lua(name, register(func), ...)
            else
                functions[tok.index or tok.mode] = func
            end
        end
    end
end
```

(End of definition for \texttt{luadef}.)

try_require  Loads a Lua module. This function loads the module similarly to the standard Lua global function \texttt{require}, with a few differences. On success, \texttt{try_require} returns \texttt{true, module}. If the module cannot be found, it returns \texttt{false, err_msg}. If the module is found, but something goes wrong when loading it, the function throws an error.

```lua
local function try_require(name)
    if package_loaded[name] then
        return true, package_loaded[name]
    end

    local failure_details = {}
    for _, searcher in ipairs(package_searchers) do
        local loader, data = searcher(name)
        if type(loader) == 'function' then
            package_loaded[name] = loader(name, data) or true
        end
    end
end
```

(End of definition for \texttt{try_require}.)
return true, package_loaded[name]

elseif type(loader) == 'string' then
  failure_details[#failure_details + 1] = loader
  end

end

return false, table_concat(failure_details, '\n')

(End of definition for try_require.)

\_\_lua\_load\_module\_p:n

Check to see if we can load a module using require. If we can load the module, then we load it immediately. Otherwise, we save the error message in \l_@@_err\_msg\_str.

local char_given = command_id'char\_given'
local c_true_bool = token_create(1, char_given)
local c_false_bool = token_create(0, char_given)
local c_str_cctab = token_create('c\_str\_cctab').mode

luacmd('\_\_lua\_load\_module\_p:n', function()
  local success, result = try_require(scan_string())
  if success then
    set_macro(c_str_cctab, '\l\_\_lua\_err\_msg\_str', '')
    put_next(c_true_bool)
  else
    set_macro(c_str_cctab, '\l\_\_lua\_err\_msg\_str', result)
    put_next(c_false_bool)
  end
end)

(End of definition for \_\_lua\_load\_module\_p:n.)

51.4 Preserving iniTeX Lua data for runs

\{@=lua\}

The Lua state is not dumped when a format is written, therefore any Lua variables filled doing format building need to be restored in order to be accessible during normal runs.

We provide some kernel-internal helpers for this. They will only be available if luatexbase is available. This is not a big restriction though, because ConTeXt (which does not use luatexbase) does not load expl3 in the format.

local register_luadata, get_luadata

if luatexbase then
  local register = token_create'@expl@luadata@bytecode'.index
  if status.ini_version then
    register_luadata
  register_luadata is only available during format generation. It accept a string which uniquely identifies the data object and has to be provided to retrieve it later. Additionally it accepts a function which is called in the pre_dump callback and which has to return a string that evaluates to a valid Lua object to be preserved.

local luadata, luadata_order = {}, {}

function register_luadata(name, func)
if luadata[name] then
    error(format("LaTeX error: data name %q already in use", name))
end
luadata[name] = func
luadata_order[#luadata_order + 1] = func and name
end

(End of definition for register_luadata.)

The actual work is done in pre_dump. The luadata_order is used to ensure that the order is consistent over multiple runs.

luatexbase.add_to_callback("pre_dump", function()
    if next(luadata) then
        local str = "return {
            for i=1, #luadata_order do
                local name = luadata_order[i]
                str = format('%s[%q]=%s, ', str, name, luadata[name]())
            end
        lua.bytecode[register] = assert(load(str .. "")")
    end
    end, "ltx.luadata")
else
    get_luadata
    get_luadata is only available if data should be restored. It accept the identifier which was used when the data object was registered and returns the associated object. Every object can only be retrieved once.

    local luadata = lua.bytecode[register]
    if luadata then
        lua.bytecode[register] = nil
        luadata = luadata()
    end
    function get_luadata(name)
        if not luadata then return end
        local data = luadata[name]
        luadata[name] = nil
        return data
    end
end

(End of definition for get_luadata.)
Chapter 52

\textbf{l3legacy implementation}

\begin{verbatim}
\legacy_if_p:n  A friendly wrapper. We need to use the \texttt{if:w} approach here, rather than testing against \texttt{iftrue/iffalse} as the latter approach fails for primitive conditionals such as \texttt{ifmode}. The \texttt{reverse_if:N} here means that we get a slightly more useful error if the name is undefined.

\prg_new_conditional:Npnn \legacy_if:n #1 { p , T , F , TF } {
  \exp_after:wN \reverse_if:N \cs:w if#1 \cs_end:
  \prg_return_false:
  \else:
  \prg_return_true:
  \fi:
}

(End of definition for \texttt{\legacy_if:nTF}. This function is documented on page 108.)

\legacy_if_set_true:n  A friendly wrapper.
\legacy_if_set_false:n
\legacy_if_gset_true:n
\legacy_if_gset_false:n

\prg_new_conditional:Nn \legacy_if_set:nn { \bool_if:nTF {#2} \legacy_if_set_true:n \legacy_if_set_false:n {#1} }

(End of definition for \texttt{\legacy_if_set_true:n} and others. These functions are documented on page 108.)

\legacy_if_set:nn  A more elaborate wrapper.
\legacy_if_gset:nn

\cs_new_protected:Nn \legacy_if_set:nn #1\{ #2 \}
\prg_new_conditional:Nn \legacy_if_set:nn { \bool_if:nTF {#2} \legacy_if_set_true:n \legacy_if_set_false:n {#1} }
\end{verbatim}
\cs_new_protected:Npn \legacy_if_gset:nn #1#2
\begin{verbatim}
{ \bool_if:nTF {#2} \legacy_if_gset_true:n \legacy_if_gset_false:n
  {#1}
}
\end{verbatim}

(End of definition for \legacy_if_set:nn and \legacy_if_gset:nn. These functions are documented on page 108.)

</package>
Chapter 53

\texttt{\texttt{\texttt{l3tl} implementation}}

A token list variable is a \TeX macro that holds tokens. By using the $\varepsilon$-\TeX primitive \texttt{\unexpanded} inside a \TeX \texttt{\edef} it is possible to store any tokens, including $\#$, in this way.

\subsection*{53.1 Functions}

\texttt{\texttt{\texttt{\_kernel\_tl\_set:Nx}} \texttt{\_kernel\_tl\_gset:Nx}} These two are supplied to get better performance for macros which would otherwise use \texttt{\tl\_set:N} or \texttt{\tl\_gset:N} internally.

\texttt{\texttt{\cs\_new\_eq:NN \_kernel\_tl\_set:Nx \cs\_set\_nopar:Npe}}\texttt{\_kernel\_tl\_gset:Nx \cs\_set\_nopar:Npe}

(End of definition for \texttt{\_kernel\_tl\_set:Nx} and \texttt{\_kernel\_tl\_gset:Nx}.)

\texttt{\texttt{\tl\_new:N}}\texttt{\_kernel\_tl\_set:Nx} Creating new token list variables is a case of checking for an existing definition and doing the definition.

\texttt{\texttt{\cs\_new\_protected:Npm \tl\_new:N \#1}}\texttt{\_kernel\_chk\_if\_free\_cs:N}\texttt{\_kernel\_tl\_gset:Nx} \texttt{\cs\_exp\_not:w \#1}\texttt{\_empty\_tl}

(End of definition for \texttt{\tl\_new:N}. This function is documented on page 110.)

\texttt{\texttt{\_kernel\_tl\_set:Nx}}\texttt{\_kernel\_tl\_gset:Nx} Constants are also easy to generate. They use \texttt{\cs\_gset\_nopar:Npe} instead of \texttt{\_kernel\_tl\_gset:Nx} so that the correct scope checking for \texttt{c}, instead of for \texttt{g}, is applied when \texttt{\debug\_on:n \{ check\_declarations \}} is used. Constant assignment functions are patched specially in \texttt{l3debug} to apply such checks.

\texttt{\texttt{\cs\_new\_protected:Npm \tl\_const:Nn \#1\#2}}\texttt{\_kernel\_chk\_if\_free\_cs:N}\texttt{\_kernel\_tl\_gset:Nx} \texttt{\cs\_gset\_nopar:Npe \#1 \{ \_kernel\_exp\_not:w \#2 \}}\texttt{\_empty\_tl}

(End of definition for \texttt{\tl\_const:Nn}. This function is documented on page 110.)
\tl_clear:N Clearing a token list variable means setting it to an empty value. Error checking is sorted out by the parent function.
\tl_clear:c
\tl_gclear:N \tl_gclear:c

Clearing a token list variable means setting it to an empty value. Error checking is sorted out by the parent function.
\tl_clear_new:N \tl_clear_new:c \tl_gclear_new:N \tl_gclear_new:c

Clearing a token list variable means setting it to an empty value. Error checking is sorted out by the parent function.
\tl_set_eq:NN \tl_set_eq:Nc \tl_set_eq:cN \tl_gset_eq:NN \tl_gset_eq:Nc \tl_gset_eq:cN

For setting token list variables equal to each other. To allow for patching, the arguments have to be explicit. In addition this ensures that a braced second argument will not cause problems.
\tl_concat:NNN \tl_concat:ccc \tl_gconcat:NNN \tl_gconcat:ccc

Concatenating token lists is easy. When checking is turned on, all three arguments must be checked: a token list #2 or #3 equal to \scan_stop: would lead to problems later on.
53.2 Constant token lists

\c_empty_tl

Never full. We need to define that constant before using \tl_new:N.

\c_novalue_tl

A special marker: as we don’t have \char_generate:nn yet, has to be created the old-fashioned way.

\c_space_tl

A space as a token list (as opposed to as a character).

53.3 Adding to token list variables

By using \exp_not:n token list variables can contain # tokens, which makes the token list registers provided by \TeX more or less redundant. The \tl_set:No version is done by hand as it is used quite a lot.
Adding to the left is done directly to gain a little performance.

\begin{verbatim}
\tl_put_left:Nn \tl_put_left:NV \tl_put_left:Nv \tl_put_left:Ne \tl_put_left:No \tl_put_left:Nx
\tl_gput_left:Nn \tl_gput_left:NV \tl_gput_left:Nv \tl_gput_left:Ne \tl_gput_left:No
\end{verbatim}

(End of definition for \tl_set:Nn and \tl_gset:Nn. These functions are documented on page 111.)
The same on the right.

\tl_put_right:Nn
\tl_put_right:NV
\tl_put_right:No
\tl_put_right:Nx
\tl_put_right:cn
\tl_put_right:cv
\tl_put_right:ce
\tl_put_right:co
\tl_put_right:cx

(End of definition for \tl_put_left:Nn and \tl_gput_left:Nn. These functions are documented on page 111.)
\begin{verbatim}
\__kernel_tl_set:Nx #1
{
  \__kernel_exp_not:w \exp_after:wN \{#1\}
  \__kernel_exp_not:w \exp_after:wN \{#2\}
}
\cs_new_protected:Npn \tl_gput_right:Nn #1#2
{ \__kernel_tl_gset:Nx #1 { \__kernel_exp_not:w \exp_after:wN { #1 #2 } } }
\cs_new_protected:Npn \tl_gput_right:NV #1#2
{ \__kernel_tl_gset:Nx #1 \exp_not:V #2 }
\cs_new_protected:Npn \tl_gput_right:Nv #1#2
{ \__kernel_tl_gset:Nx #1 \exp_not:v {#2} }
\cs_new_protected:Npn \tl_gput_right:Ne #1#2
{ \__kernel_tl_gset:Nx #1 \__kernel_exp_not:w \exp_after:wN {#1} \__kernel_exp_not:w \tex_expanded:D { {#2} } }
\cs_new_protected:Npn \tl_gput_right:No #1#2
{ \__kernel_tl_gset:Nx #1 \__kernel_exp_not:w \exp_after:wN {#1} \__kernel_exp_not:w \exp_after:wN {#2} }
\cs_generate_variant:Nn \tl_put_right:Nn { c }
\cs_generate_variant:Nn \tl_gput_right:Nn { c }
\cs_generate_variant:Nn \tl_put_right:NV { c }
\cs_generate_variant:Nn \tl_gput_right:NV { c }
\cs_generate_variant:Nn \tl_put_right:Nv { c }
\cs_generate_variant:Nn \tl_gput_right:Nv { c }
\cs_generate_variant:Nn \tl_put_right:Ne { c }
\cs_generate_variant:Nn \tl_gput_right:Ne { c }
\cs_generate_variant:Nn \tl_put_right:No { c }
\cs_generate_variant:Nn \tl_gput_right:No { c }
\cs_generate_variant:Nn \tl_gput_right:Nn { Nx, cx }
\cs_generate_variant:Nn \tl_gput_right:NV { Nx, cx }
\cs_generate_variant:Nn \tl_gput_right:Nv { Nx, cx }
\cs_generate_variant:Nn \tl_gput_right:Ne { Nx, cx }
\cs_generate_variant:Nn \tl_gput_right:No { Nx, cx }
\end{verbatim}

(End of definition for \texttt{\tl_put_right:Nn} and \texttt{\tl_gput_right:Nn}. These functions are documented on page 112.)

\section{Internal quarks and quark-query functions}

\verb|\q__tl_nil| Internal quarks.
\verb|\q__tl_mark|
\verb|\q__tl_stop|

699
53.5 Reassigning token list category codes

The rescanning code needs a special token list containing the same character (chosen here to be a colon) with two different category codes: it cannot appear in the tokens being rescanned since all colons have the same category code.

\tl_set_rescan:NNn
\tl_set_rescan:NNV
\tl_set_rescan:NNe
\tl_set_rescan:NNo
\tl_set_rescan:Nnx
\tl_set_rescan:cmn
\tl_set_rescan:cmV
\tl_set_rescan:cne
\tl_set_rescan:cnO
\tl_set_rescan:cnx
\tl_gset_rescan:NNn
\tl_gset_rescan:NNV
\tl_gset_rescan:NNe
\tl_gset_rescan:NNo
\tl_gset_rescan:Nnx
\tl_gset_rescan:cmn
\tl_gset_rescan:cmV
\tl_gset_rescan:cne
\tl_gset_rescan:cnO
\tl_gset_rescan:cnx
\tl_rescan:NNn
\tl_rescan:NNV
\tl_rescan:NNo
\tl_rescan:NNx
\_tl_rescan_aux:
\_tl_rescan:NNn
\_tl_rescan:NNV
\_tl_rescan:NNo
\_tl_rescan:NNx

In a group, after some initial setup explained below and the user setup #3 (followed by \scan_stop: to be safe), there is a call to \_tl_set_rescan:NNn. This shared auxiliary defined later distinguishes single-line and multi-line “files”. In the simplest case of multi-line files, it calls (with the same arguments) \_tl_set_rescan_multi:NNn, whose code is included here to help understand the approach. This function rescans its argument \#1, closes the group, and performs the assignment.

One difficulty when rescanning is that \scantokens treats the argument as a file, and without the correct settings a \TeX error occurs:

! File ended while scanning definition of ... 

A related minor issue is a warning due to opening a group before the \scantokens and closing it inside that temporary file; we avoid that by setting \texttt{tracingnesting}. The standard solution to the “File ended” error is to grab the rescanned tokens as a delimited argument of an auxiliary, here \_tl_rescan:NNn, that performs the assignment, then let \TeX “execute” the end of file marker. As usual in delimited arguments we use \texttt{prg_do_-nothing}: to avoid stripping an outer set braces: this is removed by using \texttt{o-expanding assignments}. The delimiter cannot appear within the rescanned token list because it contains twice the same character, with different catcodes.

For \_tl_rescan:NNn we cannot simply call \_tl_set_rescan:NNnn \texttt{prg_do_-nothing: use:n} because that would leave the end-of-file marker after the result of rescanning. If that rescanned result is code that looks further in the input stream for arguments, it would break.

For multi-line files the only subtlety is that \texttt{newlinechar} should be equal to \texttt{endlinechar} because \texttt{newlinechar} characters become new lines and then become
\linechar characters when writing to an abstract file and reading back. This equality is ensured by setting newlinechar equal to \linechar. Prior to this, \linechar is set to -1 if it was 32 (in particular true after \ExplSyntaxOn) to avoid unreasonable line-breaks at every space for instance in error messages triggered by the user setup. Another side effect of reading back from the file is that spaces (catcode 10) are ignored at the beginning of lines, and spaces and tabs (character code 32 and 9) are ignored at the end of lines.

The two if_false:... fi: are there to prevent alignment tabs to cause a change of tabular cell while rescanning. We put the “opening” one after \group_begin: so that if one accidentally f-expands \tl_set_rescan:Nnn braces remain balanced. This is essential in e-type arguments when \expanded is not available.

\begin{verbatim}
\cs_new_protected:Npn \tl_rescan:nn #1#2
\{\tl_set_rescan:Nnn \l__tl_internal_a_tl {#1} {#2}\exp_after:wN \__tl_rescan_aux:\l__tl_internal_a_tl\}
\cs_generate_variant:Nn \tl_rescan:nn { nV }
\cs_new_protected:Npn \__tl_rescan_aux: { \tl_clear:N \l__tl_internal_a_tl }
\cs_new_protected:Npn \tl_set_rescan:Nnn { \__tl_set_rescan:NNnn \tl_set:No }
\cs_new_protected:Npn \tl_gset_rescan:Nnn { \__tl_set_rescan:NNnn \tl_gset:No }
\cs_new_protected:Npn \__tl_set_rescan:NNnn #1#2#3#4
\{\group_begin:\if_false: { \fi:\int_set_eq:NN \tex_tracingnesting:D \c_zero_int\int_compare:nNnT \tex_endlinechar:D = { 32 }\{ \int_set:Nn \tex_endlinechar:D { -1 }\} \int_set_eq:NN \tex_newlinechar:D \tex_endlinechar:D \scan_stop:\exp_args:No \__tl_set_rescan:nNN \tl_to_str:n {#4} #1 #2\if_false: } \fi:\}
\cs_new_protected:Npn \__tl_set_rescan_multi:nNN #1#2#3
\{\tex_everyeof:D \exp_after:wN { \c__tl_rescan_marker_tl }\int_set_eq:NN \tex_tracingnesting:D \c_zero_int\int_compare:nNnT \tex_endlinechar:D = { 32 }\{ \int_set:Nn \tex_endlinechar:D { -1 }\} \int_set_eq:NN \tex_newlinechar:D \tex_endlinechar:D \scan_stop:\exp_args:No \__tl_set_rescan:nNN \tl_to_str:n {#4} \#1 \#2\if_false: \} \fi:\}
\cs_new_protected:Npn \__tl_set_rescan_multi:nNN #1#2#3
\{\\tex_everyeof:D \exp_after:wN { \c__tl_rescan_marker_tl }\int_set_eq:NN \tex_tracingnesting:D \c_zero_int\int_compare:nNnT \tex_endlinechar:D = { 32 }\{ \int_set:Nn \tex_endlinechar:D { -1 }\} \int_set_eq:NN \tex_newlinechar:D \tex_endlinechar:D \scan_stop:\exp_args:No \__tl_set_rescan:nNN \tl_to_str:n {#4} \#1 \#2\if_false: \} \fi:\}
\exp_args:Nno \use:nn
\{ \cs_new:Npn \__tl_set_rescan:nNN \tl_to_str:n {#4} \#1 \#2 \#3 \#4 \#1 \#2 \#3 \#4 \}
\cs_generate_variant:Nn \tl_set_rescan:Nnn { NnV , Nne , c , cnV , cne }
\cs_generate_variant:Nn \tl_set_rescan:Nnn { Nno , Nnx , cno , cnx }
\end{verbatim}
The function \_\_tl_set_rescan:nNN calls \_\_tl_set_rescan_multi:nNN or \_\_tl_set_rescan_single:nNN \{ ' \} depending on whether its argument is a single-line fragment of code/data or is made of multiple lines by testing for the presence of a newline character. If newlinechar is out of range, the argument is assumed to be a single line.

For a single line, no \endlinechar should be added, so it is set to \texttt{-1}, and spaces should not be removed. Trailing spaces and tabs are a difficult matter, as \TeX removes these at a very low level. The only way to preserve them is to rescan not the argument but the argument followed by a character with a reasonable category code. Here, 11 (letter) and 12 (other) are accepted, as these are convenient, suitable for delimiting an argument, and it is very unlikely that none of the ASCII characters are in one of these categories. To avoid selecting one particular character to put at the end, whose category code may have been modified, there is a loop through characters from \texttt{'} (ASCII 39) to ~ (ASCII 127). The choice of starting point was made because this is the start of a very long range of characters whose standard category is letter or other, thus minimizing the number of steps needed by the loop (most often just a single one). If no valid character is found (very rare), fall-back on \_\_tl_set_rescan_multi:nNN.

Otherwise, once a valid character is found (let us use \texttt{'} in this explanation) run some code very similar to \_\_tl_set_rescan_multi:nNN but with \texttt{'} added at both ends of the input. Of course, we need to define the auxiliary \_\_tl_set_rescan_single:NNww on the fly to remove the additional \texttt{'} that is just before :: [by which we mean \c\_\_tl_rescan_marker_tl]. Note that the argument must be delimited by \texttt{'} with the current catcode; this is done thanks to \char_generate:nn. Yet another issue is that the rescanned token list may contain a comment character, in which case the \texttt{'} we expected is not there. We fix this as follows: rather than just :: we set \everyeof to \texttt{:\{\texttt{\langle code1\rangle}\}} \texttt{::{\{\texttt{\langle code2\rangle}\}}} \_\_\_tl_stop. The auxiliary \_\_tl_set_rescan_single:NNww runs the \texttt{o}-expanding assignment, expanding either \texttt{\langle code1\rangle} or \texttt{\langle code2\rangle} before its the main argument \#3. In the typical case without comment character, \texttt{\langle code1\rangle} is expanded, removing the leading \texttt{'} . In the rarer case with comment character, \texttt{\langle code2\rangle} is expanded, calling \_\_tl_set_rescan_single_aux:w, which removes the trailing ::\texttt{\{\texttt{\langle code1\rangle}\}} and the leading \texttt{'}.
53.6 Modifying token list variables

All of the replace functions call \__tl_replace:NnNNnn with appropriate arguments. The first two arguments are explained later. The next controls whether the replacement function calls itself (\__tl_replace_next:w) or stops (\__tl_replace_wrap:w) after the first replacement. Next comes an e-type assignment function \tl_set:Ne or \tl_gset:Ne for local or global replacements. Finally, the three arguments \{\langle pattern\rangle\} \{\langle replacement\rangle\} provided by the user. When describing the auxiliary functions below, we denote the contents of the \langle ti var\rangle by \langle token list\rangle.

\cs_new_protected:Npn \tl_replace_once:Nnn
\cs_new_protected:Npn \tl_replace_once:NVn
\cs_new_protected:Npn \tl_replace_once:NnV
\cs_new_protected:Npn \tl_replace_once:Nen
\cs_new_protected:Npn \tl_replace_once:Nne
\cs_new_protected:Npn \tl_replace_once:Nee
\cs_new_protected:Npn \tl_replace_once:Nxn
\cs_new_protected:Npn \tl_replace_once:Nnx
\cs_new_protected:Npn \tl_greplace_once:Nnn
\cs_new_protected:Npn \tl_greplace_once:NVn
\cs_new_protected:Npn \tl_greplace_once:NnV
\cs_new_protected:Npn \tl_greplace_once:Nen
\cs_new_protected:Npn \tl_greplace_once:Nne
\cs_new_protected:Npn \tl_greplace_once:Nee
\cs_new_protected:Npn \tl_greplace_once:Nxn
\cs_new_protected:Npn \tl_greplace_once:Nnx
\cs_new_protected:Npn \tl_greplace_once:Nxx
\cs_new_protected:Npn \tl_replace_all:Nnn
\cs_new_protected:Npn \tl_replace_all:NVn
\cs_new_protected:Npn \tl_replace_all:NnV
\cs_new_protected:Npn \tl_replace_all:Nen
\cs_new_protected:Npn \tl_replace_all:Nne
\cs_new_protected:Npn \tl_replace_all:Nee
\cs_new_protected:Npn \tl_replace_all:Nxn
\cs_new_protected:Npn \tl_replace_all:Nnx
\cs_new_protected:Npn \tl_replace_all:Nxx
\cs_new_protected:Npn \tl_greplace_all:Nnn
\cs_new_protected:Npn \tl_greplace_all:NVn
\cs_new_protected:Npn \tl_greplace_all:NnV
\cs_new_protected:Npn \tl_greplace_all:Nen
\cs_new_protected:Npn \tl_greplace_all:Nne
\cs_new_protected:Npn \tl_greplace_all:Nee
\cs_new_protected:Npn \tl_greplace_all:Nxn
\cs_new_protected:Npn \tl_greplace_all:Nnx
\cs_new_protected:Npn \tl_greplace_all:Nxx

703
To implement the actual replacement auxiliary \_tl_replace_auxii:nNNnn we need a \texttt{delimiter} with the following properties:

- all occurrences of the \texttt{pattern} \#6 in \langle \texttt{token list} \rangle \langle \texttt{delimiter} \rangle belong to the \texttt{token list} and have no overlap with the \texttt{delimiter},
- the first occurrence of the \texttt{delimiter} in \langle \texttt{token list} \rangle \langle \texttt{delimiter} \rangle is the trailing \texttt{delimiter}.

We first find the building blocks for the \texttt{delimiter}, namely two tokens \langle A \rangle and \langle B \rangle such that \langle A \rangle does not appear in \#6 and \#6 is not \langle B \rangle (this condition is trivial if \#6 has more than one token). Then we consider the delimiters \langle "A" \rangle and \langle "A" \langle A \rangle \langle B \rangle \langle A \rangle \langle B \rangle \text{\ldots} \rangle, for \(n \geq 1\), where \langle A \rangle \langle A \rangle\text{\ldots} denotes \(n\) copies of \langle A \rangle, and we choose as our \texttt{delimiter} the first one which is not in the \texttt{token list}.

Every delimiter in the set obeys the first condition: \#6 does not contain \langle A \rangle hence cannot be overlapping with the \texttt{token list} and the \texttt{delimiter}, and it cannot be within the \texttt{delimiter} since it would have to be in one of the two \langle B \rangle hence be equal to this single token (or empty, but this is an error case filtered separately). Given the particular form of these delimiters, for which no prefix is also a suffix, the second condition is actually a consequence of the weaker condition that the \texttt{delimiter} we choose does not appear in the \texttt{token list}. Additionally, the set of delimiters is such that a \texttt{token list} of \(n\) tokens can contain at most \(O(n^{1/2})\) of them, hence we find a \texttt{delimiter} with at most \(O(n^{1/2})\) tokens in a time at most \(O(n^{3/2})\). Bear in mind that these upper bounds are reached only in very contrived scenarios: we include the case \langle "A" \rangle in the list of delimiters to try, so that the \texttt{delimiter} is simply \texttt{\_tl_replace_auxii:nNNnn} in the most common situation where neither the \texttt{token list} nor the \texttt{pattern} contains \texttt{\_tl_replace_auxii:nNNnn}.

Let us now ahead, optimizing for this most common case. First, two special cases: an empty \texttt{pattern} \#6 is an error, and if \#1 is absent from both the \texttt{token list} \#5
and the \langle pattern \rangle #6 then we can use it as the \langle delimiter \rangle through \_\_tl_replace_auxii:nNNNnn \{#1\}. Otherwise, we end up calling \_\_tl_replace:NnNNNnn repeatedly with the first two arguments \q__tl_mark {?}, \? {??}, \?? {???}, and so on, until #6 does not contain the control sequence #1, which we take as our \langle A \rangle. The argument #2 only serves to collect ? characters for #1. Note that the order of the tests means that the first two are done every time, which is wasteful (for instance, we repeatedly test for the emptiness of #6). However, this is rare enough not to matter. Finally, choose \langle B \rangle to be \q__tl_nil or \q__tl_stop such that it is not equal to #6.

The \_\_tl_replace_auxi:NnmNNnn auxiliary receives \{\langle A \rangle\} and \{\langle A \rangle^{n}(\langle B \rangle)^{n}\} as its arguments, initially with \(n = 1\). If \("\langle A \rangle^{n}(\langle B \rangle)^{n}\)" is in the \langle token list \rangle then increase \(n\) and try again. Once it is not anymore in the \langle token list \rangle we take it as our \langle delimiter \rangle and pass this to the auxii auxiliary.

The auxiliary \_\_tl_replace_auxii:nNNNnn receives the following arguments:

\{(delimiter)\} \(\langle \text{function} \rangle \langle \text{assignment} \rangle\) \langle \text{tl var} \rangle \{(pattern)\} \{(replacement)\}

All of its work is done between \group_align_safe_begin: and \group_align_safe_end: to avoid issues in alignments. It does the actual replacement within #3 #4 {...}, an e-expanding \langle assignment \rangle #3 to the \langle tl var \rangle #4. The auxiliary \_\_tl_replace_next:w is called, followed by the \langle token list \rangle, some tokens including the \langle delimiter \rangle #1, followed by the \langle pattern \rangle #5. This auxiliary finds an argument delimited by #5 (the presence of a trailing #5 avoids runaway arguments) and calls \_\_tl_replace_wrap:w to test whether this #5 is found within the \langle token list \rangle or is the trailing one.
If on the one hand it is found within the \textit{⟨token list⟩}, then \#1 cannot contain the \textit{(delimiter) #1} that we worked so hard to obtain, thus \_\_\_tl_replace_wrap:w gets \#1 as its own argument \#1, and protects it against the \textit{e}-expanding assignment. It also finds \textit{\exp_not:n as \#2} and does nothing to it, thus letting through \textit{\exp_not:n \langle(replacement)⟩} into the assignment. Note that \_\_\_tl_replace_next:w and \_\_\_tl_replace_wrap:w are always called followed by two empty brace groups. These are safe because no delimiter can match them. They prevent losing braces when grabbing delimited arguments, but require the use of \textit{\exp_not:o and \use_none:nn}, rather than simply \textit{\exp_not:n}. Afterwards, \_\_\_tl_replace_next:w is called to repeat the replacement, or \_\_\_tl_replace_wrap:w if we only want a single replacement. In this second case, \#1 is the \textit{(remaining tokens)} in the \textit{(token list)} and \#2 is some \textit{(ending code)} which ends the assignment and removes the trailing tokens \#5 using some \textit{\if_false: \fi:} trickery because \#5 may contain any delimiter.

If on the other hand the argument \#1 of \_\_\_tl_replace_next:w is delimited by the trailing \textit{(pattern) \#5}, then \#1 is “{ } { } \langle(token list) \langle(delimiter) \langle(ending code) \rangle \rangle \langle(replacement)⟩”, hence \_\_\_tl_replace_wrap:w finds “{ } { } \langle(token list) \rangle” as \#1 and the \textit{(ending code)} as \#2. It leaves the \langle(token list) \rangle into the assignment and unbraces the \textit{(ending code)} which removes what remains (essentially the \langle(delimiter) \rangle and \langle(replacement)⟩).

\begin{verbatim}
\cs_new_protected:Npn \_\_\_tl_replace_next:auxii:nNNnn #1#2#3#4#5#6
\group_align_safe_begin:
\cs_set:Npn \_\_\_tl_replace_wrap:w \#1 \#1 \#2
{ \_\_\_kernel_exp_not:w \exp_after:wN { \use_none:nn \#1 } \#2 }
\cs_set:Npe \_\_\_tl_replace_next:w \#1 \#1 #5
{ \exp_not:N \_\_\_tl_replace_wrap:w \#1
\exp_not:n \{ #1 \}
\exp_not:n \{ \exp_not:n \{#6 \}
\exp_not:n \{ #2 \{ \} \} \}
\group_align_safe_end:
\end{verbatim}

\begin{verbatim}
\cs_new_protected:Npn \_\_\_tl_replace_next:auxi:nNNnn #1#2#3#4#5
\group_align_safe_begin:
\cs_set:Npn \_\_\_tl_replace_wrap:w \#1 #1 #2
{ \_\_\_kernel_exp_not:w \exp_after:wN { \use_none:nn \#1 } \#2 }
\cs_set:Npe \_\_\_tl_replace_next:w \#1 #1 #5
{ \exp_after:wN \_\_\_tl_replace_next:aux:w #4 #1
\if_false: { \fi: }
\exp_after:wN \use_none:n \exp_after:wN { \if_false: } \fi:
#5 }
\group_align_safe_end:
\end{verbatim}

\begin{verbatim}
\cs_new:Npn \_\_\_tl_replace_next:aux:w { \_\_\_tl_replace_next:w \{} \}
\cs_new_eq:NN \_\_\_tl_replace_next:auxii: \_\_\_tl_replace_next:auxi: #1
\cs_new_eq:NN \_\_\_tl_replace_next:auxi: \_\_\_tl_replace_next:auxii:
\end{verbatim}

\textit{(End of definition for \_\_\_tl_replace:NNNNnn and others.)}

\texttt{\_\_\_tl_remove_once:NN}

\texttt{\_\_\_tl_remove_once:NV}

\texttt{\_\_\_tl_remove_once:Ne}

\texttt{\_\_\_tl_remove_once:cn}

\texttt{\_\_\_tl_remove_once:cv}

\texttt{\_\_\_tl_gremove_once:NN}

\texttt{\_\_\_tl_gremove_once:Ne}

\texttt{\_\_\_tl_gremove_once:cn}

\texttt{\_\_\_tl_gremove_once:cv}

Removal is just a special case of replacement.
Removal is just a special case of replacement.

<table>
<thead>
<tr>
<th>Removal function</th>
<th>Replacement function</th>
</tr>
</thead>
<tbody>
<tr>
<td>\tl_remove_all:Nn</td>
<td>\tl_remove_all:NV</td>
</tr>
<tr>
<td>\tl_remove_all:NV</td>
<td>\tl_remove_all:Ne</td>
</tr>
<tr>
<td>\tl_remove_all:Nx</td>
<td>\tl_remove_all:cn</td>
</tr>
<tr>
<td>\tl_remove_all:cn</td>
<td>\tl_remove_all:cV</td>
</tr>
<tr>
<td>\tl_remove_all:cV</td>
<td>\tl_remove_all:ce</td>
</tr>
<tr>
<td>\tl_remove_all:ce</td>
<td>\tl_remove_all:cx</td>
</tr>
<tr>
<td>\tl_gremove_all:Nn</td>
<td>\tl_gremove_all:NV</td>
</tr>
<tr>
<td>\tl_gremove_all:NV</td>
<td>\tl_gremove_all:Ne</td>
</tr>
<tr>
<td>\tl_gremove_all:Ne</td>
<td>\tl_gremove_all:cn</td>
</tr>
<tr>
<td>\tl_gremove_all:cn</td>
<td>\tl_gremove_all:cV</td>
</tr>
<tr>
<td>\tl_gremove_all:cV</td>
<td>\tl_gremove_all:ce</td>
</tr>
<tr>
<td>\tl_gremove_all:ce</td>
<td>\tl_gremove_all:cx</td>
</tr>
</tbody>
</table>

These functions check whether the token list in the argument is empty and execute the proper code from their argument(s).

```
\prg_new_conditional:Nppnn \tl_if_empty:N #1 { p , T , F , TF }
\tl_if_empty:cTF
```

The \if:w triggers the expansion of \tl_to_str:n which converts the argument to a string: this is empty if and only if the argument is. Then \if:w: \scan_stop: ... \scan_stop: is true if and only if the string ... is empty. It could be tempting to use \if:w: \scan_stop: \scan_stop: directly. But this fails on a token list expanding to anything starting with \scan_stop: leaving everything that follows in the input stream.

```
\prg_new_conditional:Npn \tl_if_empty:n #1 { p , T , F , TF }
```

53.7 Token list conditionals

These functions check whether the token list in the argument is empty and execute the proper code from their argument(s).

```
\prg_new_conditional:Nppnn \tl_if_empty:N #1 { p , T , F , TF }
\tl_if_empty:cTF
```

```
\prg_new_conditional:Nppnn \tl_if_empty:n #1 { p , T , F , TF }
```

(End of definition for \tl_remove_all:Nn and \tl_gremove_all:Nn. These functions are documented on page 124.)

(End of definition for \tl_if_empty_p:N, \tl_if_empty_p:V, \tl_if_empty_p:e, \tl_if_empty_p:N, \tl_if_empty_p:V, \tl_if_empty_p:e. These functions are documented on page 124.)
The auxiliary function \_\_tl_if_empty_if:o is for use in various token list conditionals which reduce to testing if a given token list is empty after applying a simple function to it. The test for emptiness is based on \tl_if_empty:nTF, but the expansion is hard-coded for efficiency, as this auxiliary function is used in several places. We don’t put \prg_return_true: and so on in the definition of the auxiliary, because that would prevent an optimization applied to conditionals that end with this code. Also the \_\_tl_if_empty_if:o is expanded once in \tl_if_empty:oTF for efficiency as well (and to reduce code doubling).

\begin{verbatim}
\cs_new:Npn \_\_tl_if_empty_if:o #1 {\if:w \scan_stop: \__kernel_tl_to_str:w \exp_after:wN {#1} \scan_stop: }
\exp_args:Nno \use:n { \prg_new_conditional:Npnn \tl_if_empty:o #1 { p , TF , T , F } }
{ \__tl_if_empty_if:o {#1} \prg_return_true: \else: \prg_return_false: \fi: }
\prg_generate_conditional_variant:Nnn \tl_if_empty:o { e , V , o } { p , T , F , TF }
\end{verbatim}

TEX skips spaces when reading a non-delimited arguments. Thus, a ⟨token list⟩ is blank if and only if \use_none:n ⟨token list⟩ ? is empty after one expansion. The auxiliary \_\_tl_if_empty_if:o is a fast emptiness test, converting its argument to a string (after one expansion) and using the test \if:w \scan_stop: \... \scan_stop:.

\begin{verbatim}
\exp_args:Nno \use:n { \prg_new_conditional:Npnn \tl_if_blank:n #1 { p , T , F , TF } }
{ \__tl_if_empty_if:o { \use_none:n #1 ? } \prg_return_true: \else: \prg_return_false: \fi: }
\prg_generate_conditional_variant:Nnn \tl_if_blank:n { e , V , o } { p , T , F , TF }
\end{verbatim}

\_\_tl_if_blank_p:NNw
\_\_tl_if_blank_p:N
\_\_tl_if_blank_p:o
\_\_tl_if_blank:n
\_\_tl_if_blank:V
\_\_tl_if_blank:o
\_\_tl_if_blank:cc
\_\_tl_if_blank:cc
Returns \c_true_bool if and only if the two token list variables are equal.

\begin{verbatim}
\prg_new_eq_conditional:NNn \tl_if_eq:NN { p , T , F , TF }
\prg_generate_conditional_variant:Nnn \tl_if_eq:NN { Nc , c , cc } { p , T , F , TF }
\end{verbatim}

(End of definition for \tl_if_empty:nTF. This function is documented on page 112.)

(End of definition for \tl_if_blank:nTF and \_\_tl_if_empty_if:o. This function is documented on page 112.)

(End of definition for \tl_if_eq:nNTF. This function is documented on page 112.)

708
Temporary storage.

(End of definition for \l__tl_internal_a_tl and \l__tl_internal_b_tl.)

A simple store and compare routine.

(End of definition for \tl_if_eq:nnTF. This function is documented on page 112.)

A simple store and compare routine.

(End of definition for \tl_if_eq:nnTF. This function is documented on page 113.)

See \tl_if_eq:nnTF for further comments. Here we simply expand the token list variable and pass it to \tl_if_in:nnTF.

(End of definition for \tl_if_eq:nnTF. This function is documented on page 113.)
Once more, the test relies on the emptiness test for robustness. The function \_\_\_tl_\_tmp:w removes tokens until the first occurrence of #2. If this does not appear in #1, then the final #2 is removed, leaving an empty token list. Otherwise some tokens remain, and the test is false. See \_\_\_tl_if_empty:nTF for details on the emptiness test.

Treating correctly cases like \_\_\_tl_if_in:nnTF \{a state\}\{states\}, where #1#2 contains #2 before the end, requires special care. To cater for this case, we insert {} between the two token lists. This marker may not appear in #2 because of \TeX\ limitations on what can delimit a parameter, hence we are safe. Using two brace groups makes the test work also for empty arguments. The if_false: constructions are a faster way to do \group_align_safe_begin: and \group_align_safe_end:. The \scan_stop: ensures that f-expanding \_\_\_tl_if_in:nnTF does not lead to unbalanced braces.

Expand the token list and feed it to \_\_\_tl_if_single:nTF.

Tests whether #1 matches \texttt{-NoValue-} exactly (with suitable catcodes): this is similar to \quark_if_nil:nTF. The first argument of \_\_\_tl_if_novalue:w is empty if and only if #1 starts with \texttt{-NoValue-}, while the second argument is empty if #1 is exactly \texttt{-NoValue-} or if it has a question mark just following \texttt{-NoValue-}. In this second case, however, the material after the first ?! remains and makes the emptiness test return false.

(expand token list and feed it to \_\_\_tl_if_single:nTF. This function is documented on page 113.)
This test is similar to \texttt{\tl_if_empty:nTF}. Expanding \texttt{\use_none:nn #1 ??} once yields an empty result if \#1 is blank, a single \? if \#1 has a single item, and otherwise yields some tokens ending with \??. Then, \texttt{\__kernel_tl_to_str:w} makes sure there are no odd category codes. An earlier version would compare the result to a single \? using string comparison, but the Lua call is slow in LuaTeX. Instead, \texttt{\__tl_if_single:nnw} picks the second token in front of it. If \#1 is empty, this token is the trailing \? and the \texttt{\if:w} test yields \texttt{false}. If \#1 has a single item, the token is \texttt{\scan_stop:} and the \texttt{\if:w} test yields \texttt{true}. Otherwise, it is one of the characters resulting from \texttt{\tl_to_str:n}, and the \texttt{\if:w} test yields \texttt{false}. Note that \texttt{\if:w} and \texttt{\__kernel_tl_to_str:w} are primitives that take care of expansion.

There are four cases: empty token list, token list starting with a normal token, with a brace group, or with a space token. If the token list starts with a normal token, remove it and check for emptiness. For the next case, an empty token list is not a single token. Finally, we have a non-empty token list starting with a space or a brace group. Applying f-expansion yields an empty result if and only if the token list is a single space.
53.8 Mapping over token lists

Expandable loop macro for token lists. We use the internal scan mark \s__tl_stop (defined later), which is not allowed to show up in the token list #1 since it is internal to l3tl. This allows us a very fast test of whether some ⟨item⟩ is the end-marker \s__tl_stop, namely call \_tl_use_none_delimit_by_s_stop:w ⟨item⟩ ⟨function⟩ \s__tl_stop, which calls ⟨function⟩ if the ⟨item⟩ is the end-marker. To speed up the loop even more, only test one out of eight items, and once we hit one of the eight end-markers, go more slowly through the last few items of the list using \_tl_map_function_end:w.

\cs_new:Npn \tl_map_function:nN #1#2
{ \__tl_map_function:Nnnnnnnnn #2 #1 \__tl_map_function_end:w \s__tl_stop \s__tl_stop \s__tl_stop \s__tl_stop \s__tl_stop \s__tl_stop \s__tl_stop \s__tl_stop \s__tl_stop \prg_break_point:Nn \tl_map_break: { } }

\cs_new:Npn \tl_map_function:NN { \exp_args:No \tl_map_function:nN }
\cs_generate_variant:Nn \tl_map_function:NN { c }
\cs_new:Npn \__tl_map_function:Nnnnnnnnn #1#2#3#4#5#6#7#8#9
{ \__tl_use_none_delimit_by_s_stop:w #9 \__tl_map_function_end:w \s__tl_stop #1 {#2} #1 {#3} #1 {#4} #1 {#5} #1 {#6} #1 {#7} #1 {#8} #1 {#9} \__tl_map_function:NNnnnnnnn #1 \__tl_map_function_end:w \s__tl_stop \s__tl_stop \s__tl_stop \s__tl_stop \s__tl_stop \s__tl_stop \s__tl_stop \s__tl_stop \s__tl_stop \prg_break_point:Nn \tl_map_break: { \int_gdecr:N \g__kernel_prg_map_int } }

\the\currentrmnumber

The inline functions are straight forward by now. We use a little trick with the counter \g__kernel_prg_map_int to make them nestable. We can also make use of \_tl_map_function:NNnnnnnnn from before.

\cs_new_protected:Npn \tl_map_inline:nn #1#2
{ \int_gincr:N \g__kernel_prg_map_int \cs_gset_protected:cpn \__tl_map_use_int:N \g__kernel_prg_map_int :w } \#1 \#2 \exp_args:Nc \_tl_map_function:NNnnnnnnn \int_gincr:N \g__kernel_prg_map_int \prg_break_point:Nn \tl_map_break: { \int_gdecr:N \g__kernel_prg_map_int } }

\cs_new_protected:Npn \tl_map_inline:Nn #1#2
{ \__tl_map_function:NNnnnnnnn \int_gdecr:N \g__kernel_prg_map_int } }

\cs_new_protected:Npn \tl_map_inline:cn #1#2
{ \__tl_map_function:NNnnnnnnn \int_gdecr:N \g__kernel_prg_map_int } }
Much like the function mapping.

\begin{verbatim}
\cs_generate_variant:Nn \tl_map_inline:Nn { c }
(End of definition for \tl_map_inline:nn and \tl_map_inline:NNn. These functions are documented on page 118.)
\end{verbatim}

\begin{verbatim}
\cs_new:Npn \tl_map_tokens:nn #1#2
{ \__tl_map_tokens:nnnnnnnnn {#2} #1 \s__tl_stop \s__tl_stop \s__tl_stop \s__tl_stop
\prg_break_point:Nn \tl_map_break: { } }
\cs_new:Npn \tl_map_tokens:Nn
{ \exp_args:No \tl_map_tokens:nn }
\cs_generate_variant:Nn \tl_map_tokens:Nn { c }
\cs_new:Npn \__tl_map_tokens:nnnnnnnnn #1#2#3#4#5#6#7#8#9
{ \__tl_use_none_delimit_by_s_stop:w #9 \__tl_map_tokens_end:w \s__tl_stop
\use:n {#1} {#2} \use:n {#1} {#3} \use:n {#1} {#4} \use:n {#1} {#5} \use:n {#1} {#6} \use:n {#1} {#7} \use:n {#1} {#8} \use:n {#1} {#9} \__tl_map_tokens:nnnnnnnn {#1} }
\cs_new:Npn \tl_map_variable:nNn #1#2#3
{ \tl_map_tokens:nn {#1} { \__tl_map_variable:Nnn #2 {#3} } }
\cs_new_protected:Npn \tl_map_variable:NNn
{ \exp_args:No \tl_map_variable:nNn }
\cs_generate_variant:Nn \tl_map_variable:NNn { c }
(End of definition for \tl_map_tokens:nn and others. These functions are documented on page 118.)
\end{verbatim}

\begin{verbatim}
\cs_new:Npn \tl_map_break:
{ \prg_map_break:Nn \tl_map_break: { } }
\cs_new:Npn \tl_map_break:n
{ \prg_map_break:Nn \tl_map_break: }
(End of definition for \tl_map_break: and \tl_map_break:n. These functions are documented on page 119.)
\end{verbatim}
53.9 Using token lists

These functions return the replacement text of a token list as a string.

```
\cs_new:Npn \tl_to_str:N #1 { \__kernel_tl_to_str:w \exp_after:wN {#1} }
\cs_generate_variant:Nn \tl_to_str:N { c }
```

(End of definition for \tl_to_str:N. This function is documented on page 115.)

```
\cs_new:Npn \tl_use:N #1
\cs_new:Npn \tl_use:c #1
```

Token lists which are simply not defined give a clear TeX error here. No such luck for ones equal to \texttt{\textbackslash scan\_stop}: so instead a test is made and if there is an issue an error is forced.

```
\cs_new:Npn \tl_use:N #1
\int_eval:n { 0 \tl_map_function:nN {#1} \__tl_count:n }
\cs_new:Npn \tl_use:c #1
\int_eval:n { 0 \tl_map_function:NN #1 \__tl_count:n }
```

(End of definition for \tl_use:N. This function is documented on page 115.)

53.10 Working with the contents of token lists

```
\__tl_count:n #1 { + 1 }
\cs_new:Npn \__tl_count:n #1 { + 1 }
\cs_generate_variant:Nn \__tl_count:n { V , v , e , o }
\cs_generate_variant:Nn \__tl_count:n { c }
```

(End of definition for \__tl_count:n, \_tl_count:n, and \__tl_count:n. These functions are documented on page 116.)

714
The token count is computed through an \texttt{int_eval:n} construction. Each \texttt{1+} is output to the left, into the integer expression, and the sum is ended by the \texttt{exp_end:} inserted by \texttt{\_\_tl_act_end:wn} (which is technically implemented as \texttt{\_c_zero_int}). Somewhat a hack!

\begin{verbatim}
\cs_new:Npn \tl_count_tokens:n #1
{ \int_eval:n { \__tl_act:NNNn \__tl_act_count_normal:N \__tl_act_count_group:n \__tl_act_count_space:{#1} } }
\end{verbatim}

(End of definition for \texttt{tl_count_tokens:n} and others. This function is documented on page 116.)

Reversal of a token list is done by taking one item at a time and putting it after \texttt{s\_\_tl_stop}.

\begin{verbatim}
\cs_new:Npn \tl_reverse_items:n #1
{ \__tl_reverse_items:nwNwn #1 ? \s__tl_mark \__tl_reverse_items:nwNwn \s__tl_mark \__tl_reverse_items:wn \s__tl_stop { } }
\cs_new:Npn \__tl_reverse_items:nwNwn #1 #2 \s__tl_mark #3 #4 \s__tl_stop #5
{ #3 #2 \s__tl_mark \__tl_reverse_items:nwNwn \s__tl_mark \__tl_reverse_items:wn \s__tl_stop { {#1} #5 } }
\cs_new:Npn \__tl_reverse_items:wn #1 \s__tl_stop #2
{ \__kernel_exp_not:w \exp_after:wN { \use_none:nn #2 } }
\end{verbatim}

(End of definition for \texttt{tl_reverse_items:n}, \texttt{\_\_tl_reverse_items:nwNwn}, and \texttt{\_\_tl_reverse_items:wn}. This function is documented on page 116.)

Trimming spaces from around the input is deferred to an internal function whose first argument is the token list to trim, augmented by an initial \texttt{\_\_tl_trim_mark:}, and whose second argument is a \texttt{continuation}, which receives as a braced argument \texttt{\_\_tl_trim_mark: \{trimmed token list\}}. The control sequence \texttt{\_\_tl_trim_mark:} expands to nothing in a single expansion. In the case at hand, we take \texttt{\_\_kernel_exp_not:w \exp_after:wN \use_none:nn} as our continuation, so that space trimming behaves correctly within an \texttt{e}-type or \texttt{x}-type expansion.

\begin{verbatim}
\cs_new:Npn \tl_trim_spaces:n #1
{ \__tl_trim_spaces:nn { \__tl_trim_mark: #1 } }
\end{verbatim}
Trimming spaces from around the input is done using delimited arguments and quarks, and to get spaces at odd places in the definitions, we nest those in \texttt{\_tl_tmp:w}, which then receives a single space as its argument: \#1 is \texttt{\_tl_tmp:w}. Removing leading spaces is done with \texttt{\_tl_tmp:w}, which loops until \texttt{\_tl_tmp:w} matches the end of the token list: then \#1 is the token list and \#3 \texttt{\_tl_tmp:w} puts the token list into a group, with a lingering \texttt{\_tl_tmp:w}, responsible for trimming trailing spaces. The end is reached when \texttt{\_tl_tmp:w} matches the one present in the definition of \texttt{\_tl_tmp:w}. Then \texttt{\_tl_tmp:w} puts the token list into a group, with a lingering \texttt{\_tl_tmp:w}: at the start (which will expand to nothing in one step of expansion), and feeds this to the \texttt{\_tl_tmp:w}.
53.11 The first token from a token list

Finding the head of a token list expandably always strips braces, which is fine as this is consistent with for example mapping over a list. The empty brace groups in \texttt{\tl_head:n} ensure that a blank argument gives an empty result. The result is returned within the \texttt{\unexpanded} primitive. The approach here is to use \texttt{\if_false:} to allow us to use \texttt{\}} as the closing delimiter: this is the only safe choice, as any other token would not be able to parse its own code. More detail in \url{http://tex.stackexchange.com/a/70168}.

\begin{verbatim}
\cs_new:Npn \tl_head:n #1
\{ \__kernel_exp_not:w \tex_expanded:D { \if_false: \{ \fi: \__tl_head_aux:n #1 \} } \}
\cs_new:Npn \tl_head:w #1#2 \q_stop {#1}
\cs_new:Npn \__tl_tl_head:w #1#2 \s__tl_stop {#1}
\cs_new:Npn \tl_head:N { \exp_args:No \tl_head:n }
\end{verbatim}

To correctly leave the tail of a token list, it’s important not to absorb any of the tail part as an argument. For example, the simple definition

\begin{verbatim}
\cs_new:Npn \tl_tail:n #1 \{ \tl_tail:w #1 \q_stop \}
\cs_new:Npn \tl_tail:w #1#2 \q_stop
\end{verbatim}

would give the wrong result for \texttt{\tl_tail:n \{ a \{ bc \} \}} (the braces would be stripped). Thus the only safe way to proceed is to first check that there is an item to grab \textit{i.e.} that the argument is not blank and assuming there is to dispose of the first item. As with \texttt{\tl_head:n}, the result is protected from further expansion by \texttt{\unexpanded}. While we could optimise the test here, this would leave some tokens “banned” in the input, which we do not have with this definition.
Accessing the first token of a token list is tricky in three cases: when it has category code 1 (begin-group token), when it is an explicit space, with category code 10 and character code 32, or when the token list is empty (obviously).

Forgetting temporarily about this issue we would use the following test in \tl_if_head_eq_charcode:nN. Here, \tl_head:w yields the first token of the token list, then passed to \exp_not:N. The two first special cases are detected by testing if the token list starts with an N-type token (the extra ? sends empty token lists to the true branch of this test). In those cases, the first token is a character, and since we only care about its character code, we can use \str_head:n to access it (this works even if it is a space character). An empty argument results in \tl_head:w leaving two tokens: \_ and \__tl_if_head_eq_empty_arg:w which will result in the \if_charcode:w test being false and remove \exp_not:N and #2.

For \tl_if_head_eq_catcode:nN, again we detect special cases with a \_\_tl_if_head_eq_empty_arg:w which will result in the \if_charcode:w test being false and remove \exp_not:N and #2. For empty argument, a hack is used, removing the token given by the user and leaving two tokens in the input stream which will make the \if_catcode:w test return false.
For `\tl_if_head_eq_meaning:nN`, again, detect special cases. In the normal case, use `\tl_head:w`, with no `\exp_not:N` this time, since `\if_meaning:w` causes no expansion. With an empty argument, the test is `true`, and `\use_none:nnn` removes `#2` and `\prg_return_true:`, and `\else:` (it is safe this way here as in this case `\prg_new_conditional:Npnn` didn’t optimize these two away). In the special cases, we know that the first token is a character, hence `\if_charcode:w` and `\if_catcode:w` together are enough. We combine them in some order, hopefully faster than the reverse. Tests are not nested because the arguments may contain unmatched primitive conditionals.
Both `\tl_if_head_eq_charcode:nN` and `\tl_if_head_eq_catcode:nN` will need to get the first token of their argument and apply `\exp_not:N` to it. `\__tl_head_exp_not:w` does exactly that.

```
\cs_new:Npn \__tl_head_exp_not:w #1 #2 \s__tl_stop
{ \exp_not:N #1 }
```

If the argument of `\tl_if_head_eq_charcode:nN` and `\tl_if_head_eq_catcode:nN` was empty `\__tl_if_head_eq_empty_arg:w` will be left in the input stream. This macro has to remove `\exp_not:N` and the following token from the input stream to make sure no unbalanced if-construct is created and leave tokens there which make the two tests return false.

```
\cs_new:Npn \__tl_if_head_eq_empty_arg:w \exp_not:N #1
{ ? }
```

A token list can be empty, can start with an explicit space character (catcode 10 and charcode 32), can start with a begin-group token (catcode 1), or start with an N-type argument. In the first two cases, and when `#1~` starts with `{}`~, `\__tl_if_head_is_N_type_auxi:w` receives an empty argument hence produces `f` and removes everything before the first `\scan_stop:`. In the third case (except when `#1~` starts with `{}`~), the second auxiliary removes the first copy of `#1` that was used for the space test, then expands `\token_to_str:N` which hits the leading begin-group token, leaving a single closing brace to be compared with `\scan_stop:`. In the last case, `\token_to_str:N` does not change the brace balance so that only `\scan_stop:` remains, making the character code test true. One cannot optimize by moving one of the `\scan_stop:` to the beginning: if `#1` contains primitive conditionals, all of its occurrences must be dealt with before the `\if:w` tries to skip the true branch of the conditional.
Pass the first token of #1 through \token_to_str:N, then check for the brace balance. The extra ? caters for an empty argument. This could be made faster, but we need all brace tricks to happen in one step of expansion, keeping the token list brace balanced at all times.

The auxiliary’s argument is all that is before the first explicit space in \prg_do_nothing:#1?. If that is a single \prg_do_nothing: the test yields true. Otherwise, that is more than one token, and the test yields false. The work is done within braces (with an \if_false: { \fi: \__tl_if_head_is_space:w \prg_do_nothing: #1 ? ~ } \scan_stop: \scan_stop: \__tl_if_head_is_space:w \fi: \if_true: \prg_return_true: \else: \prg_return_false: \fi: \if_true: \prg_return_true: \else: \prg_return_false: \fi: \if_true: } \__tl_if_empty_if:o {#1} \else: f \fi: \exp_after:wN \use_none:n \exp_after:wN \__tl_if_empty_if:o {#1} \else: f \fi: \exp_after:wN \use_none:n \exp_after:wN \__tl_if_empty_if:o {#1} \else: f \fi: \exp_after:wN \use_none:n \exp_after:wN { \if_false: } \fi: \exp_after:wN \use_none:n \exp_after:wN \__tl_if_empty_if:o {#1} \else: f \fi: \exp_after:wN \use_none:n \exp_after:wN { \if_false: } \fi: \exp_after:wN \use_none:n \exp_after:wN { \if_false: } \fi: \exp_after:wN \use_none:n \exp_after:wN { \if_false: } \fi: \exp_after:wN \use_none:n \exp_after:wN { \if_false: } \fi: 

(End of definition for \tl_if_head_is_group:nTF and \__tl_if_head_is_space:w. This function is documented on page 114.)
53.12 Token by token changes

The \_\_tl_act... functions may be applied to any token list. Hence, we use a private quark, to allow any token, even quarks, in the token list. Only \s\_\_tl_act_stop may not appear in the token lists manipulated by \_\_tl_act:NNNn functions.

To help control the expansion, \_\_tl_act:NNNn should always be preceded by \exp:w and ends by producing \exp_end: once the result has been obtained. This way no internal token of it can be accidentally end up in the input stream. Because \_\_tl_act_stop can’t appear without braces around it in the argument #1 of \_\_tl_act_loop:w, we can use this marker to set up a fast test for leading spaces.

We expand the definition \_\_tl_act_if_head_is_space:nTF when setting up \_\_tl_act_loop:w, so we can then undefine the auxiliary. In the loop, we check how the token list begins and act accordingly. In the “group” case, we may have reached \_\_tl_act_stop, the end of the list. Then leave \exp_end: and the result in the input stream, to terminate the expansion of \exp:w. Otherwise, apply the relevant function to the “arguments”, #3 and to the head of the token list. Then repeat the loop. The scheme is the same if the token list starts with an N-type or with a space, making sure that \_\_tl_act_space:wwNNN gobbles the space.

722
\tl_reverse:n The goal here is to reverse without losing spaces nor braces. This is done using the
general internal function \tl_act:NNN. Spaces and “normal” tokens are output on the
left of the current output. Grouped tokens are output to the left but without any
reversal within the group.
\tl_reverse:o
\tl_reverse:v
\tl_reverse:f
\tl_reverse:e
\tl_reverse_normal:nN
\tl_reverse_group_preserve:n
\tl_reverse_space:n
(End of definition for \tl_act:NNN and others.)
{#1}

\cs_generate_variant:Nn \tl_reverse:n { o , V , f , e }
\cs_new:Npn \_\_\_tl_reverse_normal:N
  \{ \_\_\_tl_act_reverse_output:n \}
\cs_new:Npn \_\_\_tl_reverse_group_preserve:n #1
  \{ \_\_\_tl_act_reverse_output:n \{ {#1} \} \}
\cs_new:Npn \_\_\_tl_reverse_space:
  \{ \_\_\_tl_act_reverse_output:n \{ - \} \}

(End of definition for \tl_reverse:n and others. This function is documented on page 116.)

\tl_reverse:N
\tl_reverse:c
\tl_greverse:N
\tl_greverse:c

This reverses the list, leaving \exp_stop_f: in front, which stops the f-expansion.

\cs_new:Npn \tl_reverse:N #1
  \{ \exp_args:No \tl_reverse:n { #1 } \}
\cs_new_protected:Npn \tl_greverse:N #1
  \{ \_\_\_kernel_tl_gset:Nx #1 \exp_args:No \tl_reverse:n { #1 } \}
\cs_generate_variant:Nn \tl_reverse:N { c }
\cs_generate_variant:Nn \tl_greverse:N { c }

(End of definition for \tl_reverse:N and \tl_greverse:N. These functions are documented on page 116.)

53.13 Using a single item

\tl_item:nn
\tl_item:Nn
\tl_item:cn
\__tl_item_aux:nn
\__tl_item:nn

The idea here is to find the offset of the item from the left, then use a loop to grab the correct item. If the resulting offset is too large, then \_\_\_\_tl_if_recursion_tail_break:nN terminates the loop, and returns nothing at all.

\cs_new:Npn \tl_item:nn #1#2
  \exp_args:Nf \__tl_item:nn
    \exp_args:Nf \__tl_item_aux:nn \{ \int_eval:n {#2} \} {#1} \}
\cs_new:Npn \tl_item:Nn { \exp_args:No \tl_item:nn }
\cs_generate_variant:Nn \tl_item:Nn { c }

(End of definition for \tl_item:nn and others. These functions are documented on page 121.)
Importantly, \tl_item:nn only evaluates its argument once.

\(13129\)
\(\text{cs}_{\text{new}}:Npn \ \tl\_rand\_item:n \ #1\)
\(13130\)
\{ \tl\_if\_blank:nF \{#1\} \)
\(13131\)
\{ \tl\_item:nn \{#1\} \\{ \int\_rand:nn \{ \int\_count:n \{#1\} \} \} \)
\(13132\)
\}
\(13133\)
\cs\_{\text{new}}:Npn \ \tl\_rand\_item:N \{ \exp\_args:No \ \tl\_rand\_item:n \}
\(13134\)
\cs\_\text{generate}\_\text{variant}:Nn \ \tl\_rand\_item:N \{ c \}

(End of definition for \tl\_rand\_item:n and \tl\_rand\_item:N. These functions are documented on page 121.)

To avoid checking for the end of the token list at every step, start by counting the number \(l\) of items and “normalizing” the bounds, namely clamping them to the interval \([0, l]\) and dealing with negative indices. More precisely, \_\_\_tl\_range\_items:nnNn receives the number of items to skip at the beginning of the token list, the index of the last item to keep, a function which is either \_\_\_tl\_range:w or the token list itself. If nothing should be kept, leave \{\}: this stops the \text{f}-\text{expansion of} \tl\_head:f and that function produces an empty result. Otherwise, repeatedly call \_\_\_tl\_range\_skip:w to delete \#1 items from the input stream (the extra brace group avoids an off-by-one shift). For the braced version \_\_\_tl\_range\_braced:w sets up \_\_\_tl\_range\_collect\_braced:w which stores items one by one in an argument after the semicolon. Depending on the first token of the tail, either just move it (if it is a space) or also decrement the number of items left to find. Eventually, the result is a brace group followed by the rest of the token list, and \tl\_head:f cleans up and gives the result in \text{exp}\_not:n.

\(13135\)
\cs\_{\text{new}}:Npn \ \_\_\_tl\_range:Nnn \{ \exp\_args:No \ \_\_\_tl\_range:nn \}
\(13136\)
\cs\_\text{generate}\_\text{variant}:Nn \ \_\_\_tl\_range:Nnn \{ c \}
\(13137\)
\cs\_{\text{new}}:Npn \ \_\_\_tl\_range:nnn \\{ \_\_\_tl\_range:w \}
\(13138\)
\cs\_{\text{new}}:Npn \ \_\_\_tl\_range:nnn \#1#2#3#4
\{ \tl\_head:f
\{ \exp\args:Na \_\_\_tl\_range:nnNn
\{ \tl\_count:n \{#2\} \} \{#3\} \{#4\} \#1 \{#2\}
\}
\}
\cs\_{\text{new}}:Npn \ \_\_\_tl\_range:nnnn \#1#2#3
\{ \exp\args:Na \_\_\_tl\_range:nnNn
\{ \exp\args:Na \_\_\_tl\_range\_\_\_normalize:nn
\{ \int\_eval:n \{ \#2 - 1 \} \} \{#1\}
\}
\{ \exp\args:Na \_\_\_tl\_range\_\_\_normalize:nn
\{ \int\_eval:n \{#3\} \} \{#1\}
\}
\}
\cs\_{\text{new}}:Npn \ \_\_\_tl\_range:nnnn \#1#2#3#4
\{ \if\_int\_compare:w \#2 > #1 \exp\_stop\_f: \else:
\exp\_after:wN \{ \exp\_after:wN \}
\fi:
\_\_tl_range:nn  This function converts an \langle index\rangle argument into an explicit position in the token list

(End of definition for \_\_tl_range:Nnn and others. These functions are documented on page 122.)
(a result of 0 denoting “out of bounds”). Expects two explicit integer arguments: the \( \langle \text{index} \rangle \) #1 and the string count #2. If #1 is negative, replace it by \#1 + \#2 + 1, then limit to the range \([0, \#2]\).

\begin{verbatim}
\cs_new:Npn \__tl_range_normalize:nn #1#2
\int_eval:n {
  \if_int_compare:w #1 < \c_zero_int
    \if_int_compare:w #1 < -#2 \exp_stop_f:
      0
    \else:
      #1 + #2 + 1
    \fi:
  \else:
    \if_int_compare:w #1 < #2 \exp_stop_f:
      #1
    \else:
      #2
    \fi:
  \fi:
}
\end{verbatim}

(End of definition for \__tl_range_normalize:nn.)

\section*{53.14 Viewing token lists}

\texttt{\tl_show:N} Showing token list variables is done after checking that the variable is defined (see \texttt{\__kernel_register_show:N}).

\begin{verbatim}
\cs_new_protected:Npn \tl_show:N { \__tl_show:NN \tl_show:n }
\cs_generate_variant:Nn \tl_show:N { c }
\cs_new_protected:Npn \tl_log:N { \__tl_show:NN \tl_log:n }
\cs_generate_variant:Nn \tl_log:N { c }
\cs_new_protected:Npn \__tl_show:NN #1#2
\__kernel_chk_defined:NT #2
\exp_args:Nf \tl_if_empty:nTF
\exp_args:Ne #1
\msg_error:nneee { kernel } { bad-type }{ \token_to_str:N #2 }{ \token_to_meaning:N #2 }{ tl }
\end{verbatim}

(End of definition for \tl_show:N, \tl_log:N, and \__tl_show:NN. These functions are documented on page 117.)
Many `show` functions are based on `tl_show:n`. The argument of `tl_show:n` is line-wrapped using `\iow_wrap:nnnN` but with a leading `>~` and trailing period, both removed before passing the wrapped text to the `\showtokens` primitive. This primitive shows the result with a leading `>~` and trailing period.

The token list `\__tl_internal_a_tl` containing the result of all these manipulations is displayed to the terminal using `\tex_showtokens:D` and an odd `\exp_after:wN` which expand the closing brace to improve the output slightly. The calls to `\__kernel_iow_with:Nnn` ensure that the `\newline` is set to 10 so that the `\iownewline:inserted` by the line-wrapping code are correctly recognized by TeX, and that `\errorcontextlines` is −1 to avoid printing irrelevant context.

```
\cs_new_protected:Npn \tl_show:n #1
\cs_generate_variant:Nn \tl_show:n { e , x }
\cs_new_protected:Npn \__tl_show:n #1
{ \tl_set:Nf \l__tl_internal_a_tl { \__tl_show:w #1 \s__tl_stop }
\__kernel_iow_with:Nnn \tex_newlinechar:D { 10 }
{ \__kernel_iow_with:Nnn \tex_errorcontextlines:D { -1 }
{ \tex_shownumbers:D \exp_after:wN \exp_after:wN \exp_after:wN
\exp_after:wN \l__tl_internal_a_tl }
}
\cs_new:Npn \__tl_show:w #1 >~ . \s__tl_stop {#2}
```

(End of definition for `\tl_show:n`, `\__tl_show:n`, and `\__tl_show:w`. This function is documented on page 117.)

`\log` is much easier, simply line-wrap. The `>~` and trailing period is there to match the output of `\tl_show:n`.

```
\cs_new_protected:Npn \tl_log:n #1
\cs_generate_variant:Nn \tl_log:n { e , x }
\cs_new_protected:Npn \__tl_log:n #1
{ \iow_wrap:nnnN { >~ \tl_to_str:n {#1} . } { } { } \iow_log:n }
```

(End of definition for `\tl_log:n`. This function is documented on page 117.)

`\__kernel_chk_tl_type:NnnT` Helper for checking that `#1` has the correct internal structure to be of a certain type. Make sure that it is defined and that it is a token list, namely a macro with no `\long` nor `\protected` prefix. Then compare `#1` to an attempt at reconstructing a valid structure of the given type using `#2` (see implementation of `\seq_show:N` for instance). If that is successful run the requested code `#4`.

```
\cs_new_protected:Npn \__kernel_chk_tl_type:NnnT #1#2#3#4
{ \__kernel_chk_defined:NT #1
{ \tl_if_empty:nTF { \cs_prefix_spec:N #1 \cs_parameter_spec:N #1 }
{ \tl_set:Ne \l__tl_internal_a_tl {#3}
\tl_if_eq:NNTF \l__tl_internal_a_tl \l__tl_internal_a_tl
{#4}
```

728
\msg_error:nneeee { kernel } { bad-type }
\token_to_str:N #1 \tl_to_str:N \l__tl_internal_a_tl
\token_to_str:N \l__tl_internal_b_tl
\token_to_meaning:N #1
\msg_error:nneeee { kernel } { bad-type }
\token_to_str:N \l__tl_internal_a_tl
\token_to_meaning:N #2
\token_to_str:N \l__tl_internal_b_tl
\token_to_meaning:N #2

(End of definition for \__kernel_chk_tl_type:NnnT.)

### 53.15 Internal scan marks

\s__tl_nil \s__tl_mark \s__tl_stop

Internal scan marks. These are defined here at the end because the code for \scan_new:N depends on some l3tl functions.

\scan_new:N \s__tl_nil
\scan_new:N \s__tl_mark
\scan_new:N \s__tl_stop

(End of definition for \s__tl_nil, \s__tl_mark, and \s__tl_stop.)

### 53.16 Scratch token lists

\g_tmpa_tl \g_tmpb_tl
\l_tmpa_tl \l_tmpb_tl

Global temporary token list variables. They are supposed to be set and used immediately, with no delay between the definition and the use because you can’t count on other macros not to redefine them from under you.

\tl_new:N \g_tmpa_tl
\tl_new:N \g_tmpb_tl

(End of definition for \g_tmpa_tl and \g_tmpb_tl. These variables are documented on page 126.)

\l_tmpa_tl \l_tmpb_tl

These are local temporary token list variables. Be sure not to assume that the value you put into them will survive for long—see discussion above.

\tl_new:N \l_tmpa_tl
\tl_new:N \l_tmpb_tl

(End of definition for \l_tmpa_tl and \l_tmpb_tl. These variables are documented on page 126.)

We finally clean up a temporary control sequence that we have used at various points to set up some definitions.

\cs_undefine:N \__tl_tmp:w

⟨/package⟩
Chapter 54

\texttt{l3tl-build} implementation

\begin{verbatim}
\tl_build_begin:N\tl_var
\tl_build_gbegin:N\tl_var
\__tl_build_begin:NN \cs_set_nopar:Npe \tl_var
\__tl_build_begin:NNN \cs_gset_nopar:Npe \tl_var

First construct the \texttt{\tl_var}: using a prime here conflicts with the usual \texttt{expl3} convention but we need a name that can be derived from \texttt{#1} without any external data such as a counter. Empty that \texttt{\tl_var} and setup the structure. The local and global versions only differ by a single function \texttt{\cs_(g)set_nopar:Npe} used for all assignments: this is important because only that function is stored in the \texttt{\tl_var} and \texttt{\tl_var} for subsequent assignments. In principle \texttt{\__tl_build_begin:NNN} could use \texttt{\tl_(g)clear_new:N} to clear \texttt{#1} and make sure it is defined, but logging the definition does not seem useful so we just do \texttt{#3 \tl_var} to clear it locally or globally as appropriate.
\end{verbatim}
Similar to \texttt{\_\_tl\_build\_put\_right:Nn}, but apply \texttt{\_\_exp:} to \texttt{#1}. Most of the time this just removes one \texttt{\_\_exp\_end:}. When there are none left, \texttt{\_\_tl\_build\_last:NNn} is expanded instead. It resets the definition of the \texttt{⟨tl\ var⟩} by ending the \texttt{\_\_exp\_not:n} and the definition early. Then it makes sure the \texttt{⟨next\ tl⟩} (its argument \texttt{#1}) is set-up and starts a new definition. Then \texttt{\_\_tl\_build\_put:nn} and \texttt{\_\_\_tl\_build\_put:nw} place the \texttt{⟨left⟩} part of the original \texttt{⟨tl\ var⟩} as appropriate for the definition of the \texttt{⟨next\ tl⟩} (the \texttt{⟨right⟩} part is left in the right place without ever becoming a macro argument). We use \texttt{\_\_exp\_after:wN} rather than some \texttt{\_\_exp\_args:No} to avoid reading arguments that are likely very long token lists. We use \texttt{\_\_cs\_gset\_nopar:Npe} rather than \texttt{\_\_tl\_gset:Ne} partly for the same reason and partly because the assignments are interrupted by brace tricks, which implies that the assignment does not simply set the token list to an \texttt{e}\texttt{-expansion} of the second argument.

\begin{verbatim}
\cs_new_protected:Npn \_\_\_\_tl\_build\_put:nn #1 \_\_\_\_tl\_build\_put:nw
\end{verbatim}

See \texttt{\_\_tl\_build\_put\_right:Nn} for all the machinery. We could easily provide \texttt{\_\_\_\_tl\_build\_put\_left\_right:Nnn}, by just adding the \texttt{⟨right⟩} material after the \texttt{⟨(left)⟩} in the \texttt{e}\texttt{-expansion} assignment.
\tl_build_end:N
\tl_build_gend:N
\__tl_build_end_loop:NN

Get the data then clear the \langle next tl \rangle recursively until finding an empty one. It is perhaps wasteful to repeatedly use \cs_to_str:N. The local/global scope is checked by \tl_set:Ne or \tl_gset:Ne.

\tl_build_get_intermediate:NN

The idea is to expand the \langle tl var \rangle then the \langle next tl \rangle and so on, all within an e-expanding assignment, and wrap as appropriate in \exp_not:n. The various \langle left \rangle parts are left in the assignment as we go, which enables us to expand the \langle next tl \rangle at
the right place. The various ⟨right⟩ parts are eventually picked up in one last \exp_not:n, with a brace trick to wrap all the ⟨right⟩ parts together.

\cs_new_protected:Npn \_tl_build_get:NNN #1#2#3
\exp_not:n {#1 #3 { \if_false: { \fi: \exp_after:wN \_tl_build_get:w #2 } } }
\cs_new:Npn \_tl_build_get:w #1 \_tl_build_last:Nn #2#3#4
\exp_not:n {#4}
\if_meaning:w \c_empty_tl #3
\exp_after:wN \_tl_build_get_end:w
\fi:
\exp_after:wN \_tl_build_get:w #3
\cs_new:Npn \_tl_build_get_end:w #1#2#3
{ \_kernel_exp_not:w \exp_after:wN \if_false: } \fi: }

(End of definition for \_tl_build_get:NNN, \_tl_build_get:w, and \_tl_build_get_end:w)

\endinput
Chapter 55

l3str implementation

55.1 Internal auxiliaries

\s__str_mark
\s__str_stop
Internal scan marks.
\scan_new:N \s__str_mark
\scan_new:N \s__str_stop
(End of definition for \s__str_mark and \s__str_stop.)

\_str_use_none_delimit_by_s_stop:v
\_str_use_i_delimit_by_s_stop:nw
Functions to gobble up to a scan mark.
\cs_new:Npn \_str_use_none_delimit_by_s_stop:w \#1 \s__str_stop { }\cs_new:Npn \_str_use_i_delimit_by_s_stop:nw \#1 \#2 \s__str_stop {#1}
(End of definition for \_str_use_none_delimit_by_s_stop:v and \_str_use_i_delimit_by_s_stop:nw.)

\q__str_recursion_tail
\q__str_recursion_stop
Internal recursion quarks.
\quark_new:N \q__str_recursion_tail
\quark_new:N \q__str_recursion_stop
(End of definition for \q__str_recursion_tail and \q__str_recursion_stop.)

\_str_if_recursion_tail_break:NN
\_str_if_recursion_tail_stop_do:Nn
Functions to query recursion quarks.
\__kernel_quark_new_test:N \_str_if_recursion_tail_break:NN
\__kernel_quark_new_test:N \_str_if_recursion_tail_stop_do:Nn
(End of definition for \_str_if_recursion_tail_break:NN and \_str_if_recursion_tail_stop_do:Nn.)
55.2 Creating and setting string variables

A string is simply a token list. The full mapping system isn’t set up yet so do things by hand.

\begin{verbatim}
\str_new:N \str_new:c \str_use:N \str_use:c \str_clear:N \str_clear:c \str_gclear:N \str_gclear:c \str_clear_new:N \str_clear_new:c \str_gclear_new:N \str_gclear_new:c \str_set_eq:NN \str_set_eq:cN \str_set_eq:Nc \str_set_eq:cc \str_gset_eq:NN \str_gset_eq:cN \str_gset_eq:Nc \str_gset_eq:cc \str_concat:NNN \str_concat:ccc \str_gconcat:NNN \str_gconcat:ccc
\end{verbatim}

A string is simply a token list. The full mapping system isn’t set up yet so do things by hand.

\begin{verbatim}
\__str_tmp:n #1 \tl_if_blank:nF {#1} \cs_new_eq:cc { str_ #1 :N } { tl_ #1 :N } \exp_args:Nc \cs_generate_variant:Nn { str_ #1 :N } { c } \__str_tmp:n \__str_tmp:n \cs_set_protected:Npn \__kernel_tl_set:Nx \__kernel_tl_to_str:w \__str_tmp:n \cs_new_eq:NN \tl_set_eq:NN \cs_new_eq:NN \tl_gset_eq:NN \cs_new_eq:NN \tl_concat:NNN \cs_new_eq:NN \tl_gconcat:NNN
\end{verbatim}

Similar to corresponding \texttt{l3tl} base functions, except that \texttt{\_kernel\_exp\_not:w} is replaced with \texttt{\_kernel\_tl\_to\_str:w}. Just like token list, string constants use \texttt{\_kernel\_exp\_not:w} instead of \texttt{\_kernel\_tl\_gset:Nx} so that the scope checking for \texttt{c} is applied when \texttt{l3debug} is used. To maintain backward compatibility, in \texttt{\str_\_gput_left:Nn} and \texttt{\str_\_gput_right:Nn}, contents of string variables are wrapped in \texttt{\_kernel\_exp\_not:w} to prevent further expansion.

\begin{verbatim}
\str_set:Nn \str_set:NV \str_set:Ne \str_set:Nx \str_set:cn \str_set:cV \str_set:ce \str_set:cx \str_gset:Nn \str_gset:NV \str_gset:Ne \str_gset:Nx \str_gset:cn \str_gset:cV \str_gset:ce \str_gset:cx \str_gset:NN \str_gset:NNc \str_gset:NNc \str_gset:NNN \str_gset:NNNc \str_gset:NNNc \str_gset:NNN
\end{verbatim}

(End of definition for \texttt{\str\_new:N} and others. These functions are documented on page 130.)
55.3 Modifying string variables

Start by applying \texttt{\textbackslash tl\_to\_str:n} to convert the old and new token lists to strings, and also apply \texttt{\textbackslash tl\_to\_str:N} to avoid any issues if we are fed a token list variable. Then the code is a much simplified version of the token list code because neither the delimiter nor the replacement can contain macro parameters or braces. The delimiter \texttt{\textbackslash s\_\textbackslash str\_mark} cannot appear in the string to edit so it is used in all cases. Some e-expansion is unnecessary. There is no need to avoid losing braces nor to protect against expansion. The ending code is much simplified and does not need to hide in braces.
\use:e
{

\exp_not:n { \__str_replace_aux:NNNnnn #1 #2 #3 }
\cs_set:Npn \__str_replace_next:w ##1 #5 { ##1 #6 #1 }
\__str_use_none_delimit_by_s_stop:w #5
\s__str_stop
}
\cs_new_protected:Npn \__str_replace_next:w ?

\str_remove_once:Nn \str_remove_once:cn \str_gremove_once:Nn \str_gremove_once:cn
\str_remove_all:Nn \str_remove_all:cn \str_gremove_all:Nn \str_gremove_all:cn

\str_if_empty_p:N \str_if_empty_p:c \str_if_empty:NTF \str_if_empty:cpN \str_if_exist_p:N \str_if_exist_p:c \str_if_exist:NTF \str_if_exist:cpN

(End of definition for \str_replace_all:Nnn and others. These functions are documented on page 138.)

Removal is just a special case of replacement.

Removal is just a special case of replacement.

(End of definition for \str_remove_once:Nn and \str_gremove_once:Nn. These functions are documented on page 138.)

More copy-paste!

55.4 String comparisons
String comparisons rely on the primitive \( \text{strcmp} \), so we define a new name for it.

\[ \text{cs_new_eq:NN} \ \text{\_str_if_eq:nn} \ \text{\_str_compare_p:nNn} \ \text{\_str_compare:eNe} \]

Simply rely on \_\_str_if_eq:nn, which expands to -1, 0 or 1. The ee version is created directly because it is more efficient.

Modern engines provide a direct way of comparing two token lists, but returning a number. This set of conditionals therefore makes life a bit clearer. The \text{nn} and \text{ee} versions are created directly as this is most efficient. Since \_\_str_if_eq:nn will expand to 0 as an explicit character with category 12 if the two lists match (and either -1 or 1 if they don’t) we can use \text{if:w} here which is faster than using \text{if_int_compare:w}.

Note that \text{str_if_eq:NNTF} is different from \text{\_\_str_if_eq:NNTF} because it needs to ignore category codes.
Everything here needs to be detokenized but beyond that it is a simple token list test. It would be faster to fine-tune the `T`, `F`, `TF` variants by calling the appropriate variant of \texttt{\str_if_in:nTF} directly but that takes more code.

The aim here is to allow the case statement to be evaluated using a known number of expansion steps (two), and without needing to use an explicit "end of recursion" marker. That is achieved by using the test input as the final case, as this is always true. The trick is then to tidy up the output such that the appropriate case code plus either the true or false branch code is inserted.
To tidy up the recursion, there are two outcomes. If there was a hit to one of the cases searched for, then #1 is the code to insert, #2 is the next case to check on and #3 is all of the rest of the cases code. That means that #4 is the true branch code, and #5 tidies up the spare \s__str_mark and the false branch. On the other hand, if none of the cases matched then we arrive here using the “termination” case of comparing the search with itself. That means that #1 is empty, #2 is the first \s__str_mark and so #4 is the false code (the true code is mopped up by #3).
End of definition for \str_case:nnTF and others. These functions are documented on page 132.

55.5 Mapping over strings

The inline and variable mappings are similar to the usual token list mappings but start out by turning the argument to an “other string”. Doing the same for the expandable function mapping would require \_\_kernel_str_to_other:n, quadratic in the string length. To deal with spaces in that case, \_\_str_map_function:w replaces the following space by a braced space and a further call to itself. These are received by \_\_str_map_function:nn, which passes the space to \#1 and calls \_\_str_map_function:w to deal with the next space. The space before the braced space allows to optimize the \q__str_recursion_tail test. Of course we need to include a trailing space (the question mark is needed to avoid losing the space when \TeX{} tokenizes the line). At the cost of about three more auxiliaries this code could get a 9 times speed up by testing only every 9-th character for whether it is \q__str_recursion_tail (also by converting 9 spaces at a time in the \str_map_function:nN case).

For the \map_variable functions we use a string assignment to store each character because spaces are made catcode 12 before the loop.
\prg_break_point:Nn \str_map_break:
\{ \int_gdecr:N \g__kernel_prg_map_int \}
\cs_new_protected:Npn \str_map_inline:NN
\{ \exp_args:No \str_map_inline:nn \}
\cs_generate_variant:Nn \str_map_inline:NN { c }
\cs_new:Npn \__str_map_inline:NN #1#2
\{ \__str_if_recursion_tail_break:NN #2 \str_map_break:
\exp_args:No \exp_no: #1 \{ \token_to_str:N \exp_arg: #2 \}
\__str_map_inline:NN #1 \}
\cs_new_protected:Npn \str_map_variable:nNn #1#2#3
\{ \use:e
\{ \exp_not:n \__str_map_variable:NNN #2 #3 \}
\__kernel_str_to_other_fast:n \exp_not_n:x \#1
\}
\__str_recursion_tail
\prg_break_point:Nn \str_map_break: \}
\cs_new_protected:Npn \str_map_variable:NNn
\{ \exp_args:No \str_map_variable:nNn \}
\cs_generate_variant:Nn \str_map_variable:NNn { c }
\cs_new:Npn \str_map_break:
\prg_break_point:Nn \str_map_break: { }
\cs_new:Npn \str_map_break:n
\prg_break_point:Nn \str_map_break: { }

(End of definition for \str_map_function:NN and others. These functions are documented on page 133.)

\str_map_tokens:Nn
\str_map_tokens:cn
\str_map_tokens:nn

Uses an auxiliary of \str_map_function:NN.

\cs_new:Npn \str_map_tokens:nn #1#2
\{ \exp_args:Nno \use:nn
\exp_not:n \__str_map_variable:NNN #2 #3 \}
\__kernel_str_to_other_fast:n \#1
\}
\__str_recursion_tail
\prg_break_point:Nn \str_map_break: { }
\cs_new:Npn \str_map_tokens:Nn { \exp_args:No \str_map_tokens:nn }
\cs_generate_variant:Nn \str_map_tokens:Nn { c }

(End of definition for \str_map_tokens:Nn and \str_map_tokens:nn. These functions are documented on page 134.)

742
55.6 Accessing specific characters in a string

First apply \texttt{tl_to_str:n}, then replace all spaces by “other” spaces, 8 at a time, storing the converted part of the string between the \texttt{s_str_mark} and \texttt{s_str_stop} markers. The end is detected when \texttt{__str_to_other_loop:w} finds one of the trailing A, distinguished from any contents of the initial token list by their category. Then \texttt{__str_to_other_end:w} is called, and finds the result between \texttt{s_str_mark} and the first A (well, there is also the need to remove a space).

\begin{verbatim}
\cs_new:Npn \__kernel_str_to_other:n #1
\exp_after:wN \__str_to_other_loop:w \tl_to_str:n {#1} ~ A ~ A ~ A ~ A ~ A ~ A ~ A ~ A ~ \s__str_mark \s__str_stop
\end{verbatim}

\begin{verbatim}
\cs_new:Npn \__str_to_other_end:w \fi: #1 \s__str_mark #2 * A #3 \s__str_stop
\end{verbatim}

(End of definition for \texttt{\__kernel_str_to_other:n}, \texttt{\__str_to_other_loop:w}, and \texttt{\__str_to_other_end:w}.)

The difference with \texttt{\__kernel_str_to_other:n} is that the converted part is left in the input stream, making these commands only restricted-expandable.

\begin{verbatim}
\cs_new:Npn \__kernel_str_to_other_fast:n #1
\end{verbatim}

\begin{verbatim}
\cs_new:Npn \__str_to_other_fast_end:w \fi: \s__str_mark \s__str_stop
\end{verbatim}

(End of definition for \texttt{\__kernel_str_to_other_fast:n}, \texttt{\__str_to_other_fast_loop:w}, and \texttt{\__str_to_other_fast_end:w}.)
The `\str_item:nn` function does not escape spaces, which are thus ignored by `\str_item:nn` since everything else is done with undelimited arguments. Evaluate the ⟨index⟩ argument #2 and count characters in the string, passing those two numbers to `\str_item:w` for further analysis. If the ⟨index⟩ is negative, shift it by the ⟨count⟩ to know the how many character to discard, and if that is still negative give an empty result. If the ⟨index⟩ is larger than the ⟨count⟩, give an empty result, and otherwise discard ⟨index⟩ − 1 characters before returning the following one. The shift by −1 is obtained by inserting an empty brace group before the string in that case: that brace group also covers the case where the ⟨index⟩ is zero.

The `\str_item:nn` hands its argument with spaces escaped to `\str_item:nn`, and makes sure to turn the result back into a proper string (with category code 10 spaces) eventually. The `\str_item_ignore_spaces:nn` function does not escape spaces, which are thus ignored by `\str_item:nn` since everything else is done with undelimited arguments. Evaluate the ⟨index⟩ argument #2 and count characters in the string, passing those two numbers to `\str_item:w` for further analysis. If the ⟨index⟩ is negative, shift it by the ⟨count⟩ to know the how many character to discard, and if that is still negative give an empty result. If the ⟨index⟩ is larger than the ⟨count⟩, give an empty result, and otherwise discard ⟨index⟩ − 1 characters before returning the following one. The shift by −1 is obtained by inserting an empty brace group before the string in that case: that brace group also covers the case where the ⟨index⟩ is zero.

```latex
\begin{verbatim}
\cs_new:Npn \__str_item:nn #1#2
{ \exp_after:wN \__str_item:w \int_value:w \int_eval:n {#2} \exp_after:wN ; \int_value:w \__str_count:n {#1} ; #1 \s__str_stop }
\cs_new:Npn \__str_item:w #1; #2;
{ \int_compare:nNnTF {#1} < 0
{ \int_compare:nNnTF {#1} < {-#2}
{ \__str_use_none_delimit_by_s_stop:w }
{ \__str_use_none_delimit_by_s_stop:wnw }
\exp:w \exp_after:wN \__str_use_i_delimit_by_s_stop:wnw
\exp:w \exp_after:wN \__str_skip_exp_end:w
\int_value:w \int_eval:n { #1 + #2 } ;
}
{ \int_compare:nNnTF {#1} > {#2}
{ \__str_use_none_delimit_by_s_stop:w }
\end{verbatim}
```
Removes \( \max(#1,0) \) characters from the input stream, and then leaves \( \exp_end: \). This should be expanded using \( \exp:w \). We remove characters 8 at a time until there are at most 8 to remove. Then we do a dirty trick: the \( \if_case:w \) construction leaves between 0 and 8 times the \( \or:w \) control sequence, and those \( \or:w \) become arguments of \( \__str_skip_end:NNNNNNNN \). If the number of characters to remove is 6, say, then there are two \( \or:w \) left, and the 8 arguments of \( \__str_skip_end:NNNNNNNN \) are the two \( \or:w \), and 6 characters from the input stream, exactly what we wanted to remove. Then close the \( \if_case:w \) conditional with \( \fi:w \); and stop the initial expansion with \( \exp_end: \) (see places where \( \__str_skip_exp_end:w \) is called).

Sanitize the string. Then evaluate the arguments. At this stage we also decrement the \( \langle \text{start index} \rangle \), since our goal is to know how many characters should be removed. Then limit the range to be non-negative and at most the length of the string (this avoids needing to check for the end of the string when grabbing characters), shifting negative numbers by the appropriate amount. Afterwards, skip characters, then keep some more, and finally drop the end of the string.
\begin{verbatim}
\cs_new:Npn \str_range:Nnn #1\{ \__kernel_str_to_other:n {#1} \} {#2} {#3}
{
\exp_args:Nf \__str_range:nnn { \tl_to_str:n {#1} } {#2} {#3}
}
\cs_new:Npn \str_range_ignore_spaces:nnn #1\{\exp_args:No \__str_range:nnn { \tl_to_str:n {#1} } \}
\cs_new:Npn \__str_range:nnn #1#2#3
{
\exp_after:wN \__str_range:w
\int_value:w \__str_count:n {#1} \exp_after:wN ;
\int_value:w \int_eval:n { (#2) - 1 } \exp_after:wN ;
\int_value:w \int_eval:n {#3} ;
#1 \s__str_stop
}
\cs_new:Npn \__str_range:w #1; #2; #3;
{
\exp_after:wN \__str_range:nnw
\__str_range_normalize:nn {#2} {#1}
\__str_range_normalize:nn {#3} {#1}
}
\cs_new:Npn \__str_range:nnw #1#2
{
\exp_after:wN \__str_collect_delimit_by_q_stop:w
\int_value:w \int_eval:n { #2 - #1 } \exp_after:wN ;
\exp:w \__str_skip_exp_end:w #1 ;
}
\end{verbatim}

(End of definition for \str_range:Nnn and others. These functions are documented on page \pageref{page}.)

\__str_range_normalize:nn This function converts an \langle index \rangle argument into an explicit position in the string (a result of 0 denoting “out of bounds”). Expects two explicit integer arguments: the \langle index \rangle #1 and the string count #2. If #1 is negative, replace it by #1 + #2 + 1, then limit to the range [0, #2].

\begin{verbatim}
\cs_new:Npn \__str_range_normalize:nn #1#2
{
\int_eval:n
\if_int_compare:w #1 < \c_zero_int
  \if_int_compare:w #1 < -#2 \exp_stop_f:
    0
  \else:
    #1 + #2 + 1
  \fi:
\else:
  #1
\fi:
\else:
  \if_int_compare:w #1 < #2 \exp_stop_f:
    #1
  \else:
    #2
  \fi:
\fi:
}
\end{verbatim}

(End of definition for \__str_range_normalize:nn.)
Collects \(\max(#1,0)\) characters, and removes everything else until \s__str_stop. This is somewhat similar to \_\_str_skip_exp_end:w, but accepts integer expression arguments. This time we can only grab 7 characters at a time. At the end, we use an \if_case:w trick again, so that the 8 first arguments of \_\_str_collect_end:nnnnnnnnw are some \or:, followed by an \fi:, followed by \#1 characters from the input stream. Simply leaving this in the input stream closes the conditional properly and the \or: disappear.

\begin{verbatim}
1388 \cs_new:Npn \__str_collect_delimit_by_q_stop:w #1; 
1389 { \__str_collect_loop:wn #1 ; { } }
1390 \cs_new:Npn \__str_collect_loop:wn #1 ;
1391 { \if_int_compare:w #1 > 7 \exp_stop_f:
1392 \exp_after:wN \__str_collect_loop:wnNNNNNNN
1393 \else: \exp_after:wN \__str_collect_end:wn
1394 \fi: #1 ;
1395 }
1396 \cs_new:Npn \__str_collect_loop:wnNNNNNNN #1; #2 #3#4#5#6#7#8#9
1397 { \exp_after:wN \__str_collect_loop:wn
1398 \int_value:w \int_eval:n { #1 - 7 } ;
1399 { #2 #3#4#5#6#7#8#9 } }
1400 \cs_new:Npn \__str_collect_end:wn #1 ;
1401 { \exp_after:wN \__str_collect_end:nnnnnnnnw
1402 \if_case:w \if_int_compare:w #1 > \c_zero_int
1403 #1 \else: 0 \fi: \exp_stop_f:
1404 \or: \or: \or: \or: \or: \or: \or: \fi:
1405 }
1406 \cs_new:Npn \__str_collect_end:nnnnnnnnw #1#2#3#4#5#6#7#8 #9 \s__str_stop
1407 { #1#2#3#4#5#6#7#8}
\end{verbatim}

(End of definition for \_\_str_collect_delimit_by_q_stop:w and others.)

55.7 Counting characters

To speed up this function, we grab and discard 9 space-delimited arguments in each iteration of the loop. The loop stops when the last argument is one of the trailing X\langle number\rangle, and that \langle number\rangle is added to the sum of 9 that precedes, to adjust the result.
To count characters in a string we could first escape all spaces using \_kernel_str_to_other:n, then pass the result to \tl_count:n. However, the escaping step would be quadratic in the number of characters in the string, and we can do better. Namely, sum the number of spaces (\str_count_spaces:n) and the result of \tl_count:n, which ignores spaces. Since strings tend to be longer than token lists, we use specialized functions to count characters ignoring spaces. Namely, loop, grabbing 9 non-space characters at each step, and end as soon as we reach one of the 9 trailing items. The internal function \_str_count:n, used in \str_item:nn and \str_range:nnn, is similar to \str_count_ignore_spaces:n but expects its argument to already be a string or a string with spaces escaped.
The first character in a string

The \_ignore\_spaces variant applies \texttt{tl\_to\_str:n} then grabs the first item, thus skipping spaces. As usual, \texttt{str\_head:N} expands its argument and hands it to \texttt{str\_head:n}. To circumvent the fact that \TeX{} skips spaces when grabbing undelimited macro parameters, \texttt{\_str\_head:w} takes an argument delimited by a space. If \texttt{#1} starts with a non-space character, \texttt{\_str\_use\_i\_delimit\_by\_s\_stop:nw} leaves that in the input stream. On the other hand, if \texttt{#1} starts with a space, the \texttt{\_str\_head:w} takes an empty argument, and the single (initially braced) space in the definition of \texttt{\_str\_head:w} makes its way to the output. Finally, for an empty argument, the (braced) empty brace group in the definition of \texttt{str\_head:n} gives an empty result after passing through \texttt{\_str\_use\_i\_delimit\_by\_s\_stop:nw}.

Getting the tail is a little bit more convoluted than the head of a string. We hit the front of the string with \texttt{reverse\_if:N} \texttt{if\_charcode:w} \texttt{scan\_stop:}. This removes the first character, and necessarily makes the test true, since the character cannot match \texttt{scan\_stop:}. The auxiliary function then inserts the required \texttt{fi:} to close the conditional, and leaves the tail of the string in the input stream. The details are such that an empty string has an empty tail (this requires in particular that the end-marker \texttt{X} be unexpandable and not a control sequence). The \_ignore\_spaces is rather simpler: after converting the input to a string, \texttt{\_str\_tail\_auxi:w} removes one undelimited argument and leaves everything else until an end-marker \texttt{\_str\_mark}. One can check that an empty (or blank) string yields an empty tail.
Case changing for programmatic reasons is done by first detokenizing input then doing a simple loop that only has to worry about spaces and everything else. The output is detokenized to allow data sharing with text-based case changing. Similarly, for 8-bit engines the multi-byte information is shared.

Case changing for programmatic reasons is done by first detokenizing input then doing a simple loop that only has to worry about spaces and everything else. The output is detokenized to allow data sharing with text-based case changing. Similarly, for 8-bit engines the multi-byte information is shared.

Case changing for programmatic reasons is done by first detokenizing input then doing a simple loop that only has to worry about spaces and everything else. The output is detokenized to allow data sharing with text-based case changing. Similarly, for 8-bit engines the multi-byte information is shared.

Case changing for programmatic reasons is done by first detokenizing input then doing a simple loop that only has to worry about spaces and everything else. The output is detokenized to allow data sharing with text-based case changing. Similarly, for 8-bit engines the multi-byte information is shared.

Case changing for programmatic reasons is done by first detokenizing input then doing a simple loop that only has to worry about spaces and everything else. The output is detokenized to allow data sharing with text-based case changing. Similarly, for 8-bit engines the multi-byte information is shared.

Case changing for programmatic reasons is done by first detokenizing input then doing a simple loop that only has to worry about spaces and everything else. The output is detokenized to allow data sharing with text-based case changing. Similarly, for 8-bit engines the multi-byte information is shared.

Case changing for programmatic reasons is done by first detokenizing input then doing a simple loop that only has to worry about spaces and everything else. The output is detokenized to allow data sharing with text-based case changing. Similarly, for 8-bit engines the multi-byte information is shared.

Case changing for programmatic reasons is done by first detokenizing input then doing a simple loop that only has to worry about spaces and everything else. The output is detokenized to allow data sharing with text-based case changing. Similarly, for 8-bit engines the multi-byte information is shared.

Case changing for programmatic reasons is done by first detokenizing input then doing a simple loop that only has to worry about spaces and everything else. The output is detokenized to allow data sharing with text-based case changing. Similarly, for 8-bit engines the multi-byte information is shared.

Case changing for programmatic reasons is done by first detokenizing input then doing a simple loop that only has to worry about spaces and everything else. The output is detokenized to allow data sharing with text-based case changing. Similarly, for 8-bit engines the multi-byte information is shared.

Case changing for programmatic reasons is done by first detokenizing input then doing a simple loop that only has to worry about spaces and everything else. The output is detokenized to allow data sharing with text-based case changing. Similarly, for 8-bit engines the multi-byte information is shared.

Case changing for programmatic reasons is done by first detokenizing input then doing a simple loop that only has to worry about spaces and everything else. The output is detokenized to allow data sharing with text-based case changing. Similarly, for 8-bit engines the multi-byte information is shared.

Case changing for programmatic reasons is done by first detokenizing input then doing a simple loop that only has to worry about spaces and everything else. The output is detokenized to allow data sharing with text-based case changing. Similarly, for 8-bit engines the multi-byte information is shared.

Case changing for programmatic reasons is done by first detokenizing input then doing a simple loop that only has to worry about spaces and everything else. The output is detokenized to allow data sharing with text-based case changing. Similarly, for 8-bit engines the multi-byte information is shared.

Case changing for programmatic reasons is done by first detokenizing input then doing a simple loop that only has to worry about spaces and everything else. The output is detokenized to allow data sharing with text-based case changing. Similarly, for 8-bit engines the multi-byte information is shared.

Case changing for programmatic reasons is done by first detokenizing input then doing a simple loop that only has to worry about spaces and everything else. The output is detokenized to allow data sharing with text-based case changing. Similarly, for 8-bit engines the multi-byte information is shared.

Case changing for programmatic reasons is done by first detokenizing input then doing a simple loop that only has to worry about spaces and everything else. The output is detokenized to allow data sharing with text-based case changing. Similarly, for 8-bit engines the multi-byte information is shared.

Case changing for programmatic reasons is done by first detokenizing input then doing a simple loop that only has to worry about spaces and everything else. The output is detokenized to allow data sharing with text-based case changing. Similarly, for 8-bit engines the multi-byte information is shared.

Case changing for programmatic reasons is done by first detokenizing input then doing a simple loop that only has to worry about spaces and everything else. The output is detokenized to allow data sharing with text-based case changing. Similarly, for 8-bit engines the multi-byte information is shared.

Case changing for programmatic reasons is done by first detokenizing input then doing a simple loop that only has to worry about spaces and everything else. The output is detokenized to allow data sharing with text-based case changing. Similarly, for 8-bit engines the multi-byte information is shared.

55.9 String manipulation
\__str_change_case_output:nw \ ( - )
\__str_change_case_loop:nw \(#1\)
\}:
\cs_new:Npn \__str_change_case_char:nN \#1#2
\{\__str_if_recursion_tail_stop_do:Nn \#2\}
\__str_change_case_end:wn \}
\__str_change_case_codepoint:nN \(#1\) \#2
\}
\if_int_compare:w 0
\cs_if_exist:NT \tex_XeTeXversion:D \ { 1 }\n\cs_if_exist:NT \tex_luatexversion:D \ { 1 }\n> 0 \exp_stop_f:
\cs_new:Npn \__str_change_case_codepoint:nN \#1#2
\{ \__str_change_case_char:fnn \{ \int_eval:n \{\'#2\} \} \{#1\} \{#2\} \}
\else:
\cs_new:Npe \__str_change_case_codepoint:nN \#1#2
\{\exp_not:N \int_compare:nNnTF \{\'#2\} \ > \ { "80 }\}
\cs_if_exist:NTF \tex_pdftexversion:D
\exp_not:N \__str_change_case_char_auxi:nN \}
\exp_not:N \int_compare:nNnTF \{\'#2\} \ > \ { "FF }\}
\exp_not:N \__str_change_case_char_auxii:nN \}
\exp_not:N \__str_change_case_char_auxi:nN \}
\}
\exp_not:N \__str_change_case_char_auxii:nN \}
\{\exp_not:N \__str_change_case_char_auxi:nN \}
\{#1\} \#2
\}
\cs_new:Npn \__str_change_case_char_auxi:nN \#1#2
\{\int_compare:nNnTF \{\'#2\} \ < \ { "E0 }\}
\cs_if_exist:NTF \tex_pdfxversion:D
\{\int_compare:nNnTF \{\'#2\} \ < \ { "F0 }\}
\{\int_compare:nNnTF \{\'#2\} \ < \ { "FF }\}
\{\int_compare:nNnTF \{\'#2\} \ < \ { "FF }\}
\}
\{#1\} \#2
\}
\cs_new:Npn \__str_change_case_char_auxii:nN \#1#2
\{\__str_change_case_char_fnn \{\int_eval:n \{\'#2\} \} \{#1\} \{#2\} \}
\cs_new:Npn \__str_change_case_char_fnn
\{ \int_eval:n \{\'#2 - "C0\} \ast "40 + \'#3 - "80 \}
\{#1\} \{#2#3\}
\}
\cs_new:Npn \__str_change_case_char_fnn
\{ \int_eval:n \{\'#2 - "C0\} \ast "40 + \'#3 - "80 \}
\{#1\} \{#2#3\}
\}:
\__str_change_case_codepoint:nNN \#1#2#3
\{ \__str_change_case_char_fnn
\{ \int_eval:n \{\'#2 - "C0\} \ast "40 + \'#3 - "80 \}
\{#1\} \{#2#3\}
\}
\__str_change_case_codepoint:nNNN \#1#2#3#4
\{ \__str_change_case_char_fnn
\{ \int_eval:n \{\'#2 - "C0\} \ast "40 + \'#3 - "80 \}
\{#1\} \{#2#3\} \{#4\}
\}
\__str_change_case_codepoint:nNNNNN \#1#2#3#4#5
\{ \__str_change_case_char_fnn
\{ \int_eval:n \{\'#2 - "C0\} \ast "40 + \'#3 - "80 \}
\{#1\} \{#2#3\} \{#4#5\}
\}
\__str_change_case_codepoint:nN \#1#2#3#4#5#6
\{ \__str_change_case_char_fnn
\{ \int_eval:n \{\'#2 - "C0\} \ast "40 + \'#3 - "80 \}
\{#1\} \{#2#3\} \{#4#5\} \{#6\}
\}
\__str_change_case_codepoint:nNN \#1#2#3#4#5#6
\{ \__str_change_case_char_fnn
\{ \int_eval:n \{\'#2 - "C0\} \ast "40 + \'#3 - "80 \}
\{#1\} \{#2#3\} \{#4#5\} \{#6\} \}
751
\int_eval:n
\{ ('#2 - "E0) * "1000 + ('#3 - "80) * "40 + ('#4 - "80 }
\}
\{#1} \{#2\#3\#4\}
\}
cs_new:Npn \__str_change_case_codepoint:nNNNN #1#2#3#4#5
\{ \__str_change_case_char:fnn
\{ \int_eval:n
\{ ('#2 - "F0) * "40000
+ ('#3 - "80) * "1000
+ ('#4 - "80) * "40
+ ('#5 - "80 }
\}
\{#1} \{#2\#3\#4\#5\}
\}
\fi:
cs_new:Npn \__str_change_case_char:nnn #1#2#3
\{ \__str_change_case_output:fw
\{ \exp_args:Ne \__str_change_case_char_aux:nnn
\{ \__kernel_codepoint_case:nn \{#2\} \{#1\} \{#1\} \{#3\}
\}
\__str_change_case_loop:nw \{#2\}
\}
cs_generate_variant:Nn \__str_change_case_char:nnn \{ f \}
cs_new:Npn \__str_change_case_char_aux:nnnn \#1#2#3#4#5
\{ \int_compare:nNnTF \{#1\} = \{#4\}
\{ \tl_to_str:n \{#5\} \}
\{ \codepoint_str_generate:n \{#1\}
\tl_if_blank:nF \{#2\}
\{ \codepoint_str_generate:n \{#2\}
\tl_if_blank:nF \{#3\}
\{ \codepoint_str_generate:n \{#3\} \}
\}
\}
\}
(End of definition for \str_casefold:n and others. These functions are documented on page 140.)
\str_mdfive_hash:n
\str_mdfive_hash:e
\cs_new:Npn \str_mdfive_hash:n #1 \{ \tex_mdfivesum:D \{ \tl_to_str:n \{#1\} \} \}
\cs_new:Npn \str_mdfive_hash:e #1 \{ \tex_mdfivesum:D \{#1\} \}

752
For all of those strings, use \cs_to_str:N to get characters with the correct category code without worries.

\c_empty_str  An empty string is simply an empty token list.

\l_tmpa_str \l_tmpb_str \g_tmpa_str \g_tmpb_str  Scratch strings.

55.10 Viewing strings

Displays a string on the terminal.

(End of definition for \str_show:n and others. These functions are documented on page 140.)
Chapter 56

l3str-convert implementation

56.1 Helpers

56.1.1 Variables and constants

\_\_str\_tmp:w  
Internal scratch space for some functions.

\l\_str\_internal\_tl  
\cs\_new\_protected:Npn \_\_str\_tmp:w { }  
\tl\_new:N \l\_str\_internal\_tl  
(End of definition for \_\_str\_tmp:w and \l\_str\_internal\_tl.)

\g\_str\_result\_tl  
The \g\_str\_result\_tl variable is used to hold the result of various internal string operations (mostly conversions) which are typically performed in a group. The variable is global so that it remains defined outside the group, to be assigned to a user-provided variable.

\tl\_new:N \g\_str\_result\_tl  
(End of definition for \g\_str\_result\_tl.)

\c\_str\_replacement\_char\_int  
When converting, invalid bytes are replaced by the Unicode replacement character "FFFD.

\int\_const:Nn \c\_str\_replacement\_char\_int { "FFFD }  
(End of definition for \c\_str\_replacement\_char\_int.)

\c\_str\_max\_byte\_int  
The maximal byte number.

\int\_const:Nn \c\_str\_max\_byte\_int { 255 }  
(End of definition for \c\_str\_max\_byte\_int.)

\s\_str  
Internal scan marks.

\scan\_new:N \s\_str  
(End of definition for \s\_str.)
\q_str_nil Internal quarks.
\quark_new:N \q_str_nil
(End of definition for \q_str_nil.)
\g_str_alias_prop To avoid needing one file per encoding/escaping alias, we keep track of those in a property list.
\prop_new:N \g_str_alias_prop
\prop_gput:Nnn \g_str_alias_prop { latin1 } \{ iso88591 \}
\prop_gput:Nnn \g_str_alias_prop { latin2 } \{ iso88592 \}
\prop_gput:Nnn \g_str_alias_prop { latin3 } \{ iso88593 \}
\prop_gput:Nnn \g_str_alias_prop { latin4 } \{ iso88594 \}
\prop_gput:Nnn \g_str_alias_prop { latin5 } \{ iso88599 \}
\prop_gput:Nnn \g_str_alias_prop { latin6 } \{ iso885910 \}
\prop_gput:Nnn \g_str_alias_prop { latin7 } \{ iso885913 \}
\prop_gput:Nnn \g_str_alias_prop { latin8 } \{ iso885914 \}
\prop_gput:Nnn \g_str_alias_prop { latin9 } \{ iso885915 \}
\prop_gput:Nnn \g_str_alias_prop { latin10 } \{ iso885916 \}
\prop_gput:Nnn \g_str_alias_prop { utf16le } \{ utf16 \}
\prop_gput:Nnn \g_str_alias_prop { utf16be } \{ utf16 \}
\prop_gput:Nnn \g_str_alias_prop { utf32le } \{ utf32 \}
\prop_gput:Nnn \g_str_alias_prop { utf32be } \{ utf32 \}
\prop_gput:Nnn \g_str_alias_prop { hexadecimal } \{ hex \}
\bool_lazy_any:nTF
{ \sys_if_engine_luatex_p: \sys_if_engine_xetex_p: }
{ \prop_gput:Nnn \g_str_alias_prop { default } \{ \}
\prop_gput:Nnn \g_str_alias_prop { default } \{ utf8 \}
}
(End of definition for \g_str_alias_prop.)
\g_str_error_bool In conversion functions with a built-in conditional, errors are not reported directly to the user, but the information is collected in this boolean, used at the end to decide on which branch of the conditional to take.
\bool_new:N \g_str_error_bool
(End of definition for \g_str_error_bool.)
\l_str_byte_flag Conversions from one \encoding/\escaping pair to another are done within e-expanding assignments. Errors are signalled by raising the relevant flag.
\flag_new:N \l_str_byte_flag
\flag_new:N \l_str_error_flag
(End of definition for \l_str_byte_flag and \l_str_error_flag.)
56.2 String conditionals

\_str\_if\_contains\_char:nnTF \{\text{token list}\} \{\text{char}\}

Expects the \textit{token list} to be an \textit{other string}: the caller is responsible for ensuring that no (too-)special catcodes remain. Loop over the characters of the string, comparing character codes. The loop is broken if character codes match. Otherwise we return “false”.

\begin{verbatim}
\prg_new_conditional:Npnn \__str_if_contains_char:Nn #1#2 { T , TF }
{ \exp_after:wN \__str_if_contains_char_aux:nn \exp_after:wN {#1} {#2}
\prg_break:n { ? \fi: } }
\prg_break_point:
\prg_return_false:
\end{verbatim}

\begin{verbatim}
\cs_new:Npn \__str_if_contains_char_aux:nn #1#2
{ \__str_if_contains_char_auxi:nN {#2} #1 }
\prg_new_conditional:Npnn \__str_if_contains_char:nn #1#2 { TF }
{ \exp_after:wN \__str_if_contains_char_true: \fi:
\__str_if_contains_char_auxi:nN {#2} #1 { \prg_break:n { ? \fi: } }
\prg_break_point:
\prg_return_false:
\end{verbatim}

\begin{verbatim}
\cs_new:Npn \__str_if_contains_char_auxi:nN #1#2
{ \if_charcode:w #1 #2 \exp_after:wN \__str_if_contains_char_true: \fi:
\__str_if_contains_char_auxi:nN {#1}
\end{verbatim}

\begin{verbatim}
\cs_new:Npn \__str_if_contains_char_true:
{ \prg_break:n { \prg_return_true: \use_none:n } }
\end{verbatim}

\textit{End of definition for} \__str\_if\_contains\_char:nnTF and others.

\_str\_octal\_use:NTF

\begin{verbatim}
\prg_new_conditional:Npnn \__str_octal_use:N #1 { TF }
{ \if_int_compare:w 1 < \token_to_str:N #1 \exp_stop_f:
\exp_after:wN \__str_if_contains_char_true:
\else:
\prg_break:n { \prg_return_false: }
\fi:
\end{verbatim}

\textit{End of definition for} \__str\_octal\_use:NTF. 

\iffalse
\begin{verbatim}
\end{verbatim}
\fi

\TeX{X}hackers note: This function will fail if the escape character is an octal digit. We are thus careful to set the escape character to a known value before using it. \TeX{} dutifully detects octal digits for us: if \#1 is an octal digit, then the right-hand side of the comparison is ‘1#1’, greater than 1. Otherwise, the right-hand side stops as ‘1’, and the conditional takes the false branch.

\begin{verbatim}
\prg_new_conditional:Npnn \__str_octal_use:N #1 { TF }
{ \if_int_compare:w 1 < \token_to_str:N #1 \exp_stop_f:
\exp_after:wN \__str_if_contains_char_true:
\else:
\prg_break:n { \prg_return_false: }
\fi:
\end{verbatim}

\textit{End of definition for} \__str\_octal\_use:NTF.
\__str_hexadecimal_use:NTF \TeX detects uppercase hexadecimal digits for us (see \__str_octal_use:NTF), but not the lowercase letters, which we need to detect and replace by their uppercase counterpart.

\begin{verbatim}
\prg_new_conditional:Npnn \__str_hexadecimal_use:N #1 { TF }
  { \if_int_compare:w 1 < "1 \token_to_str:N #1 \exp_stop_f:
    \prg_return_true:
  \else:
    \if_case:w \int_eval:n { \exp_after:wN ' \token_to_str:N #1 - 'a }
      A \or: B \or: C \or: D \or: E \or: F \else:
        \prg_return_false:
    \fi:
    \fi:
  }
\end{verbatim}

\textit{(End of definition for \__str_hexadecimal_use:NTF.)}

56.3 Conversions

56.3.1 Producing one byte or character

\c__str_byte_0_tl \c__str_byte_1_tl \c__str_byte_255_tl \c__str_byte_-1_tl

For each integer \( N \) in the range \([0, 255]\), we create a constant token list which holds three character tokens with category code other: the character with character code \( N \), followed by the representation of \( N \) as two hexadecimal digits. The value \(-1\) is given a default token list which ensures that later functions give an empty result for the input \(-1\).

\begin{verbatim}
\group_begin:
  \__kernel_tl_set:Nx \l__str_internal_tl { \tl_to_str:n { 0123456789ABCDEF } }
  \tl_map_inline:Nn \l__str_internal_tl
    \tl_map_inline:Nn \l__str_internal_tl
      \tl_const:ce { c__str_byte_\int_eval:n { "#1##1} _tl }
        \char_generate:nn { "#1##1 } { 12 } #1 ##1 }
  \tl_map_inline:Nn \l__str_internal_tl
  \tl_const:cn { c__str_byte_-1_tl } { { } \use_none:n { } }
\group_end:
\end{verbatim}

\textit{(End of definition for \c__str_byte_0_tl and others.)}

\c__str_output_byte:n \c__str_output_byte:w \c__str_output_hexadecimal:n \c__str_output_end:

Those functions must be used carefully: feeding them a value outside the range \([-1, 255]\] will attempt to use the undefined token list variable \c__str_byte\( \langle \text{number} \rangle \)_tl. Assuming that the argument is in the right range, we expand the corresponding token list, and pick either the byte (first token) or the hexadecimal representations (second and third tokens). The value \(-1\) produces an empty result in both cases.
Convert a number in the range \([0, 65535]\) to a pair of bytes, either big-endian or little-endian.

\begin{verbatim}
\cs_new:Npn \__str_output_byte_pair_be:n #1
\cs_new:Npn \__str_output_byte_pair_le:n #1
\cs_new:Npn \__str_output_byte_pair:nnN #1#2#3
\end{verbatim}

(End of definition for \__str_output_byte:n and others.)

56.3.2 Mapping functions for conversions

This maps the function \#1 over all characters in \texttt{g\_str\_result\_tl}, which should be a byte string in most cases, sometimes a native string.

\begin{verbatim}
\cs_new_protected:Npn \__str_convert_gmap:N #1
\cs_new_protected:Npn \__str_convert_gmap_loop:N
\end{verbatim}

758
This maps the function #1 over all character codes in \texttt{\g__str_result_tl}, which must be in the internal representation.

\begin{verbatim}
\cs_new_protected:Npn \__str_convert_gmap_internal:N #1
{ \__kernel_tl_gset:Nx \g__str_result_tl
  \exp_after:wN \__str_convert_gmap_internal_loop:Nww
  \exp_after:wN #1
  \g__str_result_tl \s__str \s__str_stop \prg_break: \s__str
  \prg_break_point:
}
\end{verbatim}

(End of definition for \texttt{\__str_convert_gmap_internal:N} and \texttt{\__str_convert_gmap_internal_loop:Nw}.)

\subsection*{56.3.3 Error-reporting during conversion}

When converting using the function \texttt{\str_set_convert:Nnnn}, errors should be reported to the user after each step in the conversion. Errors are signalled by raising some flag (typically \texttt{@@_error}), so here we test that flag: if it is raised, give the user an error, otherwise remove the arguments. On the other hand, in the conditional functions \texttt{\str_set_convert:NnnnTF}, errors should be suppressed. This is done by changing \texttt{\__str_if_flag_error:Nne} into \texttt{\__str_if_flag_no_error:Nne} locally.

\begin{verbatim}
\cs_new_protected:Npn \__str_if_flag_error:Nne #1
{ \flag_if_raised:NTF #1
  { \msg_error:nne { str } }
  { \use_none:nn }
}
\end{verbatim}

\begin{verbatim}
\cs_new_protected:Npn \__str_if_flag_no_error:Nne #1#2#3
{ \flag_if_raised:NT #1 { \bool_gset_true:N \g__str_error_bool } }
\end{verbatim}

(End of definition for \texttt{\__str_if_flag_error:Nne} and \texttt{\__str_if_flag_no_error:Nne}.)

\begin{verbatim}
\__str_if_flag_times:NT
\end{verbatim}

At the end of each conversion step, we raise all relevant errors as one error message, built on the fly. The height of each flag indicates how many times a given error was encountered. This function prints #2 followed by the number of occurrences of an error if it occurred, nothing otherwise.
56.3.4 Framework for conversions

Most functions in this module expect to be working with “native” strings. Strings can also be stored as bytes, in one of many encodings, for instance `utf8`. The bytes themselves can be expressed in various ways in terms of \TeX tokens, for instance as pairs of hexadecimal digits. The questions of going from arbitrary Unicode code points to bytes, and from bytes to tokens are mostly independent.

Conversions are done in four steps:

- “unescape” produces a string of bytes;
- “decode” takes in a string of bytes, and converts it to a list of Unicode characters in an internal representation, with items of the form
  \[
  \text{(bytes) } \backslash s\_\text{str} \langle \text{Unicode code point} \rangle \backslash s\_\text{str}
  \]
  where we have collected the \text{(bytes)} which combined to form this particular Unicode character, and the \text{(Unicode code point)} is in the range \([0, 10FFFF]\).
- “encode” encodes the internal list of code points as a byte string in the new encoding;
- “escape” escapes bytes as requested.

The process is modified in case one of the encoding is empty (or the conversion function has been set equal to the empty encoding because it was not found): then the unescape or escape step is ignored, and the decode or encode steps work on tokens instead of bytes. Otherwise, each step must ensure that it passes a correct byte string or internal string to the next step.

The input string is stored in `\g__str_result_tl`, then we: unescape and decode; encode and escape; exit the group and store the result in the user’s variable. The various conversion functions all act on `\g__str_result_tl`. Errors are silenced for the conditional functions by redefining `\_\_str_if_flag_error:Nne` locally.
\begin{verbatim}
{ \bool_gset_false:N \g__str_error_bool \\n  \__str_convert:nNNnnn \\n  { \cs_set_eq:NN \__str_if_flag_error:Nne \__str_if_flag_no_error:Nne } \\n  \tl_gset_eq:NN #1 {#2} {#3} {#4} \\n  \bool_if:NTF \g__str_error_bool \prg_return_false: \prg_return_true: \\n  } \\n\cs_new_protected:N \__str_convert:nNNnnn #1#2#3#4#5#6 \\n{ \\
  \group_begin: \\
  #1 \\
  \__kernel_tl_gset:Nx \g__str_result_tl { \__kernel_str_to_other_fast:n \prg_do_nothing: } \\
  \exp_after:wN \__str_convert_decode_: \\
  \exp_after:wN \__str_convert_encode_: \\
  \exp_after:wN \__str_convert_encode_decode_: \\
  \exp_after:wN \__str_convert:wwwnn \\
  { \tl_to_str:n \prg_do_nothing: } \\
  { \tl_to_str:n \use_ii_i:nn } \\
  \__kernel_tl_gset:Nx \g__str_result_tl \\
  { \tl_to_str:V \g__str_result_tl } \\
  \group_end: \\
  #2 #3 \g__str_result_tl 
\end{verbatim}

(End of definition for \str_set_convert:Nnnn and others. These functions are documented on page 144.)

The task of \_\_str_convert:wwwnn is to split \langle\text{encoding}\rangle/\langle\text{escaping}\rangle pairs into their components, \#1 and \#2. Calls to \_\_str_convert:nnn ensure that the corresponding conversion functions are defined. The third auxiliary does the main work.

- \#1 is the encoding conversion function;
- \#2 is the escaping function;
- \#3 is the escaping name for use in an error message;
- \#4 is \prg_do_nothing: for unescaping/decoding, and \use_ii_i:nn for encoding/escaping;
- \#5 is the default encoding function (either “decode” or “encode”), for which there should be no escaping.

Let us ignore the native encoding for a second. In the unescaping/decoding phase, we want to do \#2\#1 in this order, and in the encoding/escaping phase, the order should be reversed: \#4\#2\#1 does exactly that. If one of the encodings is the default (native), then the escaping should be ignored, with an error if any was given, and only the encoding, \#1, should be performed.
\__str_convert:nnn {enc} {#4} {#1}
\__str_convert:nnn {esc} {#5} {#2}
\exp_args:Ncc \__str_convert:NNnNN
\__str_convert:_#4_#1: \__str_convert:_#5_#2: {#2}
\cs_new_protected:Npn \__str_convert:NNnNN #1#2#3#4#5
{
\if_meaning:w #1 #5
\tl_if_empty:nF {#3}
{ \msg_error:nne { str } { native-escaping } {#3} }
#1
\else:
#4 #2 #1
\fi:
}

(End of definition for \__str_convert:wwnn and \__str_convert:NNnNN.)

The arguments of \__str_convert:nnn are: enc or esc, used to build filenames, the
type of the conversion (unescape, decode, encode, escape), and the encoding or escaping
name. If the function is already defined, no need to do anything. Otherwise, filter out
all non-alphanumerics in the name, and lowercase it. Feed that, and the same three
arguments, to \__str_convert:nnnn. The task is then to make sure that the conversion
function #3_#1 corresponding to the type #3 and filtered name #1 is defined, then set
our initial conversion function #3_#4 equal to that.

How do we get the #3_#1 conversion to be defined if it isn’t? Two main cases.
First, if #1 is a key in \g__str_alias_prop, then the value \l__str_internal_tl
tells us what file to load. Loading is skipped if the file was already read, i.e., if the
conversion command based on \l__str_internal_tl already exists. Otherwise, try to
load the file; if that fails, there is an error, use the default empty name instead.

Second, #1 may be absent from the property list. The \cs_if_exist:cF test is
automatically false, and we search for a file defining the encoding or escaping #1 (this
should allow third-party .def files). If the file is not found, there is an error, use the
default empty name instead.

In all cases, the conversion based on \l__str_internal_tl is defined, so we can
set the #3_#1 function equal to that. In some cases (e.g., utf16be), the #3_#1 function
is actually defined within the file we just loaded, and it is different from the \l__str_-internal_tl-based function: we mustn’t clobber that different definition.

\cs_new_protected:Npn \__str_convert:nnn #1#2#3
{\cs_if_exist:cF { \__str_convert:_#2_#3: }
{\exp_args:Ne \__str_convert:nnn
{ \__str_convert_lowercase_alphanum:n {#3} }
{#1} {#2} {#3}
}
}
\cs_new_protected:Npn \__str_convert:nnnn #1#2#3#4#5
{\cs_if_exist:cF { \__str_convert:_#3_#1: }
{\prop_get:NnNF \g__str_alias_prop {#1} \l__str_internal_tl
{ \tl_set:Nn \l__str_internal_tl {#1} }
762
This function keeps only letters and digits, with upper case letters converted to lower case.
\cs_new:Npn \__str_convert_lowercase_alphanum:n #1
\exp_after:wN \__str_convert_lowercase_alphanum_loop:N
\tl_to_str:n {#1} { ? \prg_break: }
\prg_break_point:
\cs_new:Npn \__str_convert_lowercase_alphanum_loop:N #1
\use_none:n #1
\if_int_compare:w '#1 > 'Z \exp_stop_f:
  \if_int_compare:w '#1 > 'z \exp_stop_f: \else:
    #1
  \fi:
  \fi:
\else:
  \__str_output_byte:n { '#1 + 'a - 'A }
\fi:
\fi:
\else:
  \__str_output_byte:n { '#1 + 'a - 'A }
\fi:
\fi:
\__str_convert_lowercase_alphanum_loop:N
56.3.5 Byte unescape and escape

Strings of bytes may need to be stored in auxiliary files in safe “escaping” formats. Each such escaping is only loaded as needed. By default, on input any non-byte is filtered out, while the output simply consists in letting bytes through.

In the case of 8-bit engines, every character is a byte. For Unicode-aware engines, test the character code; non-bytes cause us to raise the flag \l__str_byte_flag. Spaces have already been given the correct category code when this function is called.

The simplest unescaping method removes non-bytes from \g__str_result_tl.

(End of definition for \__str_convert_lowercase_alphanum:n and \__str_convert_lowercase_alphanum_loop:N.)

(End of definition for \__str_filter_bytes:n and \__str_filter_bytes_aux:N.)

(End of definition for \__str_convert_unescape_: and \__str_convert_unescape_bytes:)
The simplest form of escape leaves the bytes from the previous step of the conversion unchanged.

56.3.6 Native strings

Convert each character to its character code, one at a time.

The conversion from an internal string to native character tokens basically maps \texttt{char\_generate:nn} through the code-points, but in non-Unicode-aware engines we use a fall-back character ? rather than nothing when given a character code outside \([0, 255]\). We detect the presence of bad characters using a flag and only produce a single error after the e-expanding assignment.
Character code too large for this engine.

This engine only supports 8-bit characters: valid character codes are in the range \([0, 255]\). To manipulate arbitrary Unicode, use LuaTeX or XeTeX.

(End of definition for \texttt{\_\_str_convert_encode_} and \texttt{\_\_str_encode_native_char:n}.)

\section*{56.3.7 \texttt{clist}}

Convert each integer to the internal form. We first turn \texttt{\_\_str_result_tl} into a clist variable, as this avoids problems with leading or trailing commas.

\begin{verbatim}
\cs_new_protected:Npn \_\_str_convert_decode_clist:n #1
{ \_\_str_decode_clist_char:n #1 }
\end{verbatim}

(End of definition for \texttt{\_\_str_convert_decode_clist:} and \texttt{\_\_str_decode_clist_char:n}.)

\subsection*{56.3.8 8-bit encodings}

It is not clear in what situations 8-bit encodings are used, hence it is not clear what should be optimized. The current approach is reasonably efficient to convert long strings, and it scales well when using many different encodings.

The data needed to support a given 8-bit encoding is stored in a file that consists of a single function call

\begin{verbatim}
\_\_str_declare_eight_bit_encoding:n\_\_nnnn \{\langle name\rangle\} \{\langle modulo\rangle\} \{\langle mapping\rangle\} \{\langle missing\rangle\}
\end{verbatim}
This declares the encoding \langle name \rangle to map bytes to Unicode characters according to the \langle mapping \rangle, and map those bytes which are not mentioned in the \langle mapping \rangle either to the replacement character (if they appear in \langle missing \rangle), or to themselves. The \langle mapping \rangle argument is a token list of pairs \{ \langle byte \rangle \} \{ \langle Unicode \rangle \} expressed in uppercase hexadecimal notation. The \langle missing \rangle argument is a token list of \{ \langle byte \rangle \}. Every \langle byte \rangle which does not appear in the \langle mapping \rangle nor the \langle missing \rangle lists maps to itself in Unicode, so for instance the latin1 encoding has empty \langle mapping \rangle and \langle missing \rangle lists. The \langle modulo \rangle is a (decimal) integer between 256 and 558 inclusive, modulo which all Unicode code points supported by the encodings must be different.

We use two integer arrays per encoding. When decoding we only use the decode integer array, with entry \( n + 1 \) (offset needed because integer array indices start at 1) equal to the Unicode code point that corresponds to the \( n \)-th byte in the encoding under consideration, or \(-1\) if the given byte is invalid in this encoding. When encoding we use both arrays: upon seeing a code point \( n \), we look up the entry (1 plus) \( n \) modulo some \( M \) in the encode array, which tells us the byte that might encode the given Unicode code point, then we check in the decode array that indeed this byte encodes the Unicode code point we want. Here, \( M \) is an encoding-dependent integer between 256 and 558 (it turns out), chosen so that among the Unicode code points that can be validly represented in the given encoding, no pair of code points have the same value modulo \( M \).

Loop through both lists of bytes to fill in the decode array accordingly. For bytes that are invalid in the given encoding, store \(-1\) in the decode array.
\int_mod:nn { \intarray_item:Nn #1 { 1 + ##1 } }
{ \intarray_count:N #2 }

\{##1\}
\}
\}
\cs_new_protected:Npn \__str_declare_eight_bit_loop:Nnn #1#2#3
{\__str_use_none_delimit_by_s_stop:w #2 \s__str_stop
\intarray_gset:Nnn #1 { 1 + "#2 } { "#3 }
\__str_declare_eight_bit_loop:Nnn #1}
\}
\cs_new_protected:Npn \__str_declare_eight_bit_loop:Nn #1#2
{\__str_use_none_delimit_by_s_stop:w #2 \s__str_stop
\intarray_gset:Nnn #1 { 1 + "#2 } { -1 }
\__str_declare_eight_bit_loop:Nn #1}
\}
(End of definition for \__str_declare_eight_bit_encoding:nnnn and others.)

The map from bytes to Unicode code points is in the decode array corresponding to the given encoding. Define \__str_tmp:w and pass it successively all bytes in the string. It produces an internal representation with suitable \s__str inserted, and the corresponding code point is obtained by looking it up in the integer array. If the entry is \-1 then issue a replacement character and raise the flag indicating that there was an error.

\cs_set:Npe \__str_tmp:w
{\exp_not:N \__str_decode_eight_bit_aux:Nn \exp_not:c { g__str_decode_#1_intarray }
\flag_clear:N \l__str_error_flag
\__str_convert_gmap:N \__str_tmp:w
\__str_if_flag_error:Nne \l__str_error_flag { decode-8-bit } {#1}
\}
\cs_new:Npn \__str_decode_eight_bit_loop:Nn #1#2
{\__str_decode_eight_bit_aux:n #2 \s__str
\exp_args:Nf \__str_decode_eight_bit_aux:n { \intarray_item:Nn #1 { 1 + '#2 } }
\s__str
\}
\cs_new:Npn \__str_decode_eight_bit_aux:n #1
{\if_int_compare:w #1 < \c_zero_int
\flag_raise:N \l__str_error_flag
\int_value:w \c__str_replacement_char_int
\else:
#1
\fi:
}
It is not practical to make an integer array with indices in the full Unicode range, so we work modulo some number, which is simply the size of the \texttt{encode} integer array for the given encoding. This gives us a candidate byte for representing a given Unicode code point. Of course taking the modulo leads to collisions so we check in the \texttt{decode} array that the byte we got is indeed correct. Otherwise the Unicode code point we started from is simply not representable in the given encoding.

```latex
\begin{verbatim}
\int_new:N \l__str_modulo_int
\cs_new_protected:Npn \__str_convert_encode_eight_bit:n #1
{\cs_set:Npe \__str_tmp:w {\exp_not:N \__str_encode_eight_bit_aux:NNn
\exp_not:c { g__str_encode_#1_intarray }
\exp_not:c { g__str_decode_#1_intarray }}
\flag_clear:N \l__str_error_flag
\__str_convert_gmap_internal:N \__str_tmp:w
\__str_if_flag_error:Nne \l__str_error_flag { encode-8-bit } {#1}}
\cs_new:Npn \__str_encode_eight_bit_aux:NNn #1#2#3
{\exp_args:Nf \__str_encode_eight_bit_aux:nnN
{\intarray_item:Nn #1 { 1 + \int_mod:nn {#3} { \intarray_count:N #1 } }}
#2}
\cs_new:Npn \__str_encode_eight_bit_aux:nnN #1#2#3
{\int_compare:nNnTF { \intarray_item:Nn #3 { 1 + #1 } } = #2
{ \__str_output_byte:n {#1} }
{ \flag_raise:N \l__str_error_flag }}
\end{verbatim}
```

(End of definition for \verb|\__str_convert_encode_eight_bit:n|, \verb|\__str_encode_eight_bit_aux:n|, and \verb|\__str_encode_eight_bit_aux:Nn|.)

56.4 Messages

General messages, and messages for the encodings and escapings loaded by default (“native”, and “bytes”).
Since native strings do not consist in bytes, none of the escaping methods make sense. The specified escaping, "#1", will be ignored.

```
\msg_new:nnn { str } { file-not-found }
\{ File-'l3str-#1.def'-not-found. }
```

Message used when the “bytes” unescaping fails because the string given to \str_set_convert:Nnnn contains a non-byte. This cannot happen for the -8-bit engines. Messages used for other escapings and encodings are defined in each definition file.

```
\bool_lazy_any:nT
\sys_if_engine_luatex_p:
\sys_if_engine_xetex_p:
\}
\msg_new:nnnn { str } { non-byte }
\{ String-invalid-in-escaping-'#1':-it-may-only-contain-bytes. }
\{ Some-characters-in-the-string-you-asked-to-convert-are-not-8-bit-characters.-Perhaps-the-string-is-a-'native'-Unicode-string?-If-it-is,-try-using\}
\\ \}\io_indent:n
\{ \io_char:N:\str_set_convert:Nnnn \}
\\ <str-var>-\{<-string>-\}-\{<-target-encoding>-\}
```

Those messages are used when converting to and from 8-bit encodings.

```
\msg_new:nnnn { str } { decode-8-bit }
\{ Invalid-string-in-encoding-'#1'. }
\LaTeX-came-across-a-byte-which-is-not-defined-to-represent-any-character-in-the-encoding-'#1'.
```

```
\msg_new:nnnn { str } { encode-8-bit }
\{ Unicode-string-cannot-be-converted-to-encoding-'#1'. }
\LaTeX-was-asked-to-convert-a-string-to-that-encoding,-but-that-string-contains-a-character-that-'#1'-does-not-support.
```

56.5 Escaping definitions

Several of those encodings are defined by the pdf file format. The following byte storage methods are defined:

- **bytes** (default), non-bytes are filtered out, and bytes are left untouched (this is defined by default);
- hex or hexadecimal, as per the pdfTeX primitive \pdfescapehex
- name, as per the pdfTeX primitive \pdfescapename
- string, as per the pdfTeX primitive \pdfescapestring
- url, as per the percent encoding of urls.

### 56.5.1 Unescape methods

Take chars two by two, and interpret each pair as the hexadecimal code for a byte. Anything else than hexadecimal digits is ignored, raising the flag. A string which contains an odd number of hexadecimal digits gets 0 appended to it: this is equivalent to appending a 0 in all cases, and dropping it if it is alone.
\_\_str_convert_unescape_name: function replaces each occurrence of \# followed by two hexadecimal digits in \g__str_result_tl by the corresponding byte. The url function is identical, with escape character % instead of \#. Thus we define the two together. The arguments of \_\_str_tmp:w are the character code of \# or % in hexadecimal, the name of the main function to define, and the name of the auxiliary which performs the loop.

The \_\_str_convert_unescape_name: function replaces each occurrence of \# followed by two hexadecimal digits in \g__str_result_tl by the corresponding byte. The url function is identical, with escape character % instead of \#. Thus we define the two together. The arguments of \_\_str_tmp:w are the character code of \# or % in hexadecimal, the name of the main function to define, and the name of the auxiliary which performs the loop.

The looping auxiliary \#3 finds the next escape character, reads the following two characters, and tests them. The test \_\_str_hexadecimal_use:NTF leaves the uppercase digit in the input stream, hence we surround the test with \_\_str_output_byte:w " and \_\_str_output_end:. If both characters are hexadecimal digits, they should be removed before looping: this is done by use_i:nnn. If one of the characters is not a hexadecimal digit, then feed \^{#1} to \_\_str_output_byte:w to produce the escape character, raise the flag, and call the looping function followed by the two characters (remove use_i:nnn).

\cs_set_protected:Npn \_\_str_tmp:w \#1\#2\#3
\cs_new_protected:cpn { \_\_str_convert_unescape_#2: } 
\group_begin: 
\flag_clear:N \l__str_byte_flag
\flag_clear:N \l__str_error_flag
\int_set:Nn \tex_escapechar:D { 92 }
\__kernel_tl_gset:Nx \g__str_result_tl { \exp_after:wN #3 \g__str_result_tl \#1 ? { ? \prg_break: } \prg_break_point: }
\__str_if_flag_error:Nne \l__str_byte_flag { non-byte } { \#2 }
\__str_if_flag_error:Nne \l__str_error_flag { unescape-\#2 } { } \group_end: 
\cs_new:Npn \#3 \##1\#1\#2\#3
{ \_\_str_filter_bytes:n {\##1} \use_none:n \#3 \_\_str_output_byte:w " \_\_str_hexadecimal_use:NTF \#2 
{ \_\_str_hexadecimal_use:NTF \#3 
} \flag_raise:N \l__str_error_flag

772
The string escaping is somewhat similar to the name and url escapings, with escape character \. The first step is to convert all three line endings, \^^J, \^^M, and ^^^M\^^J to the common ^^^J, as per the PDF specification. This step cannot raise the flag.

Then the following escape sequences are decoded.

- \n Line feed (10)
- \r Carriage return (13)
- \t Horizontal tab (9)
- \b Backspace (8)
- \f Form feed (12)
- \( Left parenthesis
- \) Right parenthesis
- \ Backslash

\ddd (backslash followed by 1 to 3 octal digits) Byte ddd (octal), subtracting 256 in case of overflow.

If followed by an end-of-line character, the backslash and the end-of-line are ignored. If followed by anything else, the backslash is ignored, raising the error flag.
\group_begin:  
\flag_clear:N \l__str_byte_flag  
\flag_clear:N \l__str_error_flag  
\int_set:Nn \tex_escapechar:D { 92 }  
\__kernel_tl_gset:Nx \g__str_result_tl  
{  
\exp_after:wN \__str_unescape_string_newlines:wN  
\g__str_result_tl \prg_break: ^^M ?  
\prg_break_point:  
}  
\__kernel_tl_gset:Nx \g__str_result_tl  
{  
\exp_after:wN \__str_unescape_string_loop:wNNN  
\g__str_result_tl #1 ?? { ? \prg_break: }  
\prg_break_point:  
}  
\__str_if_flag_error:Nne \l__str_byte_flag { non-byte } { string }  
\__str_if_flag_error:Nne \l__str_error_flag { unescape-string } { }  
\group_end:  
\exp_args:No \__str_tmp:w { \c_backslash_str }  
\exp_last_unbraced:NNNN  
\cs_new:Npn \__str_unescape_string_loop:wNNN #1 \c_backslash_str #2#3#4  
{  
\__str_filter_bytes:n {#1}  
\use_none:n #4  
\__str_output_byte:w ,  
\__str_octal_use:NTF #2  
{  
\__str_octal_use:NTF #3  
{  
\__str_octal_use:NTF #4  
{  
\if_int_compare:w #2 > 3 \exp_stop_f:  
 - 256  
\fi:  
\__str_unescape_string_repeat:NNNNNNN  
\__str_unescape_string_repeat:NNNNNNN ? }  
\__str_unescape_string_repeat:NNNNNNN ? }  
\__str_unescape_string_repeat:NNNNNNN ? }  
\str_case_e:mmF {#2}  
{  
{ \c_backslash_str } { 134 }  
{ ( ) } { 50 }  
{ ) } { 51 }  
{ r } { 15 }  
{ f } { 14 }  
{ n } { 12 }  
{ t } { 11 }  
\group_end:
56.5.2 Escape methods

Currently, none of the escape methods can lead to errors, assuming that their input is made out of bytes.

Loop and convert each byte to hexadecimal.
\begin{verbatim}
\cs_new_protected:Npn \__str_convert_escape_hex:
{ \__str_convert_gmap:N \__str_escape_hex_char:N }
\end{verbatim}

For each byte, test whether it should be output as is, or be “hash-encoded”. Roughly, bytes outside the range \([2A,7E]\) are hash-encoded. We keep two lists of exceptions: characters in \c__str_escape_name_not_str are not hash-encoded, and characters in the \c__str_escape_name_str are encoded.

\begin{verbatim}
\str_const:Nn \c__str_escape_name_not_str { ! " $ & ' } %$
\end{verbatim}

\begin{verbatim}
\\str_const:Nn \c__str_escape_name_str { {}/<\} ] }
\end{verbatim}
Any character below (and including) space, and any character above (and including) del, are converted to octal. One backslash is added before each parenthesis and backslash.

\__str_convert_escape_string:
\__str_convert_escape_string_char:N
\__str_if_escape_string:NTF
\c_str_escape_string_str
\c_backslash_str ( )
\prg_new_protected:Npn \__str_convert_escape_string:
\cs_new:Npn \__str_escape_string_char:N #1
{ \__str_if_escape_string:NTF #1
{ \__str_if_contains_char:NnT \c__str_escape_string_str {#1}
{ \c_backslash_str }
#1
}
{ \c_backslash_str
\int_div_truncate:nn {'#1} {64}
\int_mod:nn { \int_div_truncate:nn {'#1} {8} } {8}
\int_mod:nn {'#1} {8}
}
}
\prg_new_protected:Npn \__str_escape_string_char:N #1
{ \__str_if_escape_string:NTF #1
{ \__str_if_contains_char:NnT \c__str_escape_string_str {#1}
{ \c_backslash_str }
#1
}
}
56.6 Encoding definitions

The native encoding is automatically defined. Other encodings are loaded as needed. The following encodings are supported:

- UTF-8;
- UTF-16, big-, little-endian, or with byte order mark;
- UTF-32, big-, little-endian, or with byte order mark;
- the iso 8859 code pages, numbered from 1 to 16, skipping the inexistent ISO 8859-12.

56.6.1 utf-8 support

Loop through the internal string, and convert each character to its UTF-8 representation. The representation is built from the right-most (least significant) byte to the left-most (most significant) byte. Continuation bytes are in the range [128, 191], taking 64 different values, hence we roughly want to express the character code in base 64, shifting the first digit in the representation by some number depending on how many continuation bytes there are. In the range [0, 127], output the corresponding byte directly. In the range [128, 2047], output the remainder modulo 64, plus 128 as a continuation byte, then output the quotient (which is in the range [0, 31]), shifted by 192. In the next range, [2048, 65535], split the character code into residue and quotient modulo 64, output the...
residue as a first continuation byte, then repeat; this leaves us with a quotient in the range $[0, 15]$, which we output shifted by 224. The last range, $[65536, 1114111]$, follows the same pattern: once we realize that dividing twice by 64 leaves us with a number larger than 15, we repeat, producing a last continuation byte, and offset the quotient by 240 for the leading byte.

How is that implemented? \texttt{\_\_str\_encode\_utf\_viii\_loop:wwnnw} takes successive quotients as its first argument, the quotient from the previous step as its second argument (except in step 1), the bound for quotients that trigger one more step or not, and finally the offset used if this step should produce the leading byte. Leading bytes can be in the ranges $[0, 127]$, $[192, 223]$, $[224, 239]$, and $[240, 247]$ (really, that last limit should be 244 because Unicode stops at the code point 1114111). At each step, if the quotient \#1 is less than the limit \#3 for that range, output the leading byte (\#1 shifted by \#4) and stop. Otherwise, we need one more step: use the quotient of \#1 by 64, and \#1 as arguments for the looping auxiliary, and output the continuation byte corresponding to the remainder $\#2 - 64 \#1 + 128$. The bizarre construction $- 1 + 0 \ast$ removes the spurious initial continuation byte (better methods welcome).

\begin{verbatim}
\cs_new_protected:cpn { \_\_str\_convert\_encode\_utf8: }
{ \_\_str\_convert\_gmap\_internal:N \_\_str\_encode\_utf\_viii\_char:n }
\cs_new:Npn \_\_str\_encode\_utf\_viii\_char:n #1
{ \_\_str\_encode\_utf\_viii\_loop:wwnnw #1 ; - 1 + 0 \ast ;
  { 128 } { 0 }
  { 32 } { 192 }
  { 16 } { 224 }
  { 8 } { 240 }
  \s__str\_stop }
\cs_new:Npn \_\_str\_encode\_utf\_viii\_loop:wwnnw #1; #2; #3#4 #5 \s__str\_stop
{ \if_int_compare:w #1 < #3 \exp_stop_f:
  \_\_str\_output\_byte:n { #1 + #4 }
  \exp_after:wN \_\_str\_use\_none\_delimit\_by\_s\_stop:w
  \fi:
  \exp_after:wN \_\_str\_encode\_utf\_viii\_loop:wwnnw
  \int_value:w \int_div_truncate:nn {#1} {64} ; #1 ;
  #5 \s__str\_stop
\_\_str\_output\_byte:n { #2 - 64 \ast ( #1 - 2 ) }
}
\end{verbatim}

(End of definition for \texttt{\_\_str\_convert\_encode\_utf8:}, \texttt{\_\_str\_encode\_utf\_viii\_char:n}, and \texttt{\_\_str\_encode\_utf\_viii\_loop:wwnnw}.)

When decoding a string that is purportedly in the utf-8 encoding, four different errors can occur, signalled by a specific flag for each (we define those flags using \texttt{\flag\_clear\_new:N} rather than \texttt{\flag\_new:N}, because they are shared with other encoding definition files).

- **“Missing continuation byte”:** a leading byte is not followed by the right number of continuation bytes.
- **“Extra continuation byte”:** a continuation byte appears where it was not expected, i.e., not after an appropriate leading byte.

778
• "Overlong": a Unicode character is expressed using more bytes than necessary, for instance, "C0\x80" for the code point 0, instead of a single null byte.

• "Overflow": this occurs when decoding produces Unicode code points greater than 1114111.

We only raise one \LaTeX3 error message, combining all the errors which occurred. In the short message, the leading comma must be removed to get a grammatically correct sentence. In the long text, first remind the user what a correct utf-8 string should look like, then add error-specific information.

\begin{verbatim}
\msg_new:nnnn { str } { utf8-decode }
{ \texttt{Invalid-UTF-8-string:}
  \exp_last_unbraced:Nf \use_none:n
  \begin{quote}
  \begin{itemize}
  {\item \texttt{\__str_if_flag_times:NT \l__str_missing_flag { ,missing-continuation-byte }}
  \item \texttt{\__str_if_flag_times:NT \l__str_extra_flag { ,extra-continuation-byte }}
  \item \texttt{\__str_if_flag_times:NT \l__str_overlong_flag { ,overlong-form }}
  \item \texttt{\__str_if_flag_times:NT \l__str_overflow_flag { ,code-point-too-large }}
  \end{itemize}
  \end{quote}
  \end{verbatim}

In-the-UTF-8-encoding, each Unicode character consists in 1-to-4-bytes, with the following bit-pattern: \texttt{10xxxxxx} are called continuation-bytes. Bytes of the form 10xxxxxx are called continuation-bytes.

\begin{verbatim}
\msg_new:nnnn { str } { utf8-decode }
\begin{quote}
\begin{itemize}
  \item Code-point-<128:0xxxxxxx \item Code-point-<2048:110xxxxx~10xxxxxx \item Code-point-<65536:1110xxxx~10xxxxxx~10xxxxxx \item Code-point-<1114112:11110xxx~10xxxxxx~10xxxxxx~10xxxxxx
\end{itemize}
\end{quote}
\end{verbatim}

Every-Unicode-code-point-must-be-expressed-in-the-shortest-possible-form. For instance, '-0x0C0'-'0x83'-is-not-a-valid-representation-for-the-code-point-3.
Decoding is significantly harder than encoding. As before, lower some flags, which are tested at the end (in bulk, to trigger at most one \TeX error, as explained above).

We expect successive multi-byte sequences of the form \((\text{start byte}) (\text{continuation bytes})\). The \_\_\_start auxiliary tests the first byte:

- \([0,7F]\): the byte stands alone, and is converted to its own character code;
- \([80, BF]\): unexpected continuation byte, raise the appropriate flag, and convert that byte to the replacement character \*FFFD;
- \([C0, FF]\): this byte should be followed by some continuation byte(s).

In the first two cases, \use\_none\_delimit\_by\_q\_stop: removes data that only the third case requires, namely the limits of ranges of Unicode characters which can be expressed with 1, 2, 3, or 4 bytes.

We can now concentrate on the multi-byte case and the \_\_\_continuation auxiliary. We expect \#3 to be in the range \(\text{["80,"BF]}\). The test for this goes as follows: if the character code is less than \"80, we compare it to \"C0, yielding false; otherwise to \"C0, yielding true in the range \["80,"BF\] and false otherwise. If we find that the byte is not a continuation range, stop the current slew of bytes, output the replacement character, and continue parsing with the \_\_\_start auxiliary, starting at the byte we just tested. Once we know that the byte is a continuation byte, leave it behind us in the input stream, compute what code point the bytes read so far would produce, and feed that number to the \_\_\_aux function.

The \_\_\_aux function tests whether we should look for more continuation bytes or not. If the number it receives as \#1 is less than the maximum \#4 for the current range, then we are done: check for an overlong representation by comparing \#1 with the maximum \#3 for the previous range. Otherwise, we call the \_\_\_continuation auxiliary again, after shifting the “current code point” by \#4 (maximum from the range we just checked).

Two additional tests are needed: if we reach the end of the list of range maxima and we are still not done, then we are faced with an overflow. Clean up, and again insert the code point \*FFFD for the replacement character. Also, every time we read a byte, we need to check whether we reached the end of the string. In a correct \UTF-8 string, this happens automatically when the \_\_\_start auxiliary leaves its first argument in the input stream: the end-marker begins with \prg\_break:, which ends the loop. On the other hand, if the end is reached when looking for a continuation byte, the \use\_\_none\_n \#3 construction removes the first token from the end-marker, and leaves the \_\_\_end auxiliary, which raises the appropriate error flag before ending the mapping.
\flag_clear:N \__str_error_flag
\flag_clear:N \__str_missing_flag
\flag_clear:N \__str_extra_flag
\flag_clear:N \__str_overlong_flag
\flag_clear:N \__str_overflow_flag
\kernel_tl_gset:Nx \g__str_result_tl
{
  \exp_after:wN \__str_decode_utf_viii_start:N \g__str_result_tl
  { \prg_break: \__str_decode_utf_viii_end: }
  \prg_break_point:
}
\__str_if_flag_error:Nne \l__str_error_flag { utf8-decode } { }
}
cs_new:Npn \__str_decode_utf_viii_start:N #1
{\if_int_compare:w #1 < "C0 \exp_stop_f:
  \s__str
  \if_int_compare:w #1 < "80 \exp_stop_f:
    \int_value:w #1
  \else:
    \flag_raise:N \l__str_extra_flag
    \flag_raise:N \l__str_error_flag
    \int_use:N \c__str_replacement_char_int
  \fi:
  \else:
    \exp_after:wN \__str_decode_utf_viii_continuation:wwN
    \int_value:w \int_eval:n { #1 - "C0 } \exp_after:wN
    \__str_decode_utf_viii_start:N
  \fi:
  \else:
    \exp_after:wN \__str_decode_utf_viii_continuation:w\W
    \int_value:w \int_eval:n { #1 - "80 } \s__str
    \__str_decode_utf_viii_start:N
  \fi:
}
cs_new:Npn \__str_decode_utf_viii_continuation:w\W
#1 \s__str #2 \__str_decode Utf_viii_start:N #3
{\if_int_compare:w #3 < "80 \exp_stop_f: - \fi:
  "C0 \exp_stop_f:
  #3
  \exp_after:wN \__str_decode_utf_viii_aux:w\W\W\W
  \int_value:w \int_eval:n { #1 + "40 + #3 } \exp_after:wN\W
  \else:
    \s__str
    \__str_decode_utf_viii_aux:w\W\W\W
  \fi:
}
cs_new:Npn \__str_decode_utf_viii_aux:w\W\W\W
#3
{\if_int_compare:w #3 < "80 \exp_stop_f: - \fi:
  "C0 \exp_stop_f:
  #3
  \exp_after:wN \__str_decode_utf_viii_aux:w\W\W\W
  \int_value:w \int_eval:n { #1 + "40 + #3 } \exp_after:wN\W
  \else:
    \s__str
    \__str_decode_utf_viii_aux:w\W\W\W
  \fi:
}
56.6.2 utf-16 support

The definitions are done in a category code regime where the bytes 254 and 255 used by the byte order mark have catcode 12.

When the endianness is not specified, it is big-endian by default, and we add a byte-order mark. Convert characters one by one in a loop, with different behaviours depending on the character code.

- [0, "D7FF]: converted to two bytes;
• [“D800,” “DFFF”] are used as surrogates: they cannot be converted and are replaced by the replacement character;
• [“E000,” “FFFF”]: converted to two bytes;
• [“10000,” “10FFFF”]: converted to a pair of surrogates, each two bytes. The magic ”D7C0" is "D800 – "10000/"100.

For the duration of this operation, \_\_str_temp:w is defined as a function to convert a number in the range [0, "FFFF"] to a pair of bytes (either big endian or little endian), by feeding the quotient of the division of #1 by "100, followed by #1 to \_\_str_encode_utf_xvi_be:nn or its le analog: those compute the remainder, and output two bytes for the quotient and remainder.

\cs_new_protected:cpn { \_\_str_convert_encode_utf16: }\{
\_\_str_encode_utf_xvi_aux:N \_\_str_output_byte_pair_be:n
\tl_gput_left:Ne \g__str_result_tl { ^^fe ^^ff }
\}
\cs_new_protected:cpn { \_\_str_convert_encode_utf16be: }\{
\_\_str_encode_utf_xvi_aux:N \_\_str_output_byte_pair_be:n 
\}\cs_new_protected:cpn { \_\_str_convert_encode_utf16le: }\{
\_\_str_encode_utf_xvi_aux:N \_\_str_output_byte_pair_le:n 
\}\cs_new_protected:Npn \_\_str_encode_utf_xvi_aux:N #1\{
\flag_clear:N \l__str_error_flag
\cs_set_eq:NN \__str_tmp:w #1
\__str_convert_gmap_internal:N \__str_encode_utf_xvi_char:n
\__str_if_flag_error:Nne \l__str_error_flag { utf16-encode } { }
\}
\cs_new:Npn \__str_encode_utf_xvi_char:n #1\{
\if_int_compare:w #1 < "D800 \exp_stop_f:
\__str_temp:w {#1}\n\else:
\if_int_compare:w #1 < "10000 \exp_stop_f:
\if_int_compare:w #1 < "E000 \exp_stop_f:
\flag_raise:N \l__str_error_flag
\__str_temp:w { \c__str_replacement_char_int }
\else:
\__str_temp:w {#1}\n\fi:
\else:
\exp_args:Nf \__str_temp:w { \int_div_truncate:nn {#1} {"400} + "D7C0
\exp_args:Nf \__str_temp:w { \int_mod:nn {#1} {"400} + "DC00 }
\fi:
\fi:
\}
(End of definition for \_\_str_convert_encode_utf16: and others.)

__str_missing
__str_extra
__str_end

When encoding a Unicode string to utf-16, only one error can occur: code points in the range [“D800,” “DFFF”], corresponding to surrogates, cannot be encoded. We use the all-purpose flag @\_error to signal that error.
When decoding a Unicode string which is purportedly in UTF-16, three errors can occur: a missing trail surrogate, an unexpected trail surrogate, and a string containing an odd number of bytes.

\texttt{\flag_clear_new:N \_str_missing_flag}
\texttt{\flag_clear_new:N \_str_extra_flag}
\texttt{\flag_clear_new:N \_str_end_flag}
\texttt{msg_new:nnnn \{ str \} \{ utf16-encode \}}
\texttt{\{ Unicode-string-cannot-be-expressed-in-UTF-16:-surrogate. \}}
\texttt{\{ Surrogate-code-points-(in-the-range-[U+D800,-U+DFFF])-}
\texttt{can-be-expressed-in-the-UTF-8-and-UTF-32-encodings,-}
\texttt{but-not-in-the-UTF-16-encoding. \}}
\texttt{msg_new:nnnn \{ str \} \{ utf16-decode \}}
\texttt{\{ Invalid-UTF-16-string: \}}
\texttt{\exp_last_unbraced:Nf \use_none:n}
\texttt{\{ \_str_if_flag_times:NT \_str_missing_flag \{ ,-missing-trail-surrogate \}}
\texttt{\_str_if_flag_times:NT \_str_extra_flag \{ ,extra-trail-surrogate \}}
\texttt{\_str_if_flag_times:NT \_str_end_flag \{ ,odd-number-of-bytes \}}
\texttt{\}}
\texttt{\}}
\texttt{In-the-UTF-16-encoding,-each-Unicode-character-is-encoded-as-}
\texttt{2-or-4-bytes: \}}
\texttt{\low-indent:n}
\texttt{\{ Code-point-in-[U+0000,-U+D7FF]:two-bytes \}}
\texttt{Code-point-in-[U+D800,-U+DFFF]:illegal \}}
\texttt{Code-point-in-[U+E000,-U+FFFF]:two-bytes \}}
\texttt{Code-point-in-[U+10000,-U+10FFFF]:-}
\texttt{a-lead-surrogate-and-a-trail-surrogate \}}
\texttt{Lead-surrogates-are-pairs-of-bytes-in-the-range-[0xD800,-0xDBFF],-}
\texttt{and-trail-surrogates-are-in-the-range-[0xDC00,-0xDFFF].}
\texttt{\flag_if_raised:NT \_str_missing_flag}
\texttt{\{ \}}
\texttt{A-lead-surrogate-was-not-followed-by-a-trail-surrogate.}
\texttt{\}}
\texttt{\flag_if_raised:NT \_str_extra_flag}
\texttt{\{ \}}
\texttt{LaTeX-came-across-a-trail-surrogate-when-it-was-not-expected.}
\texttt{\}}
\texttt{\flag_if_raised:NT \_str_end_flag}
\texttt{\{ \}}
\texttt{The-string-contained-an-odd-number-of-bytes.-This-is-invalid:-}
\texttt{the-basic-code-unit-for-UTF-16-is-16-bits-(2-bytes).}\texttt{\}}
As for UTF-8, decoding UTF-16 is harder than encoding it. If the endianness is unknown, check the first two bytes: if those are "FE and "FF in either order, remove them and use the corresponding endianness, otherwise assume big-endianess. The three endianness cases are based on a common auxiliary whose first argument is 1 for big-endian and 2 for little-endian, and whose second argument, delimited by the scan mark \s__str_stop, is expanded once (the string may be long; passing \g__str_result_tl as an argument before expansion is cheaper).

The \__str_decode_utf_xvi:Nw function defines \__str_tmp:w to take two arguments and return the character code of the first one if the string is big-endian, and the second one if the string is little-endian, then loops over the string using \__str_decode_utf_xvi_pair:NN described below.

Bytes are read two at a time. At this stage, \@@_tmp:w #1#2 expands to the character code of the most significant byte, and we distinguish cases depending on which range it lies in:
• [*D8,*DB] signals a lead surrogate, and the integer expression yields 1 (\texttt{\textasciitilde TpX}
rounds ties away from zero);
• [*DC,*DF] signals a trail surrogate, unexpected here, and the integer expression
yields 2;
• any other value signals a code point in the Basic Multilingual Plane, which stands
for itself, and the \texttt{\textbackslash if\_case:w} construction expands to nothing (cases other than 1 or 2),
leaving the relevant material in the input stream, followed by another call to the
\_pair auxiliary.

The case of a lead surrogate is treated by the \_quad auxiliary, whose arguments \#1, \#2, \#4
and \#5 are the four bytes. We expect the most significant byte of \#4\#5 to be in the range
[*DC,*DF] (trail surrogate). The test is similar to the test used for continuation bytes
in the \texttt{UTF-8} decoding functions. In the case where \#4\#5 is indeed a trail surrogate, leave
\#1\#2\#4\#5 \texttt{\textbackslash s\_str\{code point\}\textbackslash s\_str}, and remove the pair \#4\#5 before looping with
\texttt{\_str\_decode\_utf\_xvi\_pair:NN}. Otherwise, of course, complain about the missing
surrogate.

The magic number "D7F7 is such that "D7F7\*400 = "D800\*400+"DC00−"10000.
Every time we read a pair of bytes, we test for the end-marker \texttt{\_str\_nil}. When
reaching the end, we additionally check that the string had an even length. Also, if the
end is reached when expecting a trail surrogate, we treat that as a missing surrogate.

\texttt{\cs\_new:Npn \_str\_decode\_utf\_xvi\_pair:NN \#1\#2}

\{ \texttt{\if\meaning:w \_q\_str\_nil \#2}
\texttt{\_str\_decode\_utf\_xvi\_pair\_end:Nw \#1}
\texttt{\fi:}
\texttt{\if\case:w}
\texttt{\if\_int\_eval:n { ( \texttt{\_str\_tmp:w \#1\#2 - "D6\} / 4 \text{\scan\_stop:}}}
\texttt{\or: \exp\_after:wN \_str\_decode\_utf\_xvi\_quad:NNwNN}
\texttt{\or: \exp\_after:wN \_str\_decode\_utf\_xvi\_extra:NNw}
\texttt{\fi:}
\texttt{\#1\#2 \textbackslash s\_str}
\texttt{\_int\_eval:n { "100 * \texttt{\_str\_tmp:w \#1\#2 + \_str\_tmp:w \#2\#1 } \textbackslash s\_str}
\texttt{\_str\_decode\_utf\_xvi\_pair:NN}
\texttt{\_str\_decode\_utf\_xvi\_quad:NNwNN}
\texttt{\#1\#2 \#3 \_str\_decode\_utf\_xvi\_pair:NN \#4\#5}
\texttt{\}}
\texttt{\if\meaning:w \_q\_str\_nil \#5}
\texttt{\_str\_decode\_utf\_xvi\_error:nNN \{missing\} \#1\#2}
\texttt{\_str\_decode\_utf\_xvi\_pair:NN \#4}
\texttt{\fi:}
\texttt{\if\_int\_compare:w}
\texttt{\if\_int\_compare:w \_str\_tmp:w \#4\#5 < "DC \exp\_stop\_f:}
\texttt{0 = 1}
\texttt{\else:}
\texttt{\_str\_tmp:w \#4\#5 < "E0}
\texttt{\fi:}
\texttt{\exp\_stop\_f:}
\texttt{\#1 \#2 \#4 \#5 \textbackslash s\_str}
\texttt{\_int\_eval:n}
\texttt{\}}
Convert each integer in the comma-list \verb|\g__str_result_tl| to a sequence of four bytes. The functions for big-endian and little-endian encodings are very similar, but the \verb|\__str_output_byte:n| instructions are reversed.

\begin{verbatim}
\cs_new_protected:cpn { __str_convert_encode_utf32: }
\tl_gput_left:Ne \g__str_result_tl { ^^00 ^^00 ^^fe ^^ff }
\end{verbatim}

\subsection*{utf-32 support}

The definitions are done in a category code regime where the bytes 0, 254 and 255 used by the byte order mark have catcode “other”.

\begin{verbatim}
\char_set_catcode_other:N \^^00
\char_set_catcode_other:N \^^fe
\char_set_catcode_other:N \^^ff
\end{verbatim}
There can be no error when encoding in UTF-32. When decoding, the string may not have length $4n$, or it may contain code points larger than "10FFFF. The latter case often happens if the encoding was in fact not UTF-32, because most arbitrary strings are not valid in UTF-32.

There can be no error when encoding in UTF-32. When decoding, the string may not have length $4n$, or it may contain code points larger than "10FFFF. The latter case often happens if the encoding was in fact not UTF-32, because most arbitrary strings are not valid in UTF-32.
The length of the string is not a multiple of 4. -
Perhaps the string was truncated?
}

(End of definition for __str_overflow and __str_end.)

The structure is similar to UTF-16 decoding functions. If the endianness is not given, test
the first 4 bytes of the string (possibly \s__str_stop if the string is too short) for the
presence of a byte-order mark. If there is a byte-order mark, use that endianness, and
remove the 4 bytes, otherwise default to big-endian, and leave the 4 bytes in place. The
\__str_decode_utf_xxxii_bom:NNNN auxiliary receives 1 or 2 as its first argument indicating
endianness, and the string to convert as its second argument (expanded or not). It sets
\__str_tmp:w to expand to the character code of either of its two arguments depending
on endianness, then triggers the _loop auxiliary inside an e-expanding assignment to
\g__str_result_tl.

The _loop auxiliary first checks for the end-of-string marker \s__str_stop, calling
the _end auxiliary if appropriate. Otherwise, leave the \langle 4 bytes \rangle \s__str_behind, then
check that the code point is not overflowing: the leading byte must be 0, and the following
byte at most 16.

In the ending code, we check that there remains no byte: there should be nothing
left until the first \s__str_stop. Break the map.

\cs_new_protected:cpn { __str_convert_decode_utf32be: }
\cs_new_protected:cpn { __str_convert_decode_utf32le: }
\cs_new_protected:cpn { __str_convert_decode_utf32: }
\cs_new_protected:Npn \__str_decode_utf_xxxii_bom:NNNN #1#2#3#4
{ \str_if_eq:nnTF { #1#2#3#4 } { ^^ff ^^fe ^^00 ^^00 } { \__str_decode_utf_xxxii:Nw 1 \g__str_result_tl \s__str_stop } \__str_decode_utf_xxxii:Nw 2 \g__str_result_tl \s__str_stop } \__str_decode_utf_xxxii_bom:NNNN \g__str_result_tl \s__str_stop \s__str_stop \s__str_stop \s__str_stop \s__str_stop
\cs_new_protected:Npn \__str_decode_utf_xxxii_bom:NNNN #1#2#3#4
{ \str_if_eq:nnTF { #1#2#3#4 } { ^^00 ^^00 ^^fe ^^ff } { \__str_decode_utf_xxxii_bom:NNNN #1#2#3#4 } \__str_decode_utf_xxxii_bom:NNNN #1#2#3#4 } \__str_decode_utf_xxxii_bom:NNNN #1#2#3#4
\flag_clear:N \l__str_overflow_flag \flag_clear:N \l__str_end_flag \flag_clear:N \l__str_error_flag \cs_set:Npn \_str_tmp:w ##1 ##2 { ' ## #1 }
\__kernel_tl_gset:Nx \g__str_result_tl { \exp_after:wN \__str_decode_utf_xxxii_loop:NNNN \__str_decode_utf_xxxii_loop:NNNN \g__str_result_tl \s__str_stop \s__str_stop \s__str_stop \s__str_stop \s__str_stop

789
To convert to PDF names by expansion, we work purely on UTF-8 input. The first step is to make a string with “other” spaces, after which we use a simple token-by-token approach. In Unicode engines, we break down everything before one-byte codepoints, but for 8-bit engines there is no need to worry. Actual escaping is covered by the same code as used in the non-expandable route.
\begin{verbatim}
15465 \str_map_function:nN {#1} \_str_convert_pdfname:n }
15466 \bool_lazy_or:nnTF
15467 \sys_if_engine_luatex_p: }
15468 \sys_if_engine_xetex_p: }
15469 { \cs_new:Npn \_str_convert_pdfname:n #1
15470 { \int_compare:nNnTF { '#1 } > { "7F }
15471 { \_str_convert_pdfname_bytes:n {#1} }
15472 { \_str_escape_name_char:n {#1} }
15473 }
15474 \cs_new:Npn \_str_convert_pdfname_bytes:n #1
15475 { \exp_args:Ne \_str_convert_pdfname_bytes_aux:n
15476 { \_kernel_codepoint_to_bytes:n {'#1} }
15477 }
15478 \cs_new:Npn \_str_convert_pdfname_bytes_aux:n #1
15479 { \__kernel_codepoint_to_bytes:n {'#1} }
15480 }
15481 \cs_new:Npn \_str_convert_pdfname_bytes_aux:nnnn #1#2#3#4
15482 { \c_hash_str \exp_not:N \__str_output_hexadecimal:n {#1}
15483 \c_hash_str \exp_not:N \__str_output_hexadecimal:n {#2}
15484 \exp_not:N \tl_if_blank:nF {#3}
15485 { \c_hash_str \exp_not:N \__str_output_hexadecimal:n {#3}
15486 \exp_not:N \tl_if_blank:nF {#4}
15487 { \c_hash_str \exp_not:N \__str_output_hexadecimal:n {#4}
15488 }
15489 }
15490 }
15491 \cs_new_eq:NN \_str_convert_pdfname:n \_str_escape_name_char:n }
15492 \end{verbatim}

(End of definition for \str_convert_pdfname:n and others. This function is documented on page 144.)

56.7.1 iso 8859 support

The iso-8859-1 encoding exactly matches with the 256 first Unicode characters. For other 8-bit encodings of the iso-8859 family, we keep track only of differences, and of unassigned bytes.

\begin{verbatim}
\__str_declare_eight_bit_encoding:nnnn { iso88591 } { 256 }
\{ \}
\__str_declare_eight_bit_encoding:nnnn { iso88592 } { 399 }
\{ \}
\end{verbatim}
792
```c
    __str_declare_eight_bit_encoding:nnnn { iso88593 } { 384 }
    {
        { A1 } { 0126 }
        { A2 } { 02D8 }
        { A6 } { 0124 }
        { A9 } { 0130 }
        { AA } { 015E }
        { AB } { 011E }
        { AC } { 0134 }
        { AF } { 017B }
        { B1 } { 0127 }
        { B6 } { 0125 }
        { B9 } { 0131 }
        { BA } { 015F }
        { BB } { 011F }
        { BC } { 0135 }
        { BF } { 017C }
        { C5 } { 010A }
        { C6 } { 0108 }
        { D5 } { 0120 }
        { D8 } { 011C }
        { DD } { 016C }
        { DE } { 015C }
        { E5 } { 010B }
        { E6 } { 0109 }
        { F5 } { 0121 }
        { F8 } { 011D }
        { FD } { 016D }
        { FE } { 015D }
        { FF } { 02D9 }
    }
    {
        { A5 }
        { AE }
        { BE }
        { C3 }
        { D0 }
        { E3 }
        { FO }
    }
    }/iso88593)
```

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>A2</td>
<td>0138</td>
<td></td>
</tr>
<tr>
<td>A3</td>
<td>0156</td>
<td></td>
</tr>
<tr>
<td>A5</td>
<td>0128</td>
<td></td>
</tr>
<tr>
<td>A6</td>
<td>013B</td>
<td></td>
</tr>
<tr>
<td>A9</td>
<td>0160</td>
<td></td>
</tr>
<tr>
<td>AA</td>
<td>0112</td>
<td></td>
</tr>
<tr>
<td>AB</td>
<td>0122</td>
<td></td>
</tr>
<tr>
<td>AC</td>
<td>0166</td>
<td></td>
</tr>
<tr>
<td>AE</td>
<td>017D</td>
<td></td>
</tr>
<tr>
<td>B1</td>
<td>0105</td>
<td></td>
</tr>
<tr>
<td>B2</td>
<td>02DB</td>
<td></td>
</tr>
<tr>
<td>B3</td>
<td>0157</td>
<td></td>
</tr>
<tr>
<td>B5</td>
<td>0129</td>
<td></td>
</tr>
<tr>
<td>B6</td>
<td>013C</td>
<td></td>
</tr>
<tr>
<td>B7</td>
<td>02C7</td>
<td></td>
</tr>
<tr>
<td>B9</td>
<td>0161</td>
<td></td>
</tr>
<tr>
<td>BA</td>
<td>0113</td>
<td></td>
</tr>
<tr>
<td>BB</td>
<td>0123</td>
<td></td>
</tr>
<tr>
<td>BC</td>
<td>0167</td>
<td></td>
</tr>
<tr>
<td>BD</td>
<td>014A</td>
<td></td>
</tr>
<tr>
<td>BE</td>
<td>017E</td>
<td></td>
</tr>
<tr>
<td>BF</td>
<td>014B</td>
<td></td>
</tr>
<tr>
<td>C0</td>
<td>0100</td>
<td></td>
</tr>
<tr>
<td>C7</td>
<td>012E</td>
<td></td>
</tr>
<tr>
<td>C8</td>
<td>010C</td>
<td></td>
</tr>
<tr>
<td>CA</td>
<td>0118</td>
<td></td>
</tr>
<tr>
<td>CC</td>
<td>0116</td>
<td></td>
</tr>
<tr>
<td>CF</td>
<td>012A</td>
<td></td>
</tr>
<tr>
<td>D0</td>
<td>0110</td>
<td></td>
</tr>
<tr>
<td>D1</td>
<td>0145</td>
<td></td>
</tr>
<tr>
<td>D2</td>
<td>014C</td>
<td></td>
</tr>
<tr>
<td>D3</td>
<td>0136</td>
<td></td>
</tr>
<tr>
<td>D9</td>
<td>0172</td>
<td></td>
</tr>
<tr>
<td>DD</td>
<td>0168</td>
<td></td>
</tr>
<tr>
<td>DE</td>
<td>016A</td>
<td></td>
</tr>
<tr>
<td>E0</td>
<td>0101</td>
<td></td>
</tr>
<tr>
<td>E7</td>
<td>012F</td>
<td></td>
</tr>
<tr>
<td>E8</td>
<td>010D</td>
<td></td>
</tr>
<tr>
<td>EA</td>
<td>0119</td>
<td></td>
</tr>
<tr>
<td>EC</td>
<td>0117</td>
<td></td>
</tr>
<tr>
<td>EF</td>
<td>012B</td>
<td></td>
</tr>
<tr>
<td>F0</td>
<td>0111</td>
<td></td>
</tr>
<tr>
<td>F1</td>
<td>0146</td>
<td></td>
</tr>
<tr>
<td>F2</td>
<td>014D</td>
<td></td>
</tr>
<tr>
<td>F3</td>
<td>0137</td>
<td></td>
</tr>
<tr>
<td>F9</td>
<td>0173</td>
<td></td>
</tr>
<tr>
<td>FD</td>
<td>0169</td>
<td></td>
</tr>
<tr>
<td>FE</td>
<td>016B</td>
<td></td>
</tr>
<tr>
<td>FF</td>
<td>02D9</td>
<td></td>
</tr>
</tbody>
</table>

794
```python
__str_declare_eight_bit_encoding:nnnn { iso88595 } { 374 }
{
    { A1 } { 0401 }
    { A2 } { 0402 }
    { A3 } { 0403 }
    { A4 } { 0404 }
    { A5 } { 0405 }
    { A6 } { 0406 }
    { A7 } { 0407 }
    { A8 } { 0408 }
    { A9 } { 0409 }
    { AA } { 040A }
    { AB } { 040B }
    { AC } { 040C }
    { AE } { 040E }
    { AF } { 040F }
    { B0 } { 0410 }
    { B1 } { 0411 }
    { B2 } { 0412 }
    { B3 } { 0413 }
    { B4 } { 0414 }
    { B5 } { 0415 }
    { B6 } { 0416 }
    { B7 } { 0417 }
    { B8 } { 0418 }
    { B9 } { 0419 }
    { BA } { 041A }
    { BB } { 041B }
    { BC } { 041C }
    { BD } { 041D }
    { BE } { 041E }
    { BF } { 041F }
    { C0 } { 0420 }
    { C1 } { 0421 }
    { C2 } { 0422 }
    { C3 } { 0423 }
    { C4 } { 0424 }
    { C5 } { 0425 }
    { C6 } { 0426 }
    { C7 } { 0427 }
    { C8 } { 0428 }
    { C9 } { 0429 }
    { CA } { 042A }
    { CB } { 042B }
    { CC } { 042C }
    { CD } { 042D }
    { CE } { 042E }
    { CF } { 042F }
    { D0 } { 0430 }
    { D1 } { 0431 }
    { D2 } { 0432 }
    { D3 } { 0433 }
    { D4 } { 0434 }
    { D5 } { 0435 }
```
\__str_declare_eight_bit_encoding:nnnn \{ iso88596 \} { 344 }
{ C3 } { 0623 }
{ C4 } { 0624 }
{ C5 } { 0625 }
{ C6 } { 0626 }
{ C7 } { 0627 }
{ C8 } { 0628 }
{ C9 } { 0629 }
{ CA } { 062A }
{ CB } { 062B }
{ CC } { 062C }
{ CD } { 062D }
{ CE } { 062E }
{ CF } { 062F }
{ D0 } { 0630 }
{ D1 } { 0631 }
{ D2 } { 0632 }
{ D3 } { 0633 }
{ D4 } { 0634 }
{ D5 } { 0635 }
{ D6 } { 0636 }
{ D7 } { 0637 }
{ D8 } { 0638 }
{ D9 } { 0639 }
{ DA } { 063A }
{ E0 } { 0640 }
{ E1 } { 0641 }
{ E2 } { 0642 }
{ E3 } { 0643 }
{ E4 } { 0644 }
{ E5 } { 0645 }
{ E6 } { 0646 }
{ E7 } { 0647 }
{ E8 } { 0648 }
{ E9 } { 0649 }
{ EA } { 064A }
{ EB } { 064B }
{ EC } { 064C }
{ ED } { 064D }
{ EE } { 064E }
{ EF } { 064F }
{ F0 } { 0650 }
{ F1 } { 0651 }
{ F2 } { 0652 }
{ A1 }
{ A2 }
{ A3 }
{ A5 }
{ A6 }
{ A7 }
{ A8 }
{ A9 }
{ AA }

797
\__str_declare_eight_bit_encoding:\nnnn { iso88597 } { 498 }

\langle iso88596 \rangle

\langle* iso88597 \rangle
<table>
<thead>
<tr>
<th>Code</th>
<th>Hex Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>CB</td>
<td>039B</td>
</tr>
<tr>
<td>CC</td>
<td>039C</td>
</tr>
<tr>
<td>CD</td>
<td>039D</td>
</tr>
<tr>
<td>CE</td>
<td>039E</td>
</tr>
<tr>
<td>CF</td>
<td>039F</td>
</tr>
<tr>
<td>D0</td>
<td>03A0</td>
</tr>
<tr>
<td>D1</td>
<td>03A1</td>
</tr>
<tr>
<td>D3</td>
<td>03A3</td>
</tr>
<tr>
<td>D4</td>
<td>03A4</td>
</tr>
<tr>
<td>D5</td>
<td>03A5</td>
</tr>
<tr>
<td>D6</td>
<td>03A6</td>
</tr>
<tr>
<td>D7</td>
<td>03A7</td>
</tr>
<tr>
<td>D8</td>
<td>03A8</td>
</tr>
<tr>
<td>D9</td>
<td>03A9</td>
</tr>
<tr>
<td>DA</td>
<td>03AA</td>
</tr>
<tr>
<td>DB</td>
<td>03AB</td>
</tr>
<tr>
<td>DC</td>
<td>03AC</td>
</tr>
<tr>
<td>DD</td>
<td>03AD</td>
</tr>
<tr>
<td>DE</td>
<td>03AE</td>
</tr>
<tr>
<td>DF</td>
<td>03AF</td>
</tr>
<tr>
<td>E0</td>
<td>03B0</td>
</tr>
<tr>
<td>E1</td>
<td>03B1</td>
</tr>
<tr>
<td>E2</td>
<td>03B2</td>
</tr>
<tr>
<td>E3</td>
<td>03B3</td>
</tr>
<tr>
<td>E4</td>
<td>03B4</td>
</tr>
<tr>
<td>E5</td>
<td>03B5</td>
</tr>
<tr>
<td>E6</td>
<td>03B6</td>
</tr>
<tr>
<td>E7</td>
<td>03B7</td>
</tr>
<tr>
<td>E8</td>
<td>03B8</td>
</tr>
<tr>
<td>E9</td>
<td>03B9</td>
</tr>
<tr>
<td>EA</td>
<td>03BA</td>
</tr>
<tr>
<td>EB</td>
<td>03BB</td>
</tr>
<tr>
<td>EC</td>
<td>03BC</td>
</tr>
<tr>
<td>ED</td>
<td>03BD</td>
</tr>
<tr>
<td>EE</td>
<td>03BE</td>
</tr>
<tr>
<td>EF</td>
<td>03BF</td>
</tr>
<tr>
<td>F0</td>
<td>03C0</td>
</tr>
<tr>
<td>F1</td>
<td>03C1</td>
</tr>
<tr>
<td>F2</td>
<td>03C2</td>
</tr>
<tr>
<td>F3</td>
<td>03C3</td>
</tr>
<tr>
<td>F4</td>
<td>03C4</td>
</tr>
<tr>
<td>F5</td>
<td>03C5</td>
</tr>
<tr>
<td>F6</td>
<td>03C6</td>
</tr>
<tr>
<td>F7</td>
<td>03C7</td>
</tr>
<tr>
<td>F8</td>
<td>03C8</td>
</tr>
<tr>
<td>F9</td>
<td>03C9</td>
</tr>
<tr>
<td>FA</td>
<td>03CA</td>
</tr>
<tr>
<td>FB</td>
<td>03CB</td>
</tr>
<tr>
<td>FC</td>
<td>03CC</td>
</tr>
<tr>
<td>FD</td>
<td>03CD</td>
</tr>
<tr>
<td>FE</td>
<td>03CE</td>
</tr>
</tbody>
</table>


{ D2 }

{ iso88597 }

{ iso88598 }

\__str_declare_eight_bit_encoding:nnnn { iso88598 } { 308 }

{ AA } { 00D7 }

{ BA } { 00F7 }

{ DF } { 2017 }

{ E0 } { 05D0 }

{ E1 } { 05D1 }

{ E2 } { 05D2 }

{ E3 } { 05D3 }

{ E4 } { 05D4 }

{ E5 } { 05D5 }

{ E6 } { 05D6 }

{ E7 } { 05D7 }

{ E8 } { 05D8 }

{ E9 } { 05D9 }

{ EA } { 05DA }

{ EB } { 05DB }

{ EC } { 05DC }

{ ED } { 05DD }

{ EE } { 05DE }

{ EF } { 05DF }

{ F0 } { 05E0 }

{ F1 } { 05E1 }

{ F2 } { 05E2 }

{ F3 } { 05E3 }

{ F4 } { 05E4 }

{ F5 } { 05E5 }

{ F6 } { 05E6 }

{ F7 } { 05E7 }

{ F8 } { 05E8 }

{ F9 } { 05E9 }

{ FA } { 05EA }

{ FD } { 200E }

{ FE } { 200F }

{ A1 }

{ BF }

{ C0 }

{ C1 }

{ C2 }

{ C3 }

{ C4 }

{ C5 }

{ C6 }

{ C7 }

{ C8 }

{ C9 }

{ CA }

{ CB }

800
15995 \{ CC \}
15996 \{ CD \}
15997 \{ CE \}
15998 \{ CF \}
15999 \{ D0 \}
16000 \{ D1 \}
16001 \{ D2 \}
16002 \{ D3 \}
16003 \{ D4 \}
16004 \{ D5 \}
16005 \{ D6 \}
16006 \{ D7 \}
16007 \{ D8 \}
16008 \{ D9 \}
16009 \{ DA \}
16010 \{ DB \}
16011 \{ DC \}
16012 \{ DD \}
16013 \{ DE \}
16014 \{ FB \}
16015 \{ FC \}
16016 \}
16017 ⟨/iso88598⟩
16018 ⟨∗iso88599⟩
16019 \__str_declare_eight_bit_encoding:nnnn \{ iso88599 \} \{ 352 \}
16020 \{
16021 \{ D0 \} \{ 011E \}
16022 \{ DD \} \{ 0130 \}
16023 \{ DE \} \{ 015E \}
16024 \{ F0 \} \{ 011F \}
16025 \{ FD \} \{ 0131 \}
16026 \{ FE \} \{ 015F \}
16027 \}
16028 \{
16029 \}
16030 ⟨/iso88599⟩
16031 ⟨∗iso885910⟩
16032 \__str_declare_eight_bit_encoding:nnnn \{ iso885910 \} \{ 383 \}
16033 \{
16034 \{ A1 \} \{ 0104 \}
16035 \{ A2 \} \{ 0112 \}
16036 \{ A3 \} \{ 0122 \}
16037 \{ A4 \} \{ 012A \}
16038 \{ A5 \} \{ 0128 \}
16039 \{ A6 \} \{ 0136 \}
16040 \{ A8 \} \{ 013B \}
16041 \{ A9 \} \{ 0110 \}
16042 \{ AA \} \{ 0160 \}
16043 \{ AB \} \{ 0166 \}
16044 \{ AC \} \{ 017D \}
16045 \{ AE \} \{ 016A \}
16046 \{ AF \} \{ 014A \}
16047 \{ B1 \} \{ 0105 \}

801
<table>
<thead>
<tr>
<th>Decimal</th>
<th>Hex</th>
<th>Character</th>
</tr>
</thead>
<tbody>
<tr>
<td>0113</td>
<td>B2</td>
<td></td>
</tr>
<tr>
<td>0123</td>
<td>B3</td>
<td></td>
</tr>
<tr>
<td>012B</td>
<td>B4</td>
<td></td>
</tr>
<tr>
<td>0129</td>
<td>B5</td>
<td></td>
</tr>
<tr>
<td>0137</td>
<td>B6</td>
<td></td>
</tr>
<tr>
<td>013C</td>
<td>B8</td>
<td></td>
</tr>
<tr>
<td>0111</td>
<td>B9</td>
<td></td>
</tr>
<tr>
<td>0161</td>
<td>BA</td>
<td></td>
</tr>
<tr>
<td>0167</td>
<td>BB</td>
<td></td>
</tr>
<tr>
<td>017E</td>
<td>BC</td>
<td></td>
</tr>
<tr>
<td>2015</td>
<td>BD</td>
<td></td>
</tr>
<tr>
<td>016B</td>
<td>BE</td>
<td></td>
</tr>
<tr>
<td>014B</td>
<td>BF</td>
<td></td>
</tr>
<tr>
<td>0180</td>
<td>C0</td>
<td></td>
</tr>
<tr>
<td>012E</td>
<td>C7</td>
<td></td>
</tr>
<tr>
<td>010C</td>
<td>C8</td>
<td></td>
</tr>
<tr>
<td>0118</td>
<td>CA</td>
<td></td>
</tr>
<tr>
<td>0116</td>
<td>CC</td>
<td></td>
</tr>
<tr>
<td>0145</td>
<td>D1</td>
<td></td>
</tr>
<tr>
<td>014C</td>
<td>D2</td>
<td></td>
</tr>
<tr>
<td>0168</td>
<td>D7</td>
<td></td>
</tr>
<tr>
<td>0172</td>
<td>D9</td>
<td></td>
</tr>
<tr>
<td>0101</td>
<td>E0</td>
<td></td>
</tr>
<tr>
<td>012F</td>
<td>E7</td>
<td></td>
</tr>
<tr>
<td>010D</td>
<td>E8</td>
<td></td>
</tr>
<tr>
<td>0119</td>
<td>EA</td>
<td></td>
</tr>
<tr>
<td>0117</td>
<td>EC</td>
<td></td>
</tr>
<tr>
<td>0146</td>
<td>F1</td>
<td></td>
</tr>
<tr>
<td>014D</td>
<td>F2</td>
<td></td>
</tr>
<tr>
<td>0169</td>
<td>F7</td>
<td></td>
</tr>
<tr>
<td>0173</td>
<td>F9</td>
<td></td>
</tr>
<tr>
<td>0138</td>
<td>FF</td>
<td></td>
</tr>
</tbody>
</table>

\( /iso885910 \)

\{\texttt{iso885911}\}

\(__\texttt{str-_append_eight_bit_encoding:nnnn}\ {iso885911} \} \{ 369 \}

\{\texttt{iso885911}\}

\{ A1 \} \{ OE01 \}

\{ A2 \} \{ OE02 \}

\{ A3 \} \{ OE03 \}

\{ A4 \} \{ OE04 \}

\{ A5 \} \{ OE05 \}

\{ A6 \} \{ OE06 \}

\{ A7 \} \{ OE07 \}

\{ A8 \} \{ OE08 \}

\{ A9 \} \{ OE09 \}

\{ AA \} \{ OE0A \}

\{ AB \} \{ OE0B \}

\{ AC \} \{ OE0C \}

\{ AD \} \{ OE0D \}

\{ AE \} \{ OE0E \}

\{ AF \} \{ OE0F \}
{ EA } { OE4A }
{ EB } { OE4B }
{ EC } { OE4C }
{ ED } { OE4D }
{ EE } { OE4E }
{ EF } { OE4F }
{ F0 } { OE50 }
{ F1 } { OE51 }
{ F2 } { OE52 }
{ F3 } { OE53 }
{ F4 } { OE54 }
{ F5 } { OE55 }
{ F6 } { OE56 }
{ F7 } { OE57 }
{ F8 } { OE58 }
{ F9 } { OE59 }
{ FA } { OE5A }
{ FB } { OE5B }

{ DB }
{ DC }
{ DD }
{ DE }

{/iso885911}
{+iso885913}

\__str_declare_eight_bit_encoding:nnnn { iso885913 } { 399 }

{ A1 } { 201D }
{ A5 } { 201E }
{ A8 } { 00D8 }
{ AA } { 0156 }
{ AF } { 00C6 }
{ B4 } { 201C }
{ B8 } { 00F8 }
{ BA } { 0157 }
{ BF } { 00E6 }
{ C0 } { 0104 }
{ C1 } { 012E }
{ C2 } { 0100 }
{ C3 } { 0106 }
{ C6 } { 0118 }
{ C7 } { 0112 }
{ C8 } { 010C }
{ CA } { 0179 }
{ CB } { 0116 }
{ CC } { 0122 }
{ CD } { 0136 }
{ CE } { 012A }
{ CF } { 013B }
{ D0 } { 0180 }
{ D1 } { 0143 }
{ D2 } { 0145 }
\_\_str\_declare\_eight\_bit\_encoding::nnnn { iso885914 } { 529 } {

{ A1 } { 1E02 }
{ A2 } { 1E03 }
{ A4 } { 010A }
{ A5 } { 010B }
{ A6 } { 1E0A }
{ A8 } { 1E80 }
{ AA } { 1E82 }
{ AB } { 1E0B }
{ AC } { 1EF2 }
{ AF } { 0178 }
{ B0 } { 1E1E }
{ B1 } { 1E1F }
{ B2 } { 0120 }
{ B3 } { 0121 }
{ B4 } { 1E40 }
{ B5 } { 1E41 }

}
\_\_str\_declare\_eight\_bit\_encoding:nnnn { iso885914 } { 383 }
\_\_str\_declare\_eight\_bit\_encoding:nnnn { iso885915 } { 383 }
\_\_str\_declare\_eight\_bit\_encoding:nnnn { iso885916 } { 558 }

806
<table>
<thead>
<tr>
<th>Code</th>
<th>Unicode Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>B9</td>
<td>010D</td>
</tr>
<tr>
<td>BA</td>
<td>0219</td>
</tr>
<tr>
<td>BC</td>
<td>0152</td>
</tr>
<tr>
<td>BD</td>
<td>0153</td>
</tr>
<tr>
<td>BE</td>
<td>0178</td>
</tr>
<tr>
<td>BF</td>
<td>017C</td>
</tr>
<tr>
<td>C3</td>
<td>0102</td>
</tr>
<tr>
<td>C5</td>
<td>0106</td>
</tr>
<tr>
<td>D0</td>
<td>0110</td>
</tr>
<tr>
<td>D1</td>
<td>0143</td>
</tr>
<tr>
<td>D5</td>
<td>0150</td>
</tr>
<tr>
<td>D7</td>
<td>015A</td>
</tr>
<tr>
<td>D8</td>
<td>0170</td>
</tr>
<tr>
<td>DD</td>
<td>0118</td>
</tr>
<tr>
<td>DE</td>
<td>021A</td>
</tr>
<tr>
<td>E3</td>
<td>0103</td>
</tr>
<tr>
<td>E5</td>
<td>0107</td>
</tr>
<tr>
<td>F0</td>
<td>0111</td>
</tr>
<tr>
<td>F1</td>
<td>0144</td>
</tr>
<tr>
<td>F5</td>
<td>0151</td>
</tr>
<tr>
<td>F7</td>
<td>015B</td>
</tr>
<tr>
<td>F8</td>
<td>0171</td>
</tr>
<tr>
<td>FD</td>
<td>0119</td>
</tr>
<tr>
<td>FE</td>
<td>021B</td>
</tr>
</tbody>
</table>
Chapter 57

\texttt{l3quark} implementation

The following test files are used for this code: \texttt{m3quark001.lvt}.

\begin{verbatim}
\package

57.1 Quarks

\newcommand{\newquark}{Allocate a new quark.}
\newcommand{\newquark}[1]{\__kernel_chk_if_free_cs:N #1 \cs_gset_nopar:Npn #1 {#1}}

(End of definition for \texttt{\newquark}. This function is documented on page 147.)

\newcommand{\q_nil}{\q_mark}{\q_no_value}{\q_stop}

Some “public” quarks. \texttt{\q_stop} is an “end of argument” marker, \texttt{\q_nil} is a empty value and \texttt{\q_no_value} marks an empty argument.

\newcommand{\newq_nil}{\newq_mark}{\newq_no_value}{\newq_stop}

(End of definition for \texttt{\newq_nil} and others. These variables are documented on page 147.)

\newcommand{\q_recursion_tail}{\q_recursion_stop}

Quarks for ending recursions. Only ever used there! \texttt{\q_recursion_tail} is appended to whatever list structure we are doing recursion on, meaning it is added as a proper list item with whatever list separator is in use. \texttt{\q_recursion_stop} is placed directly after the list.

\newcommand{\newq_recursion_tail}{\newq_recursion_stop}

(End of definition for \texttt{\newq_recursion_tail} and \texttt{\newq_recursion_stop}. These variables are documented on page 148.)

\texttt{\s__quark}

Private scan mark used in \texttt{l3quark}. We don’t have \texttt{l3scan} yet, so we declare the scan mark here and add it to the scan mark pool later.

\end{verbatim}
Private quark use for some tests.

When doing recursions, it is easy to spend a lot of time testing if the end marker has been found. To avoid this, a dedicated end marker is used each time a recursion is set up. Thus if the marker is found everything can be wrapper up and finished off. The simple case is when the test can guarantee that only a single token is being tested. In this case, there is just a dedicated copy of the standard quark test. Both a gobbling version and one inserting end code are provided.

\begin{verbatim}
\cs_new:Npn \__quark_if_recursion_tail:w #1 \q_recursion_tail #2 ? #3 ?! { #1 #2 }
\cs_generate_variant:Nn \quark_if_recursion_tail_stop:n { o }
\cs_generate_variant:Nn \quark_if_recursion_tail_stop_do:nn { o }
\end{verbatim}

See \quark_if_nil:nTF for the details. Expanding \__quark_if_recursion_tail:w once in front of the tokens chosen here gives an empty result if and only if \#1 is exactly \q_recursion_tail.
Analogues of the \quark_if_recursion_tail_stop... functions. Break the mapping using \#2.
\begin{verbatim}
\cs_new:Npn \quark_if_recursion_tail_break:NN #1#2
{ \if_meaning:w \q_recursion_tail #1 \exp_after:wN #2 \fi: }
\cs_new:Npn \quark_if_recursion_tail_break:nN #1#2
{ \tl_if_empty:oT { \__quark_if_recursion_tail:w {} #1 {} ?! \q_recursion_tail ??! } {#2} }
\end{verbatim}

(End of definition for \quark_if_recursion_tail_break:NN and \quark_if_recursion_tail_break:nN. These functions are documented on page 149.)

\begin{verbatim}
\quark_if_nil_p:N \quark_if_nil:NTF
\quark_if_no_value_p:N \quark_if_no_value:NTF
\quark_if_no_value_p:c \quark_if_no_value_p:V
\quark_if_no_value:NTF \quark_if_no_value:cFF
\end{verbatim}

Here we test if we found a special quark as the first argument. We better start with \q_no_value as the first argument since the whole thing may otherwise loop if \#1 is wrongly given a string like aabc instead of a single token.\footnote{\textit{It may still loop in special circumstances however!}}
\begin{verbatim}
\prg_new_conditional:Npnn \quark_if_nil:N #1 { p, T , F , TF }
{ \if_meaning:w \q_nil #1 \prg_return_true: \else: \prg_return_false: \fi: }
\prg_new_conditional:Npnn \quark_if_no_value:N #1 { p, T , F , TF }
{ \if_meaning:w \q_no_value #1 \prg_return_true: \else: \prg_return_false: \fi: }
\prg_generate_conditional_variant:Nnn \quark_if_no_value:N { c } { p , T , F , TF }
\end{verbatim}

(End of definition for \quark_if_nil:NTF and \quark_if_no_value:NTF. These functions are documented on page 147.)

\begin{verbatim}
\quark_if_nil_p:n \quark_if_nil_p:V
\quark_if_nil_p:o \quark_if_nil_p:VTF
\quark_if_nil:nTF \quark_if_nil:VTF
\quark_if_no_value_p:n \quark_if_no_value_p:V
\quark_if_no_value:nFF \quark_if_no_value:nFF
\__quark_if_nil:w \__quark_if_no_value:w
\__quark_if_empty_if:o
\end{verbatim}

Let us explain \quark_if_nil:NTF. Expanding \__quark_if_nil:w once is safe thanks to the trailing \q_nil ??!. The result of expanding once is empty if and only if both delimited arguments \#1 and \#2 are empty and \#3 is delimited by the last tokens ?!. Thanks to the leading {}, the argument \#1 is empty if and only if the argument of \quark_if_nil:n starts with \q_nil. The argument \#2 is empty if and only if this \q_nil is followed immediately by ? or by \{\}. coming either from the trailing tokens in the definition of \quark_if_nil:n, or from its argument. In the first case, \__quark_if_nil:w is followed by \{\} \q_nil \}? \q_nil ??!, hence \#3 is delimited by the final ?!, and the test returns \textit{true} as wanted. In the second case, the result is not empty since
the first ?! in the definition of \quark_if_nil:n stop #3. The auxiliary here is the same as \tl_if_empty_if:o, with the same comments applying.

\begin{verbatim}
\prg_new_conditional:Npnn \quark_if_nil:n #1 { p, T , F , TF }
{ \__quark_if_empty_if:o
\__quark_if_nil:w {} #1 {} ? ! \q_nil ? ? ! }
\prg_return_true:
\else:
\prg_return_false:
\fi:
\end{verbatim}

\begin{verbatim}
\prg_new_conditional:Npnn \quark_if_no_value:n #1 { p, T , F , TF }
{ \__quark_if_empty_if:o
\__quark_if_no_value:w {} #1 {} ? ! \q_no_value ? ? ! }
\prg_return_true:
\else:
\prg_return_false:
\fi:
\end{verbatim}

\begin{verbatim}
\cs_new:Npn \__quark_if_nil:w #1 \q_nil #2 ? #3 ? ! { #1 #2 }
\end{verbatim}

\begin{verbatim}
\prg_generate_conditional_variant:Nnn \quark_if_nil:n
{ V , o } { p , TF , T , F }
\cs_new:Npn \__kernel_tl_to_str:w \exp_after:wN {#1} \q_nil
\end{verbatim}

(End of definition for \quark_if_nil:nTF and others. These functions are documented on page 147.)

\__kernel_quark_new_test:N
The function \__kernel_quark_new_test:N defines #1 in a similar way as \quark-if-recursion_tail... functions (as described below), using \q⟨namespace⟩-recursion_tail as the test quark and \q⟨namespace⟩-recursion_stop as the delimiter quark, where the ⟨namespace⟩ is determined as the first _-delimited part in #1.

There are six possible function types which this function can define, and which is defined depends on the signature of the function being defined:

:n gives an analogue of \quark_if_recursion_tail_stop:n

:nn gives an analogue of \quark_if_recursion_tail_stop_do:nn

:nN gives an analogue of \quark_if_recursion_tail_break:nN

:N gives an analogue of \quark_if_recursion_tail_stop:N

:Nn gives an analogue of \quark_if_recursion_tail_stop_do:Nn

:NN gives an analogue of \quark_if_recursion_tail_break:NN

Any other signature causes an error, as does a function without signature.

811
Similar to __kernel_quark_new_test:N, but defines quark branching conditionals like __kernel_quark_if_nil:nTF that test for the quark q__⟨namespace⟩_⟨name⟩. The ⟨namespace⟩ and ⟨name⟩ are determined from the conditional #1, which must take the rather rigid form __⟨namespace⟩_quark_if⟨name⟩:⟨arg spec⟩. There are only two cases for the ⟨arg spec⟩ here:

:n gives an analogue of quark_if_nil:nTF

:N gives an analogue of quark_if_nil:NTF

Any other signature causes an error, as does a function without signature. We use low-level emptiness tests as l3tl is not available yet when these functions are used; thankfully we only care about whether strings are empty so a simple \if_meaning:w \q_nil ⟨string⟩ \q_nil suffices.
These macros implement the six possibilities mentioned above, passing the right arguments to \__quark_new_test_aux_do:nNNnnnnNnn, which defines some auxiliaries, and then to \__quark_new_test_define_tl:nNnnn (:n(n) variants) or to \__quark_new_test_define_ifx:nNNnnn (:N(n)) which define the main conditionals.

(End of definition for \__quark_new_test_n:Nnnn and others.)
The control sequence names which will be used by \_\_quark_test_define_aux:NNNNnNnN, and then later by \_\_quark_new_test_define_tl:nNnnNn or \_\_quark_new_test_define_ifx:nNnnNn. The control sequences defined here are analogous to \_\_quark_if_recursion_tail:w and to \_\_use_{none|if}delimit_by_q_recursion_stop:(|n)w.

The name is composed by the name-space and the name of the quarks. Suppose \_\_kernel_quark_new_test:N was used with:

\_\_kernel_quark_new_test:N \_\_test_quark_tail:n

then the first auxiliary will be \_\_test_quark_recursion_tail:w, and the second one will be \_\_test_use_none_delimit_by_q_recursion_stop:w.

Note that the actual quarks are not defined here. They should be defined separately using \quark_new:N.

Finally, these two macros define the main conditional function using what’s been set up before.

(End of definition for \_\_quark_new_test_aux_do:nNnnnNnN and \_\_quark_test_define_aux:NNNNnNnN.)
These macros implement the two possibilities for branching quark conditionals. To avoid constructing without defining the \_\_\_\textit{\texttt{if\_quark}}\_\textit{\texttt{name}}:\texttt{w} helper, \texttt{N}-type function accepts a \texttt{prg\_do\_nothing}: as a placeholder.

(End of definition for \_\_\_\textit{\texttt{if\_quark}}\_\textit{\texttt{name}} and others.)

\_\_\textit{\texttt{module}}\_\textit{\texttt{name}}:\texttt{N} takes a control sequence and returns its \textit{\texttt{module}} name, determined as the first non-empty non-single-character word, separated by \_ or \texttt{:}. These rules give the correct result for public functions \_\_\textit{\texttt{module}}\_\textit{\texttt{...}}, private functions \_\_\_\_\_\_\textit{\texttt{module}}\_\textit{\texttt{...}}, and variables such as \_\_\textit{\texttt{\texttt{\_\_\_\_\_\texttt{module}}\_\texttt{...}}. If no valid module is found the result is an empty string. The approach is to first cut off everything after the (first) \_ if any is present, then repeatedly grab \_ -delimited words until finding one of length at least 2 (we use low-level tests as \texttt{l3tl} is not fully available when \_\_\textit{\texttt{kernel\_quark\_new\_\_\_\texttt{test}}}:\texttt{N} is first used. If no \_\_\textit{\texttt{module}} is found (such as in \_\_\texttt{\_\_\_\texttt{\texttt{\_\_\_\_\_\_\texttt{module}}\_\texttt{...}}}) we get the trailing marker \texttt{use\_none}:\texttt{n} {}, which expands to nothing.
\cs_new:Npn \__quark_module_name:N #1 \#2 \s__quark { \exp_last_unbraced:Nf \__quark_module_name:w { \cs_to_str:N ##1 } #1 \s__quark }

\cs_new:Npn \__quark_module_name:w #1 #2 \s__quark { \__quark_module_name_loop:w #1 #2 \use_none:n { } #2 \s__quark }

\cs_new:Npn \__quark_module_name_loop:w #1 #2 { \use_i_ii:nnn \if_meaning:w \prg_do_nothing: #1 \prg_do_nothing: \prg_do_nothing: \exp_after:wN \__quark_module_name_loop:w \else: \__quark_module_name_end:w #1 \fi: }

\cs_new:Npn \__quark_module_name_end:w #1 \fi: #2 \s__quark { \fi: #1 }

\exp_after:wN \__quark_tmp:w \tl_to_str:n { : _quark_if_ } \s__quark

\__quark_quark_conditional_name:N \__quark_quark_conditional_name:w \__quark_quark_conditional_name:N determines the quark name that the quark conditional function \#1 queries, as the part of the function name between \_quark_if_ and the trailing :. Again we define it through \__quark_tmp:w, which receives : as \#1 and \_quark_if_ as \#2. The auxiliary \__quark_quark_conditional_name:w returns the part between the first \_quark_if_ and the next :, and we apply this auxiliary to the function name followed by : (in case the function name is lacking a signature), and \_quark_if_: so that \__quark_quark_conditional_name:N returns an empty string if \_quark_if_ is not present.

\cs_set:Npn \__quark_tmp:w \tl_to_str:n { : _ }

\exp_after:wN \__quark_tmp:w \tl_to_str:n { : _quark_if_ } \s__quark

(End of definition for \__quark_module_name:N and others.)

\_quark_quark_conditional_name:N \_quark_quark_conditional_name:w

\_quark_quark_conditional_name:N \_quark_quark_conditional_name:w

57.2 \textbf{Scan marks}

\scan_new:N Check whether the variable is already a scan mark, then declare it to be equal to \scan_stop: globally.

\cs_new_protected:Npn \scan_new:N #1

(End of definition for \__quark_quark_conditional_name:N and \__quark_quark_conditional_name:w.)
We only declare one scan mark here, more can be defined by specific modules. Can't use \scan_new:N yet because l3tl isn't loaded, so define \s_stop by hand and add it to \g__scan_marks_tl. We also add the scan marks declared earlier to the pool here. Since they live in a different namespace, a little DocStrip cheating is necessary.

(End of definition for \s_stop and \g__scan_marks_tl. This variable is documented on page 150.)

\use_none_delimit_by_s_stop:w

Similar to \use_none_delimit_by_q_stop:w.

(End of definition for \use_none_delimit_by_s_stop:w. This function is documented on page 150.)

/package
Chapter 58

\textbf{l3seq implementation}

The following test files are used for this code: \texttt{m3seq002,m3seq003}.

\begin{verbatim}
\package
\@@=seq
A sequence is a control sequence whose top-level expansion is of the form \texttt{\s__-seq \_\_seq_item:n \{item\} ... \_\_seq_item:n \{item\}}\}, with a leading scan mark followed by \texttt{n} items of the same form. An earlier implementation used the structure \texttt{\seq_elt:w \{item\} \seq_elt_end: ... \seq_elt:w \{item\} \seq_elt_end:}. This allowed rapid searching using a delimited function, but was not suitable for items containing \{, \} and \# tokens, and also lead to the loss of surrounding braces around items

\_\_seq_item:n = \_\_seq_item:n \{item\}

The internal token used to begin each sequence entry. If expanded outside of a mapping or manipulation function, an error is raised. The definition should always be set globally.

\_\_seq_push_item_def:n
\_\_seq_push_item_def:e

Saves the definition of \_\_seq_item:n and redefines it to accept one parameter and expand to \texttt{\{code\}}. This function should always be balanced by use of \_\_seq_pop_item_def:

\_\_seq_pop_item_def:

Restores the definition of \_\_seq_item:n most recently saved by \_\_seq_push_item_def:n. This function should always be used in a balanced pair with \_\_seq_push_item_def:n.

\s__seq
This private scan mark.

\s__seq_mark
\s__seq_stop
Private scan marks.

(End of definition for \s__seq.)

(End of definition for \s__seq_mark and \s__seq_stop.)
\end{verbatim}
\_seq\_item:n\ The delimiter is always defined, but when used incorrectly simply removes its argument and hits an undefined control sequence to raise an error.

\begin{verbatim}
\cs_new:Npn \_seq\_item:n
{\msg\_expandable\_error:nn { seq } { misused }
 \use\_none:n}
\end{verbatim}

(End of definition for \_seq\_item:n.)

\_\_\_seq\_internal\_a_tl \_\_\_seq\_internal\_b_tl
Scratch space for various internal uses.

\begin{verbatim}
\tl\_new:N \l\_\_seq\_internal\_a_tl
\tl\_new:N \l\_\_seq\_internal\_b_tl
\end{verbatim}

(End of definition for \_\_\_seq\_internal\_a_tl and \_\_\_seq\_internal\_b_tl.)

\_\_\_seq\_tmp:w
Scratch function for internal use.

\begin{verbatim}
\cs\_new\_eq:NN \_\_\_seq\_tmp:w ?
\end{verbatim}

(End of definition for \_\_\_seq\_tmp:w.)

\c\_empty\_seq
A sequence with no item, following the structure mentioned above.

\begin{verbatim}
\tl\_const:Nn \c\_empty\_seq { \s\_seq }
\end{verbatim}

(End of definition for \c\_empty\_seq. This variable is documented on page 165.)

### 58.1 Allocation and initialisation

\seq\_new:N \seq\_new:c
Sequences are initialized to \c\_empty\_seq.

\begin{verbatim}
\cs\_new\_protected:Npn \seq\_new:N \#1
{\_kernel\_check\_if\_free\_cs:N \#1
 \cs\_gset\_eq:NN \#1 \c\_empty\_seq}
\end{verbatim}

(End of definition for \seq\_new:N. This function is documented on page 151.)

\seq\_clear:N \seq\_clear:c \seq\_gclear:N \seq\_gclear:c
Clearing a sequence is similar to setting it equal to the empty one.

\begin{verbatim}
\cs\_new\_protected:Npn \seq\_clear:N \#1
{\seq\_set\_eq:NN \#1 \c\_empty\_seq}
\cs\_generate\_variant:Nn \seq\_clear:N \{ c \}
\end{verbatim}

(End of definition for \seq\_clear:N and \seq\_gclear:N. These functions are documented on page 151.)

Once again we copy code from the token list functions.

\begin{verbatim}
\cs\_new\_protected:Npn \seq\_clear\_new:N \#1
{\_seq\_if\_exist:NTF \#1 \{ \seq\_clear\_new:N \#1 \} \{ \seq\_new:N \#1 \}}
\cs\_generate\_variant:Nn \seq\_clear\_new:N \{ c \}
\end{verbatim}

\begin{verbatim}
\cs\_new\_protected:Npn \seq\_gclear\_new:N \#1
{\_seq\_if\_exist:NTF \#1 \{ \seq\_gclear\_new:N \#1 \} \{ \seq\_new:N \#1 \}}
\cs\_generate\_variant:Nn \seq\_gclear\_new:N \{ c \}
\end{verbatim}

819
Copying a sequence is the same as copying the underlying token list.

Setting a sequence from a comma-separated list is done using a simple mapping.

Almost identical to \texttt{\seq\textunderscore\texttt{const}\textunderscore\texttt{from}\textunderscore\texttt{clist}:Nn}.

\begin{verbatim}
\seq_set_eq:NN \seq_set_eq:cN \seq_set_eq:Nc \seq_set_eq:cc \\
\seq_gset_eq:NN \seq_gset_eq:cN \seq_gset_eq:Nc \seq_gset_eq:cc
\end{verbatim}

\begin{verbatim}
\seq_set_from_clist:NN \seq_set_from_clist:cN \seq_set_from_clist:Nn \seq_set_from_clist:cn \\
\seq_gset_from_clist:NN \seq_gset_from_clist:cN \seq_gset_from_clist:Nn \seq_gset_from_clist:cn \\
\seq_const_from_clist:Nn \seq_const_from_clist:cn
\end{verbatim}

(End of definition for \texttt{\seq\textunderscore clear\textunderscore new:N} and \texttt{\seq\textunderscore gclear\textunderscore new:N}. These functions are documented on page 151.)
When the separator is empty, everything is very simple, just map \__seq_wrap_item:n through the items of the last argument. For non-trivial separators, the goal is to split a given token list at the marker, strip spaces from each item, and remove one set of outer braces if after removing leading and trailing spaces the item is enclosed within braces. After \tl_replace_all:Nnn, the token list \l__seq_internal_a_tl is a repetition of the pattern \__seq_set_split:Nw \prg_do_nothing: (item with spaces) \__seq_set_split_end:. Then, e-expansion causes \__seq_set_split:Nw to trim spaces, and leaves its result as \__seq_set_split:w (trimmed item) \__seq_set_split_end:. This is then converted to the \l3seq internal structure by another e-expansion. In the first step, we insert \prg_do_nothing: to avoid losing braces too early: that would cause space trimming to act within those lost braces. The second step is solely there to strip braces which are outermost after space trimming.

\begin{verbatim}
\cs_new_protected:Npn \seq_set_split:Nnn { \__seq_set_split:NNnn \__kernel_tl_set:Nx \tl_trim_spaces:n }
\cs_new_protected:Npn \seq_gset_split:Nnn { \__seq_set_split:NNnn \__kernel_tl_gset:Nx \tl_trim_spaces:n }
\cs_new_protected:Npn \seq_set_split_keep_spaces:Nnn { \__seq_set_split:NNnn \__kernel_tl_set:Nx \exp_not:n }
\cs_new_protected:Npn \seq_gset_split_keep_spaces:Nnn { \__seq_set_split:NNnn \__kernel_tl_gset:Nx \exp_not:n }
\cs_new_protected:Npn \__seq_set_split:NNNnn #1#2#3#4#5
{ \tl_if_empty:nTF {#4}
{ \tl_set:Nn \l__seq_internal_a_tl { \tl_map_function:nN {#5} \__seq_wrap_item:n } }
{ \tl_set:Nn \l__seq_internal_a_tl { \__seq_set_split:Nw #2 \prg_do_nothing: #5 \__seq_set_split_end: }
\tl_replace_all:Nnn \l__seq_internal_a_tl {#4}
{ \__seq_set_split_end: }
\__kernel_tl_set:Nx \l__seq_internal_a_tl { \l__seq_internal_a_tl } }
\__kernel_tl_set:Nx \l__seq_internal_a_tl { #1 #3 { \s__seq \l__seq_internal_a_tl } }
\cs_new:Npn \__seq_set_split:Nw #1#2 \__seq_set_split_end:
{ \exp_not:N \__seq_set_split:w \exp_args:No #1 \exp_not:N \__seq_set_split_end: }
\cs_new:Npn \__seq_set_split:w #1 \__seq_set_split_end:
{ \__seq_wrap_item:n {#1} }
\end{verbatim}

821
Similar to \texttt{\textbackslash seq_map_inline:N}, without a \texttt{prg_break_point}: because the user's code is performed within the evaluation of a boolean expression, and skipping out of that would break horribly. The \texttt{\_\_seq_wrap_item:n} function inserts the relevant \texttt{\_\_seq_item:n} without expansion in the input stream, hence in the e-expanding assignment.

When concatenating sequences, one must remove the leading \texttt{\_\_seq} of the second sequence. The result starts with \texttt{\_\_seq} (of the first sequence), which stops f-expansion.

Copies of the \texttt{cs} functions defined in \texttt{l3basics}.

When adding to the left of a sequence, remove \texttt{\_\_seq}. This is done by \texttt{\_\_seq_put_left_aux:w}, which also stops f-expansion.
8.3 Modifying sequences

This function converts its argument to a proper sequence item in an e-expansion context.

\(\text{\texttt{\_\_seq_wrap_item:n}}\) An internal sequence for the removal routines.

\(\text{\texttt{\_\_seq_remove_seq}}\) Removing duplicates means making a new list then copying it.
The idea of the code here is to avoid a relatively expensive addition of items one at a time to an intermediate sequence. The approach taken is therefore similar to that in \_\_seq_pop_right:NNN, using a “flexible” e-type expansion to do most of the work. As \_tl_if_eq:nnNT is not expandable, a two-part strategy is needed. First, the e-type expansion uses \_str_if_eq:nnNT to find potential matches. If one is found, the expansion is halted and the necessary set up takes place to use the \_tl_if_eq:nnNT test. The e-type is started again, including all of the items copied already. This happens repeatedly until the entire sequence has been scanned. The code is set up to avoid needing an intermediate scratch list: the lead-off e-type expansion (#1 #2 (##)) ensures that nothing is lost.

(End of definition for \_\_seq_remove_all:NN, \_\_seq_remove_all:NV, and \_\_seq_remove_all:Ne. These functions are documented on page 156.)
Useful to more quickly go through items.

(End of definition for \_\_seq_int_eval:w.)

The conditionals are distinguished from the \_\_seq_item:nn versions by the last argument \_\_seq_item:ii:nn vs \_\_seq_item:i:nn.

Save the item to be stored and evaluate the position and the sequence length only once. Then depending on the sign of the position, check that it is not bigger than the length (in absolute value) nor zero.

If the position is not ok, \_\_seq_set_item_false:nnNNNN calls an error or returns false (depending on the \_\_seq_item:ii:nn vs \_\_seq_item:ii:nn argument mentioned above).
If the position is ok, \_\_seq_set_item:nNnnNnn makes the assignment and returns \texttt{true} (in the case of conditionals). Here \#1 is an integer expression (position minus one), it needs to be evaluated. The sequence \#5 starts with \_\_seq (even if empty), which stops the integer expression and is absorbed by it. The \texttt{\_if_meaning:w} test is slightly faster than an integer test (but only works when testing against zero, hence the offset we chose in the position). When we are done skipping items, insert the saved item \_\_\_\_seq_internal_a_tl. For \texttt{put} functions the last argument of \_\_\_\_seq_set_item_end:w is \texttt{\use_none:nn} and it absorbs the item \#2 that we are removing: this is only useful for the \texttt{pop} functions.

(End of definition for \texttt{\_\_seq_set_item:nNnnNnn} and others. These functions are documented on page 156.)

Previously, \texttt{\_\_seq_reverse:N} was coded by collecting the items in reverse order after an \texttt{\exp_stop_f:} marker.

At first, this seems optimal, since we can forget about each item as soon as it is placed after \texttt{\exp_stop_f:}. Unfortunately, \TeX{}'s usual tail recursion does not take place in
this case: since the following \texttt{\_\_seq_re\_verse_i\_tem:nw} only reads tokens until \texttt{\exp_\texttt{-\_\_\_stop\_f}}; and never reads the \texttt{\_\_\_item:n \{#1\}} left by the previous call, \TeX{} cannot remove that previous call from the stack, and in particular must retain the various macro parameters in memory, until the end of the replacement text is reached. The stack is thus only flushed after all the \texttt{\_\_\_seq_re\_verse_i\_tem:nw} are expanded. Keeping track of the arguments of all those calls uses up a memory quadratic in the length of the sequence. \TeX{} can then not cope with more than a few thousand items.

Instead, we collect the items in the argument of \texttt{\exp_not:n}. The previous calls are cleanly removed from the stack, and the memory consumption becomes linear.

\begin{verbatim}
16939 \cs_new_protected:Npn \seq_reverse:N { \__seq_reverse:NN \__kernel_tl_set:Nx }
16940 \cs_new_protected:Npn \seq_greverse:N { \__seq_reverse:NN \__kernel_tl_gset:Nx }
16941 \cs_new_protected:Npn \__seq_reverse:NN #1 #2 
16942 { \cs_set_eq:NN \__seq_tmp:w \__seq_item:n
16943 \cs_set_eq:NN \__seq_item:n \__seq_reverse_item:nwn
16944 #1 #2 { \exp_not:n { } }
16945 \cs_set_eq:NN \__seq_item:n \__seq_tmp:w }
16946 \cs_new:Npn \__seq_reverse_item:nwn #1 #2 \exp_not:n #3 
16947 { #2
16948 \exp_not:n { \__seq_item:n {#1} #3 }
16949 }
16950 \cs_generate_variant:Nn \seq_reverse:N { c }
16951 \cs_generate_variant:Nn \seq_greverse:N { c }
\end{verbatim}

\begin{tt}
\seq_sort:Nn
\seq_sort:cn
\seq_gsort:Nn
\seq_gsort:cn
\end{tt}

Implemented in \texttt{l3sort}.

\begin{tt}
\seq_shuffle:N
\seq_shuffle:c
\seq_gshuffle:N
\seq_gshuffle:c
\end{tt}

\section*{58.4 Sequence conditionals}

Similar to token lists, we compare with the empty sequence.

\begin{verbatim}
16957 \prg_new_conditional:Npnn \seq_if_empty:N #1 { p , T , F , TF } 
16958 { \if_meaning:w #1 \c_empty_seq
16959 \prg_return_true:
16960 \else:
16961 \prg_return_false:
16962 \fi:
16963 \prg_generate_conditional_variant:Nnn \seq_if_empty:N 
16964 { c } { p , T , F , TF }
\end{verbatim}

\begin{tt}
\seq_if_empty:p:N
\seq_if_empty:p:c
\seq_if_empty:NTF
\seq_if_empty:cTF
\end{tt}

\begin{tt}
\seq_shuffle:N
\seq_shuffle:c
\seq_gshuffle:N
\seq_gshuffle:c
\__seq_shuffle:NN
\__seq_shuffle_item:n
\g__seq_internal_seq
\end{tt}

\begin{tt}
\seq_sort:Nn
\seq_sort:cn
\seq_gsort:Nn
\seq_gsort:cn
\end{tt}

\begin{tt}
\seq_shuffle:N
\seq_shuffle:c
\seq_gshuffle:N
\seq_gshuffle:c
\__seq_shuffle:NN
\__seq_shuffle_item:n
\g__seq_internal_seq
\end{tt}

\begin{verbatim}
\end{verbatim}

\begin{verbatim}
\end{verbatim}
there are more possible permutations than possible seeds ($13! > 2^{28}$) so the question of uniformity is somewhat moot. The integer variables are declared in \texttt{l3int: load-order issues}.

\begin{verbatim}
\cs_new:N \g__seq_internal_seq
\cs_new_protected:Npn \seq_shuffle:N { \__seq_shuffle:NN \seq_set_eq:NN }
\cs_new_protected:Npn \seq_gshuffle:N { \__seq_shuffle:NN \seq_gset_eq:NN }
\cs_new_protected:Npn \__seq_shuffle:NN #1#2
{ \int_compare:nNnTF { \seq_count:N #2 } > \c_max_register_int
{ \msg_error:nne { seq } { shuffle-too-large } \{ \token_to_str:N #2 \}
{ \group_begin:
\int_zero:N \l__seq_internal_a_int
\__seq_push_item_def:
\cs_gset_eq:NN \__seq_item:n \__seq_shuffle_item:n #2
\__seq_pop_item_def:
\seq_gclear:N \g__seq_internal_seq
\int_step_inline:nn \l__seq_internal_a_int
{ \seq_gput_right:Ne \g__seq_internal_seq \{ \tex_the:D \tex_toks:D ##1 \}
\group_end:
#1 #2 \g__seq_internal_seq
\seq_gclear:N \g__seq_internal_seq
\group_end:
#1 #2 \g__seq_internal_seq
\seq_gclear:N \g__seq_internal_seq
\cs_new_protected:Npn \__seq_shuffle_item:n
{ \int_incr:N \l__seq_internal_a_int
\int_set:Nn \l__seq_internal_b_int
{ 1 + \tex_uniformdeviate:D \l__seq_internal_a_int }
\tex_toks:D \l__seq_internal_a_int
= \tex_toks:D \l__seq_internal_b_int
\tex_toks:D \l__seq_internal_b_int }
\cs_generate_variant:Nn \seq_shuffle:N { c }
\cs_generate_variant:Nn \seq_gshuffle:N { c }
\end{verbatim}

(End of definition for \texttt{\seq_shuffle:N} and others. These functions are documented on page 157.)

The approach here is to define \texttt{\__seq_item:n} to compare its argument with the test sequence. If the two items are equal, the mapping is terminated and \texttt{\group_end: \prg_return_true:} is inserted after skipping over the rest of the recursion. On the other hand, if there is no match then the loop breaks, returning \texttt{\prg_return_false:}. Everything is inside a group so that \texttt{\_\_seq_item:n} is preserved in nested situations.

\begin{verbatim}
\prg_new_protected_conditional:Nn \seq_if_in:Nn \{ T , F , TF \}
\prg_new_protected_conditional:Nn \seq_if_in:NV \{ T , F , TF \}
\prg_new_protected_conditional:Nn \seq_if_in:cv \{ T , F , TF \}
\prg_new_protected_conditional:Nn \seq_if_in:cVF \{ T , F , TF \}
\prg_new_protected_conditional:Nn \seq_if_in:CTF \{ T , F , TF \}
\prg_new_protected_conditional:Nn \seq_if_in:coTF \{ T , F , TF \}
\prg_new_protected_conditional:Nn \seq_if_in:cxTF \{ T , F , TF \}
\end{verbatim}

828
58.5 Recovering data from sequences

The two pop functions share their emptiness tests. We also use a common emptiness test for all branching get and pop functions.

\begin{verbatim}
\cs_new_protected:Npn \_seq_pop:NNNN \_seq_pop:TF:NNNN
\cs_new_protected:Npn \_seq_pop:NNNN #1#2#3#4
{\if_meaning:w #3 \c_empty_seq
 \tl_set:Nn #4 { \q_no_value }
\else:
 #1#2#3#4
\fi:}
\cs_new_protected:Npn \_seq_pop:TF:NNNN #1#2#3#4
{\if_meaning:w #3 \c_empty_seq
 % \tl_set:Nn #4 { \q_no_value }
 \prg_return_false:
\else:
 #1#2#3#4
 \prg_return_true:
\fi:}
\end{verbatim}

\begin{verbatim}
\_seq_get_left:NN \_seq_get_left:cN \_seq_get_left:wnw
\_seq_get_left:NN
Getting an item from the left of a sequence is pretty easy: just trim off the first item after \_seq_item:n at the start. We append a \q_no_value item to cover the case of an empty sequence
\begin{verbatim}
\cs_new_protected:Npn \seq_get_left:NN \seq_get_left:cN \seq_get_left:wnw
{\_kernel_tl_set:Nx \_seq_get_left:wnw
 \exp_after:wN \_seq_get_left:wnw
\end{verbatim}

(End of definition for \_seq_pop:NNNN and \_seq_pop:TF:NNNN.)

(End of definition for \seq_if_in:MnTF and \_seq_if_in:. This function is documented on page 157.)
The approach to popping an item is pretty similar to that to get an item, with the only difference being that the sequence itself has to be redefined. This makes it more sensible to use an auxiliary function for the local and global cases.

First remove \texttt{\_seq} and prepend \texttt{q_no_value}. The first argument of \texttt{\_seq_get_right_loop:nw} is the last item found, and the second argument is empty until the end of the loop, where it is code that applies \texttt{\exp_not:n} to the last item and ends the loop.

(End of definition for \texttt{\seq_pop_left:NN} and \texttt{\seq_pop_left:cN}. This function is documented on page 154.)
The approach to popping from the right is a bit more involved, but does use some of the same ideas as getting from the right. What is needed is a “flexible length” way to set a token list variable. This is supplied by the \if_false: \fi: \ldots if \false: \{} \fi: construct. Using an e-type expansion and a “non-expanding” definition for \__seq_item:n, the left-most \( n - 1 \) entries in a sequence of \( n \) items are stored back in the sequence. That needs a loop of unknown length, hence using the strange \if_false: way of including braces. When the last item of the sequence is reached, the closing brace for the assignment is inserted, and \tl_set:Nn \#3 is inserted in front of the final entry. This therefore does the pop assignment. One more iteration is performed, with an empty argument and \use_none:nn, which finally stops the loop.

\begin{verbatim}
\cs_new_protected:Npn \seq_pop_right:NN { \__seq_pop:NNNN \__seq_pop_right:NNN \__kernel_tl_set:Nx }
\cs_new_protected:Npn \seq_gpop_right:NN { \__seq_pop:NNNN \__seq_pop_right:NNN \__kernel_tl_gset:Nx }
\cs_new_protected:Npn \__seq_pop_right:NNN #1#2#3
{ \cs_set_eq:NN \__seq_tmp:w \__seq_item:n
\cs_set_eq:NN \__seq_item:n \scan_stop:
#1 #2{ \if_false: } \fi: \s__seq
\exp_after:wN \use_i:nnn \exp_after:wN \__seq_pop_right_loop:nn
#2{ } \use_none:nn
\cs_set_eq:NN \__seq_item:n \__seq_tmp:w
}
\cs_new:Npn \__seq_pop_right_loop:nn #1#2
{ #2{ \exp_not:n {#1} } \__seq_pop_right_loop:nn
}
\cs_generate_variant:Nn \seq_pop_right:NN { c }
\cs_generate_variant:Nn \seq_gpop_right:NN { c }
\end{verbatim}

(End of definition for \seq_pop_right:NN and others. These functions are documented on page 154.)

Getting from the left or right with a check on the results. The first argument to \__seq_pop_TF:NNNN is left unused.

\begin{verbatim}
\prg_new_protected_conditional:Npnn \seq_get_left:NNT { T , F , TF }
\prg_generate_conditional_variant:Nnn \seq_get_left:NNT { c } { T , F , TF }
\end{verbatim}

(End of definition for \seq_get_left:NNT and \seq_get_right:NNF. These functions are documented on page 155.)
More or less the same for popping.

\prg_new_protected_conditional:Nnn \seq_pop_left:NN \seq_pop_left:cTF \seq_pop_left:NNTF \seq_pop_left:NNTF \seq_gpop_left:NN \seq_gpop_left:cTF \seq_gpop_left:NNTF \seq_gpop_left:NNTF \seq_pop_right:NN \seq_pop_right:cTF \seq_pop_right:NNTF \seq_gpop_right:NN \seq_gpop_right:cTF \seq_gpop_right:NNTF \seq_gpop_right:NNTF

(End of definition for \seq_pop_left:NNTF and others. These functions are documented on page 155.)

The idea here is to find the offset of the item from the left, then use a loop to grab the correct item. If the resulting offset is too large, then the argument delimited by \seq_item:n is \prg_break: instead of being empty, terminating the loop and returning nothing at all.

\cs_new:Npn \seq_item:Nn #1
{ \exp_after:wN \__seq_item:wNn #1 \s__seq_stop #1 }

\cs_new:Npn \__seq_item:wNn \s__seq #1#2#3
{ \exp_args:Nf \__seq_item:nwn { \int_eval:n {#3} } #2 #1 \prg_break: \__seq_item:n { } \prg_break_point: }

\cs_new:Npn \__seq_item:nwn #1#2 \__seq_item:n #3
{ #2 \int_compare:nNnTF {#1} = 1 { \prg_break:n { \exp_not:n {#3} } } { \exp_args:Nf \__seq_item:nwn { \int_eval:n { #1 - 1 } } } }

(End of definition for \seq_item:Nn and others. This function is documented on page 154.)
Importantly, \seq_item:Nn only evaluates its argument once.

\begin{verbatim}
\cs_new:Npn \seq_rand_item:N #1
  \{ \seq_if_empty:NF #1
  \{ \seq_item:Nn \#1 \{ \int_rand:nn \{ 1 \} \{ \seq_count:N \#1 \} \}
  \}
\cs_generate_variant:Nn \seq_rand_item:N \c
\end{verbatim}

(End of definition for \seq_rand_item:N. This function is documented on page 155.)

58.6 Mapping over sequences

\begin{verbatim}
\seq_map_break:
\seq_map_break:n
\begin{verbatim}
\cs_new:Npn \seq_map_break:
  \prg_map_break:Nn \seq_map_break: { }
\end{verbatim}
\end{verbatim}

(End of definition for \seq_map_break: and \seq_map_break:n. These functions are documented on page 159.)

\begin{verbatim}
\seq_map_function:NN
\seq_map_function:cN
\__seq_map_function:Nw
\begin{verbatim}
\cs_new:Npn \seq_map_function:NN \#1#2
  \exp_after:wN \use_i_ii:nnn
  \exp_after:wN \__seq_map_function:Nw
  \exp_after:wN \#2 \#1 \prg_break:
  \__seq_item:n \#1 \__seq_item:n \#1 \__seq_item:n \#1 \__seq_item:n \#1
  \prg_break_point:
  \prg_break_point:Nn \seq_map_break: { }
\end{verbatim}
\end{verbatim}

The idea here is to apply the code of \#2 to each item in the sequence without altering the definition of \_\_\seq_item:n. The even-numbered arguments of \_\_\seq_map_function:Nw delimited by \_\_\seq_item:n are almost always empty, except at the end of the loop where it is \prg_break:. This allows to break the loop without needing to do a (relatively-expensive) quark test.

\begin{verbatim}
\cs_new:Npn \seq_map_function:NN \#1\#2
  \{ \exp_after:wN \use_i_ii:nnn
  \exp_after:wN \__seq_map_function:Nw
  \exp_after:wN \#2 \#1 \prg_break:
  \_\_\seq_item:n \#1 \_\_\seq_item:n \#1 \_\_\seq_item:n \#1 \_\_\seq_item:n \#1
  \prg_break_point:
  \prg_break_point:Nn \seq_map_break: { }
\}
\end{verbatim}

\end{verbatim}

833
The definition of \_\_seq_item:n needs to be saved and restored at various points within the mapping and manipulation code. That is handled here: as always, this approach uses global assignments.

\begin{verbatim}
\cs_new_protected:Npn \__seq_push_item_def:n
\cs_gset:Npn \__seq_item:n ##1
\cs_new_protected:Npn \__seq_push_item_def:e
\cs_gset:Npe \__seq_item:n ##1
\cs_new_protected:Npn \__seq_push_item_def:
\int_gincr:N \g__kernel_prg_map_int
\cs_gset_eq:cN { \_\_seq_map_ \int_use:N \g__kernel_prg_map_int :w }
\__seq_item:n
\cs_new_protected:Npn \__seq_pop_item_def:
\cs_gset_eq:Nc \__seq_item:n { \_\_seq_map_ \int_use:N \g__kernel_prg_map_int :w }
\int_gdecr:N \g__kernel_prg_map_int
\end{verbatim}

The idea here is that \_\_seq_item:n is already “applied” to each item in a sequence, and so an in-line mapping is just a case of redefining \_\_seq_item:n.

\begin{verbatim}
\cs_new_protected:Npn \seq_map_inline:Nn #1#2
\__seq_push_item_def:n {#2}
#1
\prg_break_point:Nn \seq_map_break: { \__seq_pop_item_def: }
\cs_generate_variant:Nn \seq_map_inline:Nn { c }
\end{verbatim}

This is based on the function mapping but using the same tricks as described for \prop_map_tokens:Nn. The idea is to remove the leading \_\_seq and apply the tokens such that they are safe with the break points, hence the \use:n.

\begin{verbatim}
\cs_new:Npn \seq_map_tokens:Nn #1#2
\\exp_last_unbraced:Nno
\use_i:nn { \_\_seq_map_tokens:nw {#2} } #1
\prg_break:
\_\_seq_item:n \__seq_item:n \__seq_item:n \__seq_item:n
\prg_break_point:Nn \seq_map_break: { }
\end{verbatim}

(End of definition for \seq_map_function:NN and \_\_seq_map_function:Nw. This function is documented on page 157.)
\seq_map_variable:NNn
\seq_map_variable:Ncn
\seq_map_variable:cNn
\seq_map_variable:ccn

This is just a specialised version of the in-line mapping function, using an e-type expansion for the code set up so that the number of # tokens required is as expected.

\cs_new_protected:Npn \seq_map_variable:NNn \#1\#2\#3
{
  \__seq_push_item_def:e
  \tl_set:Nn \exp_not:N \#2 {##1}
  \exp_not:n \#3
}
\prg_break_point:Nn \seq_map_break: { \__seq_pop_item_def: }

\cs_new_protected:Npn \seq_map_variable:Ncn\cs_new_protected:Npn \seq_map_variable:cNn\cs_new_protected:Npn \seq_map_variable:ccn

(End of definition for \seq_map_variable:NNn. This function is documented on page 158.)

\seq_map_indexed_function:NN
\seq_map_indexed_inline:Nn
\__seq_map_indexed:nNN
\__seq_map_indexed:Nw

Similar to \seq_map_function:NN but we keep track of the item index as a ; delimited argument of \__seq_map_indexed:Nw.

\cs_new:Npn \seq_map_indexed_function:NN \#1\#2
{
  \__seq_map_indexed:NN \#1\#2
  \prg_break_point:Nn \seq_map_break: { }
}
\cs_new_protected:Npn \seq_map_indexed_inline:Nn \#1\#2
{
  \int_gincr:N \g__kernel_prg_map_int
  \cs_gset_protected:cpn
  \{ \__seq_map_ \int_use:N \g__kernel_prg_map_int :w \} \#1\#2 \{#2\}
  \exp_args:NNc \__seq_map_indexed:NN \#1
  \prg_break_point:Nn \seq_map_break:
  \{ \int_gdecr:N \g__kernel_prg_map_int \}
}
\cs_new:Npn \__seq_map_indexed:NN \#1\#2
{
  \int_gincr:N \g__kernel_prg_map_int
  \cs_gset_protected:cpn
  \{ \__seq_map_ \int_use:N \g__kernel_prg_map_int :w \} \#1\#2 \{#2\}
  \exp_args:NNc \__seq_map_indexed:NN \#1
  \prg_break_point:Nn \seq_map_break:
  \{ \int_gdecr:N \g__kernel_prg_map_int \}
}
\cs_new:Npn \__seq_map_indexed:NN \#1\#2
{
}

835
The idea is to first expand both sequences, adding the usual \{ ? \prg_break: \} \} to the end of each one. This is most conveniently done in two steps using an auxiliary function. The mapping then throws away the first tokens of #2 and #5, which for items in both sequences are \s__seq \__seq_item:n. The function to be mapped are then be applied to the two entries. When the code hits the end of one of the sequences, the break material stops the entire loop and tidy up. This avoids needing to find the count of the two sequences, or worrying about which is longer.

Very similar to \seq_set_filter:NNn. We could actually merge the two within a single function, but it would have weird semantics.
\cs_new_protected:Npn \seq_set_map_e:NNn
\{ \_\_seq_set_map_e:NNNn \_\_kernel_tl_set:Nx \}
\cs_new_protected:Npn \seq_gset_map_e:NNn
\{ \_\_seq_set_map_e:NNNn \_\_kernel_tl_gset:Nx \}
\cs_new_protected:Npn \__seq_set_map_e:NNNn #1#2#3#4
\{
\_\_seq_push_item_def:n \{ \exp_not:N \_\_seq_item:n \{#4\} \}
\_\_seq_pop_item_def:
\}

(End of definition for \seq_set_map_e:NNn, \seq_gset_map_e:NNn, and \__seq_set_map_e:NNNn. These functions are documented on page 160.)

\seq_set_map:NNn \seq_gset_map:NNn \__seq_set_map:NNNn
Similar to \seq_set_map_e:NNn, but prevents expansion of the <inline function>.
\cs_new_protected:Npn \seq_set_map:NNn
\seq_gset_map:NNn
\__seq_set_map:NNNn
\{ \_\_seq_set_map:NNNn \_\_kernel_tl_set:Nx \}
\cs_new_protected:Npn \seq_gset_map:NNn
\_\_seq_set_map:NNNn \_\_kernel_tl_gset:Nx
\cs_new_protected:Npn \__seq_set_map:NNNn #1#2#3#4
\{
\_\_seq_push_item_def:n \{ \exp_not:n \_\_seq_item:n \{#4\} \}
\_\_seq_pop_item_def:
\}

(End of definition for \seq_set_map:NNn, \seq_gset_map:NNn, and \__seq_set_map:NNNn. These functions are documented on page 159.)

\seq_count:N \seq_count:c \__seq_count:w \__seq_count_end:w
Since counting the items in a sequence is quite common, we optimize it by grabbing 8 items at a time and correspondingly adding 8 to an integer expression. At the end of the loop, #9 is \__seq_count_end:w instead of being empty. It removes 8+ and instead places the number of \_\_seq_item:n that \__seq_count:w grabbed before reaching the end of the sequence.
\cs_new:Npn \seq_count:N #1
\{
\int_eval:n
\{\exp_after:wN \use_i:nn \exp_after:wN \_\_seq_count:w
\_\_seq_count_end:w \_\_seq_item:n 7
\_\_seq_count_end:w \_\_seq_item:n 6
\_\_seq_count_end:w \_\_seq_item:n 5
\_\_seq_count_end:w \_\_seq_item:n 4
\_\_seq_count_end:w \_\_seq_item:n 3
\_\_seq_count_end:w \_\_seq_item:n 2
\_\_seq_count_end:w \_\_seq_item:n 1
\_\_seq_count_end:w \_\_seq_item:n 0
\prg_break_point:
\}
\}
\cs_new:Npn \_\_seq_count:w
\_\_seq_item:n 1 \_\_seq_item:n 2 \_\_seq_item:n 3 \_\_seq_item:n 4 \_\_seq_item:n

837
58.7 Using sequences

See \clist_use:Nnnn for a general explanation. The main difference is that we use \_\seq_item:n as a delimiter rather than commas. We also need to add \_\seq_item:n at various places, and \s\_seq.

(End of definition for \seq_count:N, \_\seq_count:w, and \_\seq_count_end:w. This function is documented on page 160.)
58.8 Sequence stacks

The same functions as for sequences, but with the correct naming.

```
\seq_push:Nn
\seq_push:NV
\seq_push:Ne
\seq_push:Nx
\seq_push:cn
\seq_gpush:Nn
```

Pushing to a sequence is the same as adding on the left.

```
\cs_new_eq:NN \seq_push:Nn \seq_put_left:Nn
\cs_generate_variant:Nn \seq_push:Nn { NV , Nv , Ne , c , cV , cv , ce }
\cs_generate_variant:Nn \seq_push:Nn { No , Nx , co , cx }
```

(End of definition for \seq_push:Nn and \seq_gpush:Nn. These functions are documented on page 162.)

In most cases, getting items from the stack does not need to specify that this is from the left. So aliases are provided.

```
\cs_new_eq:NN \seq_get:NN \seq_get_left:NN
\cs_new_eq:NN \seq_get:cN \seq_get_left:cN
\cs_new_eq:NN \seq_pop:NN \seq_pop_left:NN
```

(End of definition for \seq_get:NN, \seq_pop:NN, and \seq_gpop:NN. These functions are documented on page 161.)

```
\prg_new_eq_conditional:NNn \seq_get:NN \seq_get_left:NN { T , F , TF }
\prg_new_eq_conditional:NNn \seq_get:cN \seq_get_left:cN { T , F , TF }
\prg_new_eq_conditional:NNn \seq_pop:NN \seq_pop_left:NN { T , F , TF }
```

(End of definition for \seq_get:NN, \seq_pop:NN, and \seq_gpop:NN. These functions are documented on page 161.)

58.9 Viewing sequences

```
\seq_show:N
\seq_show:c
\seq_log:N
```

Apply the general \_kernel_chk_tl_type:NnT.

```
\__seq_show:N { seq }
```

More copies.

```
\prg_new_eq_conditional:NNn \seq_get:NN \seq_get_left:NN { T , F , TF }
\prg_new_eq_conditional:NNn \seq_get:cN \seq_get_left:cN { T , F , TF }
```

(End of definition for \seq_get:NN, \seq_pop:NN, and \seq_gpop:NN. These functions are documented on page 161.)
(End of definition for \seq_show:N and others. These functions are documented on page 164.)

\section{Scratch sequences}

\begin{itemize}
\item \texttt{l_tmpa_seq} Temporary comma list variables.
\item \texttt{l_tmpb_seq}
\item \texttt{g_tmpa_seq}
\item \texttt{g_tmpb_seq}
\end{itemize}

(End of definition for \texttt{l_tmpa_seq} and others. These variables are documented on page 164.)

\end{document}
Chapter 59

l3int implementation

The following test files are used for this code: m3int001,m3int002,m3int03.

\c_max_register_int

Done in l3basics.

(End of definition for \c_max_register_int. This variable is documented on page 178.)

\__int_to_roman:w \if_int_compare:w

Done in l3basics.

(End of definition for \__int_to_roman:w and \if_int_compare:w. This function is documented on page 179.)

\or:

Done in l3basics.

(End of definition for \or:. This function is documented on page 179.)

\int_value:w \__int_eval:w \__int_eval_end: \if_int_odd:w \if_case:w

Here are the remaining primitives for number comparisons and expressions.

\cs_new_eq:NN \int_value:w \tex_number:D
\cs_new_eq:NN \__int_eval:w \tex_numexpr:D
\cs_new_eq:NN \__int_eval_end: \tex_relax:D
\cs_new_eq:NN \if_int_odd:w \tex_ifodd:D
\cs_new_eq:NN \if_case:w \tex_ifcase:D

(End of definition for \int_value:w and others. These functions are documented on page 179.)

\s__int_mark \s__int_stop

Scan marks used throughout the module.

\scan_new:N \s__int_mark
\scan_new:N \s__int_stop

(End of definition for \s__int_mark and \s__int_stop.)

\__int_use_none_delimit_by_s_stop:w

Function to gobble until a scan mark.

\cs_new:Npn \__int_use_none_delimit_by_s_stop:w #1 \s__int_stop { }

(End of definition for \__int_use_none_delimit_by_s_stop:w.)

\q__int_recursion_tail \q__int_recursion_stop

Quarks for recursion.

\quark_new:N \q__int_recursion_tail
\quark_new:N \q__int_recursion_stop
59.1 Integer expressions

Wrapper for \_\_int_eval:w: can be used in an integer expression or directly in the input stream. It is very slightly faster to use \the rather than \number to turn the expression to a number. When debugging, we introduce parentheses to catch early termination (see \l3debug).

```
\cs_new:Npn \int_eval:n #1
\cs_new:Npn \int_eval:w { \tex_the:D \__int_eval:w #1 \__int_eval_end: }
```

See \int_abs:n. Evaluate the expression once (and when debugging is enabled, check that the expression is well-formed), then test the first character to determine the sign. This is wrapped in \int_value:w...\exp_stop_f: to ensure a fixed number of expansions and to avoid dealing with closing the conditionals.

```
\cs_new:Npn \int_abs:n #1
\cs_new:Npn \__int_abs:N #1
\cs_new:Npn \int_max:nn #1#2
\cs_new:Npn \int_min:nn #1#2
\cs_new:Npn \__int_maxmin:wwN #1
```

Functions for min, max, and absolute value with only one evaluation. The absolute value is obtained by removing a leading sign if any. All three functions expand in two steps.
As \texttt{\_\_int_eval:w} rounds the result of a division we also provide a version that truncates the result. We use an auxiliary to make sure numerator and denominator are only evaluated once: this comes in handy when those are more expressions are expensive to evaluate (e.g., \texttt{\_\_int_eval:w} \texttt{\_\_int_div_truncate:NwWw}). If the numerator \texttt{\_\_int_eval:w} \texttt{\_\_int_div_truncate:NwWw} is 0, then we divide 0 by the denominator (this ensures that \texttt{0/0} is correctly reported as an error). Otherwise, shift the numerator \texttt{\_\_int_eval:w} \texttt{\_\_int_div_truncate:NwWw} towards 0 by \texttt{(#3#4 - 1)/2}, which we round away from zero. It turns out that this quantity exactly compensates the difference between \LaTeX's rounding and the truncating behaviour that we want. The details are thanks to Heiko Oberdiek: getting things right in all cases is not so easy.
For the sake of completeness:

```latex
\cs_new:Npn \int_div_round:nn #1#2
\{ \int_value:w \__int_eval:w ( #1 ) / ( #2 ) \__int_eval_end: \}
```

Finally there's the modulus operation.

```latex
\cs_new:Npn \int_mod:nn #1#2
\{ \int_value:w \__int_eval:w \exp_after:wN \__int_mod:ww
\int_value:w \__int_eval:w #1 \exp_after:wN ;
\int_value:w \__int_eval:w #2 ;
\__int_eval_end:
\}
```

(End of definition for \int_div_truncate:nn and others. These functions are documented on page 168.)

\begin{verbatim}
\__kernel_int_add:nnn
\end{verbatim}

Equivalent to \int_eval:n \{#1+#2+#3\} except that overflow only occurs if the final result overflows \([−2^{31} + 1, 2^{31} − 1]\). The idea is to choose the order in which the three numbers are added together. If \#1 and \#2 have opposite signs (one is in \([−2^{31} + 1, −1]\) and the other in \([0, 2^{31} − 1]\)) then \#1+\#2 cannot overflow so we compute the result as \#1+\#2+\#3. If they have the same sign, then either \#3 has the same sign and the order does not matter, or \#3 has the opposite sign and any order in which \#3 is not last will work. We use \#1+\#3+\#2.

```latex
\cs_new:Npn \__kernel_int_add:nnn #1#2#3
\{ \int_value:w \__int_eval:w \__int_eval:w \exp_after:wN \__int_div_truncate:NwNw #1 \exp_after:wN #2 \exp_after:wN ;
\__int_eval_end:
\}
```

(End of definition for \__kernel_int_add:nnn.)

\section{Creating and initialising integers}

\texttt{\int_new:N} and \texttt{\int_new:c} Two ways to do this: one for the format and one for the \LaTeX{} package. In plain \TeX{}, \newcount (and other allocators) are \texttt{outer}: to allow the code here to work in “generic” mode this is therefore accessed by name. (The same applies to \newbox, \newdimen and so on.)

```latex
\cs_new_protected:Npn \int_new:N #1
\{ \__kernel_chk_if_free_cs:N #1 \cs:w newcount \cs_end: #1 \}
```

(End of definition for \int_new:N. This function is documented on page 168.)
As stated, most constants can be defined as `\chardef` or `\mathchardef` but that’s engine
dependent. As a result, there is some set up code to determine what can be done. No full
engine testing just yet so everything is a little awkward. We cannot use `\int_gset:Nn`
because (when `\check-declarations` is enabled) this runs some checks that constants
would fail.

```
\cs_new_protected:Npm \int_const:Nn \#1#2
  \{ \__int_const:eN \{ \int_eval:n {#2} \} #1 \}
\cs_generate_variant:Nn \int_const:Nn { c }
\cs_new_protected:Npm \__int_const:nN \#1#2
  \{ \int_compare:nNnTF {#1} < \c_zero_int
    \{ \int_new:N #2 \text_global:D \}
  \}
\cs_generate_variant:Nn \int_compare:nNnTF { e }
\if_int_odd:w 0
  \cs_if_exist:NT \tex_luatexversion:D { 1 }
  \cs_if_exist:NT \tex_omathchardef:D { 1 }
  \cs_if_exist:NTF \tex_omathchardef:D
    { \cs_new_eq:NN \__int_constdef:Nw \tex_omathchardef:D }
    { \cs_new_eq:NN \__int_constdef:Nw \tex_mathchardef:D }
  \else:
    \cs_new_eq:NN \__int_constdef:Nw \tex_mathchardef:D
    \text_global:D \__int_constdef:Nw \c__int_max_constdef_int
    \c__int_max_constdef_int 1114111
  \fi:
```

(End of definition for `\int_const:Nn` and others. This function is documented on page 168.)

```
\int_zero:N
\int_zero:c
\int_gzero:N
\int_gzero:c
\int_zero_new:N
\int_zero_new:c
\int_gzero_new:N
\int_gzero_new:c
```

Functions that reset an ⟨integer⟩ register to zero.

```
\cs_new_protected:Npm \int_zero:N \#1 \{ \#1 = \c_zero_int \}
\cs_new_protected:Npm \int_gzero:N \#1 \{ \text_global:D \#1 = \c_zero_int \}
\cs_generate_variant:Nn \int_zero:N { c }
\cs_generate_variant:Nn \int_gzero:N { c }
```

(End of definition for `\int_zero:N` and `\int_gzero:N`. These functions are documented on page 169.)

```
\int_zero_new:N
\int_zero_new:c
\int_gzero_new:N
\int_gzero_new:c
```

Create a register if needed, otherwise clear it.
59.3 Setting and incrementing integers

Adding and subtracting to and from a counter. Including here the optional by would slow down these operations by a few percent.

Incrementing and decrementing of integer registers is done with the following functions.

(Copies of the cs functions defined in l3basics.

(End of definition for \int_if_exist:NTF. This function is documented on page 169.)

(End of definition for \int_set_eq:NN and \int_gset_eq:NN. These functions are documented on page 169.)

(End of definition for \int_zero_new:N and \int_gzero_new:N. These functions are documented on page 169.)

(End of definition for \int_add:Nn and others. These functions are documented on page 169.)

(End of definition for \int_if_exist:NTF #1 { \int_zero:N #1 } { \int_new:N #1 } )
\cs_new_protected:Npn \int_zero_new:N #1
\cs_new_protected:Npn \int_gzero_new:N #1
\cs_generate_variant:Nn \int_zero_new:N { c }
\cs_generate_variant:Nn \int_gzero_new:N { c }
\cs_generate_variant:Nn \int_zero_new:N { c , Nc , cc }
\cs_generate_variant:Nn \int_gzero_new:N { c , Nc , cc }
\cs_generate_variant:Nn \int_add:Nn { c }
\cs_generate_variant:Nn \int_gadd:Nn { c }
\cs_generate_variant:Nn \int_sub:Nn { c }
\cs_generate_variant:Nn \int_gsub:Nn { c }
\cs_new_protected:Npn \int_incr:N #1
\cs_new_protected:Npn \int_decr:N #1
\cs_new_protected:Npn \int_gincr:N #1
\cs_new_protected:Npn \int_gdecr:N #1
\prg_new_eq_conditional:NNn \int_if_exist:N \cs_if_exist:N { TF , T , F , p }
\prg_new_eq_conditional:NNn \int_if_exist:c \cs_if_exist:c { TF , T , F , p }
\int_set_eq:NN \int_set_eq:cN \int_set_eq:Nc \int_set_eq:cc
\int_gset_eq:NN \int_gset_eq:cN \int_gset_eq:Nc \int_gset_eq:cc
As integers are register-based \TeX{} issues an error if they are not defined. While the =
sign is optional, this version with = is slightly quicker than without, while adding the
optional space after = slows things down minutely.

\begin{verbatim}
\cs_new_protected:Npn \int_set:Nn #1#2 { \_\_int_eval:w #2 \_\_int_eval_end: }
\cs_new_protected:Npn \int_gset:Nn #1#2 { \tex_global:D #1 = \_\_int_eval:w #2 \_\_int_eval_end: }
\end{verbatim}

(End of definition for \int_set:Nn and \int_gset:Nn. These functions are documented on page 169.)

59.4 Using integers

\begin{verbatim}
\cs_new_eq:NN \int_use:N \tex_the:D
\cs_new:Npn \int_use:c #1 { \tex_the:D \cs:w #1 \cs_end: }
\end{verbatim}

(End of definition for \int_use:N. This function is documented on page 170.)

59.5 Integer expression conditionals

Those functions are used for comparison tests which use a simple syntax where only
one set of braces is required and additional operators such as != and >= are supported.
The tests first evaluate their left-hand side, with a trailing \_\_int_compare_error:. This
marker is normally not expanded, but if the relation symbol is missing from the
test’s argument, then the marker inserts = (and itself) after triggering the relevant \TeX
error. If the first token which appears after evaluating and removing the left-hand side is
not a known relation symbol, then a judiciously placed \_\_int_compare_error:Nw gets
expanded, cleaning up the end of the test and telling the user what the problem was.

\begin{verbatim}
\cs_new_protected:Np \_\_int_compare_error: {
  \if_int_compare:w \c_zero_int \c_zero_int \fi:
  \_\_int_compare_error:
}\cs_new_protected:Np \_\_int_compare_error:Nw #1#2 \s__int_stop {
  \_\_int_compare_error:
\msg_expandable_error:nnn { kernel } { unknown-comparison } {#1}
\end{verbatim}

847
Comparisons tests using a simple syntax where onle onle set of braces is required, aoditional operators such as != and >= are supported, and multiple comparisons can be performed at once, for instance 0 < 5 <= 1. The idea is to loop through the argument, finding one operand at a time, and comparing it to the previous one. The looping auxiliary \__int_compare:Nw reads one ('operand') and one ('comparison') symbol, and leaves roughly

\begin{verbatim}
\reverse_if:N \if_int_compare:w \langle operand \rangle \langle comparison \rangle
\__int_compare:Nw
\end{verbatim}

in the input stream. Each call to this auxiliary provides the second operand of the last call's if_int_compare:w. If one of the ('comparisons') is false, the true branch of the \TeX{} conditional is taken (because of \reverse_if:N), immediately returning false as the result of the test. There is no \TeX{} conditional waiting the first operand, so we add an if_false: and expand by hand with \int_value:w, thus skipping \prg_return_false: on the first iteration.

Before starting the loop, the first step is to make sure that there is at least one relation symbol. We first let \TeX{} evaluate this left hand side of the (in)equality using \__int_eval:w. Since the relation symbols <, >, = and ! are not allowed in integer expressions, they would terminate the expression. If the argument contains no relation symbol, \__int_compare_error: is expanded, inserting = and itself after an error. In all cases, \__int_compare:w receives as its argument an integer, a relation symbol, and some more tokens. We then setup the loop, which is ended by the two odd-looking items e and \{=nd\}, with a trailing \s__int_stop used to grab the entire argument when necessary.

The goal here is to find an ('operand') and a ('comparison'). The ('operand') is already evaluated, but we cannot yet grab it as an argument. To access the following relation symbol, we remove the number by applying \__int_to_roman:w, after making sure that the argument becomes non-positive: its roman numeral representation is then empty. Then probe the first two tokens with \__int_compare:Nw to determine the relation symbol, building a control sequence from it (\token_to_str:N gives better errors if \#1 is not a character). All the extended forms have an extra = hence the test for that as a second token. If the relation symbol is unknown, then the control sequence is turned by \TeX{} into \scan_stop:, ignored thanks to \unexpanded, and \__int_compare_error:Nw raises an error.
When the last ⟨operand⟩ is seen, \_\_int_compare:NNw receives e and =nd as arguments, hence calling \_\_int_compare_end_:NNw to end the loop: return the result of the last comparison (involving the operand that we just found). When a normal relation is found, the appropriate auxiliary calls \_\_int_compare:nnN where #1 is \if_int_compare:w or \reverse_if:N \if_int_compare:w, #2 is the ⟨operand⟩, and #3 is one of <, =, or >. As announced earlier, we leave the ⟨operand⟩ for the previous conditional. If this conditional is true the result of the test is known, so we remove all tokens and return false. Otherwise, we apply the conditional #1 to the ⟨operand⟩ #2 and the comparison #3, and call \_\_int_compare:Nw to look for additional operands, after evaluating the following expression.

The actual comparisons are then simple function calls, using the relation as delimiter for a delimited argument and discarding \_\_int_compare_error:Nw ⟨token⟩ responsible for error detection.
More efficient but less natural in typing.

```
\begin{verbatim}
\prg_new_conditional:Npnn \int_compare:nNnTF #1 #2 #3 { p , T , F , TF }
\{
    \if_int_compare:w \__int_eval:w #1 \_ \__int_eval:w #2 \_ \__int_eval:w #3 \_ \__int_eval_end:
    \prg_return_true:
    \else:
    \prg_return_false:
    \fi:
\}
\end{verbatim}
```

(End of definition for \texttt{\int_compare:nNnTF} and others. This function is documented on page 171.)

```
\begin{verbatim}
\prg_new_conditional:Npnn \int_if_zero:nTF #1 { p , T , F , TF }
\{
    \if_int_compare:w \__int_eval:w #1 = \c_zero_int
    \prg_return_true:
    \else:
    \prg_return_false:
    \fi:
\}
\end{verbatim}
```

(End of definition for \texttt{\int_if_zero:nTF}. This function is documented on page 170.)

For integer cases, the first task to fully expand the check condition. The overall idea is then much the same as for \texttt{\str_case:nnTF} as described in l3str.

```
\begin{verbatim}
\cs_new:Npn \__int_case:nnTF #1 #2 #3 #4
\{
    \__int_case:nw {#1} #2 {#1} { } \s__int_mark {#3} \s__int_mark {#4} \s__int_stop
\}
\end{verbatim}
```

```
\begin{verbatim}
\cs_new:Npn \__int_case:nw #1 #2 #3
\{
    \if_int_compare:w \__int_eval:w #1 #2 \__int_eval:n \c_zero_int
    \prg_return_true:
    \else:
    \prg_return_false:
    \fi:
\}
\end{verbatim}
```

(End of definition for \texttt{\int_if_zero:nTF}. This function is documented on page 172.)
A predicate function.

\begin{verbatim}
\prg_new_conditional:Npnn \int_if_odd:n #1 { p , T , F , TF} 
{ \if_int_odd:w \__int_eval:w #1 \__int_eval_end: 
\prg_return_true: 
\else: 
\prg_return_false: 
\fi: }
\prg_new_conditional:Npnn \int_if_even:n #1 { p , T , F , TF} 
{ \reverse_if:N \if_int_odd:w \__int_eval:w #1 \__int_eval_end: 
\prg_return_true: 
\else: 
\prg_return_false: 
\fi: }
\end{verbatim}

(End of definition for \texttt{int\_if\_odd:nTF} and \texttt{int\_if\_even:nTF}. These functions are documented on page 172.)

59.6 Integer expression loops

These are quite easy given the above functions. The \texttt{while} versions test first and then execute the body. The \texttt{do\_while} does it the other way round.

\begin{verbatim}
\cs_new:Npn \int\_while\_do:nn \int\_until\_do:nn \int\_do\_while:nn \int\_do\_until:nn 
{ \int\_compare:nT \{#1\} 
{ #2 
 \int\_while\_do:nn \{#1\} \{#2\} 
} 
\cs_new:Npn \int\_until\_do:nn \int\_do\_while:nn \int\_do\_until:nn 
{ \int\_compare:nF \{#1\} 
{ #2 
 \int\_until\_do:nn \{#1\} \{#2\} 
} 
\cs_new:Npn \int\_do\_while:nn \int\_do\_until:nn 
{ #2 
\end{verbatim}
59.7 Integer step functions

Before all else, evaluate the initial value, step, and final value. Repeating a function by steps first needs a check on the direction of the steps. After that, do the function for the start value then step and loop around. It would be more symmetrical to test for a step size of zero before checking the sign, but we optimize for the most frequent case (positive step).

(End of definition for \texttt{\int\_while\_do:nn} and others. These functions are documented on page 173.)
The approach here is to build a function, with a global integer required to make the nesting safe (as seen in other in line functions), and map that function using \texttt{\int_step_function:nnn}. We put a \texttt{\prg_break_point:Nn} so that \texttt{\map_break} functions from other modules correctly decrement \texttt{\g__kernel_prg_map_int} before looking for their own break point. The first argument is \texttt{\scan_stop:}, so that no breaking function recognizes this break point as its own.

\begin{verbatim}
\cs_new_protected:Npn \int_step_inline:nnn #1 #2 #3 #4
\{
  \int_gincr:N \g__kernel_prg_map_int
  \exp_args:NNc \__int_step:NNnnnn
  \cs_gset_protected:Npn
\}
\end{verbatim}

853
59.8 Formatting integers

For conversion of integers to arbitrary symbols the method is in general as follows. The input number (#1) is compared to the total number of symbols available at each place (#2). If the input is larger than the total number of symbols available then the modulus is needed, with one added so that the positions don’t have to number from zero. Using an f-type expansion, this is done so that the system is recursive. The actual conversion function therefore gets a ‘nice’ number at each stage. Of course, if the initial input was small enough then there is no problem and everything is easy.
\int_to_alph:n\int_to_Alph:n These both use the above function with input functions that make sense for the alphabet in English.

\int_to_alph:n\int_to_Alph:n These both use the above function with input functions that make sense for the alphabet in English.
Converting from base ten (\#1) to a second base (\#2) starts with computing \#1: if it is a complicated calculation, we shouldn’t perform it twice. Then check the sign, store it, either \texttt{=} or \texttt{\c_empty_tl}, and feed the absolute value to the next auxiliary function.

\begin{verbatim}
\int_to_base:nn
\int_to_Base:nn
\__int_to_base:nn
\__int_to_Base:nn
\__int_to_letter:n
\__int_to_Letter:n
\end{verbatim}

Here, the idea is to provide a recursive system to deal with the input. The output is built up after the end of the function. At each pass, the value in \#1 is checked to see if it is less than the new base (\#2). If it is, then it is converted directly, putting the sign back in front. On the other hand, if the value to convert is greater than or equal to the new
base then the modulus and remainder values are found. The modulus is converted to a
symbol and put on the right, and the remainder is carried forward to the next round.

\cs_new:Npn \__int_to_base:nnN #1#2#3
\{\int_compare:nNnTF {#1} < {#2}
\{\exp_last_unbraced:Nf #3 \{ \__int_to_letter:n {#1} \} \}
\{ \exp_args:Nf \__int_to_base:nnnN
\{ \__int_to_letter:n \{ \int_mod:nn {#1} {#2} \} \}
\{#1\}
\{#2\}
\#3\}
\}
\cs_new:Npn \__int_to_base:nnnN #1#2#3#4
\{\exp_args:Nf \__int_to_base:nnN
\{ \int_div_truncate:nn {#2} {#3} \}
\{#3\}
\#4\}
\#1
\}
\cs_new:Npn \__int_to_Base:nnN #1#2#3
\{\int_compare:nNnTF {#1} < {#2}
\{\exp_last_unbraced:Nf #3 \{ \__int_to_Letter:n {#1} \} \}
\{ \exp_args:Nf \__int_to_Base:nnnN
\{ \__int_to_Letter:n \{ \int_mod:nn {#1} {#2} \} \}
\{#1\}
\{#2\}
\#3\}
\}
\cs_new:Npn \__int_to_base:nnnN #1#2#3#4
\{\exp_args:Nf \__int_to_base:nnN
\{ \int_div_truncate:nn {#2} {#3} \}
\{#3\}
\#4\}
\#1
\}
\cs_new:Npn \__int_to_Base:nnnN #1#2#3#4
\{\exp_args:Nf \__int_to_Base:nnN
\{ \int_div_truncate:nn {#2} {#3} \}
\{#3\}
\#4\}
\#1\}
\}
\cs_new:Npn \__int_to_letter:n #1
\{\exp_after:wN \exp_after:wN
\if_case:w \__int_eval:w #1 - 10 \__int_eval_end:\a
\}

Convert to a letter only if necessary, otherwise simply return the value unchanged. It
would be clearer to use \int_case:nn, but in our case, the cases are contiguous, so it
is forty times faster to use the \if_case:w primitive. The first \exp_after:wN expands
the conditional, jumping to the correct case, the second one expands after the resulting
character to close the conditional. Since #1 might be an expression, and not directly a
single digit, we need to evaluate it properly, and expand the trailing \fi:.

\cs_new:Npn \__int_to_letter:n #1
\{\exp_after:wN \exp_after:wN
\if_case:w \__int_eval:w #1 - 10 \__int_eval_end:\a
\}
\int_to_bin:n \int_to_hex:n \int_to_Hex:n \int_to_oct:n

Wrappers around the generic function.

\int_to_roman:n \int_to_Roman:n

The \_\_int_to_roman:w primitive creates tokens of category code 12 (other). Usually, what is actually wanted is letters. The approach here is to convert the output of the primitive into letters using appropriate control sequence names. That keeps everything expandable. The loop is terminated by the conversion of the Q.

\Begin{Verbatim}{
\cs_new:Npn \int_to_roman:n #1
\{ \exp_after:wN \__int_to_roman:N \__int_to_roman:w \int_eval:n {#1} Q \}
\cs_new:Npn \__int_to_roman:N #1
\{ \use:c { __int_to_roman_ #1 :w } \__int_to_roman:N \}
\cs_new:Npn \__int_to_roman_i:w { i }
\cs_new:Npn \__int_to_roman_v:w { v }
\cs_new:Npn \__int_to_roman_x:w { x }
\cs_new:Npn \__int_to_roman_l:w { l }
\cs_new:Npn \__int_to_roman_c:w { c }
\cs_new:Npn \__int_to_roman_d:w { d }
\cs_new:Npn \__int_to_roman_m:w { m }
\cs_new:Npn \__int_to_roman_Q:w #1 { }
\End{Verbatim}
59.9 Converting from other formats to integers

Called as \_\_int_pass_signs:wn \langle signs and digits \rangle \_s\_int_stop \{\langle code\rangle\}, this function leaves in the input stream any sign it finds, then inserts the \langle code\rangle before the first non-sign token (and removes \_s\_int_stop). More precisely, it deletes any + and passes any – to the input stream, hence should be called in an integer expression.

\cs_new:Npn \_\_int_pass_signs:wn #1
\begin{verbatim}
  \if:w + \if:w - \exp_not:N #1 + \fi: \exp_not:N #1
  \exp_after:wN \_\_int_pass_signs:wn
\else:
  \exp_after:wN \_\_int_pass_signs_end:wn
  \exp_after:wN #1
  \fi:
\end{verbatim}

\cs_new:Npn \_\_int_pass_signs_end:wn #1 \s__int_stop #2 { #2 #1 }

(End of definition for \_\_int_pass_signs:wn and \_\_int_pass_signs_end:wn.)

First take care of signs then loop through the input using the recursion quarks. The \_\_int_from_alph:nn auxiliary collects in its first argument the value obtained so far, and the auxiliary \_\_int_from_alph:NN converts one letter to an expression which evaluates to the correct number.

\cs_new:Npn \_\_int_from_alph:nn #1
\begin{verbatim}
  \__int_if_recursion_tail_stop_do:Nn #2 {#1}
  \exp_args:Nf \_\_int_from_alph:nn
  { \int_eval:n { #1 * 26 + \_\_int_from_alph:NN { 0 } } }
\end{verbatim}

\cs_new:Npn \_\_int_from_alph:NN #1#2
\begin{verbatim}
  #1 - \int_compare:nNnTF { #1 } < { 91 } { 64 } { 96 }
\end{verbatim}

(End of definition for \_\_int_from_alph:n, \_\_int_from_alph:nn, and \_\_int_from_alph:NN. This function is documented on page 176.)

860
Leave the signs into the integer expression, then loop through characters, collecting the value found so far in the first argument of \_\_int_from_base:NN. To convert a single character, \_\_int_from_base:N checks first for digits, then distinguishes lower from upper case letters, turning them into the appropriate number. Note that this auxiliary does not use \int_eval:n, hence is not safe for general use.

\begin{code}
\cs_new:Npn \int_from_base:nn #1#2
\{%
  \int_eval:n
  \{\exp_after:wN \__int_pass_signs:wn \tl_to_str:n {#1}\}
  \s__int_stop { \__int_from_base:nnN { 0 } {#2} }\}
\}
\end{code}

\begin{code}
\cs_new:Npn \__int_from_base:nnN #1#2#3
\{
  \__int_if_recursion_tail_stop_do:Nn #3 {#1}
  \exp_args:Nf \__int_from_base:nnN
  { \int_eval:n { #1 * #2 + \__int_from_base:N #3 } }
  {#2}
}\}
\cs_new:Npn \__int_from_base:N #1
\{
  \int_compare:nNnTF { '#1 } < { 58 } {#1}
  { '#1 - \int_compare:nNnTF { '#1 } < { 91 } { 55 } { 87 } }
\}
\end{code}

\begin{code}
\\end{code}

\begin{code}
\cs_new:Npn \\_\_int_from_base:nnN \_\_\_int_from_base:N \_\_\_int_from_base:nn
\end{code}

Wrappers around the generic function.

\begin{code}
\cs_new:Npn \int_from_bin:n #1
\{ \int_from_base:nn {#1} { 2 } \}
\cs_new:Npn \int_from_hex:n #1
\{ \int_from_base:nn {#1} { 16 } \}
\cs_new:Npn \int_from_oct:n #1
\{ \int_from_base:nn {#1} { 8 } \}
\end{code}

Constants used to convert from Roman numerals to integers.

\begin{code}
\\c__int_from_roman_i_int\c__int_from_roman_v_int
\\c__int_from_roman_x_int\c__int_from_roman_c_int
\\c__int_from_roman_d_int\c__int_from_roman_m_int
\\c__int_from_roman_I_int\c__int_from_roman_V_int
\\c__int_from_roman_X_int\c__int_from_roman_L_int
\\c__int_from_roman_C_int\c__int_from_roman_D_int
\\c__int_from_roman_M_int
\end{code}
\int_from_roman:n
\__int_from_roman:NN
\__int_from_roman_error:w

The method here is to iterate through the input, finding the appropriate value for each letter and building up a sum. This is then evaluated by \TeX. If any unknown letter is found, skip to the closing parenthesis and insert *0-1 afterwards, to replace the value by $-1$.

\begin{verbatim}
\cs_new:Npn \int_from_roman:n #1
\{\int_eval:n\{\int_if_exist:cF { c__int_from_roman_ #1 _int } \{ \__int_from_roman_error:w \} \int_if_exist:cF { c__int_from_roman_ #2 _int } \{ \__int_from_roman_error:w \} \int_compare:nNnTF { \use:c { c__int_from_roman_ #1 _int } } {<} { \use:c { c__int_from_roman_ #2 _int } } { + \use:c { c__int_from_roman_ #2 _int } \int_compare:nNnTF { \use:c { c__int_from_roman_ #1 _int } } {<} { \use:c { c__int_from_roman_ #1 _int } \__int_from_roman:NN } } \} \cs_new:Npn \__int_from_roman_error:w #1 \q__int_recursion_stop #2 \{ #2 * 0 - 1 \}
\end{verbatim}

(End of definition for \c__int_from_roman_i_int and others.)

This function is documented on page 177.)

\section*{59.10 Viewing integer}

\int_show:N
\int_show:c
\__int_show:nN

Diagnostics.
\texttt{\textbackslash int\_show:N} We don’t use the \TeX{} primitive \texttt{\textbackslash showthe} to show integer expressions: this gives a more unified output.

\texttt{\textbackslash int\_log:N} Diagnostics.

\texttt{\textbackslash int\_log:c} The zero is defined in \texttt{l3basics}.

\texttt{\textbackslash c\_zero\_int} The zero is defined in \texttt{l3basics}.

\texttt{\textbackslash c\_one\_int} The largest number allowed is $2^{31} - 1$

\texttt{\textbackslash c\_max\_int} The largest number allowed is $2^{31} - 1$

\texttt{\textbackslash c\_max\_char\_int} The largest character code is \texttt{\textbackslash lccode} are restricted to $[0, 255]$. In many places \texttt{m\TeX} and \texttt{u\TeX} support larger character codes but for instance the values of \texttt{\textbackslash lccode} are restricted to $[0, 255]$.
59.13 Scratch integers

We provide two local and two global scratch counters, maybe we need more or less.

\l_tmpa_int \l_tmpb_int \g_tmpa_int \g_tmpb_int

(End of definition for \l_tmpa_int and others. These variables are documented on page 178.)

59.14 Integers for earlier modules

\l__int_internal_a_int \l__int_internal_b_int

(End of definition for \l__int_internal_a_int and \l__int_internal_b_int.)

\l__int_internal_a_int \l__int_internal_b_int

\l__int_internal_a_int \l__int_internal_b_int

(End of definition for \l__int_internal_a_int and \l__int_internal_b_int.)
Chapter 60

l3flag implementation

The following test files are used for this code: m3flag001.

60.1 Protected flag commands

The height \( h \) of a flag (which is initially zero) is stored by setting control sequences of the form \`\( \text{flag name} \langle \text{integer} \rangle \) \` to \`relax\' for \( 0 \leq \langle \text{integer} \rangle < h \). These control sequences are produced by \`\text{cs:w (flag var) (integer) \text{cs_end;}},\` namely the \`(flag var)\' is actually a (protected) macro expanding to its own csname.

\begin{verbatim}
\flag_new:N \flag_new:c
\flag_new:N \l_tmpa_flag \l_tmpb_flag
\flag_clear:N \flag_clear:c \_flag_clear:wN
\end{verbatim}

Evaluate the csname of \#1 for use in constructing the various indexed macros.

\begin{verbatim}
\flag_new:N \l_tmpa_flag \l_tmpb_flag
\_flag_clear:wN
\end{verbatim}

Two flag variables for scratch use.

\begin{verbatim}
\flag_clear:N \flag_clear:c \_flag_clear:wN
\end{verbatim}

Undefine control sequences, starting from the 0 flag, upwards, until reaching an undefined control sequence. We don’t use \`\text{cs:undefine:c}\' because that would act globally.
\cs_set_eq:cN \#2 \#1 \tex_undefined:D
\exp_after:wN \__flag_clear:wN
\int_value:w \int_eval:w \c_one_int + \#1 ; \#2
\}
(End of definition for \flag_clear:N and \__flag_clear:wN. This function is documented on page 182.)
\flag_clear_new:N
\flag_clear_new:c
As for other datatypes, clear the \langle flag var \rangle or create a new one, as appropriate.
\cs_new_protected:Npn \flag_clear_new:N #1
\cs_generate_variant:Nn \flag_clear_new:N { c }
(End of definition for \flag_clear_new:N. This function is documented on page 182.)
\flag_show:N
\flag_show:c
\flag_log:N
\flag_log:c
\__flag_show:NN
Show the height (terminal or log file) using appropriate l3msg auxiliaries.
\cs_new_protected:Npn \flag_show:N { \__flag_show:NN \tl_show:n }
\cs_generate_variant:Nn \flag_show:N { c }
\cs_new_protected:Npn \flag_log:N { \__flag_show:NN \tl_log:n }
\cs_generate_variant:Nn \flag_log:N { c }
\cs_new_protected:Npn \__flag_show:NN #1#2
\__kernel_chk_defined:NT #2
{ \exp_args:Ne #1 { \tl_to_str:n { \#2 height } = \flag_height:N \#2 } }
\__kernel_chk_defined:NT \#2
{ \exp_args:Ne \#1 { \tl_to_str:n { \#2 height } = \flag_height:N \#2 } }
\__kernel_chk_defined:NT \#2
(End of definition for \flag_show:N, \flag_log:N, and \__flag_show:NN. These functions are documented on page 182.)
60.2 Expandable flag commands
\flag_if_exist_p:N
\flag_if_exist_p:c
\flag_if_exist:NTF
\flag_if_exist:cTF
Copies of the cs functions defined in l3basics.
\prg_new_eq_conditional:NNn \flag_if_exist:N \cs_if_exist:N
\prg_new_eq_conditional:NNn \flag_if_exist:c \cs_if_exist:c
\prg_generate_conditional_variant:Nnn \flag_if_exist:N { c } { p , T , F , TF }
(End of definition for \flag_if_exist:NTF. This function is documented on page 182.)
\flag_if_raised_p:N
\flag_if_raised_p:c
\flag_if_raised:NTF
\flag_if_raised:cTF
Test if the flag has a non-zero height, by checking the 0 control sequence.
\prg_new_conditional:Nnn \flag_if_raised:N \#1 \{ \c_p \, \c_T , \c_F , \c_TF \}
{ \if_cs_exist:w \#1 \cs_end:
\prg_return_true:
\else:
\prg_return_false:
\fi:
}
\prg_generate_conditional_variant:Nnn \flag_if_raised:N { c } \{ \c_p , \c_T , \c_F , \c_TF \}
(End of definition for \flag_if_raised:NTF. This function is documented on page 182.)
Extract the value of the flag by going through all of the control sequences starting from 0.

\cs_new:Npn \flag_height:N #1 { \__flag_height_loop:wN 0; #1 }
\cs_new:Npn \__flag_height_loop:wN #1 ; #2
{ \if_cs_exist:w #2 #1 \cs_end: \else: \exp_after:wN \__flag_height_end:wN \fi: \exp_after:wN \__flag_height_loop:wN \int_value:w \int_eval:w \c_one_int + #1 ; #2 }
\cs_new:Npn \__flag_height_end:wN #1 + #2 ; #3 { #2 }
\cs_generate_variant:Nn \flag_height:N { c }

(End of definition for \flag_height:N, \__flag_height_loop:wN, and \__flag_height_end:wN. This function is documented on page 182.)

Change the appropriate control sequence to \relax by expanding a \cs:w ... \cs_end: construction, then pass it to \use_none:n to avoid leaving anything in the input stream.
\cs_new:Npn \flag_raise:N #1 { \exp_after:wN \use_none:n \cs:w #1 \flag_height:N #1 \cs_end: }
\cs_generate_variant:Nn \flag_raise:N { c }

(End of definition for \flag_raise:N. This function is documented on page 182.)

Pass the control sequence with name ⟨flag name⟩0 to \use_none:n. Constructing the control sequence ensures that it changes from being undefined (if it was so) to being \relax.
\cs_new:Npn \flag_ensure_raised:N #1 { \exp_after:wN \use_none:n \cs:w #1 0 \cs_end: }
\cs_generate_variant:Nn \flag_ensure_raised:N { c }

(End of definition for \flag_ensure_raised:N. This function is documented on page 182.)

### 60.3 Old n-type flag commands

Here we keep the old flag commands since our policy is to no longer delete deprecated functions. The idea is to simply map ⟨flag name⟩ to \l_⟨flag name⟩_flag. When the debugging code is activated, it checks existence of the N-type flag variables that result.
\cs_new:Npn \flag_raise:n #1 { \flag_raise:c { l_#1_flag } }
\cs_new:Npn \flag_ensure_raised:n #1 { \flag_ensure_raised:c { l_#1_flag } }

(End of definition for \flag_new:n and others.)

\flag_show:n
\flag_log:n
\__flag_show:Nn

To avoid changing the output here we mostly keep the old code.

\cs_new_protected:Npn \flag_show:n { \__flag_show:Nn \tl_show:n }
\cs_new_protected:Npn \flag_log:n { \__flag_show:Nn \tl_log:n }
\cs_new_protected:Npn \__flag_show:Nn #1#2
{\exp_args:Nc \__kernel_chk_defined:NT { l_#2_flag }
{ \exp_args:Ne #1
{ \tl_to_str:n { flag~#2~height } = \flag_height:n {#2} }
}
}

(End of definition for \flag_show:n, \flag_log:n, and \__flag_show:Nn.)

{/package}
Chapter 61
I3clist implementation

The following test files are used for this code: m3clist002.
\begin{verbatim}
\{+package\}
\{@@=clist\}
\end{verbatim}
\begin{verbatim}
\c_empty_clist
An empty comma list is simply an empty token list.
\end{verbatim}
\begin{verbatim}
\cs_new_eq:NN \c_empty_clist \c_empty_tl
(End of definition for \c_empty_clist. This variable is documented on page 194.)
\end{verbatim}
\begin{verbatim}
\l__clist_internal_clist
Scratch space for various internal uses. This comma list variable cannot be declared as
such because it comes before \clist_new:N
\end{verbatim}
\begin{verbatim}
\tl_new:N \l__clist_internal_clist
(End of definition for \l__clist_internal_clist.)
\end{verbatim}
\begin{verbatim}
\s__clist_mark \s__clist_stop
Internal scan marks.
\end{verbatim}
\begin{verbatim}
\scan_new:N \s__clist_mark
\scan_new:N \s__clist_stop
(End of definition for \s__clist_mark and \s__clist_stop.)
\end{verbatim}
\begin{verbatim}
\__clist_use_none_delimit_by_s_mark:w \__clist_use_none_delimit_by_s_stop:w \__clist_use_i_delimit_by_s_stop:nw
Functions to gobble up to a scan mark.
\begin{verbatim}
\cs_new:Npm \__clist_use_none_delimit_by_s_mark:w \s__clist_mark { }
\cs_new:Npm \__clist_use_none_delimit_by_s_stop:w \s__clist_stop { }
\cs_new:Npm \__clist_use_i_delimit_by_s_stop:nw \s__clist_stop +#1 \s__clist_stop #1
(End of definition for \__clist_use_none_delimit_by_s_mark:w, \__clist_use_none_delimit_by_s_stop:w, and \__clist_use_i_delimit_by_s_stop:nw.)
\end{verbatim}
\begin{verbatim}
\__clist_tmp:w
A temporary function for various purposes.
\end{verbatim}
\begin{verbatim}
\cs_new_protected:Npm \__clist_tmp:w { }
(End of definition for \__clist_tmp:w.)
\end{verbatim}

869
61.1 Removing spaces around items

\__clist_trim_next:w  Called as \exp:w \__clist_trim_next:w \prg_do_nothing:\langle comma list\rangle \ldots it expands to \{\langle trimmed item\rangle\} where the \langle trimmed item\rangle is the first non-empty result from removing spaces from both ends of comma-delimited items in the \langle comma list\rangle. The \prg_do_nothing: marker avoids losing braces. The test for blank items is a somewhat optimized \tl_if_empty:oTF construction; if blank, another item is sought, otherwise trim spaces.

\cs_new:Npn \__clist_trim_next:w #1 ,
{ \tl_if_empty:oTF { \use_none:nn #1 ? } { \__clist_trim_next:w \prg_do_nothing: } { \tl_trim_spaces_apply:oN {#1} \exp_end: } }

(End of definition for \__clist_trim_next:w.)

\__clist_sanitize:n \__clist_sanitize:Nn  The auxiliary \__clist_sanitize:Nn receives a delimiter (\c_empty_tl the first time, afterwards a comma) and that item as arguments. Unless we are done with the loop it calls \__clist_wrap_item:w to unbrace the item (using a comma delimiter is safe since \#2 came from removing spaces from an argument delimited by a comma) and possibly re-brace it if needed.

\cs_new:Npn \__clist_sanitize:n #1
{ \exp_after:wN \__clist_sanitize:Nn \exp_after:wN \c_empty_tl \exp:w \__clist_trim_next:w \prg_do_nothing: #1 , \s__clist_stop \prg_break: , \prg_break_point: }

\cs_new:Npn \__clist_sanitize:Nn #1#2
{ \__clist_use_none_delimit_by_s_stop:w #2 \s__clist_stop #1 \__clist_wrap_item:w #2 , \exp_after:wN \__clist_sanitize:Nn \exp_after:wN , \exp:w \__clist_trim_next:w \prg_do_nothing: #1 }

(End of definition for \__clist_sanitize:n and \__clist_sanitize:Nn.)

\__clist_if_wrap:nTF \__clist_if_wrap:w  True if the argument must be wrapped to avoid getting altered by some clist operations. That is the case whenever the argument

- starts or end with a space or contains a comma,
- is empty, or
- consists of a single braced group.

If the argument starts or ends with a space or contains a comma then one of the three arguments of \__clist_if_wrap:w will have its end delimiter (partly) in one of the three copies of \#1 in \__clist_if_wrap:nTF; this has a knock-on effect meaning that the result of the expansion is not empty; in that case, wrap. Otherwise, the argument is safe unless it starts with a brace group (or is empty) and it is empty or consists of a single \n-type argument.

\prg_new_conditional:Npnn \__clist_if_wrap:n #1 { TF }
\cs_new:Npn \__clist_if_wrap:w #1 \s__clist_mark ? ~ #2 ~ \s__clist_mark #3 , { }

(End of definition for \__clist_if_wrap:nTF and \__clist_if_wrap:w.)

\__clist_wrap_item:w
Safe items are put in \exp_not:n, otherwise we put an extra set of braces.
\cs_new:Npn \__clist_wrap_item:w #1 , { \__clist_if_wrap:nTF {#1} { \exp_not:n { {#1} } } { \exp_not:n {#1} } }

(End of definition for \__clist_wrap_item:w.)

61.2 Allocation and initialisation

\clist_new:N
\clist_new:c
Internally, comma lists are just token lists.
\cs_new_eq:NN \clist_new:N \tl_new:N
\cs_new_eq:NN \clist_new:c \tl_new:c

(End of definition for \clist_new:N. This function is documented on page 185.)

\clist_const:Nn
\clist_const:Nx
\clist_const:cn
\clist_const:ce
\clist_const:cx
Creating and initializing a constant comma list is done by sanitizing all items (stripping spaces and braces).
\cs_new_protected:Npn \clist_const:Nn #1#2
{ \tl_const:Ne #1 { \__clist_sanitize:n {#2} } }
\cs_generate_variant:Nn \clist_const:Nn { Ne , c , ce }
\cs_generate_variant:Nn \clist_const:cn { Nx , cx }

(End of definition for \clist_const:Nn. This function is documented on page 185.)

\clist_clear:N
\clist_clear:c
\clist_gclear:N
\clist_gclear:c
Clearing comma lists is just the same as clearing token lists.
\cs_new_eq:NN \clist_clear:N \tl_clear:N
\cs_new_eq:NN \clist_clear:c \tl_clear:c
\cs_new_eq:NN \clist_gclear:N \tl_gclear:N
\cs_new_eq:NN \clist_gclear:c \tl_gclear:c
\clist_clear_new:N\clist_clear_new:c\clist_gclear_new:N\clist_gclear_new:c

Once again a copy from the token list functions.
\begin{verbatim}
\cs_new_eq:NN \clist_clear_new:N \tl_clear_new:N
\cs_new_eq:NN \clist_clear_new:c \tl_clear_new:c
\cs_new_eq:NN \clist_gclear_new:N \tl_gclear_new:N
\cs_new_eq:NN \clist_gclear_new:c \tl_gclear_new:c
\end{verbatim}

(End of definition for \clist_clear:N and \clist_gclear:N. These functions are documented on page 185.)

\clist_set_eq:NN\clist_set_eq:cN\clist_set_eq:Nc\clist_set_eq:cc\clist_gset_eq:NN\clist_gset_eq:cN\clist_gset_eq:Nc\clist_gset_eq:cc

Once again, these are simple copies from the token list functions.
\begin{verbatim}
\cs_new_eq:NN \clist_set_eq:NN \tl_set_eq:NN
\cs_new_eq:NN \clist_set_eq:Nc \tl_set_eq:Nc
\cs_new_eq:NN \clist_set_eq:cN \tl_set_eq:cN
\cs_new_eq:NN \clist_set_eq:cc \tl_set_eq:cc
\cs_new_eq:NN \clist_gset_eq:NN \tl_gset_eq:NN
\cs_new_eq:NN \clist_gset_eq:Nc \tl_gset_eq:Nc
\cs_new_eq:NN \clist_gset_eq:cN \tl_gset_eq:cN
\cs_new_eq:NN \clist_gset_eq:cc \tl_gset_eq:cc
\end{verbatim}

(End of definition for \clist_set_eq:NN and \clist_gset_eq:NN. These functions are documented on page 185.)

\clist_set_from_seq:NN\clist_set_from_seq:cN\clist_set_from_seq:Nc\clist_set_from_seq:cc\clist_gset_from_seq:NN\clist_gset_from_seq:cN\clist_gset_from_seq:Nc\clist_gset_from_seq:cc

Setting a comma list from a comma-separated list is done using a simple mapping. Safe items are put in \exp_not:n, otherwise we put an extra set of braces. The first comma must be removed, except in the case of an empty comma-list.
\begin{verbatim}
\cs_new_protected:Npn \clist_set_from_seq:NN { \__clist_set_from_seq:NNNN \clist_clear:N \__kernel_tl_set:Nx }
\cs_new_protected:Npn \clist_gset_from_seq:NN { \__clist_set_from_seq:NNNN \clist_gclear:N \__kernel_tl_gset:Nx }
\cs_new_protected:Npn \__clist_set_from_seq:NNNN #1#2#3#4 { \seq_if_empty:NTF #4 \#1 \#3 }
\cs_generate_variant:Nn \clist_set_from_seq:NN { Nc }
\cs_generate_variant:Nn \clist_set_from_seq:NN { cc }
\cs_generate_variant:Nn \clist_gset_from_seq:NN { Nc }
\cs_generate_variant:Nn \clist_gset_from_seq:NN { c , cc }
\end{verbatim}

872
Concatenating comma lists is not quite as easy as it seems, as there needs to be the correct addition of a comma to the output. So a little work to do.

\begin{verbatim}
\cs_new_protected:Npn \clist_concat:NNN { \__clist_concat:NNNN \__kernel_tl_set:Nx }
\cs_new_protected:Npn \clist_gconcat:NNN { \__clist_concat:NNNN \__kernel_tl_gset:Nx }
\__clist_concat:NNNN #1#2#3#4 {
#1 #2 
{ \exp_not:o #3 
\clist_if_empty:NF #3 { \clist_if_empty:NF #4 { , } }
\exp_not:o #4 }
}
\cs_generate_variant:Nn \clist_concat:NNN { ccc }
\cs_generate_variant:Nn \clist_gconcat:NNN { ccc }
\end{verbatim}

(End of definition for \clist_concat:NNN, \clist_gconcat:NNN, and \__clist_concat:NNNN. These functions are documented on page 186.)

\begin{verbatim}
\cs_new_protected:Npn \clist_if_exist:N { \cs_if_exist:N 
{T F , T , F , p} }
\cs_new_protected:Npn \clist_if_exist:c { \cs_if_exist:c 
{T F , T , F , p} }
\end{verbatim}

(End of definition for \clist_if_exist:NTF. This function is documented on page 186.)

\subsection{Adding data to comma lists}

\begin{verbatim}
\cs_new_protected:Npm \clist_set:Nn \clist_set:NV \clist_set:Ne \clist_set:No \clist_set:Nx \clist_set:cn \clist_set:cV \clist_set:ce \clist_set:co \clist_set:cx 
\end{verbatim}

(End of definition for \clist_set:Nn and \clist_set:NV. These functions are documented on page 186.)

\begin{verbatim}
\cs_new_protected:Npm \clist_gset:Nn \clist_gset:NV \clist_gset:Ne \clist_gset:No \clist_gset:Nx \clist_gset:cn \clist_gset:cV \clist_gset:ce \clist_gset:co \clist_gset:cx 
\end{verbatim}

Everything is based on concatenation after storing in \l__clist_internal_clist. This avoids having to worry here about space-trimming and so on.

\begin{verbatim}
\cs_new_protected:Npm \clist_put_left:Nn \clist_put_left:NV \clist_put_left:Nv \clist_put_left:Ne \clist_put_left:No \clist_put_left:Nx \clist_put_left:cn \clist_put_left:cV \clist_put_left:ce \clist_put_left:co \clist_put_left:cx 
\end{verbatim}

(End of definition for \clist_put_left:Nn and \clist_put_left:NV. These functions are documented on page 186.)
61.4 Comma lists as stacks

Getting an item from the left of a comma list is pretty easy: just trim off the first item using the comma. No need to trim spaces as comma-list variables are assumed to have “cleaned-up” items. (Note that grabbing a comma-delimited item removes an outer pair of braces if present, exactly as needed to uncover the underlying item.)
An empty clist leads to \q_no_value, otherwise grab until the first comma and assign to the variable. The second argument of \__clist_pop:wwNNN is a comma list ending in a comma and \s__clist_mark, unless the original clist contained exactly one item: then the argument is just \s__clist_mark. The next auxiliary picks either \exp_not:n or \use_none:n as \#2, ensuring that the result can safely be an empty comma list.

The same, as branching code: very similar to the above.

(End of definition for \clist_pop:NN and others. These functions are documented on page 192.)
Pushing to a comma list is the same as adding on the left.
\cs_new_eq:NN \clist_push:Nn \clist_put_left:Nn
\cs_generate_variant:Nn \clist_push:Nn { NV , No , Nx , c , cV , co , cx }
\cs_new_eq:NN \clist_gpush:Nn \clist_gput_left:Nn
\cs_generate_variant:Nn \clist_gpush:Nn { NV , No , Nx , c , cV , co , cx }

(End of definition for \clist_push:Nn and \clist_gpush:Nn. These functions are documented on page 193.)

61.5 Modifying comma lists

An internal comma list and a sequence for the removal routines.
\l__clist_internal_remove_clist
\l__clist_internal_remove_seq

(End of definition for \l__clist_internal_remove_clist and \l__clist_internal_remove_seq.)

Removing duplicates means making a new list then copying it.
\cs_new_protected:Npn \clist_remove_duplicates:N
\cs_new_protected:Npn \clist_gremove_duplicates:N
{ \__clist_remove_duplicates:NN \clist_set_eq:NN }
\cs_new_protected:Npn \__clist_remove_duplicates:NN #1#2
{ \clist_clear:N \l__clist_internal_remove_clist
\clist_map_inline:Nn #2
{ \clist_if_in:NnF \l__clist_internal_remove_clist {##1}
{ \tl_put_right:Ne \l__clist_internal_remove_clist
{ \clist_if_empty:NF \l__clist_internal_remove_clist { , }
\__clist_if_wrap:nTF {##1} { \exp_not:n { {##1} } } { \exp_not:n {##1} } }
\exp_not:n {##1} }
} }
\cs_generate_variant:Nn \clist_remove_duplicates:N { c }
\cs_generate_variant:Nn \clist_gremove_duplicates:N { c }

(End of definition for \clist_remove_duplicates:N, \clist_gremove_duplicates:N, and \__clist_remove_duplicates:NN. These functions are documented on page 187.)
The method used here for safe items is very similar to \texttt{\tl_replace_all:Nnn}. However, if the item contains commas or leading/trailing spaces, or is empty, or consists of a single brace group, we know that it can only appear within braces so the code would fail; instead just convert to a sequence and do the removal with \texttt{l3seq} code (it involves somewhat elaborate code to do most of the work expandably but the final token list comparisons non-expandably).

For “safe” items, build a function delimited by the \texttt{⟨item⟩} that should be removed, surrounded with commas, and call that function followed by the expanded comma list, and another copy of the \texttt{⟨item⟩}. The loop is controlled by the argument grabbed by \texttt{\_\_clist_remove_all:w}: when the item was found, the \texttt{\_\_clist_mark} delimiter used is the one inserted by \texttt{\_\_clist_tmp:w}, and \texttt{\_\_clist_use_none_delimit_by_s_stop:w} is deleted. At the end, the final \texttt{⟨item⟩} is grabbed, and the argument of \texttt{\_\_clist_tmp:w} contains \texttt{\_\_clist_mark}: in that case, \texttt{\_\_clist_remove_all:w} removes the second \texttt{\_\_clist_mark} (inserted by \texttt{\_\_clist_tmp:w}), and lets \texttt{\_\_clist_use_none_delimit_by_s_stop:w} act.

No brace is lost because items are always grabbed with a leading comma. The result of the first assignment has an extra leading comma, which we remove in a second assignment. Two exceptions: if the clist lost all of its elements, the result is empty, and we shouldn’t remove anything; if the clist started up empty, the first step happens to turn it into a single comma, and the second step removes it.
\begin{Verbatim}{}
\cs_new:Npn \__clist_remove_all: {}
{ \exp_after:wN \__clist_remove_all:w \__clist_tmp:w , }
\cs_new:Npn \__clist_remove_all:w #1 , \s__clist_mark , #2 , { \exp_not:n {#1} }
\cs_generate_variant:Nn \clist_remove_all:Nn { c , NV , cV }
\cs_generate_variant:Nn \clist_gremove_all:Nn { c , NV , cV }
\end{Verbatim}

(End of definition for \clist_remove_all:Nn and others. These functions are documented on page 187.)

\begin{Verbatim}{}
\clist_reverse:N
\clist_reverse:c
\clist_greverse:N
\clist_greverse:c
\end{Verbatim}

Use \clist_reverse:n in an \texttt{e}-expanding assignment. The extra work that \clist_reverse:n does to preserve braces and spaces would not be needed for the well-controlled case of N-type comma lists, but the slow-down is not too bad.

\begin{Verbatim}{}
\cs_new_protected:Npn \clist_reverse:N #1 {}
{ \__kernel_tl_set:Nx #1 { \exp_args:No \clist_reverse:n {#1} } }
\cs_new_protected:Npn \clist_greverse:N #1 {}
{ \__kernel_tl_gset:Nx #1 { \exp_args:No \clist_reverse:n {#1} } }
\cs_generate_variant:Nn \clist_reverse:N { c }
\cs_generate_variant:Nn \clist_greverse:N { c }
\end{Verbatim}

(End of definition for \clist_reverse:N and \clist_greverse:N. These functions are documented on page 187.)

\begin{Verbatim}{}
\clist_reverse:n \__clist_reverse:wwNww \__clist_reverse_end:ww
\end{Verbatim}

The reversed token list is built one item at a time, and stored between \s__clist_stop and \s__clist_mark, in the form of \texttt{?} followed by zero or more instances of \texttt{\langle item\rangle}. We start from a comma list \texttt{\langle item1\rangle,...,\langle item_n\rangle}. During the loop, the auxiliary \_\clist_reverse:wwNww receives \texttt{\langle item_i\rangle} as \#1, \texttt{\langle item_{i+1}\rangle,...,\langle item_n\rangle} as \#2, \_\clist_reverse:wwNww as \#3, what remains until \s__clist_stop as \#4, and \texttt{\langle item_{i-1}\rangle,...,\langle item_1\rangle} as \#5. The auxiliary moves \#1 just before \#5, with a comma, and calls itself (\#3). After the last item is moved, \_\clist_reverse:wwNww receives \texttt{\langle s__clist_mark \_\clist_reverse:wwNww !} as its argument \#1, thus \_\clist_reverse_end:ww as its argument \#3. This second auxiliary cleans up until the marker \texttt{!}, removes the trailing comma (introduced when the first item was moved after \s__clist_stop), and leaves its argument \#1 within \exp_not:n. There is also a need to remove a leading comma, hence \exp_not:o and \use_none:n.

\begin{Verbatim}{}
\cs_new:Npn \clist_reverse:n #1 {}
{ \__clist_reverse:wwNww ? #1 , \s__clist_mark \__clist_reverse:wwNww ! , \s__clist_mark \__clist_reverse_end:ww \s__clist_stop ? \s__clist_mark }
\end{Verbatim}

\begin{Verbatim}{}
\cs_new:Npn \__clist_reverse:wwNww \s__clist_mark \__clist_reverse:wwNww ! \s__clist_mark \__clist_reverse_end:ww \s__clist_stop \s__clist_mark }
\end{Verbatim}

(End of definition for \clist_reverse:n, \_\clist_reverse:wwNww, and \_\clist_reverse_end:ww. This function is documented on page 187.)

\begin{Verbatim}{}
\clist_sort:Nn
\clist_sort:cn
\clist_gsort:Nn
\clist_gsort:cn
\end{Verbatim}

Implemented in \l3sort.

(End of definition for \clist_sort:Nn and \clist_gsort:Nn. These functions are documented on page 188.)
61.6 Comma list conditionals

Simple copies from the token list variable material.

\clist_if_empty_p:N \clist_if_empty_p:c
\clist_if_empty:N \clist_if_empty:c
\clist_if_empty:NTF \clist_if_empty:TF
\clist_if_empty:n \clist_if_empty:n\__clist_if_empty_n:w
\clist_if_empty:n:W
\clist_if_in:Nn \clist_if_in:NV
\clist_if_in:No \clist_if_in:cn
\clist_if_in:cV \clist_if_in:co
\clist_if_in:nn \clist_if_in:nV
\clist_if_in:no
\__clist_if_in_return:nnN

As usual, we insert a token (here ?) before grabbing any argument: this avoids losing braces. The argument of \tl_if_empty:oTF is empty if #1 is ? followed by blank spaces (besides, this particular variant of the emptiness test is optimized). If the item of the comma list is blank, grab the next one. As soon as one item is non-blank, exit: the second auxiliary grabs \prg_return_false: as #2, unless every item in the comma list was blank and the loop actually got broken by the trailing \s__clist_mark \prg_return_false: item.

\prg_new_conditional:Npnn \clist_if_empty:n #1 { p , T , F , TF }
\prg_new_protected_conditional:Npnn \clist_if_in:Nn #1#2 { T , F , TF }
\prg_new_protected_conditional:Npnn \clist_if_in:nn #1#2 { T , F , TF }
\prg_new_protected Conditional:Npnn \__clist_if_in_return:nnN "safe" items, we simply surround the comma list, and the item, with commas, then use the same code as for \tl_if_in:Nn. For “unsafe” items we follow the same route as \seq_if_in:Nn, mapping through the list a comparison function. If found, return true and remove \prg_return_false:.

(End of definition for \clist_if_empty:NTF. This function is documented on page 188.)

(End of definition for \clist_if_empty:nTF, \__clist_if_empty_n:v, and \__clist_if_empty_n:wW. This function is documented on page 188.)

879
61.7 Mapping over comma lists

If the variable is empty, the mapping is skipped (otherwise, that comma-list would be seen as consisting of one empty item). Then loop over the comma-list, grabbing eight comma-delimited items at a time. The end is marked by \s__clist_stop, which may not appear in any of the items. Once the last group of eight items has been reached, we go through them more slowly using \__clist_map_function_end:w. The auxiliary function \__clist_map_function:Nw is also used in some other clist mappings.

```latex
\clist_map_function:NN \clist_map_function:cN \__clist_map_function:Nw \__clist_map_function_end:w
{ \exp_not:N \tl_if_eq:nnT {##1} \exp_not:n
 \exp_not:n
 \{ (##2)
 \{ \clist_map_break:n { \prg_return_true: \use_none:n } \}
 \}
 \clist_map_function:NN #3 \__clist_tmp:w \prg_return_false:
 \}
{ \cs_set:Npn \__clist_tmp:w ##1 ,#2, { } \tl_if_empty:oTF
 { \__clist_tmp:w ,#1, {} {} ,#2, } { \prg_return_false: } { \prg_return_true: }
}
\prg_generate_conditional_variant:Nnn \clist_if_in:Nn { NV , No , c , cV , co } { T , F , TF }
\prg_generate_conditional_variant:Nnn \clist_if_in:nn { nV , no } { T , F , TF }
(End of definition for \clist_if_in:NnTF, \clist_if_in:nnTF, and \__clist_if_in_return:nnN. These functions are documented on page 188.)
```
The n-type mapping function is a bit more awkward, since spaces must be trimmed from each item. Space trimming is again based on \clist_trim_next:w. The auxiliary \clist_map_function_n:Nn receives as arguments the function, and the next non-empty item (after space trimming but before brace removal). One level of braces is removed by \clist_map_unbrace:wn.

\cs_new:Npn \clist_map_function:nN #1#2
\exp_after:wN \__clist_map_function_n:Nn \exp_after:wN #2
\exp:w \__clist_trim_next:w \prg_do_nothing: #1 ,
\s__clist_stop \clist_map_break: ,
\prg_break_point:Nn \clist_map_break: { } \}
\cs_generate_variant:Nn \clist_map_function:nN { e }
\cs_new:Npn \__clist_map_function_n:Nn #1 #2
\__clist_use_none_delimit_by_s_stop:w #2 \s__clist_stop
\__clist_map_unbrace:wn #2 , #1
\exp_after:wN \__clist_map_function_n:Nn \exp_after:wN #1
\exp:w \__clist_trim_next:w \prg_do_nothing: \}
\cs_new:Npn \clist_map_unbrace:wn #1, #2 { #2 {#1} }

Inline mapping is done by creating a suitable function “on the fly”: this is done globally to avoid any issues with \TeX{}’s groups. We use a different function for each level of nesting.

Since the mapping is non-expandable, we can perform the space-trimming needed by the n version simply by storing the comma-list in a variable. We don’t need a different comma-list for each nesting level: the comma-list is expanded before the mapping starts.

\cs_new_protected:Npn \clist_map_inline:Nn #1\#2
\clist_if_empty:NFN #1
\int_gincr:N \g__kernel_prg_map_int 
\cs_gset_protected:cpn
\{ \__clist_map_unbrace:wn \g__kernel_prg_map_int \}
\exp_last_unbraced:Nco \clist_map_function:NN \g__kernel_prg_map_int
\int_gdecr:N \g__kernel_prg_map_int 
\{
\cs_gset_protected:cpn
\{ \__clist_map_unbrace:wn \g__kernel_prg_map_int \}
\}

\clist_map_inline:Nn \clist_map_inline:cn \clist_map_inline:nn

(End of definition for \clist_map_function:nN, \clist_map_function:Nn, and \clist_map_unbrace:wn. This function is documented on page 189.)
The \texttt{N}-type version is a straightforward application of \texttt{clist_map_tokens:Nn}, calling \texttt{clist_map_variable:Nnn} for each item to assign the variable and run the user’s code. The \texttt{N}-type version is \textit{not} implemented in terms of the \texttt{n}-type function \texttt{clist_map_tokens:Nn}, because here we are allowed to clean up the \texttt{n}-type comma list non-expandably.

\begin{verbatim}
\cs_new_protected:Npn \clist_map_variable:NNn #1#2#3
{ \clist_map_tokens:Nn #1 { \__clist_map_variable:Nnn #2 {#3} } }
\cs_generate_variant:Nn \clist_map_variable:NNn { c }
\cs_new_protected:Npn \__clist_map_variable:Nnn #1#2#3
{ \tl_set:Nn #1 {#3} #2 }
\cs_new_protected:Npn \clist_map_variable:nNn #1
{
\clist_set:Nn \l__clist_internal_clist {#1}
\clist_map_variable:NNn \l__clist_internal_clist
}
\end{verbatim}

(End of definition for \texttt{clist_map_variable:NNn}, \texttt{clist_map_variable:nNn}, and \texttt{__clist_map_variable:Nnn}. These functions are documented on page 189.)

\begin{verbatim}
\cs_new:Npn \clist_map_tokens:Nn #1#2
{ \clist_if_empty:NF #1
\exp_last_unbraced:Nno \__clist_map_tokens:nw {#2} #1 ,
\__clist_stop , \__clist_stop , \__clist_stop , \__clist_stop ,
\prg_break_point:Nn \clist_map_break: { } }
\cs_new:Npn \__clist_map_tokens:nw #1 #2, #3, #4, #5, #6, #7, #8, #9,
{ \__clist_use_none_delimit_by_s_stop:w #9 \__clist_map_tokens_end:w \s__clist_stop
\use:n {#1} {#2} \use:n {#1} {#3} \use:n {#1} {#4} \use:n {#1} {#5}
\use:n {#1} {#6} \use:n {#1} {#7} \use:n {#1} {#8} \use:n {#1} {#9}
\__clist_map_tokens:nw {#1} }
\cs_new:Npn \__clist_map_tokens_end:w \s__clist_stop \use:n #1#2
{ \__clist_use_none_delimit_by_s_stop:w \s__clist_stop \use:n \s__clist_stop
\__clist_map_tokens_end:w \s__clist_stop \use:n #1#2 }
\end{verbatim}

Essentially a copy of \texttt{clist_map_function:NN} with braces added.
Similar to \clist_map_function:nN but with a different way of grabbing items because we cannot use \exp_after:wN to pass the ⟨code⟩.

\cs_new:Npn \clist_map_tokens:nn #1#2
{ \__clist_map_tokens_n:nw {#2} \prg_do_nothing: #1, \s__clist_stop \clist_map_break: , \prg_break_point:Nn \clist_map_break: { } }

\cs_new:Npn \__clist_map_tokens_n:nw #1#2,
{ \tl_if_empty:oF { \use_none:nn #2 ? } { \__clist_use_none_delimit_by_s_stop:w #2 \s__clist_stop \tl_trim_spaces_apply:oN {#2} \use_ii_i:nn \__clist_map_unbrace:wn , {#1} } \__clist_map_tokens_n:nw {#1} \prg_do_nothing: }

The break statements use the general \prg_map_break:Nn mechanism.

\cs_new:Npn \clist_map_break:
{ \prg_map_break:Nn \clist_map_break: { } }

\cs_new:Npn \clist_map_break:n
{ \prg_map_break:Nn \clist_map_break: }

Counting the items in a comma list is done using the same approach as for other token count functions: turn each entry into a +1 then use integer evaluation to actually do the mathematics. In the case of an n-type comma-list, we could of course use \clist_map_:function:nN, but that is very slow, because it carefully removes spaces. Instead, we loop manually, and skip blank items (but not \}, hence the extra spaces).

\cs_new:Npn \clist_count:N #1
{ \int_eval:n { \clist_map_function:NN #1 \__clist_count:n } }

\cs_generate_variant:Nn \clist_count:N { c }

\cs_new:Npn \__clist_count:n #1 \__clist_count:n
{ \int_eval:n { \clist_map_function:NN #1 \__clist_count:n } }

\cs_set_protected:Npn \__clist_tmp:w #1
{}
\cs_new:Npn \clist_count:n #1
\int_eval:n
\__clist_count:w #1
\#1, \s__clist_stop \prg_break:, \prg_break_point:
\}
\}
\__clist_use_none_delimit_by_s_stop:w ##1 \s__clist_stop
\tl_if_blank:nF {##1} { + 1 }
\__clist_count:w #1
\}
\}
\exp_args:No \__clist_tmp:w \c_space_tl
\cs_generate_variant:Nn \clist_count:n { e }
\endinput

(End of definition for \clist_count:N and others. These functions are documented on page 190.)

61.8 Using comma lists

First check that the variable exists. Then count the items in the comma list. If it has none, output nothing. If it has one item, output that item, brace stripped (note that space-trimming has already been done when the comma list was assigned). If it has two, place the \langle separator between two\rangle in the middle.

Otherwise, \__clist_use:nwwwwnwn takes the following arguments: 1: a \langle separator\rangle, 2, 3, 4: three items from the comma list (or quarks), 5: the rest of the comma list, 6: a \langle continuation\rangle function (use_ii or use_iii with its \langle separator\rangle argument), 7: junk, and 8: the temporary result, which is built in a brace group following \s__clist_stop. The \langle separator\rangle and the first of the three items are placed in the result, then we use the \langle continuation\rangle, placing the remaining two items after it. When we begin this loop, the three items really belong to the comma list, the first \s__clist_mark is taken as a delimiter to the use_ii function, and the continuation is use_ii itself. When we reach the last two items of the original token list, \s__clist_mark is taken as a third item, and now the second \s__clist_mark serves as a delimiter to use_ii, switching to the other \langle continuation\rangle, use_iii, which uses the \langle separator between final two\rangle.

\exp_args:No \__clist_tmp:w \c_space_tl
\cs_generate_variant:Nn \clist_count:n { e }

\cs_new:Npn \clist_use:Nnnn #1#2#3#4
\clist_if_exist:NTF #1
\int_case:nnF { \clist_count:N #1 }
\{ 0 \} { }\}
\{ 1 \} { \exp_after:wN \__clist_use:wn #1 \#1, \{ \} }
\{ 2 \} { \exp_after:wN \__clist_use:wn #1 \#2\} }
\}
\}
\exp_after:wN \__clist_use:nwwwn
\exp_after:wN \__clist_use:nwwwwwn
\exp_after:wN { \exp_after:wN } #1,
\s__clist_stop, \{ \__clist_use:nwwwwwn \#3 \}

884
Items are grabbed by \clist_use:Nw, which detects blank items with a \tl_if_empty:OTF test (in which case it recurses). Non-blank items are either the end of the list, in which case the argument #1 of \clist_use:Nw is used to properly end the list, or are normal items, which must be trimmed and properly unbraced. As we find successive items, the long list of \clist_use:Nw calls gets shortened and we end up calling \clist_use_more:w once we have found 3 items. This auxiliary leaves the first-found item and the general separator, and calls \clist_use:Nw to find more items. A subtlety is that we use \clist_use_end:w both in the case of a two-item list and for the last two items of a general list: to get the correct separator, \clist_use_more:w replaces the separator-of-two by the last-separator when called, namely as soon as we have found three items.
61.9 Using a single item

To avoid needing to test the end of the list at each step, we first compute the (length) of the list. If the item number is 0, less than −(length), or more than (length), the result is empty. If it is negative, but not less than −(length), add (length) + 1 to the item number before performing the loop. The loop itself is very simple, return the item if the counter reached 1, otherwise, decrease the counter and repeat.

\[\text{\texttt{clist_item:Nn \ __clist_item:nnN \ __clist_item:ffoN \ __clist_item:ffnN \ __clist_item_N_loop:nw}}\]

(End of definition for \texttt{clist_use:nnn} and others. These functions are documented on page 192.)
This starts in the same way as \clist_item:Nn by counting the items of the comma list. The final item should be space-trimmed before being brace-stripped, hence we insert a couple of odd-looking \prg_do_nothing: to avoid losing braces. Blank items are ignored.
61.10 Viewing comma lists

Apply the general \_kernel_chk_tl_type:NnnT with \exp_not:o #2 serving as a dummy code to prevent a check performed by this auxiliary.

```
\cs_new_protected:Npn \clist_show:N { \__clist_show:NN \msg_show:nneeee }
\cs_new_protected:Npn \clist_log:N { \__clist_show:NN \msg_log:nneeee }
\cs_new_protected:Npn \__clist_show:NN #1#2
{ \__kernel_chk_tl_type:NnnT #2 { clist } { \exp_not:o #2 }
  \int_compare:nNnTF { \clist_count:N #2 }
  { \exp_args:No \clist_count:n #2 }
  { \msg_error:nnee { clist } { non-clist } }
  { \token_to_str:N #2 } { \tl_to_str:N #2 }
}
```

(End of definition for \clist_show:N, \clist_log:N, and \__clist_show:NN. These functions are documented on page 193.)

A variant of the above: no existence check, empty first argument for the message.

```
\cs_new_protected:Npm \clist_show:n { \__clist_show:NN \msg_show:nneeee }
\cs_new_protected:Npm \clist_log:n { \__clist_show:NN \msg_log:nneeee }
\cs_new_protected:Npm \__clist_show:Nn #1#2
{ \__kernel_chk_tl_type:NnnT #2 { clist } { \exp_not:o #2 }
  \int_compare:nNnTF { \clist_count:N #2 }
  { \exp_args:No \clist_count:n #2 }
  { \msg_error:nnee { clist } { non-clist } }
  { \token_to_str:N #2 } { \tl_to_str:N #2 }
}
```

(End of definition for \clist_show:n, \clist_log:n, and \__clist_show:Nn. These functions are documented on page 194.)
61.11 Scratch comma lists

\l_tmpa_clist  Temporary comma list variables.
\l_tmpb_clist  \clist_new:N \l_tmpa_clist
\g_tmpa_clist  \clist_new:N \l_tmpb_clist
\g_tmpb_clist  \clist_new:N \g_tmpa_clist
\clist_new:N \g_tmpb_clist

(End of definition for \l_tmpa_clist and others. These variables are documented on page 194.)

\langle/package\rangle
Chapter 62

\l3token implementation

62.1 Internal auxiliaries

\s__char_stop  Internal scan mark.
\scan_new:N \s__char_stop
(End of definition for \s__char_stop.)

\q__char_no_value  Internal recursion quarks.
\quark_new:N \q__char_no_value
(End of definition for \q__char_no_value.)

\__char_quark_if_no_value_p:N \__char_quark_if_no_value:N  Functions to query recursion quarks.
\__kernel_quark_new_conditional:Nn \__char_quark_if_no_value:N { TF }
(End of definition for \__char_quark_if_no_value:N.)

62.2 Manipulating and interrogating character tokens

\char_set_catcode:nn  \char_value_catcode:n  \char_show_value_catcode:n
Simple wrappers around the primitives.
\cs_new_protected:Npn \char_set_catcode:nn #1#2
\cs_new:Npn \char_value_catcode:n #1
\cs_new_protected:Npn \char_show_value_catcode:n #1
(End of definition for \char_set_catcode:nn, \char_value_catcode:n, and \char_show_value_catcode:n.
These functions are documented on page 198.)
(End of definition for \char_set_catcode_escape:N and others. These functions are documented on page 197.)
(End of definition for \char_set_catcode_escape:n and others. These functions are documented on page 198.)

Pretty repetitive, but necessary!

Two sequences for dealing with special characters. The first is characters which may be active, the second longer list is for “special” characters more generally. Both lists are escaped so that for example bulk code assignments can be carried out. In both cases, the order is by ASCII character code (as is done in for example \ExplSyntaxOn).

(End of definition for \char_set_catcode_math_subscript:n and others. These functions are documented on page 199.)

Two sequences for dealing with special characters. The first is characters which may be active, the second longer list is for “special” characters more generally. Both lists are escaped so that for example bulk code assignments can be carried out. In both cases, the order is by ASCII character code (as is done in for example \ExplSyntaxOn).

\seq_new:N \l_char_special_seq
\seq_set_split:Nnn \l_char_special_seq { }
62.3 Creating character tokens

Four simple functions with very similar definitions, so set up using an auxiliary. These are similar to Lua\TeX’s \texttt{\letcharcode} primitive.

\begin{verbatim}
\char_set_active_eq:NN \char_set_active_eq:Nc
\char_gset_active_eq:NN \char_gset_active_eq:Nc
\char_set_active_eq:nN \char_set_active_eq:nc
\char_gset_active_eq:nN \char_gset_active_eq:nc
\group_begin:
  \char_set_catcode_active:N \^^@ \\
  \cs_set_protected:Npn \__char_tmp:nN #1#2 \\
  { \cs_new_protected:cpn { #1 :nN } ##1 \\
    \group_begin:
    \char_set_lccode:nn { \^^@ } { ##1 } \\
    \tex_lowercase:D { \group_end: #2 ^^@ } \\
  } \cs_new_protected:cpe { #1 :NN } ##1 \\
  \exp_not:c { #1 : nN } { '##1 } \\
\__char_tmp:nN \char_set_active_eq:NN \__char_tmp:nN \char_gset_active_eq:NN \\
\group_end:
\cs_generate_variant:Nn \char_set_active_eq:NN { Nc }
\cs_generate_variant:Nn \char_gset_active_eq:NN { Nc }
\cs_generate_variant:Nn \char_set_active_eq:nN { nc }
\cs_generate_variant:Nn \char_gset_active_eq:nN { nc }
\end{verbatim}

(End of definition for \texttt{\char_set_active_eq:NN} and others. These functions are documented on page 196.)

\__char_int_to_roman:w

For efficiency in 8-bit engines, we use the faster primitive approach to making roman numerals.

\begin{verbatim}
\cs_new_eq:NN \__char_int_to_roman:w \tex_romannumeral:D
\end{verbatim}

(End of definition for \texttt{\__char_int_to_roman:w}.)

\char_generate:nn
\__char_generate_aux:nn
\__char_generate_auxx:nnw
\__char_generate_auxii:nnw
\l__char_tmp_tl
\char_generate:invalid:catcode:

The aim here is to generate characters of (broadly) arbitrary category code. Where possible, that is done using engine support (Xet\TeX, Lua\TeX). There are though various issues which are covered below. At the interface layer, turn the two arguments into integers up-front so this is only done once.

\begin{verbatim}
\cs_new:Npn \char_generate:nn #1#2 \\
  { \exp:w \exp_after:wN \__char_generate_aux:w \\
    \int_value:w \int_eval:n {#1} \exp_after:wN ; \\
    \int_value:w \int_eval:n {#2} ; }
\end{verbatim}
Before doing any actual conversion, first some special case filtering. Spaces are out here as LuaTeX emulation only makes normal (charcode 32 spaces). However, ^@ is filtered out separately as that can’t be done with macro emulation either, so is treated separately. That done, hand off to the engine-dependent part.

\cs_new:Npn \__char_generate_aux:w #1 ; #2 ;
\nl
\tl_new:N \l__char_tmp_tl

Engine-dependent definitions are now needed for the implementation. Recent (u)p\TeX{} and the Unicode engines Lua\TeX{} and Xe\TeX{} have engine-level support for expandable character creation. pdf\TeX{} and older (u)p\TeX{} releases do not. The branching here if low-level to avoid fixing the category code of the null character used in the false branch. The final level is the basic definition at the engine level: the arguments here are integers so there is no need to worry about them too much. Older versions of Xe\TeX{} cannot generate active characters so we filter that: at some future stage that may change: the slightly odd ordering of auxiliaries reflects that.

\group_begin:
\char_set_catcode_active:N \^^L
\cs_set:Npn ^^L { }
\if_cs_exist:N \tex_Ucharcat:D
\cs_new:Npn \__char_generate_aux:nnw #1#2#3 \exp_end:
\nl
#3
\exp_after:wN \exp_end:
\tex_Ucharcat:D #1 \exp_stop_f: #2 \exp_stop_f:
\else:
\nl
\exp_end: \l__char_tmp_tl

For engines where \Ucharcat{} isn’t available or emulated, we have to work in macros, and cover only the 8-bit range. The first stage is to build up a \tl{} containing \^^@ with each category code that can be accessed in this way, with an error set up for the other
cases. This is all done such that it can be quickly accessed using a \if_case:w low-level conditional. The list is done in reverse as this puts the case of an active token first: that’s needed to cover the possibility that it is \outer. Getting the braces into the list is done using some standard \if_false: manipulation, while all of the \exp_not:N are required as there is an expansion in the setup.

For making spaces, there needs to be an o-type expansion of a \use:n (or some other tokenization) to avoid dropping the space.

Convert the above temporary list into a series of constant token lists, one for each character code, using \tex_lowercase:D to convert \^@ in each case. The e-type expansion ensures that \text_lowercase:D receives the contents of the token list.

As \TeX{} is very unhappy if it finds an alignment character inside a primitive \halign even when skipping false branches, some precautions are required. \TeX{} is happy if the token is hidden between braces within \if_false: ... \fi:. The rather low-level approach here expands in one step to the \target_token \or: ... then \exp_after:wN (target token) \or: ... expands in one step to \target_token. This means that \exp_not:N is applied to a potentially-problematic active token.
\texttt{\cs_new:Npn \_char_generate_aux:nw \#1\#2\#3 \exp_end:}
\begin{verbatim}
    \#3
    \if_false: { \fi:
    \exp_after:wN \exp_after:wN \exp_after:wN \exp_end:
    \exp_after:wN \exp_after:wN \exp_after:wN \if_case:w \tex_numexpr:D 13 - \#2
    \exp_after:wN \exp_after:wN \exp_after:wN \exp_after:wN
    \exp_after:wN \exp_after:wN \exp_after:wN \exp_after:wN \exp_not:N
    \cs:w c__char_ \_char_int_to_roman:w \#1 _tl \cs_end:
    \fi:
    \fi:
\end{verbatim}
\texttt{\group_end:}

(End of definition for \char_generate:nn and others. This function is documented on page 196.)

\texttt{\c_catcode_active_space_tl}

While \char_generate:nn can produce active characters in some engines it cannot in general. It would be possible to simply change the catcode of space but then the code would need to avoid all spaces, making it quite unreadable. Instead we use the primitive \texttt{\tex_lowercase:D} trick.
\begin{verbatim}
\group_begin:
\char_set_catcode_active:N *
\char_set_lccode:nn { ‘* } { ‘ } \tex_lowercase:D \tl_const:Nn \c_catcode_active_space_tl { * }
\group_end:
\end{verbatim}

(End of definition for \c_catcode_active_space_tl. This variable is documented on page 196.)

\texttt{\c_catcode_other_space_tl}

Create a space with category code 12: an “other” space.
\begin{verbatim}
\tl_const:Ne \c_catcode_other_space_tl \{ \char_generate:nn { ‘ } { 12 } \}
\end{verbatim}

(End of definition for \c_catcode_other_space_tl. This function is documented on page 197.)

\section*{62.4 \label{sec:62.4} \sloppy\bfseries Generic tokens}

\texttt{\_token_mark \_token_stop}

Internal scan marks.
\begin{verbatim}
\scan_new:N \_token_mark
\scan_new:N \_token_stop
\end{verbatim}

(End of definition for \_token_mark and \_token_stop.)

\texttt{\token_to_meaning:N \token_to_meaning:c \token_to_str:N \token_to_str:c}

These are all defined in \texttt{l3basics}, as they are needed “early”. This is just a reminder!

(End of definition for \token_to_meaning:N and \token_to_str:N. These functions are documented on page 201.)
The macro works by comparing the input token with \texttt{\if_catcode:w} with all valid category codes. Since the most common tokens in an average argument list are of category 11 or 12 those are tested first. And since a space and braces are no ordinary N-type arguments, and only control sequences let to those categories can match them they are tested last.

```latex
\cs_new:Npn \token_to_catcode:N
\__token_to_catcode:N
{ \int_value:w \group_align_safe_begin: \__token_to_catcode:N #1 \exp_stop_f: }
\cs_new:Npn \__token_to_catcode:N #1
{\if_catcode:w \exp_not:N #1 \c_catcode_letter_token 11
  \else: \if_catcode:w \exp_not:N #1 \c_catcode_other_token 12
    \else: \if_catcode:w \exp_not:N #1 \c_math_toggle_token 3
      \else: \if_catcode:w \exp_not:N #1 \c_alignment_token 4
        \else: \if_catcode:w \exp_not:N #1 \c_math_superscript_token 7
          \else: \if_catcode:w \exp_not:N #1 \c_math_subscript_token 8
            \else: \if_catcode:w \exp_not:N #1 \c_group_begin_token 1
              \else: \if_catcode:w \exp_not:N #1 \c_group_end_token 2
                \else: \if_catcode:w \exp_not:N #1 \c_space_token 10
                  \else: \token_if_cs:NTF #1 { 16 } { 13 }
            \fi:
          \fi:
        \fi:
      \fi:
    \fi:
  \fi:
}\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\group_align_safe_end:
\exp_stop_f:
```
We define these useful tokens. For the brace and space tokens things have to be done by hand: the formal argument spec. for \cs_new_eq:NN does not cover them so we do things by hand. (As currently coded it would \emph{work} with \cs_new_eq:NN but that’s not really a great idea to show off: we want people to stick to the defined interfaces and that includes us.) So that these few odd names go into the log when appropriate there is a need to hand-apply the \_\_kernel_chk_if_free_cs:N check.

\begin{verbatim}
\group_begin:
\__kernel_chk_if_free_cs:N \c_group_begin_token
\tex_global:D \tex_let:D \c_group_begin_token { \__kernel_chk_if_free_cs:N \c_group_end_token
\tex_global:D \tex_let:D \c_group_end_token }
\char_set_catcode_math_toggle:N \*
\cs_new_eq:NN \c_math_toggle_token *
\char_set_catcode_alignment:N \*
\cs_new_eq:NN \c_alignment_token *
\cs_new_eq:NN \c_parameter_token #
\cs_new_eq:NN \c_math_superscript_token ^
\char_set_catcode_math_subscript:N \*
\cs_new_eq:NN \c_math_subscript_token *
\__kernel_chk_if_free_cs:N \c_space_token
\use:n { \tex_global:D \tex_let:D \c_space_token = ~ }
\cs_new_eq:NN \c_catcode_letter_token a
\cs_new_eq:NN \c_catcode_other_token 1
\group_end:
\end{verbatim}

(End of definition for \c_group_begin_token and others. These functions are documented on page 200.)

\c_catcode_active_tl

Not an implicit token!

\begin{verbatim}
\group_begin:
\char_set_catcode_active:N \*
\tl_const:Nn \c_catcode_active_tl { \exp_not:N * }
\group_end:
\end{verbatim}

(End of definition for \c_catcode_active_tl. This variable is documented on page 200.)

## 62.5 Token conditionals

\token_if_group_begin_p:N \token_if_group_begin:N\text{T}F

Check if token is a begin group token. We use the constant \c_group_begin_token for this.

\begin{verbatim}
\prg_new_conditional:Nppnn \token_if_group_begin:N \#1 { p , T , F , TF }
{ \if_catcode:w \exp_not:N \#1 \c_group_begin_token
 \prg_return_true: \else: \prg_return_false: \fi:
}
\end{verbatim}

(End of definition for \token_if_group_begin:N\text{T}F. This function is documented on page 201.)

\token_if_group_end_p:N \token_if_group_end:N\text{T}F

Check if token is an end group token. We use the constant \c_group_end_token for this.

\begin{verbatim}
\prg_new_conditional:Nppnn \token_if_group_end:N \#1 { p , T , F , TF }
{ \if_catcode:w \exp_not:N \#1 \c_group_end_token
 \prg_return_true: \else: \prg_return_false: \fi:
}
\end{verbatim}

898
Check if token is a math shift token. We use the constant \c_math_toggle_token for this.

\prg_new_conditional:Npnn \token_if_math_toggle:N #1 { p , T , F , TF }
\{ 
  \if_catcode:w \exp_not:N #1 \c_math_toggle_token 
  \prg_return_true: \else: \prg_return_false: \fi:
\}

(End of definition for \token_if_math_toggle:NTF. This function is documented on page 201.)

Check if token is an alignment tab token. We use the constant \c_alignment_token for this.

\prg_new_conditional:Npnn \token_if_alignment:N #1 { p , T , F , TF }
\{ 
  \if_catcode:w \exp_not:N #1 \c_alignment_token 
  \prg_return_true: \else: \prg_return_false: \fi:
\}

(End of definition for \token_if_alignment:NTF. This function is documented on page 202.)

Check if token is a parameter token. We use the constant \c_parameter_token for this. We have to trick \TeX a bit to avoid an error message: within a group we prevent \c_parameter_token from behaving like a macro parameter character. The definitions of \prg_new_conditional:Npnn are global, so they remain after the group.

\group_begin:
\cs_set_eq:NN \c_parameter_token \scan_stop:
\prg_new_conditional:Npnn \token_if_parameter:N #1 { p , T , F , TF }
\{ 
  \if_catcode:w \exp_not:N #1 \c_parameter_token 
  \prg_return_true: \else: \prg_return_false: \fi:
\}
\group_end:

(End of definition for \token_if_parameter:NTF. This function is documented on page 202.)

Check if token is a math superscript token. We use the constant \c_math_superscript_token for this.

\prg_new_conditional:Npnn \token_if_math_superscript:N #1 
\{ p , T , F , TF \}
\{ 
  \if_catcode:w \exp_not:N #1 \c_math_superscript_token 
  \prg_return_true: \else: \prg_return_false: \fi:
\}

(End of definition for \token_if_math_superscript:NTF. This function is documented on page 202.)
Check if token is a math subscript token. We use the constant \texttt{c_math_subscript_token} for this.

```latex
\prg_new_conditional:Nppnn \token_if_math_subscript:N \#1 \{ \texttt{p} , \texttt{T} , \texttt{F} , \texttt{TF} \}
\begin{Verbatim}
\if_catcode:w \exp_not:N \#1 \texttt{c_math_subscript_token}
\prg_return_true: \else: \prg_return_false: \fi:
\end{Verbatim}
```

(End of definition for \texttt{\token_if_math_subscript:NTF}. This function is documented on page 202.)

Check if token is a space token. We use the constant \texttt{c_space_token} for this.

```latex
\prg_new_conditional:Nppnn \token_if_space:N \#1 \{ \texttt{p} , \texttt{T} , \texttt{F} , \texttt{TF} \}
\begin{Verbatim}
\if_catcode:w \exp_not:N \#1 \texttt{c_space_token}
\prg_return_true: \else: \prg_return_false: \fi:
\end{Verbatim}
```

(End of definition for \texttt{\token_if_space:NTF}. This function is documented on page 202.)

Check if token is a letter token. We use the constant \texttt{c_catcode_letter_token} for this.

```latex
\prg_new_conditional:Nppnn \token_if_letter:N \#1 \{ \texttt{p} , \texttt{T} , \texttt{F} , \texttt{TF} \}
\begin{Verbatim}
\if_catcode:w \exp_not:N \#1 \texttt{c_catcode_letter_token}
\prg_return_true: \else: \prg_return_false: \fi:
\end{Verbatim}
```

(End of definition for \texttt{\token_if_letter:NTF}. This function is documented on page 202.)

Check if token is an other char token. We use the constant \texttt{c_catcode_other_token} for this.

```latex
\prg_new_conditional:Nppnn \token_if_other:N \#1 \{ \texttt{p} , \texttt{T} , \texttt{F} , \texttt{TF} \}
\begin{Verbatim}
\if_catcode:w \exp_not:N \#1 \texttt{c_catcode_other_token}
\prg_return_true: \else: \prg_return_false: \fi:
\end{Verbatim}
```

(End of definition for \texttt{\token_if_other:NTF}. This function is documented on page 202.)

Check if token is an active char token. We use the constant \texttt{c_catcode_active_tl} for this. A technical point is that \texttt{c_catcode_active_tl} is in fact a macro expanding to \texttt{\exp_not:N *}, where \texttt{*} is active.

```latex
\prg_new_conditional:Nppnn \token_if_active:N \#1 \{ \texttt{p} , \texttt{T} , \texttt{F} , \texttt{TF} \}
\begin{Verbatim}
\if_catcode:w \exp_not:N \#1 \texttt{c_catcode_active_tl}
\prg_return_true: \else: \prg_return_false: \fi:
\end{Verbatim}
```

(End of definition for \texttt{\token_if_active:NTF}. This function is documented on page 202.)

Check if the tokens \#1 and \#2 have same meaning.

```latex
\prg_new_eq_conditional:NNn \token_if_eq_meaning:NN \cs_if_eq:NN \#1 \{ \texttt{p} , \texttt{T} , \texttt{F} , \texttt{TF} \}
\begin{Verbatim}
\cs_if_eq:NN \#1 \#2 \texttt{\cs_if_eq:NN}
\prg_return_true: \else: \prg_return_false: \fi:
\end{Verbatim}
```

(End of definition for \texttt{\token_if_eq_meaning:NNTF}. This function is documented on page 203.)
Check if the tokens \#1 and \#2 have same category code.
\token_if_eq_catcode_p:NN \token_if_eq_catcode:NN \#1\#2 \{ p , T , F , TF \}
\{ \if_catcode:w \exp_not:N \#1 \exp_not:N \#2 \prg_return_true: \else: \prg_return_false: \fi: \}

(End of definition for \token_if_eq_catcode:NTFF. This function is documented on page 202.)

Check if the tokens \#1 and \#2 have same character code.
\token_if_eq_charcode_p:NN \token_if_eq_charcode:NN \#1\#2 \{ p , T , F , TF \}
\{ \if_charcode:w \exp_not:N \#1 \exp_not:N \#2 \prg_return_true: \else: \prg_return_false: \fi: \}

(End of definition for \token_if_eq_charcode:NTFF. This function is documented on page 203.)

When a token is a macro, \token_to_meaning:N always outputs something like \long macro:\#1->\#1 so we could naively check to see if the meaning contains ->. However, this can fail the five \...mark primitives, whose meaning has the form \...mark:{user material}. The problem is that the {user material} can contain ->.

However, only characters, macros, and marks can contain the colon character. The idea is thus to grab until the first :, and analyse what is left. However, macros can have any combination of \long, \protected or \outer (not used in \LaTeX3) before the string macro:. We thus only select the part of the meaning between the first ma and the first following :. If this string is cro, then we have a macro. If the string is rk, then we have a mark. The string can also be cro parameter character for a colon with a weird category code (namely the usual category code of \#). Otherwise, it is empty.

This relies on the fact that \long, \protected, \outer cannot contain ma, regardless of the escape character, even if the escape character is m...

Both ma and : must be of category code 12 (other), so are detokenized.
\token_if_macro_p:N \token_if_macro:NTF \_token_if_macro_p:w
\use:e
\{ \prg_new_conditional:Nppnn \exp_not:N \token_if_macro:N \#1 \{ p , T , F , TF \}
\{ \exp_not:N \exp_after:wN \exp_not:N \_token_if_macro_p:w \exp_not:N \token_to_meaning:N \#1 \tl_to_str:n \{ ma : \} \s__token_stop \}
\cs_new:Npn \exp_not:N \_token_if_macro_p:w \#1 \tl_to_str:n \{ ma \} \#2 \_c_colon_str \#3 \s__token_stop \}
\{ \str_if_eq:nnTF \{ \#2 \} \{ cro \} \{ \prg_return_true: \} \{ \prg_return_false: \} \}

(End of definition for \token_if_macro:NTF and \_token_if_macro_p:w. This function is documented on page 203.)
\token_if_cs_p:N \token_if_cs:N

Check if token has same catcode as a control sequence. This follows the same pattern as for \token_if_letter:N etc. We use \scan_stop: for this.

\prg_new_conditional:Npnn \token_if_cs:N #1 { p , T , F , TF }
\begin{verbatim}
  \if_catcode:w \exp_not:N #1 \scan_stop:
  \prg_return_true: \else: \prg_return_false: \fi:
\end{verbatim}

(End of definition for \token_if_cs:NTF. This function is documented on page 203.)

\token_if_expandable_p:N \token_if_expandable:N

Check if token is expandable. We use the fact that \TeX temporarily converts \exp_not:N \langle token \rangle into \scan_stop: if \langle token \rangle is expandable. An undefined token is not considered as expandable. No problem nesting the conditionals, since the third #1 is only skipped if it is non-expandable (hence not part of \TeX’s conditional apparatus).

\prg_new_conditional:Npnn \token_if_expandable:N #1 { p , T , F , TF }
\begin{verbatim}
  \exp_after:wN \if_meaning:w \exp_not:N #1 #1
  \prg_return_false:
  \else:
  \if_cs_exist:N #1
  \prg_return_true:
  \else:
  \prg_return_false:
  \fi:
  \fi:
\end{verbatim}

(End of definition for \token_if_expandable:NTF. This function is documented on page 203.)

\__token_delimit_by_char":w \__token_delimit_by_count:w \__token_delimit_by_dimen:w \__token_delimit_by_font:w \__token_delimit_by_macro:w \__token_delimit_by_muskip:w \__token_delimit_by_skip:w \__token_delimit_by_toks:w

These auxiliary functions are used below to define some conditionals which detect whether the \meaning of their argument begins with a particular string. Each auxiliary takes an argument delimited by a string, a second one delimited by \s__token_stop, and returns the first one and its delimiter. This result is eventually compared to another string. Note that the “font” auxiliary is delimited by a space followed by “font”. This avoids an unnecessary check for the \font primitive below.

\group_begin:
\cs_set_protected:Npn \__token_tmp:w #1
\begin{verbatim}
  \use:e
  \cs_new:Npn \exp_not:c { \__token_delimit_by_ #1 :w }
    ##1 \tl_to_str:n {#1} \s__token_stop
    { ##1 \tl_to_str:n {#1} }
\end{verbatim}
\__token_tmp:w { char" }
\__token_tmp:w { count }
\__token_tmp:w { dimen }
\__token_tmp:w { - font }
\__token_tmp:w { macro }
\__token_tmp:w { muskip }
\__token_tmp:w { skip }
\__token_tmp:w { toks }
\group_end:
Each of these conditionals tests whether its argument’s `meaning` starts with a given string. This is essentially done by having an auxiliary grab an argument delimited by the string and testing whether the argument was empty. Of course, a copy of this string must first be added to the end of the `meaning` to avoid a runaway argument in case it does not contain the string. Two complications arise. First, the escape character is not fixed, and cannot be included in the delimiter of the auxiliary function (this function cannot be defined on the fly because tests must remain expandable): instead the first argument of the auxiliary (plus the delimiter to avoid complications with trailing spaces) is compared using `str_if_eq:eeTF` to the result of applying `token_to_str:N` to a control sequence.

Second, the `meaning` of primitives such as `dimen` or `dimendef` starts in the same way as registers such as `dimen123`, so they must be tested for.

Characters used as delimiters must have catcode 12 and are obtained through `tl_to_str:n`. This requires doing all definitions within e-expansion. The temporary function `_token_tmp:w` used to define each conditional receives three arguments: the name of the conditional, the auxiliary’s delimiter (also used to name the auxiliary), and the string to which one compares the auxiliary’s result. Note that the `meaning` of a protected long macro starts with `protected\long macro`, with no space after `protected` but a space after `long`, hence the mixture of `token_to_str:N` and `tl_to_str:n`.

For the first six conditionals, `cs_if_exist:cT` turns out to be `false` (thanks to the leading space for `font`), and the code boils down to a string comparison between the result of the auxiliary on the `meaning` of the conditional’s argument, and `#3`. Both are evaluated at run-time, as this is important to get the correct escape character.

The other five conditionals have additional code that compares the argument to two \TeX primitives which would wrongly be recognized as registers otherwise. Despite using \TeX’s primitive conditional construction, this does not break when `###1` is itself a conditional, because branches of the conditionals are only skipped if `###1` is one of the two primitives that are tested for (which are not \TeXX conditionals).

19490 \group_begin:
19491 \cs_set_protected:Npn \__token_tmp:w #1#2#3
19492 { \use:e
19493 \prg_new_conditional:Npnn \exp_not:c { token_if_ #1 :N } ##1
19494 { p , T , F , TF }
19495 \cs_if_exist:cT { tex_ #2 :D }
19496 { \exp_not:N \if_meaning:w ##1 \exp_not:c { tex_ #2 :D }
19497 \exp_not:N \prg_return_false:
19498 \exp_not:N \else:
19499 \exp_not:N \if_meaning:w ##1 \exp_not:c { tex_ #2 def:D }
19500 \exp_not:N \prg_return_false:
19501 \exp_not:N \else:
19502 }
19503 \exp_not:N \str_if_eq:eeTF
19504 { \exp_not:N \exp_after:wN
19505 \exp_not:c { __token_delimit_by_ #2 :w }
19506 \exp_not:N \token_to_meaning:N #1
19507 \exp_not:N \tl_to_str:n #2 \s__token_stop
19508 903
We filter out macros first, because they cause endless trouble later otherwise.

Primitives are almost distinguished by the fact that the result of \token_to_-meaning:N is formed from letters only. Every other token has either a space (e.g., the letter A), a digit (e.g., \count123) or a double quote (e.g., \char“A”).

Ten exceptions: on the one hand, \tex_undefined:D is not a primitive, but its meaning is undefined, only letters; on the other hand, \space, \italiccorr, \hyphen, \firstmark, \topmark, \botmark, \splitfirstmark, \splitbotmark, and \nullfont are primitives, but have non-letters in their meaning.

We start by removing the two first (non-space) characters from the meaning. This removes the escape character (which may be nonexistent depending on \endlinechar), and takes care of three of the exceptions: \space, \italiccorr and \hyphen, whose meaning is at most two characters. This leaves a string terminated by some :; and \s__token_stop.

The meaning of each one of the five \...mark primitives has the form ⟨letters⟩:(user material). In other words, the first non-letter is a colon. We remove everything after the first colon.

We are now left with a string, which we must analyze. For primitives, it contains only letters. For non-primitives, it contains either ”, or a space, or a digit. Two exceptions remain: \tex_undefined:D, which is not a primitive, and \nullfont, which is a primitive.

Spaces cannot be grabbed in an undelimited way, so we check them separately. If there is a space, we test for \nullfont. Otherwise, we go through characters one by one, and stop at the first character less than ‘A (this is not quite a test for “only letters”,}
but is close enough to work in this context). If this first character is : then we have a
primitive, or \text{\texttt{\ttfamily{\texttt{D}}} and if it is " or a digit, then the token is not a primitive.

For \LaTeX{} we use a different implementation which just looks at the command
code for the token and compares it to a list of non-primitives. Again, \texttt{\texttt{\ttfamily{\ttfamily{nullfont}}} is a
special case because it is the only primitive with the normally non-primitive \texttt{\texttt{\ttfamily{\ttfamily{set_font}}} command code.

In \MetaTeX{} some of the command names are different, so we check for both
versions. The first one is always the \LaTeX{} version.

\begin{verbatim}
\sys_if_engine_luatex:TF
{\langle /tex \rangle \langle */lua \rangle \*lua}
\sys_if_engine_luatex:TF
local get_next = token.get_next
local get_command = token.get_command
local get_index = token.get_index
local get_mode = token.get_mode or token.get_index
local cmd = command_id
local set_font = cmd'get_font'
local biggest_char = token.biggest_char and token.biggest_char()
or status.getconstants().max_character_code
local mode_below_biggest_char = {}
local index_not_nil = {}
local mode_not_null = {}
local non_primitive = {
    [cmd'left_brace'] = true,
    [cmd'right_brace'] = true,
    [cmd'math_shift'] = true,
    [cmd'mac_param' or cmd'parameter'] = mode_below_biggest_char,
    [cmd'sup_mark' or cmd'superscript'] = true,
    [cmd'sub_mark' or cmd'subscript'] = true,
    [cmd'endv' or cmd'ignore'] = true,
    [cmd'spacer'] = true,
    [cmd'letter'] = true,
    [cmd'other_char'] = true,
    [cmd'tab_mark' or cmd'alignment_tab'] = mode_below_biggest_char,
    [cmd'char_given'] = true,
    [cmd'math_given' or 'math_char_given'] = true,
    [cmd'xmath_given' or 'math_char_xgiven'] = true,
    [cmd'set_font'] = mode_not_null,
    [cmd'undefined_cs'] = true,
    [cmd'call'] = true,
    [cmd'long_call' or cmd'protected_call'] = true,
    [cmd'outer_call' or cmd'tolerant_call'] = true,
    [cmd'long_outer_call' or cmd'tolerant_protected_call'] = true,
    [cmd'assign_glue' or cmd'register_glue'] = index_not_nil,
    [cmd'assign_mu_glue' or cmd'register_mu_glue' or cmd'register_muglue'] = index_not_nil,
    [cmd'assign_toks' or cmd'register_toks'] = index_not_nil,
    [cmd'assign_int' or cmd'register_int' or cmd'register_integer'] = index_not_nil,
    [cmd'assign_attr' or cmd'register_attribute'] = true,
    [cmd'assign_dimen' or cmd'register_dimen' or cmd'register_dimension'] = index_not_nil,
}\end{verbatim}
luacmd("__token_if_primitive_lua:N", function()
  local tok = get_next()
  local is_non_primitive = non_primitive[get_command(tok)]
  return put_next(
    is_non_primitive == true
      and false_tok
    or is_non_primitive == nil
      and true_tok
    or is_non_primitive == mode_not_null
      and (get_mode(tok) == 0 and true_tok or false_tok)
    or is_non_primitive == index_not_nil
      and (get_index(tok) and false_tok or true_tok)
    or is_non_primitive == mode_below_biggest_char
      and (get_mode(tok) > biggest_char and true_tok or false_tok)
  end, "global")
end

\prg_new_conditional:Npnn \token_if_primitive:N #1 { p , T , F , TF }
  { \__token_if_primitive_lua:N #1 }
\end

\prg_new_conditional:Npnn \exp_not:N \token_if_primitive:N #1 { p , T , F , TF }
  { \exp_not:N \__token_if_primitive:NNw #1 #2 #3 \c_colon_str #4 \s__token_stop }
\end

\cs_new:Npn \__token_if_primitive_space:w #1 ~ { }
\cs_new:Npn \__token_if_primitive_nullfont:N #1 { }
The aim here is to allow the case statement to be evaluated using a known number of expansion steps (two), and without needing to use an explicit “end of recursion” marker. That is achieved by using the test input as the final case, as this is always true. The trick is then to tidy up the output such that the appropriate case code plus either the \texttt{true} or \texttt{false} branch code is inserted.

(End of definition for \texttt{token_if_primitive:NTF} and others. This function is documented on page 204.)
To tidy up the recursion, there are two outcomes. If there was a hit to one of the cases searched for, then #1 is the code to insert, #2 is the next case to check on and #3 is all of the rest of the cases code. That means that #4 is the true branch code, and #5 tidies up the spare \s__token_mark and the false branch. On the other hand, if none of the cases matched then we arrive here using the “termination” case of comparing the search with itself. That means that #1 is empty, #2 is the first \s__token_mark and so #4 is the false code (the true code is mopped up by #3).

(End of definition for \token_case_catcode:NnTF and others. These functions are documented on page 205.)

62.6 Peeking ahead at the next token

Peeking ahead is implemented using a two part mechanism. The outer level provides a defined interface to the lower level material. This allows a large amount of code to be shared. There are four cases:

1. peek at the next token;
2. peek at the next non-space token;
3. peek at the next token and remove it;
4. peek at the next non-space token and remove it.
\l_peek_token  Storage tokens which are publicly documented: the token peeked.
\g_peek_token
\l_peek_search_token  The token to search for as an implicit token: cf. \l_peek_search_tl.
\l_peek_search_tl  The token to search for as an explicit token: cf. \l_peek_search_token.
\__peek_true:w  Functions used by the branching and space-stripping code.
\__peek_true_aux:w
\__peek_false:w
\__peek_tmp:w
\s__peek_mark
\s__peek_stop  Internal scan marks.
\__peek_use_none_delimit_by_s_stop:w  Functions to gobble up to a scan mark.
\peek_after:Nw  Simple wrappers for \futurelet: no arguments absorbed here.
\peek_gafter:Nw
\__peek_true_remove:w  A function to remove the next token and then regain control.
Repeatedly use \_\_\_peek_true_remove:w to remove a space and call \_\_\_peek_true_aux:w.

\cs_new_protected:Npn \peek_remove_spaces:n #1
\cs_set:Npe \_\_\_peek_false:w { \exp_not:n {#1} }
\group_align_safe_begin:
\cs_set:Npn \_\_\_peek_true_aux:w { \peek_after:Nw \_\_\_peek_remove_spaces: }
\_\_\_peek_true_aux:w
\group_align_safe_end:
\exp_after:wN \_\_\_peek_false:w
\fi:
\}

(End of definition for \peek_remove_spaces:n and \_\_\_peek_remove_spaces:. This function is documented on page 206.)

Here we expand the input, removing spaces and \scan_stop: tokens until we reach a non-expandable token. At that stage we re-insert the payload. To deal with the problem of \& tokens, we have to put the align-safe group in the correct place.

\cs_new_protected:Npn \peek_remove_filler:n #1
\cs_set:Npn \_\_\_peek_true_aux:w { \_\_\_peek_remove_filler:w }
\cs_set:Npe \_\_\_peek_false:w
\group_align_safe_end:
\exp_after:wN \_\_\_peek_false:w
\else:
\\if_meaning:w \l_peek_token \c_space_token
\exp_after:wN \_\_\_peek_true_remove:w
\else:
\\exp_after:wN \exp_after:wN \exp_after:wN
\_\_\_peek_true_remove:w
\else:
\exp_after:wN \exp_after:wN \exp_after:wN
\fi:
\}

Here we can nest conditionals as \l_peek_token is only skipped over in the nested one if it's a space: no problems with conditionals or outer tokens.
To deal with undefined control sequences in the same way \TeX{} does, we need to check for expansion manually.

\begin{verbatim}
\cs_new_protected:Npn \__peek_remove_filler_expand:w
\{
|\exp_after:wN \if_meaning:w \exp_not:N \l_peek_token \l_peek_token
|\exp_after:wN \__peek_false:w
|\else:
|\exp_after:wN \__peek_remove_filler:w
|\fi:
\}
\end{verbatim}

(End of definition for \peek_remove_filler:n and others. This function is documented on page 207.)

The generic functions store the test token in both implicit and explicit modes, and the \texttt{true} and \texttt{false} code as token lists, more or less. The two branches have to be absorbed here as the input stream needs to be cleared for the peek function itself. Here, \texttt{#1} is \__peek_true_remove:w when removing the token and \__peek_true_aux:w otherwise.

\begin{verbatim}
\cs_new_protected:Npn \__peek_token_generic_aux:NNNTF #1#2#3#4#5
\{
|\group_align_safe_begin:
|\cs_set_eq:NN \l__peek_search_token #3
|\tl_set:Nn \l__peek_search_tl {#3}
|\cs_set:Npe \__peek_true_aux:w
\{
|\exp_not:N \group_align_safe_end:
|\exp_not:n {#4}
\}
|\cs_set_eq:NN \__peek_true:w #1
|\cs_set:Npe \__peek_false:w
\{
|\exp_not:N \group_align_safe_end:
|\exp_not:n {#5}
\}
|\peek_after:Nw #2
\}
\end{verbatim}

(End of definition for \__peek_token_general_aux:NNNTF.)

For token removal there needs to be a call to the auxiliary function which does the work.

\begin{verbatim}
\__peek_token_remove_generic:NNTF
\{ \__peek_token_remove_generic:NNTF \}
\end{verbatim}
\_peek_execute_branches_meaning: The meaning test is straightforward.

\_peek_execute_branches_catcode: \_peek_execute_branches_charcode: \_peek_execute_branches_catcode_aux: \_peek_execute_branches_catcode_auxii:N \_peek_execute_branches_catcode_auxiii: The catcode and charcode tests are very similar, and in order to use the same auxiliaries we do something a little bit odd, firing \texttt{\textbackslash if\_catcode:w} and \texttt{\textbackslash if\_charcode:w} before finding the operands for those tests, which are only given in the auxii:N and auxiii: auxiliaries. For our purposes, three kinds of tokens may follow the peeking function:

- control sequences which are not equal to a non-active character token (\textit{e.g.}, macro, primitive);
- active characters which are not equal to a non-active character token (\textit{e.g.}, macro, primitive);
- explicit non-active character tokens, or control sequences or active characters set equal to a non-active character token.

The first two cases are not distinguishable simply using \TeX{}’s \texttt{\textbackslash futurelet}, because we can only access the \texttt{\textbackslash meaning} of tokens in that way. In those cases, detected thanks to a comparison with \texttt{\textbackslash scan\_stop:}, we grab the following token, and compare it explicitly with the explicit search token stored in \texttt{\_peek\_search\_token}. The \texttt{\textbackslash exp\_not:N} prevents outer macros (coming from non-\texttt{\LaTeX} code) from blowing up. In the third case, \texttt{\_peek\_token} is good enough for the test, and we compare it again with the explicit search token. Just like the peek token, the search token may be of any of the three types above, hence the need to use the explicit token that was given to the peek function.

\_peek_execute_branches_catcode: \_peek_execute_branches_charcode: \_peek_execute_branches_catcode_aux: \_peek_execute_branches_catcode_auxii:N \_peek_execute_branches_catcode_auxiii:
\peek_catcode:NNTF
\peek_catcode_remove:NNTF
\peek_charcode:NNTF
\peek_charcode_remove:NNTF
\peek_meaning:NNTF
\peek_meaning_remove:NNTF

The public functions themselves cannot be defined using \prg_new_conditional:Nnn. Instead, the TF, T, F variants are defined in terms of corresponding variants of \__peek_token_generic:NNTF or \__peek_token_remove_generic:NNTF, with first argument one of \__peek_execute_branches_catcode:, \__peek_execute_branches_charcode:, or \__peek_execute_branches_meaning:.

\peek_N_type:TF
\__peek_execute_branches_N_type: \__peek_N_type: \__peek_N_type_aux:nnw

All tokens are \texttt{N} type tokens, except in four cases: \begin{itemize}
    \item begin-group tokens, end-group tokens, space tokens with character code \texttt{32},
    \item outer tokens. Since \texttt{\_peek_token} might be outer, we cannot use the convenient \bool_if:NTF function, and must resort to the old trick of using \texttt{ifodd} to expand a set of tests. The \texttt{false} branch of this test is taken if the token is one of the first three kinds of non-\texttt{N} type tokens (explicit or implicit), thus we call \__peek_false:w. In the \texttt{true} branch, we must detect outer tokens, without impacting performance too much for non-outer tokens. The first filter is to search for \texttt{outer} in the \meaning of \texttt{\_peek_token}. If that is absent, \__peek_use_none_delimit_by_s_stop:w cleans up, and we call \__peek_true:w. Otherwise, the token can be a non-outer macro or a primitive mark whose parameter or replacement text contains \texttt{outer}, it can be the primitive \texttt{\_outer}, or it can be an outer token. Macros and marks would have
ma in the part before the first occurrence of outer; the meaning of \outer has nothing after outer, contrarily to outer macros; and that covers all cases, calling \_\_peek_true:w or \_\_peek_false:w as appropriate. Here, there is no \{search token\}, so we feed a dummy \scan_stop: to the \_\_peek_token_generic:NNTF function.

\begin{verbatim}
\group_begin:
\cs_set_protected:Npn \_\_peek_tmp:w #1 \s__peek_stop
{ \cs_new_protected:Npn \_\_peek_execute_branches_N_type:
{ \if_int_odd:w \if_catcode:w \exp_not:N \l_peek_token { \c_zero_int \fi:
\if_catcode:w \exp_not:N \l_peek_token } \c_zero_int \fi:
\if_meaning:w \l_peek_token \c_space_token \c_zero_int \fi:
\c_one_int \exp_after:wN \_\_peek_N_type:w \token_to_meaning:N \l_peek_token \s__peek_mark \_\_peek_N_type_aux:nnw
\group_end:
\cs_new_protected:Npn \_\_peek_N_type:w \tl_to_str:n \outer \s__peek_stop
{ \_\_peek_N_type:w \tl_if_in:noTF {#1} { \tl_to_str:n {ma} } { \_\_peek_true:w } { \tl_if_empty:nTF {#2} { \_\_peek_true:w } { \_\_peek_false:w } }
\cs_new_protected:Npn \_\_peek_N_type:T
{ \_\_peek_token_generic:NNT \_\_peek_execute_branches_N_type: \scan_stop:
} \cs_new_protected:Npn \_\_peek_N_type:F
{ \_\_peek_token_generic:NNF \_\_peek_execute_branches_N_type: \scan_stop:
} \end{verbatim}

(End of definition for \_\_peek_N_type:TF and others. This function is documented on page 207.)
Chapter 63

\texttt{l3prop} implementation

The following test files are used for this code: \texttt{m3prop001}, \texttt{m3prop002}, \texttt{m3prop003}, \texttt{m3prop004}, \texttt{m3show001}.

With the (default) flat data storage, a property list is a macro whose top-level expansion is of the form

\[
\s__prop \__prop_chk:w \__prop_pair:wn \langle \text{key}_1 \rangle \s__prop \{\langle \text{value}_1 \rangle\}
\]

\[
\ldots
\]

\[
\__prop_pair:wn \langle \text{key}_n \rangle \s__prop \{\langle \text{value}_n \rangle\}
\]

where \texttt{\_\_prop} is a scan mark (equal to \texttt{\scan_stop:}), \texttt{\_\_prop_chk:w} produces a suitable error if the property list is used directly in the input stream, and \texttt{\_\_prop\_pair:wn} can be used to map through the property list.

With the linked data storage, each property list entry \texttt{\langle \text{key}_i \rangle-\langle \text{value}_i \rangle} is stored into a token list \texttt{\_\_prop \langle \text{prefix} \rangle \langle \text{key}_i \rangle}. The \langle \text{prefix} \rangle is one or more characters (no spaces), constructed automatically only once, when the property list is initially declared. The control sequence name does not conform to standard naming for variables because (1) this is an internal control sequence, not really a expl3 variable; (2) keeping track of the scope \texttt{l} or \texttt{g} throughout all functions would be a pretty big mess, especially if users accidentally mix local and global use (we would have to always check for such mistakes, rather than only checking when suitable debug options are set); (3) shorter control sequence names use less memory and are quicker in case of hash collisions, which may matter since we are using many control sequences.

We need to enable mapping through such a property list, but without storing a list of all entries anywhere: this is achieved by making each of these token lists also store a pointer to the next entry. To enable efficient deletion, the token lists also store a pointer to the previous entry. This means we have a doubly-linked list. To avoid having to special-case the two ends of the doubly-linked list when deleting entries, we include as a zeroth entry in the doubly-linked list the property list variable itself, and we include as an \texttt{(n + 1)-th} entry in the doubly-linked list an end-pointer \texttt{\_\_prop \langle \text{prefix} \rangle} (no trailing space, so it differs from an empty key). The space before \langle \text{prefix} \rangle ensures there is no collision with other \texttt{l3prop} internal functions, even if we have very many linked property lists being defined.
The property list variable itself is a token list of the form
\_\_prop\_flatten:w \_\_prop \langle prefix \rangle \s\_\_prop \{ \langle prefix \rangle \} \_\_prop \langle prefix \rangle \langle key \rangle
Here, \_\_prop\_flatten:w serves as an efficiently recognized marker, and when \textit{f}-expanded it is tasked with fully unpacking the property list into the same form as the default data storage so as to ease conversion. The \langle prefix \rangle is used when looking up an entry. The token list \_\_prop \langle prefix \rangle (see below) contains a pointer to the last key to help insert a new entry. The pointer to \langle key \rangle is needed to start a mapping. The token list labeled by \langle key \rangle is of the form
\use\_none:n \_\_prop \langle prefix \rangle \langle key \rangle \_\_prop \langle prefix \rangle \langle key \rangle \_\_prop \langle prefix \rangle \langle key \rangle \_\_prop \langle prefix \rangle \langle key \rangle \_\_prop \langle prefix \rangle \langle key \rangle \_\_prop \langle prefix \rangle \langle key \rangle
where the pointer to \langle key \rangle is needed when deleting the \langle key \rangle. Expanding this will run \_\_prop\_pair:wn on the \langle key \rangle–\langle value \rangle pair (for speed, \langle key \rangle is kept as explicit tokens rather than slowly extracting it from a control sequence name), then move on to the next key, thus mapping through the whole list. The mapping is ended upon expanding \_\_prop \langle prefix \rangle, which is the token list
\use\_none:n \_\_prop \langle prefix \rangle \langle key \rangle
Let us think about deleting the \langle key \rangle. We need to update the \langle key_{i-1} \rangle and \langle key_{i+1} \rangle to point to each other instead of \langle key_{i} \rangle. To edit the corresponding token lists, it is important that \_\_prop \langle prefix \rangle \langle key \rangle be at the “same place” in the token lists also in the boundary cases \textit{i} = 1 \textit{or} \textit{i} = \textit{n}, namely as the second token, or as the second argument after \s\_\_prop.

\section{Internal auxiliaries}
\_\_prop\_tmp:w
Scratch macro, defined as needed, for instance to save \_\_prop\_pair:wn when concatenating.
\_\_prop\_internal_tl
Token list used in various places: for the prefix; when converting from flat to linked props; and to store the new key–value pair inserted by \texttt{\prop\_put:Nnn}.
\_\_prop\_mark
Internal scan marks.
\_\_prop\_stop
Internal recursion quarks.
\_\_prop\_recursion\_tail
\_\_prop\_recursion\_stop
Internal recursion quarks.

916
Functions to query recursion quarks.

19911 \_tell_quark_new_test:N \_prop_if_recursion_tail_stop:n
19912 \cs_generate_variant:Nn \_prop_if_recursion_tail_stop:n { o }

(End of definition for \_prop_if_recursion_tail_stop:n and \_prop_if_recursion_tail_stop:o.)

63.2 Structure of a property list

\s__prop

A private scan mark is used as a marker after each key, and at the very beginning of the property list.

19913 \scan_new:N \s__prop

(End of definition for \s__prop.)

\__prop_chk:w
\__prop_chk_loop:nw
\__prop_chk_get:nw

This removes the flat property list from the input stream and complains about a bad use of a property list. Since property lists do not have an end-marker, we slowly peek ahead in a loop. Speed does not matter since this is for an error situation. While \__prop_pair:wn does not keep a fixed definition, it always includes the internal \s__prop in its argument specification, so that there is no risk of accidentally picking up a public token instead of \__prop_pair:wn when doing a meaning test. We collect the keys and values to produce a more useful error message.

19913 \cs_new_protected:Npn \__prop_chk:w { \__prop_chk_loop:nw { } }
19914 \cs_new_protected:Npn \__prop_chk_loop:nw #1
19915 {\peek_meaning:NTF \__prop_pair:wn
19916 { \__prop_chk_get:nw {#1} }
19917 { \msg_error:nne { prop } { misused } {#1} }
19920 }
19921 \cs_new_protected:Npn \__prop_pair:wn #1 \s__prop #2 { }

(End of definition for \__prop_chk:w, \__prop_chk_loop:nw, and \__prop_chk_get:nw.)

\__prop_pair:wn

Used as \__prop_pair:wn (key) \s__prop (item) for both storage types, this internal token starts each key–value pair in the property list. This default definition is changed globally by any mapping function, so there is not much point trying to make it an error. Instead, the error is produced by \__prop_chk:w.

19922 \cs_new:Npn \__prop_pair:wn #1 \s__prop #2 { }

(End of definition for \__prop_pair:wn.)

\__prop_flatten:w

We implement here the fact that f-expanding a linked property list gives a flat property list. Leaving a linked property list in the input stream will turn it into a flat property list so that the error implemented by \__prop_con:wn will correctly be triggered.

19923 \cs_new_protected:Npn \__prop_flatten:w #1 \s__prop #2
19924 { \use:e { \__prop_flatten_aux:N #3 } }

(End of definition for \__prop_flatten:w.)
The main function \texttt{\_prop_flatten:N} receives a linked property list and flattens it. The auxiliary \texttt{\_prop_flatten_aux:N} receives a pointer to the first key and flattens the linked property list into a flat property list. This is only restricted-expandable as it involves mapping through all of the property list’s entries starting from \{\texttt{key1}\}.

The looping function \texttt{\_prop_flatten_loop:w} removes \texttt{\use_none:n} and a backwards pointer \#2, leaves the key–value pair for \texttt{\use:e} to receive, and calls itself again after expanding the next key’s token list. Its argument \#3 is empty, except at the end where it is the \texttt{\use_none:nnnn} appearing in the definition of \texttt{\_prop_flatten_aux:N}, which ends the loop.

\begin{verbatim}
\cs_new:Npn \_prop_flatten:N #1 { \exp_after:wN \_prop_flatten_aux:w #1 }
\cs_new:Npn \_prop_flatten_aux:w #1 \s__prop #2 { \_prop_flatten_aux:N }
\cs_new:Npn \_prop_flatten_loop:w #1#2#3 \__prop_pair:wn #4 \s__prop #5 {
  \exp_not:n { \__prop_pair:wn #4 \s__prop {#5} }
  \exp_after:wN \_prop_flatten_loop:w
}
\end{verbatim}

(End of definition for \texttt{\_prop_flatten:N} and others.)

\texttt{\_prop_prefix_int} Used to assign prefixes for each linked property list. It is converted to base \texttt{\_prop BASIS_INT}, then each digit is converted to a character, starting at \texttt{!} (the character after space).

\begin{verbatim}
\int_new:N \g__prop_prefix_int
\int_const:Nn \c__prop_basis_int { \c_max_char_int - '\!' }
\end{verbatim}

(End of definition for \texttt{\_prop_prefix_int} and \texttt{\_prop BASIS_INT}.)

\texttt{\_prop_next_prefix:} \texttt{\_prop_to_prefix:n} Store in \texttt{\_prop_internal_tl} the conversion of \texttt{\_prop_prefix_int} to characters, and increment this integer for use in the next linked property list. No need to optimize since this is only used when declaring the property list the first time. The aim here is to make this string as short as we can, given the range of distinct characters available. This speeds up the work of \texttt{\cs:w \cs_end:} that looks up keys in the hash table.

\begin{verbatim}
\cs_new_protected:Npn \_prop_next_prefix: { \tl_set:Ne \l__prop_internal_tl { \_prop_to_prefix:n \g__prop_prefix_int } \int_gincr:N \g__prop_prefix_int }
\end{verbatim}

\texttt{\_prop_internal_tl}
63.3 Allocation and initialisation

\c_empty_prop An empty flat prop.

\prop_new:N Flat property lists are initialized with the value \c_empty_prop.

\prop_new_linked:N The auxiliary is used in \prop_make_linked:N. For linked property lists, get a new prefix in \l__prop_internal_tl, then use it to set up the internal structure: the last token in #1 is usually a pointer to the first key, which is here the end-pointer. That end-pointer has a pointer to the previous key (usually the last key), which is the variable #1 itself that begins the doubly-linked list.

(End of definition for \__prop_next_prefix: and \__prop_to_prefix:n.)
Clearing a flat property list is like declaring it anew, simply setting it equal to \c_empty_prop. For linked property lists we must clear all of the variables storing individual keys, which requires a loop. At each step of the loop, \_\_prop_clear_loop:Nw receives \cs_gset_eq:NN, \use_none:n, the backwards pointer, an empty #4 (except at the end of the loop), and the key–value pair #5=#6 which we disregard. The looping auxiliary undefines the previous key’s token list (this includes the main token list, but that is fine because it is restored at the end) and calls itself after expanding the next key’s token list. The loop ends when #4 is \use_none:nnnn. After the loop, \_\_prop_clear:wNNN correctly sets up the main variable #6 and the end-pointer #1. Importantly, this is done using \cs_gset_nopar:Npe and \exp_not:n because the almost-equivalent \tl_set:Nn would complain in debug mode about the fact that the main variable is undefined at this stage. Importantly, \_\_prop_clear_entries:NN is used in the implementation of \prop_set_eq:NN.
A simple variation of the token list functions.

\prop_set_eq:NN
\prop_set_eq:cN
\prop_set_eq:Nc
\prop_set_eq:cc
\prop_gset_eq:NN
\prop_gset_eq:cN
\prop_gset_eq:Nc
\prop_gset_eq:cc
\__prop_set_eq:NNNN
\__prop_set_eq:wNNNN
\__prop_set_eq:nNnNN
\__prop_set_eq_loop:NNnw
\__prop_set_eq_end:w

If both variables are accidentally the same variable (or equal flat property lists, as it turns out) we do nothing, otherwise the following code would lose all entries. If the target variable #3 is a flat prop, either copy directly or flatten before copying. If it is a linked prop, we must clear it, then go through the entries in #4 to add them to #3.

(End of definition for \prop_clear:N and others. These functions are documented on page 214.)
We have used that $f$-expanding either type of prop gives a flat prop. At this stage \_\_prop_set_eq:nNNNN receives the second variable as a flat prop, the end-pointer, the prefix, the suitable $\texttt{\textbackslash set nopar:}$ assignment, and the first variable itself. Remove the leading $\texttt{\textbackslash s\textunderscore prop}$ and $\texttt{\textbackslash prop\_chk:w}$ with $\texttt{\use_i:nnn}$, then start the loop.

The looping function receives the current pointer $\#1$ (initially the variable itself), the defining function $\#2$ and the prefix $\#3$, then a partial definition $\#4$ (which in later stages includes the backwards pointer), followed by the current value as $\texttt{\textbackslash s\textunderscore prop} \{\#5\}$. It seeks the next key $\#7$ to construct in $\l__prop\_internal\_tl$ the next pointer $\__prop$ $\langle$prefix$\rangle$ $\langle$next key$\rangle$ (the argument $\#6$ is empty, except at the end of the loop, where it is $\texttt{\use none:n}$ in such a way as to delete the $\langle$space$\rangle$ and $\langle$next key$\rangle$). Then the token list (current pointer) $\#1$ is set-up to contain the partial definition and current value, as well as the newly constructed next pointer. After a line responsible for correctly ending the loop with $\__prop\_set\_eq\_end:w$, we loop, setting up the next definition, which starts with $\texttt{\use none:n}$ and a backwards pointer to $\#1$ followed by the $\langle$next key$\rangle$ $\#7$ and so on.

The end-code picks up what is needed to correctly assign the last token list (the end pointer), which is simply $\texttt{\use none:n}$ $\__prop\langle$prefix$\rangle$ $\langle$space$\rangle$ $\langle$key$_n$$\rangle$.

(End of definition for $\texttt{\prop\_set\_eq:NN}$ and others. These functions are documented on page 214.)
the outermost group level, and \cs_set_eq:NN is very slightly faster than its global version. Then store the contents (expanded preventively by \exp_args:NNf) with an assignment \cs_set_nopar:Npe that does not perform \l3debug checks.

\begin{verbatim}
\cs_new_protected:Npn \prop_make_flat:N #1
  { \int_compare:nNnTF { \tex_currentgrouplevel:D } = 0
    { \__prop_if_flat:NTF #1 { } { \exp_args:NNf \__prop_make_flat:Nn #1 {#1} } }
    { \msg_error:nnee { prop } { inner-make }
      { \token_to_str:N \prop_make_flat:N } { \token_to_str:N #1 } }
  }
\cs_generate_variant:Nn \prop_make_flat:N { c }
\cs_new_protected:Npn \__prop_make_flat:Nn #1#2
  { \exp_after:wN \__prop_clear:wNNN #1 \cs_set_eq:NN \cs_set_nopar:Npe #1 \exp_not:n {#2} }
\end{verbatim}

(End of definition for \prop_make_flat:N and \prop_make_flat:c \__prop_make_flat:Nn. This function is documented on page 215.)

The only interesting case is when given a flat prop. We expand the contents for later use. Then \__prop_new_linked:N disregards that previous value of #1 and initializes the linked prop. We can then use an auxiliary \__prop_set_eq:wNNNN underlying \prop_set_eq:NN, with the prop contents saved as \l__prop_internal_tl. That step is a bit unsafe, as \l__prop_internal_tl (really, a flat prop here) is used within \__prop_set_eq:wNNNN itself, but it is in fact expanded early enough to be ok.

\begin{verbatim}
\cs_new_protected:Npn \prop_make_linked:N #1
  { \int_compare:nNnTF { \tex_currentgrouplevel:D } = 0
    { \__prop_if_flat:NTF #1 { } { \exp_args:NNo \__prop_make_linked:Nn #1 {#1} } } { }
    { \msg_error:nnee { prop } { inner-make }
      { \token_to_str:N \prop_make_linked:N } { \token_to_str:N #1 } }
  }
\cs_generate_variant:Nn \prop_make_linked:N { c }
\cs_new_protected:Npn \__prop_make_linked:Nn #1#2
  { \__prop_new_linked:N #1 \tl_set:Nn \l__prop_internal_tl {#2} \exp_after:wN \__prop_set_eq:wNNNN #1 \cs_set_eq:NN \cs_set_nopar:Npe #1 \l__prop_internal_tl }
\end{verbatim}

(End of definition for \prop_make_linked:N and \__prop_make_linked:Nn. This function is documented on page 215.)
We can now initialize the scratch variables.

\prop_new:N \l_tmpa_prop
\prop_new:N \l_tmpb_prop
\prop_new:N \g_tmpa_prop
\prop_new:N \g_tmpb_prop

(End of definition for \l_tmpa_prop and others. These variables are documented on page 222.)

The basic strategy is to copy the first variable into the target, then loop through the second variable, calling \prop_(g)put:Nnn on each item. To avoid running the l3debug scope checks on each of these steps, we use the auxiliaries that underly \prop_set_eq:NN and \prop_put:Nnn, whose syntax is a bit unwieldy. We work directly with the target prop #3 as a scratch space, because copying over from a temporary variable to #3 would be slow in the linked case. If #5 is #3 itself we have to be careful not to lose the data, and we even take the opportunity to skip the copying step completely. To keep the correct version of the duplicate keys we use the code underlying \prop_put_if_not_in:Nnn, which involves passing \use_none:nnn to the auxiliary instead of nothing. There is no need to check for the case where #3 is equal to #4 because in that case \prop_(g)set_eq:NN #3 #4 (or rather the underlying auxiliary) is correctly set up to do no needless work.

\cs_new_protected:Npn \prop_concat:NNN { \__prop_concat:NNNNN \cs_set_eq:NN \cs_set_nopar:Npe }
\cs_generate_variant:Nn \prop_concat:NNN { ccc }
\cs_new_protected:Npn \prop_gconcat:NNN { \__prop_concat:NNNNN \cs_gset_eq:NN \cs_gset_nopar:Npe }
\cs_generate_variant:Nn \prop_gconcat:NNN { ccc }
\cs_new_protected:Npn \__prop_concat:NNNNN #1#2#3#4#5
{\cs_if_eq:NNTF #3 #5 { \__prop_concat:nNNN \use_none:nnn #2 #3 #4 } { \__prop_set_eq:NNNN #1 #2 #3 #4 \__prop_concat:nNNN { } #2 #3 #5 } }
\cs_new_protected:Npn \__prop_concat:nNNN #1#2#3#4
{ \cs_gset_eq:NN \__prop_tmp:w \__prop_pair:wn \exp_last_unbraced:Nf \use_none:nn #4 \cs_gset_eq:NN \__prop_pair:wn \__prop_tmp:w }

(End of definition for \prop_concat:NNN and others. These functions are documented on page 216.)

The core is a call to \keyval_parse:nnn, with an error message \__prop_missing_eq:n for entries without =, and a call to (essentially) \prop_(g)put:Nnn for valid key–value pairs. To avoid repeated scope checks (and errors) when l3debug is active, we instead use the auxiliary underlying \prop_put:Nnn. Because blank keys are valid here, in contrast to l3keys, we set and restore \l__kernel_keyval_allow_blank_keys_bool. The key–value argument may be quite large so we avoid reading it until it is really necessary.
\prop_gset_from_keyval:Nn
\prop_gset_from_keyval:cn
\prop_gput_from_keyval:Nn
\prop_gput_from_keyval:cn

Just empty the prop (with the auxiliary underlying \prop_clear:N to avoid \l3debug problems) and push key–value entries using \prop\_(g)put_from_keyval:Nn.

\prop_const_from_keyval:Nn
\prop_const_from_keyval:cn
\prop_const_linked_from_keyval:Nn
\prop_const_linked_from_keyval:cn

For both flat and linked constant props, we create #1 then use the same auxiliary as for \prop_gput_from_keyval:Nn. It is most natural to use the already packaged \prop\_gput:Nnn, but that would mean doing an assignment on a supposedly constant property list. To avoid errors when \l3debug is activated, we use the auxiliary underlying \prop\_gput:Nnn.
63.4 Accessing data in property lists

Accessing/deleting/adding entries is mostly done by \texttt{\_\_prop_split:NnTFn}, which must be fast because it is used in many \texttt{l3prop} functions. Its syntax is as follows.

\begin{verbatim}
\_\_prop_split:NnTFn \{ \\langle \text{property list} \rangle \} \{ \langle \text{true code} \rangle \} \{ \langle \text{false code} \rangle \} \{ \langle \text{link code} \rangle \}
\end{verbatim}

If the \langle \text{property list} \rangle uses the linked data storage, then it runs the \langle \text{link code} \rangle, otherwise it does as follows.

It splits the \langle \text{property list} \rangle at the \langle \text{key} \rangle, giving three token lists: the \langle \text{entries before} \rangle the \langle \text{key} \rangle, the \langle \text{value} \rangle associated with the \langle \text{key} \rangle and the \langle \text{entries after} \rangle the \langle \text{key} \rangle. Both the \langle \text{entries before} \rangle and the \langle \text{entries after} \rangle can be empty or consist of some number of consecutive entries \texttt{\_\_prop_pair:wn \langle \text{key} \rangle \s__prop \langle \text{value} \rangle}. If the \langle \text{key} \rangle is present in the \langle \text{property list} \rangle then the \langle \text{true code} \rangle is left in the input stream, with \#1, \#2, and \#3 replaced by the \langle \text{entries before} \rangle, \langle \text{value} \rangle, and \langle \text{entries after} \rangle. If the \langle \text{key} \rangle is not present in the \langle \text{property list} \rangle then the \langle \text{false code} \rangle is left in the input stream. Only the \langle \text{true code} \rangle is used in the replacement text of a macro defined internally, which requires \#\# doubling.

The aim is to distinguish four cases: a flat prop that contains the given \langle \text{key} \rangle, a flat prop that does not contain it, a linked prop, and an invalid prop. The last case includes those that are set to \texttt{\relax} by \texttt{c}-expansion, as well as unrelated token list variables since these unfortunately used to “work” in earlier implementations. In the first three cases we run the \texttt{T}, \texttt{F}, and \texttt{n} arguments, and in the last case we raise an error, set the variable to a known state (empty prop), and run the \texttt{F} code (some conditionals such as \texttt{\prop_pop:NnTFn} otherwise blow up pretty badly).

The first distinction between these cases is done by \texttt{\_\_prop_split_test:wn}, which looks for the argument after \texttt{\_s\_prop}. For a flat prop it will be \texttt{\_\_prop_chk:w}, which leads to running \texttt{\_\_prop_split_flat:w}, explained below. For a linked prop it is the prefix, consisting of characters, so we end up running \texttt{\_\_prop_split_linked:w}, which cleans up and selects the aforementioned \texttt{n} argument. For invalid props, or rather, variables that do not contain \texttt{\_s\_prop}, the argument includes \texttt{\fi:}, and we end up calling \texttt{\_\_prop_split_wrong:Nw}, which calls \texttt{\prop_show:N} to raise a detailed error stating how the variable is wrong.

Let us return to \texttt{\_\_prop_split_flat:w}. This function is defined dynamically as

\begin{verbatim}
\_\_prop_split:NNTFn \_\_prop_split_test:wn \_\_prop_split_flat:w \_\_prop_split_linked:w \_\_prop_split_wrong:Nw
\end{verbatim}
Its job is to seek the \textlangle key\textrangle in the property list (known to be flat at this stage) by using an argument \#1 delimited essentially by that key. If indeed the variable contained the \textlangle key\textrangle, then \#1 is the \textlangle extract1\textrangle before the key–value pair, \#2 is the \{value\} associated with the \langle key\rangle, \#3 is the \textlangle extract2\textrangle after the key–value pair, \#4 is \textuse_i:nnn, and we run \textuse_i:nnn \{\texttrue code\}\{\textfalse code\}\{\textlink code\}, selecting the \texttrue code. Otherwise, the whole property list together with \textprop_mark \textuse_i:nnn is taken in as \#1, then \#2 is some tokens \texttt{\textfi: \textprop_split_wrong:Nw (variable)} that were only useful in the case of invalid props, \#3 is empty, and most importantly \#4 is \textuse_ii:nnn. This command selects the \textfalse code.

Note that we define \textprop_split_flat:w in all cases even though it is only used in the flat case. Indeed, to avoid taking in the whole property list (which may be large) as an argument more than strictly necessary, we would have to keep the \texttrue code positioned before the expansion of the prop variable in order to use it in the definition. The only way to do that is to store it using an assignment so we might as well just perform the assignment that we can actually use in the flat case.

\begin{verbatim}
\cs_new_protected:Npn \__prop_split:NnTFn #1#2
  { \exp_after:wN \__prop_split_aux:nNTFn \exp_after:wN \tl_to_str:n {#2} #1 } #1
\cs_new_protected:Npn \__prop_split_aux:nNTFn #1#2#3
  { \cs_set:Npn \__prop_split_flat:w \__prop_split_linked:w ##1 \__prop_pair:wn #1 \s__prop ##2 ##3 \s__prop_mark ##4 \s__prop_stop \{ ##4 {#3} \} \exp_after:wN \__prop_split_test:wn #2 \s__prop_mark \use_i:nnn \__prop_pair:wn #1 \s__prop \{ ? \fi: \__prop_split_wrong:Nw #2 \} \s__prop_mark \use_ii:nnn \s__prop_stop } \cs_new:Npn \__prop_split_flat:w { } \cs_new_protected:Npn \__prop_split_test:wn #1 \s__prop #2 \exp_after:wN \__prop_split_flat:w \fi: \__prop_split_linked:w \cs_set:Npn \__prop_split_linked:w \__prop_split_linked_aux:w \cs_set:Npn \__prop_split_linked_aux:w \__prop_split_linked_wrong:Nw #1#2 \__prop_stop \exp_after:wN \__prop_get:NnN \prop_get:NVN \prop_get:NvN \prop_get:NeN \prop_get:NoN \prop_get:NxN \prop_get:cnN \prop_get:cVN \prop_get:cvN \prop_get:ceN \prop_get:coN \prop_get:cxN \prop_get:cnc
\end{verbatim}

Here we implement both \prop_get:NnN and its branching version through \__prop_get:NnTF. It receives the prop and key, followed by an assignment used when the value is found, \texttt{(true code)} to run after the assignment, and some fall-back \texttt{(false code)} for absent values. It relies on \textlangle ...\rangle. For a flat prop, the first four arguments of \__prop_split:NnTF are used, and run either the assignment \#3{\#3} and \texttt{(true code)} \#4, or the \texttt{(false code)} \#5.
For a linked prop we must work a bit: \__prop_get_linked:w is followed by the expansion of the prop, then by four brace groups: the key #4, the assignment code #5, \langle true code \rangle #6, and \langle false code \rangle #7. If the key is present, its value is stored in the token list \__prop_#2~#4. If that token list exists, \__prop_get_linked_aux:w gets called followed by the expansion of that token list and we grab as #2 the value associated to that key, which we feed to the assignment code and follow-up code. If the key is absent the token list can be \undefined or \relax. In both cases \__prop_get_linked_aux:w finds an empty brace group as #2, \use_none:n as #4 and the \langle false code \rangle as #5. Note that we made \__prop_get_linked:w and subsequent auxiliaries expandable, because they are also used in \prop_item:Nn.

(End of definition for \prop_get:NnN and others. These functions are documented on page 217.)
Getting the value corresponding to a key in a flat property list in an expandable fashion simply uses \prop_map_tokens:Nn to go through the property list. The auxiliary \__prop_item:nnn receives the search string #1, the key #2 and the value #3 and returns as appropriate.

\cs_new:Npn \prop_item:Nn #1#2
\__prop_if_flat:NTF #1
\exp_args:NNo \prop_map_tokens:Nn #1
\exp_after:wN \__prop_item:nnn
\exp_after:wN { \tl_to_str:n {#2} }
\exp_after:wN \__prop_get_linked:w #1 {#2} \use:n { } { }
\end{definition}

\__prop_pop:NnNNnTF
\__prop_pop_linked:wnNNnTF
\__prop_pop_linked:NNNn
\__prop_pop_linked:w
\__prop_pop_linked_prev:w
\__prop_pop_linked_next:w

This auxiliary is used by both the \prop_pop family and the \prop_remove family of functions. It receives a \prop and a \{key\}, three assignment functions (\tl_set:Nn \cs_set_eq:NN \cs_set_nopar:Npe or their global versions), then \{code\} \{true code\} \{false code\}.

For a flat prop, split it. If the \{key\} is there, reconstruct the rest of the prop from the two extracts #2 #4 and assign using \tl_(g)set:Nn, then run \{code\} \{value\} with the \{value\} found, and run the \{true code\}. If the \{key\} is absent, run the \{false code\}.

For a linked prop, the removal is done by \__prop_pop_linked:wnNNnTF, which removes the key–value pair from the doubly-linked list and runs its last three arguments \{code\} \{true code\} \{false code\} depending on whether the key–value is found, in the same way as for flat props.

\cs_new_protected:Npn \__prop_pop:NNNNnTF #1#2#3#4#5#6#7
\__prop_split:NnF \#1 \#2
\#4 \#1 \{ \exp_not:n { \s__prop \__prop_chk:w \#1 \#3 } \} \#5 \#2\#6\#7
\{\#7\}
\exp_after:wN \__prop_pop_linked:wnNNnTF \#1 \#2 \#3 \#4 \#5 \#6 \#7

63.5 Removing data from property lists
The next auxiliary \texttt{\__prop\_pop\_linked:wnNNnTF}, together with the \texttt{NNNn} auxiliary, checks if the key is present in the \texttt{\langle linked prop \rangle}, then the corresponding value (if present) is passed as a braced argument to the \texttt{\langle code \rangle} and the \texttt{\langle true code \rangle} or \texttt{\langle false code \rangle} is run as appropriate. Before that, there are also three assignments: the token lists for the previous key and next key are made to point to each other, cf. \texttt{\__prop\_pop\_linked:w}, and the token list for the given key is made undefined.

\begin{verbatim}
\cs_new_protected:Npn \__prop_pop_linked:wnNNnTF
\__prop_flatten:w #1 \s__prop #2#3#4#5#6#7
{\if_cs_exist:w \__prop ~ #2 ~ \tl_to_str:n {#4} \cs_end:
\exp_after:wN \__prop_pop_linked:NNNn
\cs:w \__prop ~ #2 ~ \tl_to_str:n {#4} \cs_end: #5 #6 {#7}
\else:
\exp_after:wN \use_iii:nnn
\fi:
}\use_i:nn
}\cs_new_protected:Npn \__prop_pop_linked:NNNn #1#2#3#4
{\if_meaning:w \scan_stop: #1
\exp_after:wN \exp_after:wN \exp_after:wN \use_iii:nnn
\else:
\exp_after:wN \__prop_pop_linked:w #1 #1 #2 #3 {#4}
\fi:
}\cs_new_protected:Npn \__prop_pop_linked:w
\use_none:n #1#2 \s__prop #3#4#5#6#7#8
{\#6 #5 \tex_undefined:D
\#7 #1
{\exp_after:wN \__prop_pop_linked_prev:w #1
\exp_not:N #4}
\#7 #4
{\exp_not:n { \use_none:n #1 }
\exp_not:f { \exp_after:wN \__prop_pop_linked_next:w #4 }
}\#8 (#3}
}\cs_new:Npn \__prop_pop_linked_prev:w #1 \s__prop #2#3#4
{\exp_not:n { \#1 \#2 \s__prop {#2} }
}\cs_new:Npn \__prop_pop_linked_next:w \use_none:n #1 \{ \exp_stop_f: }
\end{verbatim}

Deleting from a property relies on \texttt{\__prop\_pop\_linked:wnNNnTF}. The three assignment functions are suitably local or global. The last three arguments are \texttt{\use_none:n} and two empty brace groups: if the key is found we get \texttt{\use_none:n \{\langle key\rangle\}\{empty\}}, which ex-
pands to nothing, and otherwise we just get (empty). The auxiliary takes care of actually
removing the entry from the prop.

\begin{verbatim}
\cs_new_protected:Npn \prop_remove:Nn #1#2
\{\__prop_pop:NnNNnTF #1 {#2} \cs_set_eq:NN \cs_set_nopar:Npe \use_none:n \} \}
\cs_new_protected:Npn \prop_gremove:Nn #1#2
\{\__prop_pop:NnNNnTF #1 {#2} \cs_gset_eq:NN \cs_gset_nopar:Npe \use_none:n \} \}
\cs_generate_variant:Nn \prop_remove:Nn { NV , Ne , c , cV , ce }
\cs_generate_variant:Nn \prop_gremove:Nn { NV , Ne , c , cV , ce }
\end{verbatim}

(End of definition for \prop_remove:Nn and \prop_gremove:Nn. These functions are documented on
page 218.)

\prop_pop:NnN\prop_pop:NVN\prop_pop:NoN\prop_pop:cnN\prop_pop:cVN\prop_pop:coN\prop_gpop:NnN\prop_gpop:NVN\prop_gpop:NoN\prop_gpop:cnN\prop_gpop:cVN\prop_gpop:coN Popping a value is almost the same, but the value found is kept. For the non-branching
version, we additionally set the target token list to \q_no_value, while for the branching
version we must produce \prg_return_true: or \prg_return_false:.

\begin{verbatim}
\cs_new_protected:Npn \prop_pop:NnN #1#2#3
\{\__prop_pop:NnNNnTF #1 {#2} \cs_set_eq:NN \cs_set_nopar:Npe \{ \tl_set:Nn #3 \} \prg_return_true: \prg_return_false: \}
\cs_new_protected:Npn \prop_gpop:NnN #1#2#3
\{\__prop_pop:NnNNnTF #1 {#2} \cs_gset_eq:NN \cs_gset_nopar:Npe \{ \tl_set:Nn #3 \} \prg_return_true: \prg_return_false: \}
\end{verbatim}

\prg_generate_conditional_variant:Nnn \prop_pop:NnN { NV , No , c , cV , co } { T , F , TF }\prg_generate_conditional_variant:Nnn \prop_gpop:NnN { NV , No , c , cV , co } { T , F , TF }

931
63.6 Adding data to property lists

All of the $\prop_{(g)put(_if_new)}:\text{Nnn}$ functions are based on the same auxiliary, which receives \code{key} and an “assignment”, followed by \prop{((key)) \{new value\}}. The assignment $\cs_{(g)set\_npar}\text{Npe}$ is the primitive assignment without any checking: in the case of linked props it is applied to individual pieces of the linked prop, which are typically not yet defined. Debugging the scope of the variable is done at a higher level by letting \l3debug change $\prop_{put\_if\_not\_in}$ and friends. This allows other \l3prop commands to directly call the underlying auxiliary to skip this checking step and avoid getting multiple error messages for the same error. The \code{empty} for $\prop_{put\_if\_not\_in\_in}$ is placed before the assignment in cases where the key is already present, in order to suppress the assignment in the $\prop_{put\_if\_not\_in\_in}$ case.

\begin{verbatim}
\prop_gput:NNn \prop_gput:NNv \prop_gput:NNn \prop_gput:NNv \prop_gput:NNn \prop_gput:NNn \prop_gput:NNn \prop_gput:NNn \prop_gput:NNn \prop_gput:NNn \prop_gput:NNn \prop_gput:NNn \prop_gput:NNn \prop_gput:NNn \prop_gput:NNn \prop_gput:NNn \prop_gput:NNn \prop_gput:NNn \prop_gput:NNn \prop_gput:NNn \prop_gput:NNn \prop_gput:NNn \prop_gput:NNn \prop_gput:NNn \prop_gput:NNn \prop_gput:NNn \prop_gput:NNn \prop_gput:NNn \prop_gput:NNn \prop_gput:NNn \prop_gput:NNn \prop_gput:NNn \prop_gput:NNn \prop_gput:NNn \prop_gput:NNn \prop_gput:NNn \prop_gput:NNn \prop_gput:NNn
\end{verbatim}

(End of definition for $\prop_{put\_if\_not\_in\_in\_in}$ and others. These functions are documented on page 217.)
Since the true branch of \__prop_split:NnTFn is used as the replacement text of an internal macro, and since the \langle key\rangle and new \langle value\rangle may contain arbitrary tokens, it is not safe to include them in the argument of \__prop_split:NnTFn. We thus start by storing in \l__prop_internal_tl tokens which (after x-expansion) encode the key–value pair. This variable can safely be used in \__prop_split:NnTFn. For a flat prop, if the \langle key\rangle was absent, append the new key–value to the list; otherwise concatenate the extracts \#2 and \#4 with the new key–value pair \l__prop_internal_tl. The updated entry is placed at the same spot as the original \langle key\rangle in the property list, preserving the order of entries. For a linked prop, call \__prop_put_linked:wnNN, which constructs the control sequence in which we will place the new value. If it matches \scan_stop: then the key was not yet there and we add it using \__prop_put_linked_new:w, otherwise it was already there and we use \__prop_put_linked_old:w.

\cs_new_protected:Npn \__prop_put:nNNnn #1#2#3#4#5
\tl_set:Nn \l__prop_internal_tl
{ \exp_not:N \__prop_pair:wn \tl_to_str:n {#4} \s__prop \exp_not:n {#5} }
\__prop_split:NnTFn #3 {#4}
{ #1 #2 #3
\s__prop \__prop_chk:w \exp_not:n {##1} \l__prop_internal_tl \exp_not:n {##3}
}
{ #2 #3 { \exp_not:o {#3} \l__prop_internal_tl } }
{ \exp_after:wN \__prop_put_linked:wnnN #3 {#4} {#1} #2 }
\cs_new_protected:Npn \__prop_put_linked:NNnN #1#2#3#4
{ \__prop_flatten:w #1 \s__prop \tl_to_str:n {#4} \cs_end: #1 #3 #4

\cs_new_protected:Npn \__prop_put_linked:NNnN #1#2#3#4
{ \exp_after:wN \__prop_put_linked:NNnN \cs:w __prop ~ #2 ~ \tl_to_str:n {#4} \cs_end: #1 #3 #4

933
To add a new entry, \_\_prop_put_linked_new:w receives the expansion of the end-pointer, namely \use_none:n ⟨last key pointer⟩, followed by the new key pointer #2, the end pointer #3, and an assignment function #4. Set up the doubly-linked list in the order #1, #2, #3, placing the key–value pair \l__prop_internal_tl in #2. To replace an old entry, \_\_prop_put_linked_old:w receives the expansion of that entry, and it reassigns it (#5) using the assignment #6, by simply replacing the payload #2 \s__prop #3 by \l__prop_internal_tl.

\cs_new_protected:Npn \_\_prop_put_linked_new:w
\use_none:n #1#2#3#4
{ #4 #1
  { \exp_after:wN \_\_prop_pop_linked_prev:w #1
    \exp_not:N #2
  }
  #4 #2
  { \exp_not:n { \use_none:n #1 }
    \l__prop_internal_tl
    \exp_not:N #3
  }
  #4 #3 { \exp_not:n { \use_none:n #2 } }
}
\cs_new_protected:Npn \_\_prop_put_linked_old:w
\use_none:n #1#2 \s__prop #3#4#5
{ #5
  { \exp_not:n { \use_none:n #1 }
    \l__prop_internal_tl
    \exp_not:N #4
  }
}

(End of definition for \prop_put:Nnn and others. These functions are documented on page 216.)

63.7 Property list conditionals

\prop_if_exist_p:N \prop_if_exist_p:c
\prop_if_exist:NTF \prop_if_exist:cTF

Copies of the cs functions defined in l3basics.

\prg_new_eq_conditional:NNn \prop_if_exist:N \cs_if_exist:N
\prg_new_eq_conditional:NNn \prop_if_exist:c \cs_if_exist:c
{ TF , T , F , p }

(End of definition for \prop_if_exist:NTF. This function is documented on page 218.)
A flat property list is empty if it matches `\c_empty_prop`. A linked property list is empty if its second token (the end pointer) and last token (the first key pointer) are equal. There cannot be false positives because the end pointer takes the form `\use_none:n ⟨pointer⟩` while the other pointers have more elaborate structure. The subtle code branch here is when a non-empty flat property list is given: then `\_\_prop_if_empty:w` reads the whole property list as `#1 #2 #3 #4` are 2, 3, 4, respectively.

\begin{verbatim}
\prg_new_conditional:Npnn \prop_if_empty:N #1 \{ p , T , F , TF \}
\begin{verbatim}
\if_meaning:w #1 \c_empty_prop
\prg_return_true:
\else:
\exp_after:wN \_\_prop_if_empty_return:w #1
\_\_prop_flatten:w 2 \s__prop 34 \s__prop_stop
\fi:
\end{verbatim}
\end{verbatim}
\cs_new:Npn \_\_prop_if_empty_return:w #1 \_\_prop_flatten:w #2 \s__prop #3 #4 #5 \s__prop_stop
\begin{verbatim}
\if_meaning:w #2 #4
\prg_return_true:
\else:
\prg_return_false:
\fi:
\end{verbatim}
\end{verbatim}
\prg_generate_conditional_variant:Nnn \prop_if_empty:N \{ c \} \{ p , T , F , TF \}
\end{verbatim}

(End of definition for `\prop_if_empty:NTF` and `\_\_prop_if_empty_return:w`. This function is documented on page 219.)

For a linked prop, use `\_\_prop_get_linked:w` to look up whether the control sequence constructed from the prefix and the sought-after key exists; this auxiliary calls `\use_-none:n \{⟨value⟩\} \prg_return_true:` if the key is found, and otherwise `\prg_return-_false:`. For a flat prop, testing expandably if a key is there requires to go through the key–value pairs one by one. This is rather slow, and a faster test would be

\begin{verbatim}
\_\_prop_split:NnTFn #1 {#2}
\{ \prg_return_true: \}
\{ \prg_return_false: \}
\{ ... \}
\end{verbatim}

but `\_\_prop_split:NnTFn` is non-expandable. Instead, we use `\prop_map_tokens:Nn` to compare the search key to each key in turn using `\str_if_eq:ee`, which is expandable.

\begin{verbatim}
\prg_new_conditional:Npnn \prop_if_in:Nn #1 #2 \{ p , T , F , TF \}
\begin{verbatim}
\_\_prop_if_in_flat:nnn \prop_if_in:Nn \#1 \#2
\end{verbatim}
\end{verbatim}

935
63.8 Mapping over property lists

We first \texttt{f}-expand to flatten \#1 in case it was a linked list. The \texttt{use_i:nn} removes the leading \texttt{\_\_prop \_\_prop_chk:w} of the flattened prop. The even-numbered arguments of \texttt{\_\_prop_map_function:Nw} are keys, hence have string catcodes, except at the end where they are \texttt{\_\_prop_map_break:}. The \texttt{\_\_prop_map_break:} construction and we jump out of the loop. No need for any quark test.

(End of definition for \texttt{\_\_prop_map_function:Nw} and \texttt{\_\_prop_map_function:NN}. This function is documented on page 219.)
Mapping in line requires a nesting level counter. Store the current definition of \_\_prop_pair:wn, and define it anew. At the end of the loop, revert to the earlier definition. Note that besides pairs of the form \_\_prop_pair:wn \langle key \rangle \s\_\_prop \{ \langle value \rangle \}, there are a leading and a trailing tokens, but both are equal to \scan_stop:, hence have no effect in such inline mapping. Such \scan_stop: could have affected ligatures if they appeared during the mapping.

\begin{verbatim}
\cs_new_protected:Npn \prop_map_inline:Nn #1#2
\begin{verbatim}
\cs_gset_eq:cN \__prop_map_ \int_use:N \g__kernel_prg_map_int :wn \_\_prop_pair:wn
\int_gincr:N \g__kernel_prg_map_int
\cs_gset_protected:Npn \__prop_pair:wn ##1 \s__prop ##2 {#2}
\exp_last_unbraced:Nf \use_none:nn #1
\prg_break_point:Nn \prop_map_break: \{
\int_gdecr:N \g__kernel_prg_map_int
\cs_gset_eq:Nc \__prop_pair:wn \{ \_\_prop_map_ \int_use:N \g__kernel_prg_map_int :wn \}
\}
\}
\cs_generate_variant:Nn \prop_map_inline:Nn { c }
\end{verbatim}
(End of definition for \prop_map_inline:Nn. This function is documented on page 220.)
\end{verbatim}

The mapping is very similar to \prop_map_function:NN. The odd construction \use:n \{#1\} allows #1 to contain any token without interfering with \prop_map_break:. The loop stops when the \langle key \rangle between \_\_prop_pair:wn and \s\_\_prop is \fi: \prop_map_break: instead of being a string.

\begin{verbatim}
\cs_new:Npn \prop_map_tokens:Nn #1#2
\begin{verbatim}
\exp_last_unbraced:Nnf \use_i:nnn { \__prop_map_tokens:nw {#2} } #1
\_\_prop_pair:wn \fi: \prop_map_break: \s\_\_prop \{ \}
\_\_prop_pair:wn \fi: \prop_map_break: \s\_\_prop \{ \}
\_\_prop_pair:wn \fi: \prop_map_break: \s\_\_prop \{ \}
\_\_prop_pair:wn \fi: \prop_map_break: \s\_\_prop \{ \}
\_\_prop_pair:wn \fi: \prop_map_break: \s\_\_prop \{ \}
\prg_break_point:Nn \prop_map_break: \{
\}
\cs_new:Npn \__prop_map_tokens:nw #1
\begin{verbatim}
\__prop_pair:wn #2 \s__prop #3
\__prop_pair:wn #4 \s__prop #5
\__prop_pair:wn #6 \s__prop #7
\__prop_pair:wn #8 \s__prop #9
\{\if_false: #2 \fi: \use:n \{#1\} \{#2\} \{#3\}
\if_false: #4 \fi: \use:n \{#1\} \{#4\} \{#5\}
\if_false: #6 \fi: \use:n \{#1\} \{#6\} \{#7\}
\if_false: #8 \fi: \use:n \{#1\} \{#8\} \{#9\}
\_\_prop_map_tokens:nw \{#1\}
\}
\}
\cs_generate_variant:Nn \prop_map_tokens:Nn { c }
(End of definition for \prop_map_tokens:Nn and \__prop_map_tokens:nw. This function is documented on page 220.)
\end{verbatim}

937
The break statements are based on the general \prg_map_break:Nn.

\begin{verbatim}
\cs_new:Npn \prop_map_break: 
{ \prg_map_break:Nn \prop_map_break: { } }
\cs_new:Npn \prop_map_break:n
{ \prg_map_break:Nn \prop_map_break: }
\end{verbatim}

(End of definition for \prop_map_break: and \prop_map_break:n. These functions are documented on page 221.)

### 63.9 Uses of mapping over property lists

Counting the key–value pairs in a property list is done using the same approach as for other count functions: turn each entry into a +1 then use integer evaluation to actually do the mathematics.

\begin{verbatim}
\cs_new:Npn \prop_count:N #1
{ \int_eval:n { 0 \prop_map_function:NN #1 \__prop_count:nn } }
\cs_new:Npn \__prop_count:nn #1#2 { + 1 }
\cs_generate_variant:Nn \prop_count:N { c }
\end{verbatim}

(End of definition for \prop_count:N and \__prop_count:nn. This function is documented on page 218.)

Each property name and value pair will be returned in the form \{⟨name⟩=⟨value⟩\}. As one of the main use cases for this macro is to pass the \langle property list \rangle on to a key–value parser, we have to make sure that the behaviour is as good as possible. Using a space before the opening brace we get the correct brace stripping behaviour for most of the key–value parsers available in \LaTeX. Iterate over the \langle property list \rangle and remove the leading comma afterwards. Only the value has to be protected in \__kernel_exp_not:w as the property name is always a string. After the loop the leading comma is removed by \use_none:n and afterwards \__kernel_exp_not:w eventually finds the opening brace of its argument.

\begin{verbatim}
\cs_new:Npn \prop_to_keyval:N #1
{ \__kernel_exp_not:w \prop_if_empty:NTF #1 { {} } }
\cs_new:Npn \__prop_to_keyval:nnw
{ \__kernel_exp_not:w \exp_after:wN \exp_after:wN \exp_after:wN \tex_expanded:D }
\cs_new:Npn \__kernel_exp_not:w
{ \exp_not:N \use_none:n \prop_map_function:NN #1 \__prop_to_keyval:nnw }
\end{verbatim}

938
63.10 Viewing property lists

Experience shows one source of problems is very hard to debug: when a data structure such as a seq or prop gets corrupted. In the past, \prop_show:N would in some cases happily show items of such a prop, even though other more demanding \l3prop functions would choke. It is thus best to make \prop_show:N check very thoroughly the structure and flag issues, even though that is very painful for linked props. Throughout the code below, we strive to remain as safe as possible, but in the explanations we only state what the arguments are when the prop is correctly formed, rather than saying at every step that various arguments can be arbitrary junk, made safe by using \tl_to_str:n generously.

The general \__kernel_chk_tl_type:NnnT checks that its first argument is a token list, and if it is, then it \texttt{e}-expands its second argument and compares with the contents of its first argument. Thus, within this \texttt{e}-expansion it is safe to use \__prop_if_flat:NTF to check if the prop is flat or linked. In the flat case we simply reconstruct the expected structure using \__prop_show_flat:w, which loops through the prop and correctly turns all keys to strings for instance. In the linked case, we use \__prop_show_linked:w, which ensures the form \__prop_flatten:w \__prop ⟨prefix⟩ \s__prop {⟨prefix⟩} ⟨rest⟩, where ⟨prefix⟩ is made into a string and ⟨rest⟩ cannot be a brace group or multiple tokens since \__prop_show_linked:w would in such cases give a different result from the original token list.
For flat props we are done by using `\msg_show:nneee` or `\msg_log:nneee`. The auxiliary `\__prop_show_finally:NNn` is eventually also used in the linked case after some more tests. To avoid having to bring along the message function and the property list, we store them into `\l__prop_internal_tl`.

For linked props, we now know they have a reasonable form so that we are calling `\__prop_show_prepare:w \__prop_flatten:w \__prop ⟨prefix⟩ \langle token⟩ ⟨property list⟩`, and the task is to loop through the linked list and check integrity. We first set things up: the auxiliary `\__prop_tmp:w` will be in charge of checking that various tokens start with `\__prop ⟨prefix⟩` (in the sense of string representations), and calling one of `\__prop_show_loop_key:NNN, \__prop_show_end:NNN, \__prop_show_bad_name:NNN`.

For linked props, we now know they have a reasonable form so that we are calling `\__prop_show_prepare:w \__prop_flatten:w \__prop ⟨prefix⟩ \langle token⟩ ⟨property list⟩`, and the task is to loop through the linked list and check integrity. We first set things up: the auxiliary `\__prop_tmp:w` will be in charge of checking that various tokens start with `\__prop ⟨prefix⟩` (in the sense of string representations), and calling one of `\__prop_show_loop_key:NNN, \__prop_show_end:NNN, \__prop_show_bad_name:NNN`.
The loop will consist of calls to \__prop_show_loop:NNw \__prop ⟨prefix⟩⟨token⟩⟨expansion⟩, where ⟨token⟩ is one of the items in the list, specifically the key container for ⟨key⟩ (starting at \textit{i} = 1 with the property list variable itself), and ⟨expansion⟩ stands for the expansion of that token, which has already been checked, and takes the form ⟨junk⟩ \s__prop {⟨value⟩} \__prop ⟨prefix⟩⟨key⟩. Thus, the loop auxiliary receives the prefix command as \#1, and the \((i - 1)\)-th and \(i\)-th key containers as \#2 and \#5. Then \__prop_tmp:w checks that the name of the \(i\)-th key container is valid.

If the \(i\)-th key container has the wrong name we get \__prop_show_bad_name:NNN \__prop ⟨prefix⟩⟨previous container⟩⟨current container with bad name⟩.

If the \(i\)-th key container has the name \__prop ⟨prefix⟩ (without space), it is the trailing one. We check that it is the right kind of macro to be a token list, and that it has the right contents \use_none:n ⟨previous container⟩. If so, we are done checking everything, and we display the property list using the message function and property list name stored in \l__prop_internal_tl. Note that we also use this \l__prop_internal_tl in the type argument of \__kernel_chk_tl_type:NnnT, to build up the name “⟨property list⟩ prop entry” used in error messages.

If the \(i\)-th container has a name \__prop ⟨prefix⟩ ⟨key⟩ (with a space before the key), then we have a call to \__prop_show_loop_key:WNnW ⟨key⟩ ⟨junk1⟩ ⟨junk2⟩ \__prop ⟨prefix⟩⟨previous container⟩ ⟨current container⟩ (with an \textit{f}-expansion to eliminate the space). The first argument is the ⟨key⟩ without a leading space, thanks to a judicious \textit{f}-expansion earlier on. We check that the ⟨current container⟩ is a token list with the expected structure \use_none:n ⟨previous container⟩ \__prop_pair:wn ⟨string⟩ \s__prop {⟨anything⟩} ⟨single token⟩. The auxiliary \__prop_show_flat:w
is reused to produce the \_\_prop\_pair:wn part, and the last token is produced by \tl\_\item:Nn (we don’t waste a specialized auxiliary to speed that up). If the check succeed, move on to the next item.

\cs_new_protected:Npn \__prop\_show\_loop\_key:wNNN \#1\#2\#3\#4\#5\#6
\begin{Verbatim}
\texttt{\_\_kernel\_chk\_tl\_type:NnnT \#6}
\texttt{\{ \tl\_tail:N \_\_prop\_internal\_tl prop\_entry \}}
\texttt{\{ \exp\_not:n \{ use\_none:n \#5 \}}
\texttt{\exp\_after:wN \_\_prop\_show\_flat:w \#6 \_\_prop \{}
\texttt{\_\_prop\_pair:wn \_\_prop\_recursion\_tail \_\_prop \{}
\texttt{\_\_prop\_recursion\_stop}
\texttt{\tl\_item:Nn \#6 \{-1\}}
\end{Verbatim}

\texttt{\exp\_last\_unbraced:NNNo \_\_prop\_show\_loop:NNw \#4 \#6 \#6}

\texttt{\}}

(End of definition for \prop\_show:N and others. These functions are documented on page 221.)
Chapter 64

l3skip implementation

64.1 Length primitives renamed

Primitives renamed.

\if_dim:w \__dim_eval:w \__dim_eval_end: Primitives renamed.
\cs_new_eq:NN \if_dim:w \tex_ifdim:D
\cs_new_eq:NN \__dim_eval:w \tex_dimexpr:D
\cs_new_eq:NN \__dim_eval_end: \tex_relax:D

(End of definition for \if_dim:w, \__dim_eval:w, and \__dim_eval_end:. This function is documented on page 238.)

64.2 Internal auxiliaries

Internal scan marks.
\s__dim_mark \s__dim_stop Internal scan marks.
\scan_new:N \s__dim_mark
\scan_new:N \s__dim_stop

(End of definition for \s__dim_mark and \s__dim_stop.)

\__dim_use_none_delimit_by_s_stop:w Functions to gobble up to a scan mark.
\cs_new:Npn \__dim_use_none_delimit_by_s_stop:w #1 \s__dim_stop { }

(End of definition for \__dim_use_none_delimit_by_s_stop:w.)

64.3 Creating and initialising dim variables

Allocating (dim) registers ...
\dim_new:N \dim_new:c

\cs_new_protected:Nmp \dim_new:N #1
\{ 
\__kernel_chk_if_free_cs:N #1
\cs:w newdimen \cs_end: #1
\}
\cs_generate_variant:Nn \dim_new:N \dim_new:N \dim_new:N \dim_new:N \dim_new:N { c }
Contrarily to integer constants, we cannot avoid using a register, even for constants. We cannot use \texttt{\dim_gset:Nn} because debugging code would complain that the constant is not a global variable. Since \texttt{\dim_const:Nn} does not need to be fast, use \texttt{\dim_eval:n} to avoid needing a debugging patch that wraps the expression in checking code.

\begin{Verbatim}
\cs_new_protected:Npm \dim_const:Nn #1#2
\{
\dim_new:N #1
\tex_global:D #1 = \dim_eval:n {#2} \scan_stop:
\}
\cs_generate_variant:Nn \dim_const:Nn { c }
\end{Verbatim}

Reset the register to zero. Using \texttt{\c_zero_skip} deals with the case where the variable passed is incorrectly a skip (for example a \LaTeX\ length). Besides, these functions are then simply copied for \texttt{\skip_zero:N} and related functions.

\begin{Verbatim}
\cs_new_protected:Npm \dim_zero:N #1 { #1 = \c_zero_skip }
\cs_new_protected:Npm \dim_gzero:N #1 { \tex_global:D #1 = \c_zero_skip }
\cs_generate_variant:Nn \dim_zero:N { c }
\cs_generate_variant:Nn \dim_gzero:N { c }
\end{Verbatim}

Create a register if needed, otherwise clear it.

\begin{Verbatim}
\cs_new_protected:Npm \dim_zero_new:N #1
\{
\dim_if_exist:NTF #1 { \dim_zero:N #1 } { \dim_new:N #1 }
\}
\cs_new_protected:Npm \dim_gzero_new:N #1
\{
\dim_if_exist:NTF #1 { \dim_gzero:N #1 } { \dim_new:N #1 }
\}
\cs_generate_variant:Nn \dim_zero_new:N { c }
\cs_generate_variant:Nn \dim_gzero_new:N { c }
\end{Verbatim}

Copies of the \texttt{cs} functions defined in \texttt{l3basics}.

\begin{Verbatim}
\prg_new_eq_conditional:NNn \dim_if_exist:N \cs_if_exist:N { TF , T , F , p }
\prg_new_eq_conditional:NNn \dim_if_exist:c \cs_if_exist:c { TF , T , F , p }
\end{Verbatim}

Setting dimensions is easy enough but when debugging we want both to check that the variable is correctly local/global and to wrap the expression in some code. The \texttt{\scan_stop:} deals with the case where the variable passed is a skip (for example a \LaTeX\ length).

\begin{Verbatim}
\cs_new_protected:Npm \dim_set:Nn \dim_set:cn \dim_gset:Nn \dim_gset:cn
\{
\dim_new:N \#1
\tex_global:D \#1 = \dim_eval:n {#2} \scan_stop:
\}
\end{Verbatim}
\loadlang{en}

64.5 Utilities for dimension calculations

\cs_new_protected:Npn \dim_gset:Nn #1#2
\{ \tex_global:D #1 = \__dim_eval:w #2 \__dim_eval_end: \scan_stop: \}
\cs_generate_variant:Nn \dim_set:Nn { c }
\cs_generate_variant:Nn \dim_gset:Nn { c }

(End of definition for \dim_set:Nn and \dim_gset:Nn. These functions are documented on page 224.)

\dim_set_eq:NN All straightforward, with a \scan_stop: to deal with the case where #1 is (incorrectly)
\dim_set_eq:cN a skip.
\dim_set_eq:cc
\dim_gset_eq:NN
\dim_gset_eq:cN
\dim_gset_eq:cc

(End of definition for \dim_set_eq:NN and \dim_gset_eq:NN. These functions are documented on page 224.)

\dim_add:Nn Using by here would slow things down just to detect nonsensical cases such as passing
\dim_add:cn \dimen 123 as the first argument. Using \scan_stop: deals with skip variables. Since
\dim_add:cc debugging checks that the variable is correctly local/global, the global versions cannot
be defined as \tex_global:D followed by the local versions.
\dim_gadd:Nn
\dim_gadd:cn
\dim_gadd:cc
\dim_sub:Nn
\dim_sub:cn
\dim_sub:cc
\dim_gsub:Nn
\dim_gsub:cn
\dim_gsub:cc

(End of definition for \dim_add:Nn and others. These functions are documented on page 224.)

\dim_abs:n Functions for min, max, and absolute value with only one evaluation. The absolute value
\dim_abs:N is evaluated by removing a leading - if present.
\__dim_abs:N \dim_max:nn
\dim_min:nn
\__dim_maxmin:wwN

945
With dimension expressions, something like 10 pt * ( 5 pt / 10 pt ) does not work. Instead, the ratio part needs to be converted to an integer expression. Using \int_value:w forces everything into sp, avoiding any decimal parts.

\cs_new:Npn \dim_ratio:nn #1#2 { \__dim_ratio:n {#1} / \__dim_ratio:n {#2} }
\cs_new:Npn \__dim_ratio:n #1 { \int_value:w \__dim_eval:w (#1) \__dim_eval_end: }

(End of definition for \dim_abs:n and others. These functions are documented on page 224.)

64.6 Dimension expression conditionals

Simple comparison.

\prg_new_conditional:Npnn \dim_compare:nNnTF { p, T, F, TF } { \if_dim:w \__dim_eval:w #1 #2 \__dim_eval:w #3 \__dim_eval_end: \prg_return_true: \else: \prg_return_false: \fi: }

(End of definition for \dim_compare:nNnTF. This function is documented on page 225.)

This code is adapted from the \int_compare:nTF function. First make sure that there is at least one relation operator, by evaluating a dimension expression with a trailing \__dim_compare_error:. Just like for integers, the looping auxiliary \__dim_compare:w closes a primitive conditional and opens a new one. It is actually easier to
grab a dimension operand than an integer one, because once evaluated, dimensions all
end with pt (with category other). Thus we do not need specific auxiliaries for the three
“simple” relations <, =, and >.

\begin{verbatim}
\prg_new_conditional:Npnn \dim_compare:n #1 { p , T , F , TF }
\prg_return_false: \prg_return_true: \fi: \fi:
\__dim_compare_end:nw \__dim_compare:error:
}(End of definition for \dim_compare:TF and others. This function is documented on page 226.)
\end{verbatim}

For dimension cases, the first task to fully expand the check condition. The over all idea
is then much the same as for \str_case:nnTF as described in l3basics.
\begin{verbatim}
{ \exp:w \exp_args:Nf \__dim_case:nnTF { \dim_eval:n {#1} } {#2} {#3} { } }
\cs_new:Npn \dim_case:nnF #1#2
{ \exp:w \exp_args:Nf \__dim_case:nnTF { \dim_eval:n {#1} } {#2} { } }
\cs_new:Npn \dim_case:nn #1#2
{ \exp:w \exp_args:Nf \__dim_case:nnTF { \dim_eval:n {#1} } {#2} { } { } }
\cs_new:Npn \__dim_case:nnTF #1#2#3#4
{ \__dim_case:nw {#1} #2 {#1} { } \s__dim_mark {#3} \s__dim_mark {#4} \s__dim_stop }
\cs_new:Npn \__dim_case:nw #1#2#3
{ \dim_compare:nNnTF {#1} = {#2}
{ \__dim_case_end:nw {#3} }
{ \__dim_case:nw {#1} } }
\cs_new:Npn \__dim_case_end:nw #1#2#3 \s__dim_mark #4#5 \s__dim_stop
{ \exp_end: #1 #4 }
\end{verbatim}

(End of definition for \dim_case:nnTF and others. This function is documented on page 227.)

### 64.7 Dimension expression loops

\begin{verbatim}
\dim_while_do:nn \dim_until_do:nn \dim_do_while:nn \dim_do_until:nn
\end{verbatim}

while\_do and do\_while functions for dimensions. Same as for the int type only the names have changed.

\begin{verbatim}
\cs_new:Npn \dim_while_do:nn #1#2
{ \dim_compare:nT {#1}
{ \dim_while_do:nn {#1} {#2} }
}
\cs_new:Npn \dim_until_do:nn #1#2
{ \dim_compare:nF {#1}
{ \dim_until_do:nn {#1} {#2} }
}
\cs_new:Npn \dim_do_while:nn #1#2
{ \dim_compare:nT {#1}
{ \dim_do_while:nn {#1} {#2} }
}
\cs_new:Npn \dim_do_until:nn #1#2
{ \dim_compare:nF {#1}
{ \dim_do_until:nn {#1} {#2} }
}
\end{verbatim}
\texttt{\textbackslash cs\_new:Npn \textbackslash dim\_do\_until:n} \texttt{#1#2}
\begin{verbatim}
  \exp_after:wN \__dim\_step:wwwN
  \tex\_the:D \__dim\_eval:w #1 \exp_after:wN ;
\end{verbatim}

\texttt{\textbackslash dim\_while\_do:n} \texttt{\textbackslash dim\_until\_do:n} \texttt{\textbackslash dim\_do\_while:n} \texttt{\textbackslash dim\_do\_until:n} \texttt{while\_do and do\_while functions for dimensions. Same as for the int type only the names have changed.}
\begin{verbatim}
\cs_new:Npn \dim\_while\_do:nNnn #1#2#3#4
  \exp_after:wN \__dim\_step:wwwN
  \tex\_the:D \__dim\_eval:w #1 \exp_after:wN ;
\end{verbatim}

\texttt{\textbackslash dim\_do\_while:n} \texttt{\textbackslash dim\_do\_until:n} \texttt{Dimension step functions}

Before all else, evaluate the initial value, step, and final value. Repeating a function by steps first needs a check on the direction of the steps. After that, do the function for the start value then step and loop around. It would be more symmetrical to test for a step size of zero before checking the sign, but we optimize for the most frequent case (positive step).
\begin{verbatim}
\cs_new:Npn \dim\_step\_function:nnnN #1#2#3
  \exp_after:wN \__dim\_step:wwwN
  \tex\_the:D \__dim\_eval:w #1 \exp_after:wN ;
\end{verbatim}

(End of definition for \texttt{\textbackslash dim\_while\_do:n} \texttt{\textbackslash dim\_until\_do:n} \texttt{\textbackslash dim\_do\_while:n} \texttt{\textbackslash dim\_do\_until:n} and others. These functions are documented on page 228.)
The approach here is to build a function, with a global integer required to make
the nesting safe (as seen in other in line functions), and map that function using
\text{\textbackslash dim\_step\_function:nnnN}. We put a \texttt{prg\_break\_point:Nn} so that \texttt{map\_break} functions
from other modules correctly decrement \texttt{\g__kernel\_prg\_map\_int} before looking for
their own break point. The first argument is \texttt{\scan\_stop:}, so that no breaking function
recognizes this break point as its own.

\begin{verbatim}
\begin{Verbatim}
\# dim_step_inline:nnnn
\# dim_step_variable:nnnnN
\# dim_step:Nnnnn

(End of definition for \texttt{\dim\_step\_function:nnnN}, \texttt{\dim\_step\_wwwnN}, and \texttt{\dim\_step\_NnnnnN}. This
function is documented on page 228.)
\end{Verbatim}
\end{verbatim}
64.9 Using \texttt{dim} expressions and variables

\texttt{dim_eval:n} Evaluating a dimension expression expandably.

\begin{verbatim}
\cs_new:Npn \dim_eval:n #1
{ \dim_use:N \__dim_eval:w #1 \__dim_eval_end: }
\end{verbatim}

(End of definition for \texttt{dim_eval:n}. This function is documented on page 229.)

\texttt{dim_sign:n} See \texttt{dim_abs:n}. Contrarily to \texttt{int_sign:n} the case of a zero dimension cannot be distinguished from a positive dimension by looking only at the first character, since 0.2pt and 0pt start the same way. We need explicit comparisons. We start by distinguishing the most common case of a positive dimension.

\begin{verbatim}
\cs_new:Npn \dim_sign:n #1
{ \int_value:w \exp_after:wN \__dim_sign:Nw \dim_use:N \__dim_eval:w #1 \__dim_eval_end: ; \exp_stop_f: }
\end{verbatim}

(End of definition for \texttt{dim_sign:n} and \texttt{__dim_sign:Nw}. This function is documented on page 229.)

\texttt{dim_use:N} Accessing a \texttt{(dim)}. We hand-code the \texttt{c} variant for some speed gain.

\begin{verbatim}
\cs_new_eq:NN \dim_use:N \tex_the:D
\cs_new:Npn \dim_use:c #1 { \tex_the:D \cs:w #1 \cs_end: }
\end{verbatim}

(End of definition for \texttt{dim_use:N}. This function is documented on page 229.)

\texttt{dim_to_decimal:n} A function which comes up often enough to deserve a place in the kernel. Evaluate the dimension expression \#1 then remove the trailing \texttt{pt}. When debugging is enabled, the argument is put in parentheses as this prevents the dimension expression from terminating early and leaving extra tokens lying around. This is used a lot by low-level manipulations.
64.10 Conversion of dim to other units

The conversion from pt or sp to other units is complicated by the fact that \TeX’s conversion to sp involves rounding and hard-coded ratios. In order to give re-entrant outcomes, we therefore need to do quite a bit of work: see https://github.com/latex3/latex3/issues/954 for detailed discussion. After dealing with the trivial case, we therefore have some work to do. The code to do this is contributed by Ruixi Zhang.

\dim_to_decimal_in_sp:n

The one easy case: the only requirement here is that we avoid an overflow.

\dim_to_decimal_in_bp:n
\dim_to_decimal_in_cc:n
\dim_to_decimal_in_dd:n
\dim_to_decimal_in_mm:n
\dim_to_decimal_in_pc:n
\dim_to_decimal_aux:w

We first set up a helper macro \_dim_tmp:w which takes two arguments. The first argument is one of the following engine-defined units: in, pc, cm, mm, bp, dd, cc, nd, and nc. The second argument is $\delta^{-1}$ in reduced fraction, where $\delta > 1$ is the engine-defined conversion factor for each unit. Note that $\delta$ must be strictly larger than 1 for the following algorithm to work.

Here is how the algorithm works: Suppose that a user inputs a non-negative dimension in a unit that has conversion factor $\delta > 1$. Then this dimension is internally represented as $X \text{ sp}$, where $X = \lfloor N \delta \rfloor$ for some integer $N \geq 0$. We then seek a formula to express this $N$ using $X$. The \dim_to_decimal_in_<unit>:n functions shall return the number $N/2^{16}$ in decimal. This way, we guarantee the returned decimal followed by the original unit will parse to exactly $X \text{ sp}$.

So how do we get $N$ from $X$? Well, since $X = \lfloor N \delta \rfloor$, we have $X \leq N \delta < X + 1$ and $X \delta^{-1} \leq N < (X + 1) \delta^{-1}$. Let’s focus on the midpoint of this bounding interval for $N$. The midpoint is $(X + \frac{1}{2}) \delta^{-1}$. The fact $\delta > 1$ implies that the bounding interval is shorter than 1 in length. Thus, (1) midpoint $+ \frac{1}{2} > N$ and (2) midpoint $+ \frac{1}{2} < N + 1$. In other words, $N = \lfloor \text{midpoint} + \frac{1}{2} \rfloor$. As long as we can rewrite the midpoint as the result of
a “scaling operation” of $\varepsilon$-\TeX, the $\lfloor \ldots + \frac{1}{2} \rfloor$ part will follow naturally. Indeed we can:

\[
\text{midpoint} = (2X + 1) \times (\frac{1}{2} \delta^{-1}).
\]

Addendum: If $\delta \geq 2$, then the bounding interval for $N$ is at most $\frac{1}{2}$ wide in length. In this case, the leftpoint $X\delta^{-1}$ suffices as $N = |X\delta^{-1} + \frac{1}{2}|$. Six out of the nine units listed above can be handled in this way, which is much simpler than using midpoint. But three remaining units have $1 < \delta < 2$; they are bp ($\delta = 7227/7200$), nd ($\delta = 685/642$), and dd ($\delta = 1238/1157$), and these three must be handled using midpoint. For consistency, we shall use the midpoint approach for all nine units.

\begin{verbatim}
\group_begin:
\cs_set_protected:Npn \__dim_tmp:w #1#2
  { \cs_new:cpn { dim_to_decimal_in_ #1 :n } ##1
    { \exp_after:wN \__dim_to_decimal_aux:w \int_value:w \__dim_eval:w ##1 \__dim_eval_end: ; #2 ; }
  }
\__dim_tmp:w { in } { 50 / 7227 } % delta = 7227/100
\__dim_tmp:w { pc } { 1 / 24 } % delta = 12/1
\__dim_tmp:w { cm } { 127 / 7227 } % delta = 7227/254
\__dim_tmp:w { mm } { 1270 / 7227 } % delta = 7227/2540
\__dim_tmp:w { bp } { 400 / 803 } % delta = 7227/7200
\__dim_tmp:w { dd } { 1157 / 2476 } % delta = 1238/1157
\__dim_tmp:w { cc } { 1157 / 29712 } % delta = 14856/1157
\group_end:
\end{verbatim}

Conversions to other units are now coded. Consult the pdf\TeX source for each conversion factor $\delta$. Each factor $\frac{1}{2} \delta^{-1}$ is hand-coded for accuracy (and speed). As the units nc and nd are not supported by X\TeX or (u)p\TeX, they are not included here.

\begin{verbatim}
\__dim_tmp:w { in } { 50 / 7227 } % delta = 7227/100
\__dim_tmp:w { pc } { 1 / 24 } % delta = 12/1
\__dim_tmp:w { cm } { 127 / 7227 } % delta = 7227/254
\__dim_tmp:w { mm } { 1270 / 7227 } % delta = 7227/2540
\__dim_tmp:w { bp } { 400 / 803 } % delta = 7227/7200
\__dim_tmp:w { dd } { 1157 / 2476 } % delta = 1238/1157
\__dim_tmp:w { cc } { 1157 / 29712 } % delta = 14856/1157
\group_end:
\end{verbatim}

The tokens after \__dim_to_decimal_aux:w shall have the following form: <number>; half of delta inverse; sp, where <number> represents the input dimension in sp unit. If $<$number$>$ is positive, then $\#1$ is its leading digit and $\#2$ (possibly empty) is all the remaining digits; If $<$number$>$ is zero, then $\#1$ is 0 and $\#2$ is empty; If $<$number$>$ is negative, then $\#1$ is its sign $- 1$ and $\#2$ is all its digits. In all three cases, $\#1$ is the original $<$number$>$. We can use $\#1$ to decide whether to use the $-1$ formula or the $+1$ formula.

\begin{verbatim}
\cs_new:Npn \__dim_to_decimal:n
  { \dim_to_decimal:n \#1\#2 ; \#3 ; }
\end{verbatim}

We need different formulae depending on whether the user input dimension is negative or not. For negative dimension (internally represented as X sp), the formula is $(2X - 1) \times (\frac{1}{2} \delta^{-1})$. For non-negative dimension, the formula is $(2X + 1) \times (\frac{1}{2} \delta^{-1})$. The intermediate step doubles the dimension $X$. To avoid overflow, we must invoke \int_eval:n.

\begin{verbatim}
\int_eval:n
  { \int_eval:n \#1\#2 \if:w #1 - - \else: + \fi: 1 ) * \#3 }
\end{verbatim}

Now we append sp to finish the dimension specification.

\begin{verbatim}
sp
}
\end{verbatim}

(End of definition for \dim_to_decimal_in_bp:n and others. These functions are documented on page 230.)

953
\newcommand{\dim_to_decimal_in_unit}{nn}

\begin{verbatim}
\cs_new:Npn \dim_to_decimal_in_unit:nn #1#2
\begin{Verbatim}
\exp_after:wN \__dim_chk_unit:w
\int_value:w \__dim_eval:w #2 \__dim_eval_end: ; {#1}
\end{Verbatim}
\end{verbatim}

(End of definition for \dim_to_decimal_in_unit:nn. This function is documented on page 231.)

\__dim_chk_unit:w

The tokens after \__dim_chk_unit:w shall have the following form: \texttt{<number2>;\{<dimexpr1>\}}, where \texttt{<number2>} represents \texttt{<dimexpr2>} in \texttt{sp} unit. If \texttt{#1} is 012, the “unit” \texttt{<dimexpr2>} must also be zero. So we throw out a “division by zero” error message at this point. Otherwise, if \texttt{#1} is -12, we shall negate both \texttt{<dimexpr1>} and \texttt{<dimexpr2>} for later procedures.

\begin{verbatim}
\cs_new:Npn \__dim_chk_unit:w #1#2;#3
\begin{Verbatim}
\token_if_eq_charcode:NNTF #1 0 { \msg_expandable_error:nn { dim } { zero-unit } }
{ \exp_after:wN \__dim_branch_unit:w
\int_value:w \if:w #1 - - \fi: \__dim_eval:w #3 \exp_after:wN ;
\int_value:w \if:w #1 - - \fi: #1#2 ;
\end{Verbatim}
\end{verbatim}

(End of definition for \__dim_chk_unit:w.)

\__dim_branch_unit:w

The tokens after \__dim_branch_unit:w shall have the following form: \texttt{<number1>;\{<number2>\};}, where \texttt{<number1>} represents \texttt{<dimexpr1>} in \texttt{sp} unit (whose sign is taken care of) and \texttt{<number2>} represents the absolute value of \texttt{<dimexpr2>} in \texttt{sp} unit (which is strictly positive).

As explained, the formulae \((2X\pm\delta)\times(\frac{1}{2}\delta^{-1})\) work if and only if \(\delta = \frac{<\texttt{number2}>}{65536} > 1\). This corresponds to \texttt{<dimexpr2>} strictly larger than 1 pt in absolute value. In this case, we simply call \__dim_to_decimal_aux:w and supply \(\frac{1}{2}\delta^{-1} = 32768/<\texttt{number2}>\) as \texttt{<half of delta inverse>}. Otherwise if \texttt{<number2>} = 65536, then \texttt{<dimexpr2>} is 1 pt in absolute value and we call \dim_to_decimal:n directly.

Otherwise \(0 < <\texttt{number2}> < 65536\) and we shall proceed differently.

For unit less than 1 pt, write \(n = <\texttt{number2}>\), then \(\delta = n/65536 < 1\). The midpoint formulae are not optimal. Let’s go back to the inequalities \(X\delta^{-1} \leq N < (X + 1)\delta^{-1}\). Since now \(\delta < 1\), the bounding interval is wider than 1 in length. Consider the ceiling integer \(M = \lceil X\delta^{-1}\rceil\), then \(X\delta^{-1} \leq M < (X + 1)\delta^{-1}\), or equivalently \(X \leq M\delta < X + 1\), and thus \(\lceil M\delta \rceil = X\). The key point here is that we don’t need to solve for \(N\); in fact, any integer that can reproduce \(X\) (such as \(M\)) is good enough. So the algorithm goes like this: (1) Compute rounding of \(X\delta^{-1}\), i.e., \(M' = X\delta^{-1} + \frac{1}{2}\); this \(M'\) could be either \(M\) or \(M - 1\). (2) Check if \(\lceil M'\delta \rceil = X\), i.e., whether our candidate \(M'\) can reproduce \(X\). If so, then this \(M'\) is good enough; if not, then we add one to \(M'\).

But when \(0 < n < 65536\), we cannot delay the problem of overflow any more. For \(X\delta^{-1} = X \times 65536/n\), where \(X\) can go up to \(2^{30} - 1\) and \(n\) can be as small as 1, the result is well over \(2^{31} - 1\) (largest integer allowed within \numexpr). For example, \dim_to_decimal_in_unit:nn { \maxdimen } { 1sp }. Here, all inputs are legal, so we should be able to output 1073741823 without causing arithmetic overflow.

954
As a workaround, let’s write $X = qn + r$ with some $q \geq 0$ and $0 \leq r < n$. Then $X^{-1} = 65536q + 65536r/n$, and so $M' = 65536q + \lfloor 65536r/n + \frac{1}{2}\rfloor = 65536q + R'$. Computing $R'$ will never overflow. If this $R'$ can reproduce $r$, then it is good enough; otherwise we add one to $R'/65536$ in decimal.

Note: $q = \lfloor X/n \rfloor = \lfloor 2X - n/2 \rfloor$ represents the “integer” part, while $0 \leq R' \leq 65536$ represents the “fractional” part. (Can $R'$ be 65536 really happen? Didn’t investigate.)
\__dim_get_remainder:w This is trivial. We compute $R' = \lfloor 65536r/n + \frac{1}{2} \rfloor$, then leave \__dim_test_candidate:w followed by $R';r;n;q;<\text{sign of } X>;$.

\__dim_convert_remainder:w This is trivial. We compute $R' = \lfloor 65536r/n + \frac{1}{2} \rfloor$, then leave \__dim_test_candidate:w followed by $R';r;n;q;<\text{sign of } X>;$.

\__dim_test_candidate:w Now the fun part: We take $R'$, $r$ and $n$ to test whether $r = \lfloor R'\delta \rfloor$. This is done as a dimension comparison. The left-hand side, $r$, is simply $r$ \text{sp}. The right-hand side, $\lfloor R'\delta \rfloor$, is exactly $<R' \text{ as decimal}>\text{dimen = n sp}$. If the result is true, then we've found $R'$; otherwise we add one to $R'$. After this step, $r$ and $n$ are no longer needed. We should then have \__dim_parse_decimal:w followed by $R';q;<\text{sign of } X>;$.

\__dim_parse_decimal:w The Grand Finale: We sum $q$ and $R'/65536$ together, and negate the result if necessary. These are all done expandably. If $0 < R'/65536 < 1$, the integer summation is naturally terminated at the decimal point. If $R'/65536 = 0$ (or 1?), the summation is terminated at the semicolon. The auxiliary function \_dim_parse_decimal_aux:w takes care of both cases.
64.11 Viewing dim variables

\dim_show:N Diagnostics.
\dim_show:c
\cs_new_eq:NN \dim_show:N \__kernel_register_show:N
\cs_generate_variant:Nn \dim_show:N { c }
(End of definition for \dim_show:N. This function is documented on page 231.)

\dim_show:n Diagnostics. We don’t use the TeX primitive \showthe to show dimension expressions: this gives a more unified output.
\cs_new_protected:Npn \dim_show:n { \__kernel_msg_show_eval:Nn \dim_eval:n }
(End of definition for \dim_show:n. This function is documented on page 231.)

\dim_log:N \dim_log:c \dim_log:n Diagnostics. Redirect output of \dim_show:n to the log.
\cs_new_eq:NN \dim_log:N \__kernel_register_log:N
\cs_new_eq:NN \dim_log:c \__kernel_register_log:c
\cs_new_protected:Npn \dim_log:n { \__kernel_msg_log_eval:Nn \dim_eval:n }
(End of definition for \dim_log:N and \dim_log:n. These functions are documented on page 231.)

64.12 Constant dimensions

\c_zero_dim \c_max_dim Constant dimensions.
\dim_const:Nn \c_zero_dim { 0 pt }
\dim_const:Nn \c_max_dim { 16383.99999 pt }
(End of definition for \c_zero_dim and \c_max_dim. These variables are documented on page 232.)

64.13 Scratch dimensions

\l_tmpa_dim \l_tmpb_dim \g_tmpa_dim \g_tmpb_dim We provide two local and two global scratch registers, maybe we need more or less.
\dim_new:N \l_tmpa_dim
\dim_new:N \l_tmpb_dim
\dim_new:N \g_tmpa_dim
\dim_new:N \g_tmpb_dim
(End of definition for \l_tmpa_dim and others. These variables are documented on page 232.)

64.14 Creating and initialising skip variables

\s__skip_stop Internal scan marks.
\scan_new:N \s__skip_stop
(End of definition for \s__skip_stop.)
Allocation of a new internal registers.

Contrarily to integer constants, we cannot avoid using a register, even for constants. See \skip_zero:N and \skip_gzero:N. These functions are documented on page 292.

Reset the register to zero.

Create a register if needed, otherwise clear it.

Copies of the cs functions defined in l3basics.

(End of definition for \skip_zero:N and \skip_gzero:N. These functions are documented on page 292.)
64.15 Setting skip variables

\skip_set:Nn \skip_set:cn \skip_gset:Nn \skip_gset:cn

Much the same as for dimensions.

\cs_new_protected:Npn \skip_set:Nn \skip_set:cn \skip_gset:Nn \skip_gset:cn
21207 \{ \#1 = \tex_glueexpr:D \#2 \scan_stop: \}
21210 \{ \tex_global:D \#1 = \tex_glueexpr:D \#2 \scan_stop: \}
21211 \cs_generate_variant:Nn \skip_set:Nn \{ c \}
21212 \cs_generate_variant:Nn \skip_gset:Nn \{ c \}

(End of definition for \skip_set:Nn and \skip_gset:Nn. These functions are documented on page 233.)

\skip_set_eq:NN \skip_set_eq:cN \skip_set_eq:Nc \skip_set_eq:cc
\skip_gset_eq:NN \skip_gset_eq:cN \skip_gset_eq:Nc \skip_gset_eq:cc

All straightforward.

\cs_new_protected:Npn \skip_set_eq:NN \skip_set_eq:cN \skip_set_eq:Nc \skip_set_eq:cc
21213 \{ \#1 = \#2 \}
21214 \cs_generate_variant:Nn \skip_set_eq:NN \{ c , Nc , cc \}
21215 \cs_new_protected:Npn \skip_gset_eq:NN \skip_gset_eq:cN \skip_gset_eq:Nc \skip_gset_eq:cc
21216 \{ \tex_global:D \#1 = \#2 \}
21217 \cs_generate_variant:Nn \skip_gset_eq:NN \{ c , Nc , cc \}

(End of definition for \skip_set_eq:NN and \skip_gset_eq:NN. These functions are documented on page 233.)

\skip_add:Nn \skip_add:cn \skip_gadd:Nn \skip_gadd:cn
\skip_sub:Nn \skip_sub:cn \skip_gsub:Nn \skip_gsub:cn

Using by here deals with the (incorrect) case \skip123.

\cs_new_protected:Npn \skip_add:Nn \skip_add:cn \skip_gadd:Nn \skip_gadd:cn
21217 \{ \tex_advance:D \#1 \tex_glueexpr:D \#2 \scan_stop: \}
21218 \{ \tex_global:D \tex_advance:D \#1 \tex_glueexpr:D \#2 \scan_stop: \}
21219 \cs_generate_variant:Nn \skip_add:Nn \{ c \}
21220 \cs_generate_variant:Nn \skip_gadd:Nn \{ c \}

(End of definition for \skip_add:Nn and others. These functions are documented on page 233.)

64.16 Skip expression conditionals

\skip_if_eq_p:nn \skip_if_eq:nn \skip_if_eq:nnTF

Comparing skips means doing two expansions to make strings, and then testing them.

As a result, only equality is tested.

\prg_new_conditional:Npnn \skip_if_eq:nn \skip_if_eq:nnTF
21229 \{ \str_if_eq:eeTF \{ \skip_eval:n \#1 \} \{ \skip_eval:n \#2 \}
21322 \{ \prg_return_true: \}
21323 \{ \prg_return_false: \}

(End of definition for \skip_if_eq:nnTF. This function is documented on page 234.)
With \TeX, we have an easy access to the order of infinities of the stretch and shrink components of a skip. However, to access both, we either need to evaluate the expression twice, or evaluate it, then call an auxiliary to extract both pieces of information from the result. Since we are going to need an auxiliary anyways, it is quicker to make it search for the string \texttt{fil} which characterizes infinite glue.

\begin{verbatim}
\cs_set_protected:Npn \__skip_tmp:w #1
{ \prg_new_conditional:Npn \skip_if_finite:n \#1 \p, \T, \F, \TF 
{ \exp_after:wN \__skip_if_finite:wwNw \skip_use:N \tex_glueexpr:D \#1 ; \prg_return_false: 
\#1 ; \prg_return_true: \s__skip_stop 
}\}
\exp_args:No \__skip_tmp:w \tl_to_str:n \{ \texttt{fil} \}
\end{verbatim}

(End of definition for \texttt{\skip_if_finite:nTF} and \texttt{\__skip_if_finite:wwNw}. This function is documented on page 234.)

\section{Using skip expressions and variables}

\subsection{Evaluating a skip expression expandably}

\begin{verbatim}
\cs_new:Npn \skip_eval:n #1
{ \skip_use:N \tex_glueexpr:D #1 \scan_stop: }
\end{verbatim}

(End of definition for \texttt{\skip_eval:n}. This function is documented on page 234.)

\subsection{Accessing a \langle skip\rangle}

\begin{verbatim}
\cs_new_eq:NN \skip_use:N \dim_use:N
\cs_new_eq:NN \skip_use:c \dim_use:c
\end{verbatim}

(End of definition for \texttt{\skip_use:N}. This function is documented on page 234.)

\section{Inserting skips into the output}

\begin{verbatim}
\cs_new_eq:NN \skip_horizontal:N \tex_hskip:D
\cs_new:Npn \skip_horizontal:n #1
{ \skip_horizontal:N \tex_glueexpr:D #1 \scan_stop: }
\cs_generate_variant:Nn \skip_horizontal:N { c }
\end{verbatim}

(End of definition for \texttt{\skip_horizontal:N} and others. These functions are documented on page 235.)
64.19 Viewing skip variables

\skip_show:N \skip_show:c
\cs_new_eq:NN \skip_show:N \__kernel_register_show:N
\cs_generate_variant:Nn \skip_show:N { c }

(End of definition for \skip_show:N. This function is documented on page 234.)

\skip_show:n
\cs_new_protected:Npm \skip_show:n
\{ \__kernel_msg_show_eval:Nn \skip_eval:n \}

(End of definition for \skip_show:n. This function is documented on page 234.)

\skip_log:N \skip_log:c \skip_log:n
\cs_new_eq:NN \skip_log:N \__kernel_register_log:N
\cs_new_eq:NN \skip_log:c \__kernel_register_log:c
\cs_new_protected:Npn \skip_log:n
\{ \__kernel_msg_log_eval:Nn \skip_eval:n \}

(End of definition for \skip_log:N and \skip_log:n. These functions are documented on page 235.)

64.20 Constant skips

\c_zero_skip \c_max_skip
\skip_const:Nn \c_zero_skip { \c_zero_dim }
\skip_const:Nn \c_max_skip { \c_max_dim }

(End of definition for \c_zero_skip and \c_max_skip. These functions are documented on page 235.)

64.21 Scratch skips

\l_tmpa_skip \l_tmpb_skip \g_tmpa_skip \g_tmpb_skip
\skip_new:N \l_tmpa_skip
\skip_new:N \l_tmpb_skip
\skip_new:N \g_tmpa_skip
\skip_new:N \g_tmpb_skip

(End of definition for \l_tmpa_skip and others. These variables are documented on page 235.)

64.22 Creating and initialising muskip variables

\muskip_new:N \muskip_new:c
\cs_new_protected:Npm \muskip_new:N \#1
\{ \__kernel_chk_if_free_cs:N \#1 \cs:w newmuskip \cs_end: \#1 \}
\cs_generate_variant:NN \muskip_new:N \{ c \}

(End of definition for \muskip_new:N. This function is documented on page 236.)
\muskip_const:Nn \skip_const:Nn
See \skip_const:Nn.

\muskip_zero:N \muskip_zero:c \muskip_gzero:N \muskip_gzero:c
Reset the register to zero.

\muskip_zero_new:N \muskip_zero_new:c \muskip_gzero_new:N \muskip_gzero_new:c
Create a register if needed, otherwise clear it.

\muskip_if_exist_p:N \muskip_if_exist_p:c \muskip_if_exist:N \muskip_if_exist:c
Copies of the \cs functions defined in l3basics.

\muskip_set:Nn \muskip_set:cn \muskip_gset:Nn \muskip_gset:cn
This should be pretty familiar.

64.23 Setting \muskip variables

(End of definition for \muskip const:N. This function is documented on page 236.)

(End of definition for \muskip_zero:N and \muskip_gzero:N. These functions are documented on page 236.)

(End of definition for \muskip_zero_new:N and \muskip_gzero_new:N. These functions are documented on page 236.)

(End of definition for \muskip_if_exist:NTF. This function is documented on page 236.)

(End of definition for \muskip_set:N and \muskip_gset:N. These functions are documented on page 237.)
All straightforward.

\cs_new_protected:Npn \muskip_set_eq:NN #1#2 { #1 = #2 }
\cs_generate_variant:Nn \muskip_set_eq:NN { c , Nc , cc }
\cs_new_protected:Npn \muskip_gset_eq:NN #1#2 { \tex_global:D #1 = #2 }
\cs_generate_variant:Nn \muskip_gset_eq:NN { c , Nc , cc }

(End of definition for \muskip_set_eq:NN and \muskip_gset_eq:NN. These functions are documented on page 237.)

Using by here deals with the (incorrect) case \muskip123.
\cs_new_protected:Npn \muskip_add:Nn #1#2 {
\tex_advance:D #1 \tex_muexpr:D #2 \scan_stop: }
\cs_new_protected:Npn \muskip_gadd:Nn #1#2 {
\tex_global:D \tex_advance:D #1 \tex_muexpr:D #2 \scan_stop: }
\cs_generate_variant:Nn \muskip_add:Nn { c }
\cs_generate_variant:Nn \muskip_gadd:Nn { c }
\cs_new_protected:Npn \muskip_sub:Nn #1#2 {
\tex_advance:D #1 - \tex_muexpr:D #2 \scan_stop: }
\cs_new_protected:Npn \muskip_gsub:Nn #1#2 {
\tex_global:D \tex_advance:D #1 - \tex_muexpr:D #2 \scan_stop: }
\cs_generate_variant:Nn \muskip_sub:Nn { c }
\cs_generate_variant:Nn \muskip_gsub:Nn { c }

(End of definition for \muskip_add:Nn and others. These functions are documented on page 236.)

\muskip_eval:n Evaluating a muskip expression expandably.
\cs_new:Npn \muskip_eval:n #1 {
\muskip_use:N \tex_muexpr:D #1 \scan_stop: }

(End of definition for \muskip_eval:n. This function is documented on page 237.)

Accessing a ⟨muskip⟩.
\cs_new_eq:NN \muskip_use:N \dim_use:N
\cs_new_eq:NN \muskip_use:c \dim_use:c

(End of definition for \muskip_use:N. This function is documented on page 237.)

\muskip_show:N Diagnostics.
\cs_new_eq:NN \muskip_show:N \__kernel_register_show:N
\cs_generate_variant:Nn \muskip_show:N { c }

(End of definition for \muskip_show:N. This function is documented on page 237.)

\muskip_show:n Diagnostics. We don’t use the \TeX{} primitive \showthe to show muskip expressions: this gives a more unified output.
\cs_new_protected:Npn \muskip_show:n {
\__kernel_mag_show_eval:N \muskip_eval:n }

(End of definition for \muskip_show:n. This function is documented on page 238.)
Diagnostics. Redirect output of \musrshow to the log.

\muskip_log:N \muskip_log:c \muskip_log:n
\cs_new_eq:NN \muskip_log:N \__kernel_register_log:N
\cs_new_eq:NN \muskip_log:c \__kernel_register_log:c
\cs_new_protected:Npn \muskip_log:n
{ \__kernel_msg_log_eval:Nn \muskip_eval:n }

(End of definition for \muskip_log:N and \muskip_log:n. These functions are documented on page 238.)

### 64.26 Constant muskips

\c_zero_muskip \c_max_muskip
\muskip_const:Nn \c_zero_muskip { 0 mu }
\muskip_const:Nn \c_max_muskip { 16383.99999 mu }

(End of definition for \c_zero_muskip and \c_max_muskip. These functions are documented on page 238.)

### 64.27 Scratch muskips

\l_tmpa_muskip \l_tmpb_muskip \g_tmpa_muskip \g_tmpb_muskip
\muskip_new:N \l_tmpa_muskip
\muskip_new:N \l_tmpb_muskip
\muskip_new:N \g_tmpa_muskip
\muskip_new:N \g_tmpb_muskip

(End of definition for \l_tmpa_muskip and others. These variables are documented on page 238.)
Chapter 65

l3keys implementation

65.1 Low-level interface

The low-level key parser’s implementation is based heavily on expkv. Compared to keyval it adds a number of additional “safety” requirements and allows to process the parsed list of key–value pairs in a variety of ways. The net result is that this code needs around one and a half the amount of time as keyval to parse the same list of keys. To optimise speed as far as reasonably practical, a number of lower-level approaches are taken rather than using the higher-level expl3 interfaces.

The general behavior of the l3keys module is to throw an error on blank key names. However to support the usage of \keyval_parse:nnn in the l3prop module we allow this error to be switched off temporarily and just ignore blank names.

This temporary macro will be used since some of the definitions will need an active comma or equals sign. Inside of this macro #1 will be the active comma and #2 will be the active equals sign.

The main function starts the first of two loops. The outer loop splits the key–value list at active commas, the inner loop will do so at other commas. The use of \s__keyval_mark here prevents loss of braces from the key argument.
\cs_new:Npn \keyval_parse:nnn ##1 ##2 ##3
{
  \__kernel_exp_not:w \tex_expanded:D
  {
    \__keyval_loop_active:nnw {##1} {##2}
    \s__keyval_mark ##3 #1 \s__keyval_tail #1
  }
}
\cs_new_eq:NN \keyval_parse:NNn \keyval_parse:nnn

(End of definition for \keyval_parse:nnn and \keyval_parse:NNn. These functions are documented on page 253.)

\__keyval_loop_active:nnw First a fast test for the end of the loop is done, it'll gobble everything up to a \s__keyval_tail. The loop ending macro will gobble everything to the last comma in this definition. If the end isn't reached yet, start the second loop splitting at other commas, the next iteration of this first loop will be inserted by the end of \__keyval_loop_other:nnw.
\cs_new:Npn \__keyval_loop_active:nnw ##1 ##2 ##3 #1
{
  \__keyval_if_recursion_tail:w ##3
  \__keyval_end_loop_active:w \s__keyval_tail
  \__keyval_loop_other:nnw {##1} {##2} ##3 , \s__keyval_tail ,
  \s__keyval_nil
}
(End of definition for \__keyval_loop_active:nnw.)

\__keyval_split_other:w These two macros allow to split at the first equals sign of category 12 or 13. At the same time they also execute branching by inserting the first token following \s__keyval_mark that followed the equals sign. Hence they also test for the presence of such an equals sign simultaneously.
\cs_new:Npn \__keyval_split_other:w ##1 = ##2 \s__keyval_mark ##3
{ \s__keyval_stop \s__keyval_mark ##2 }
\cs_new:Npn \__keyval_split_active:w ##1 #2 ##2 \s__keyval_mark ##3
{ \s__keyval_stop \s__keyval_mark ##2 }
(End of definition for \__keyval_split_other:w and \__keyval_split_active:w.)

\__keyval_loop_other:nnw The second loop uses the same test for its end as the first loop, next it splits at the first active equals sign using \__keyval_split_active:w. The \s__keyval_nil prevents accidental brace stripping and acts as a delimiter in the next steps. First testing for an active equals sign will reduce the number of necessary expansion steps for the expected average use case of other equals signs and hence perform better on average.
\cs_new:Npn \__keyval_loop_other:nnw ##1 #2 ##3 ,
{
  \__keyval_if_recursion_tail:w ##3
  \__keyval_end_loop_other:w \s__keyval_tail
  \__keyval_split_active:w ##1 #2 \s__keyval_mark ##3
  \s__keyval_nil
  \s__keyval_mark \__keyval_split_active_auxi:w
  {##1} {##2}
  \s__keyval_mark
}

966
After `\__keyval_split_active:w` the following will only be called if there was at least one active equals sign in the current key–value pair. Therefore this is the execution branch for a key–value pair with an active equals sign. `#1` will be everything up to the first active equals sign. First it tests for other equals signs in the key name, which will eventually throw an error via `\__keyval_misplaced_equal_after_active_error:w`. If none was found we forward the key to `\__keyval_split_active_auxii:w`.

```latex
\cs_new:Npn \__keyval_split_active_auxi:w #1 \s__keyval_stop
\begin{verbatim}
  \__keyval_split_other:w #1 \s__keyval_nil
  \s__keyval_mark \__keyval_misplaced_equal_after_active_error:w
  = \s__keyval_mark \__keyval_split_active_auxii:w
\end{verbatim}
```

`\__keyval_split_active_auxii:w` gets the correct key name with a leading `\s__-keyval_mark` as `##1`. It has to sanitise the remainder of the previous test and trims the key name which will be forwarded to `\__keyval_split_active_auxiii:w`.

```latex
\cs_new:Npn \__keyval_split_active_auxii:w #1 \s__keyval_nil \s__keyval_mark \__keyval_misplaced_equal_in_split_error:w #2 \s__keyval_mark \__keyval_split_active_auxii:w {##1}
```

Next we test for a misplaced active equals sign in the value, if none is found `\__keyval_-_split_active_auxii:w` will be called.

```latex
\cs_new:Npn \__keyval_split_active_auxiii:w #1 #2 \s__keyval_nil
\begin{verbatim}
  \__keyval_split_active:w #2 \s__keyval_nil
  \s__keyval_mark \__keyval_misplaced_equal_in_split_error:w
  #2 \s__keyval_mark \__keyval_split_active_auxiv:w
  {#1}
\end{verbatim}
```

This runs the last test after sanitising the remainder of the previous one. This time test for a misplaced equals sign of category 12 in the value. Finally the last auxiliary macro will be called.

```latex
\cs_new:Npn \__keyval_split_active_auxiv:w #1 \s__keyval_nil \s__keyval_mark \__keyval_misplaced_equal_in_split_error:w
\begin{verbatim}
  \__keyval_split_active:w #2 \s__keyval_nil
  \s__keyval_mark \__keyval_misplaced_equal_in_split_error:w
  #2 \s__keyval_mark \__keyval_split_active_auxiv:w
  {#1}
\end{verbatim}
```

This last macro in this execution branch sanitis the last test, trims the value and passes it to `\__keyval_pair:nnnn`.

```latex
\cs_new:Npn \__keyval_split_active_auxv:w #1 \s__keyval_nil \s__keyval_mark \__keyval_misplaced_equal_in_split_error:w
\begin{verbatim}
  \__keyval_split_other:w #1 \s__keyval_nil
  \s__keyval_mark \__keyval_misplaced_equal_in_split_error:w
  = \s__keyval_mark \__keyval_split_active_auxv:w
\end{verbatim}
```

(End of definition for `\__keyval_split_active_auxi:w` and others.)
The following is the branch taken if the key–value pair doesn’t contain an active equals sign. The remainder of that test will be cleaned up by \_\_keyval_clean_up_active:w which will then split at an equals sign of category other.

\cs_new:Npn \_\_keyval_clean_up_active:w
\__\_\_keyval_clean_up_active:w
\#1 \s__keyval_nil \s__keyval_mark \_\_keyval_split_active_auxi:w \s__keyval_stop
\{
\_\_keyval_split_other:w \#1 \s__keyval_nil
\s__keyval_mark \_\_keyval_split_other_auxi:w
\s__keyval_stop \s__keyval_mark \_\_keyval_clean_up_other:w
\}

(End of definition for \_\_\_\_keyval_clean_up_active:w)

This is executed if the key–value pair doesn’t contain an active equals sign but at least one other. \#1 of \_\_\_\_keyval_split_other_auxi:w will contain the complete key name, which is trimmed and forwarded to the next auxiliary macro.

\cs_new:Npn \_\_\_\_keyval_split_other_auxi:w \#1 \s__keyval_stop
\{ \_\_\_\_keyval_trim:nN \{ \#1 \} \_\_\_\_keyval_split_other_auxii:w \}

We know that the value doesn’t contain misplaced active equals signs but we have to test for others. Also we need to sanitise the previous test, which is done here and not earlier to avoid superfluous argument grabbing.

\cs_new:Npn \_\_\_\_keyval_split_other_auxii:w \#1 \#2 \s__keyval_nil \s__keyval_mark \_\_keyval_clean_up_other:w
\{ \_\_\_\_keyval_split_other:w \#2 \s__keyval_nil
\s__keyval_mark \_\_\_\_keyval_misplaced_equal_in_split_error:w
\s__keyval_stop \s__keyval_mark
\{ \#1 \}
\}

\_\_\_\_keyval_split_other_auxiii:w sanitises the test for other equals signs, trims the value and forwards it to \_\_\_\_keyval_pair:nnnn.

\cs_new:Npn \_\_\_\_keyval_split_other_auxiii:w
\#1 \s__keyval_nil \s__keyval_mark \_\_\_\_keyval_misplaced_equal_in_split_error:w
\s__keyval_stop \s__keyval_mark
\{ \_\_\_\_keyval_trim:nN \{ \#1 \} \_\_\_\_keyval_pair:nnnn \}

(End of definition for \_\_\_\_keyval_split_other_auxi:w, \_\_\_\_keyval_split_other_auxii:w, and \_\_\_\_keyval_split_other_auxiii:w)

\_\_\_\_keyval_clean_up_other:w \_\_\_\_keyval_clean_up_other:w is the last branch that might exist. It is called if no equals sign was found, hence the only possibilities left are a blank list element, which is to be skipped, or a lonely key. If it’s no empty list element this will trim the key name and forward it to \_\_\_\_keyval_key:nn.

\cs_new:Npn \_\_\_\_keyval_clean_up_other:w
\__\_\_\_\_keyval_clean_up_other:w
\#1 \s__keyval_nil \s__keyval_mark \_\_\_\_keyval_split_other_auxi:w \s__keyval_stop \s__keyval_mark
\{ \_\_\_\_keyval_if_blank:w \#1 \s__keyval_nil \s__keyval_mark \_\_\_\_keyval_blank_true:w
\s__keyval_stop \s__keyval_mark \_\_\_\_keyval_clean_up_other:w
\s__keyval_stop \s__keyval_mark \_\_\_\_keyval_key:nn
\}

(End of definition for \_\_\_\_keyval_clean_up_other:w)

968
All these two macros do is gobble the remainder of the current other loop execution and throw an error. Afterwards they have to insert the next loop iteration.

All that's left for the parsing loops are the macros which end the recursion. Both just gobble the remaining tokens of the respective loop including the next recursion call. \__keyval_end_loop_other:w also has to insert the next iteration of the active loop.

The parsing loops are done, so here ends the definition of \__keyval_tmp:w, which will finally set up the macros.

These macros will be called on the parsed keys and values of the key–value list. All arguments are completely trimmed. They test for blank key names and call the func-
tions passed to \keyval_parse:nnn inside of \exp_not:n with the correct arguments. Afterwards they insert the next iteration of the other loop.

(End of definition for __keyval_pair:nnnn and __keyval_key:nn)

\__keyval_if_empty:w
\__keyval_if_blank:w
\__keyval_if_recursion_tail:w

All these tests work by gobbling tokens until a certain combination is met, which makes them pretty fast. The test for a blank argument should be called with an arbitrary token following the argument. Each of these utilize the fact that the argument will contain a leading \s__keyval_mark.

(End of definition for __keyval_if_empty:w, __keyval_if_blank:w, and __keyval_if_recursion_tail:w)

\__keyval_blank_true:w
\__keyval_blank_key_error:w

These macros will be called if the tests above didn’t gobble them, they execute the branching.

(End of definition for __keyval_blank_true:w and __keyval_blank_key_error:w)

Two messages for the low level parsing system.

970
And an adapted version of \_\_tl\_trim\_spaces:nn which is a bit faster for our use case, as it can strip the braces at the end. This is pretty much the same concept, so I won’t comment on it here. The speed gain by using this instead of \tl\_trim\_spaces\_apply:nn is about 10% of the total time for \keyval\_parse:NNn with one key and one key–value pair, so I think it’s worth it.

This is the one macro which differs from the original definition.
65.2 Constants and variables

Various storage areas for the different data which make up keys.

\c__keys_code_root_str
\c__keys_check_root_str
\c__keys_default_root_str
\c__keys_groups_root_str
\c__keys_inherit_root_str
\c__keys_type_root_str

The prefix for storing properties.

\c__keys_props_root_str

Publicly accessible data on which choice is being used when several are generated as a set.

\l_keys_choice_int
\l_keys_choice_tl

Used for storing and recovering the list of groups which apply to a key: set as a comma list but at one point we have to use this for a token list recovery.

\l__keys_groups_clist

For normalisation.

\l__keys_module_str

The name of a key itself: needed when setting keys.

\l_keys_key_str

The tl version is deprecated but has to be handled manually.

\l_keys_key_tl

The module for an entire set of keys.

\l__keys_module_str
\l__keys_no_value_bool A marker is needed internally to show if only a key or a key plus a value was seen: this is recorded here.

\bool_new:N \l__keys_no_value_bool
(End of definition for \l__keys_no_value_bool.)

\l__keys_only_known_bool Used to track if only “known” keys are being set.

\bool_new:N \l__keys_only_known_bool
(End of definition for \l__keys_only_known_bool.)

\l__keys_path_str The “path” of the current key is stored here: this is available to the programmer and so is public.

\str_new:N \l__keys_path_str
(End of definition for \l__keys_path_str. This variable is documented on page 249.)

\l__keys_path_tl The older version is deprecated but has to be handled manually.

\tl_new:N \l__keys_path_tl
(End of definition for \l__keys_path_tl.)

\l__keys_inherit_str

\str_new:N \l__keys_inherit_str
(End of definition for \l__keys_inherit_str.)

\l__keys_relative_tl The relative path for passing keys back to the user. As this can be explicitly no-value, it must be a token list.

\tl_new:N \l__keys_relative_tl
\tl_set:Nn \l__keys_relative_tl \{ \q__keys_no_value \}
(End of definition for \l__keys_relative_tl.)

\l__keys_property_str The “property” begin set for a key at definition time is stored here.

\str_new:N \l__keys_property_str
(End of definition for \l__keys_property_str.)

\l__keys_selective_bool \l__keys_exclude_bool Two booleans for using key groups: one to indicate that “selective” setting is active, a second to specify which type (“opt-in” or “opt-out”).

\bool_new:N \l__keys_selective_bool
\bool_new:N \l__keys_exclude_bool
(End of definition for \l__keys_selective_bool and \l__keys_exclude_bool.)

\l__keys_selective_clist The list of key groups being filtered in or out during selective setting.

\seq_new:N \l__keys_selective_clist
(End of definition for \l__keys_selective_clist.)

\l__keys_tmp_clist Scratch space used as a data dump.

\seq_new:N \l__keys_tmp_clist
(End of definition for \l__keys_tmp_clist.)
\__keys_unused_clist Used when setting only some keys to store those left over.

```
\clist_new:N \__keys_unused_clist
(End of definition for \__keys_unused_clist.)
```

\__keys_value_tl The value given for a key: may be empty if no value was given.

```
\tl_new:N \__keys_value_tl
(End of definition for \__keys_value_tl. This variable is documented on page 249.)
```

\__keys_tmpbool \__keys_tmpa_tl \__keys_tmpb_tl Scratch space.

```
\bool_new:N \__keys_tmpbool
\tl_new:N \__keys_tmpa_tl
\tl_new:N \__keys_tmpb_tl
(End of definition for \__keys_tmpbool, \__keys_tmpa_tl, and \__keys_tmpb_tl.)
```

\__keys_precompilebool \__keys_precompile_tl For digesting keys.

```
\bool_new:N \__keys_precompilebool
\tl_new:N \__keys_precompile_tl
(End of definition for \__keys_precompilebool and \__keys_precompile_tl.)
```


```
\prop_new:N \__keys_usage_load_prop
\prop_new:N \__keys_usage_preamble_prop
(End of definition for \__keys_usage_load_prop and \__keys_usage_preamble_prop. These variables are documented on page 248.)
```

65.2.1 Internal auxiliaries

\s__keys_nil \s__keys_mark \s__keys_stop Internal scan marks.

```
\scan_new:N \s__keys_nil
\scan_new:N \s__keys_mark
\scan_new:N \s__keys_stop
(End of definition for \s__keys_nil, \s__keys_mark, and \s__keys_stop.)
```

\q__keys_no_value Internal quarks.

```
\quark_new:N \q__keys_no_value
(End of definition for \q__keys_no_value.)
```

\__keys_quark_if_no_value:NTF Branching quark conditional.

```
\__kernel_quark_new_conditional:Nn \__keys_quark_if_no_value:N { TF }
(End of definition for \__keys_quark_if_no_value:NTF.)
```

\__keys_precompile:n An auxiliary to allow cleaner showing of code.

```
\cs_new_protected:Npn \__keys_precompile:n #1
{\bool_if:NTF \__keys_precompilebool
\tl_put_right:Nn \__keys_precompile_tl }\use:n
{\tl_put_right:Nn \__keys_precompile_tl }\use:n
{\tl_put_right:Nn \__keys_precompile_tl }
{\tl_put_right:Nn \__keys_precompile_tl }
```

974
The key defining mechanism

The public function for definitions is just a wrapper for the lower level mechanism, more or less. The outer function is designed to keep a track of the current module, to allow safe nesting. The module is set removing any leading / (which is not needed here).

The outer functions here record whether a value was given and then converge on a common internal mechanism. There is first a search for a property in the current key name, then a check to make sure it is known before the code hands off to the next step.
Searching for a property means finding the last \texttt{.} in the input, and storing the text before and after it. Everything is first turned into strings, so there is no problem using \texttt{\cs_set_nopar:Nne} instead of \texttt{\str_set:Ne} to set \texttt{\l_keys_path_str}. To gain further speed, brace tricks are used and \texttt{\__keys_property_find_auxiv:w} is defined as expandable. Since spaces will already be trimmed from the module we can omit it from the argument to \texttt{\__keys_trim_spaces:n}.

\begin{verbatim}
\cs_new_protected:Npn \__keys_property_find:n #1
  \exp_after:wN \__keys_property_find_auxi:w \tl_to_str:n {#1}
  \s__keys_nil \__keys_property_find_auxii:w
  \s__keys_nil \__keys_property_find_err:w
\}
\cs_new:Npn \__keys_property_find_auxi:w #1 . #2 \s__keys_nil #3
  #3 #1 \s__keys_mark #2 \s__keys_nil \__keys_property_find_auxi:w
\cs_new_protected:Npn \__keys_property_find_auxii:w #1 \s__keys_mark #2 \s__keys_nil \__keys_property_find_auxii:w . \s__keys_nil
  \__keys_property_find_auxi:w
\cs_new_protected:Npn \__keys_property_find_auxiii:w #1 \s__keys_mark #2 \s__keys_nil \__keys_property_find_auxiii:w . \s__keys_nil
  \__keys_property_find_auxiv:w
\cs_new_protected:Npn \__keys_property_find_auxiv:w #1 \s__keys_nil \__keys_property_find_auxiv:w
  \s__keys_mark \s__keys_nil \__keys_property_find_auxiv:w
\cs_set_nopar:Npe \l__keys_module_str \l__keys_module_str / \fi:
\exp_after:wN \__keys_trim_spaces:n \tex_expanded:D {{
  #1
  \if_false: {{{ \fi: }}}
\__keys_property_find_auxi:w \s__keys_nil \__keys_property_find_auxi:w . \s__keys_nil
  \s__keys_nil \__keys_property_find_auxiv:w . \s__keys_nil
\exp_after:wN \__keys_property_find_auxiv:w \tl_to_str:n {\l__keys_module_str / }
\exp_after:wN \__keys_trim_spaces:n \tex_expanded:D {{
  \if_false: {{{ \fi: }}}
\cs_set_nopar:Npe \l__keys_path_str \l__keys_path_str \fi:
\str_if_empty:NF \l__keys_module_str \\l__keys_module_str /
\end{verbatim}
Two possible cases. If there is a value for the key, then just use the function. If not, then a check to make sure there is no need for a value with the property. If there should be one then complain, otherwise execute it. There is no need to check for a : as if it was missing the earlier tests would have failed.

\cs_new_protected:Npn \__keys_define_code:n #1
\__keys_define_code:w

\cs_new_protected:Npn \__keys_set:Nn #1#2
\cs_generate_variant:Nn \__keys_set:Nn { c }
\cs_new_protected:Npn \__keys_set_inverse:Nn #1#2
\cs_generate_variant:Nn \__keys_set_inverse:Nn { c }
\cs_new_protected:Nnnn #1#2#3#4

\msg_error:nnn { keys } { no-property } {#1}

(End of definition for \__keys_property_find:n and others.)

65.4 Turning properties into actions

Boolean keys are really just choices, but all done by hand. The second argument here is the scope: either empty or \texttt{g} for global.
To make a choice from a key, two steps: set the code, and set the unknown key. As
multichoices and choices are essentially the same bar one function, the code is given
together.

Auto-generating choices means setting up the root key as a choice, then defining each
choice in turn.
Setting the code for a key first logs if appropriate that we are defining a new key, then saves the code.

```
cs_new_protected:Npn \__keys_cmd_set:nn #1#2
{ \__keys_cmd_set_direct:nn {#1} { \__keys_precompile:n {#2} } }
```

(End of definition for \__keys_cmd_set:nn and \__keys_cmd_set_direct:nn.)

Creating control sequences is a bit more tricky than other cases as we need to pick up the `p` argument. To make the internals look clearer, the trailing `n` argument here is just for appearance.

```
cs_new_protected:Npn \__keys_cs_set:NNpn #1#2#3#
{ \cs_set_protected:cpe { \c__keys_code_root_str \l_keys_path_str } ##1
{ \__keys_precompile:n { #1 \exp_not:N #2 \exp_not:n {#3} {##1} } }
\use_none:n
}
cs_generate_variant:Nn \__keys_cs_set:Ncpn { Nc }
```

(End of definition for \__keys_cs_set:NNpn and \__keys_cs_set:Ncpn.)

Setting a default value is easy. These are stored using `\cs_set_nopar:cpe` as this avoids any worries about whether a token list exists.

```
cs_new_protected:Npn \__keys_default_set:n #1
{ \tl_if_empty:nTF {#1}
{ \__keys_cs_undefine:c \c__keys_default_root_str \l_keys_path_str }
\use_none:n
}
cs_generate_variant:Nn \__keys_default_set:Nhp { Nh }
```

(End of definition for \__keys_default_set:n and \__keys_default_set:Nhp.)
Assigning a key to one or more groups uses comma lists. As the list of groups only exists if there is anything to do, the setting is done using a scratch list. For the usual grouping reasons we use the low-level approach to undefining a list. We also use the low-level approach for the other case to avoid tripping up the check-declarations code.

\__keys_groups_set:n

\__keys_inherit:n  Inheritance means ignoring anything already said about the key: zap the lot and set up.

\__keys_initialise:n  A set up for initialisation: just run the code if it exists. We need to set the key string here, using the deprecated \tl as a piece of scratch space.
\_keys\_execute:no \l_keys\_path\_str \l_keys\_value\_tl
}
}

(End of definition for \_keys\_initialize:n.)

\_keys\_legacy\_if\_set:nn Much the same as expl3 booleans, except we assume that the switch exists.
}\cs\new\protected:Npn \_keys\_legacy\_if\_set:nn #1#2
{ \_keys\_legacy\_if\_set:nnnn {#1} {#2} { true } { false } }
\cs\new\protected:Npn \_keys\_legacy\_if\_set\_inverse:nn #1#2
{ \_keys\_legacy\_if\_set:nnnn {#1} {#2} { false } { true } }
\cs\new\protected:Npn \_keys\_legacy\_if\_set:nnnn #1#2#3#4
{ \_keys\_choice\_make: \_keys\_cmd\_set:ne { \l_keys\_path\_str / true }
{ \exp\_not:c { legacy\_if\_#2 set\_ #3 :n } { \exp\_not:n {#1} } }
\_keys\_cmd\_set:ne { \l_keys\_path\_str / false }
{ \exp\_not:c { legacy\_if\_#2 set\_ #4 :n } { \exp\_not:n {#1} } }
\_keys\_cmd\_set:nn { \l_keys\_path\_str / unknown }
{ \msg\_error:nne { keys } { boolean\_values\_only }
\l_keys\_path\_str }
}\_keys\_default\_set:n { true }
\cs\if\exist:cF { if#1 }
{ \cs:w new\if \exp\after:wN \cs\end:
 \cs:w if#1 \cs\end:
}

(End of definition for \_keys\_legacy\_if\_set:nn, \_keys\_legacy\_if\_inverse:nn, and \_keys\_legacy\_if\_inverse:nnnn.)

\_keys\_meta\_make:n To create a meta-key, simply set up to pass data through. The internal function is used
here as a meta key should respect the prevailing filtering, etc.
\cs\new\protected:Npn \_keys\_meta\_make:n #1
{ \exp\_args:NVo \_keys\_cmd\_set\_direct:nn \l_keys\_path\_str
{ \exp\after:wN \_keys\_set:nn \exp\after:wN
{ \_l_keys\_module\_str } {#1} }
}
\cs\new\protected:Npn \_keys\_meta\_make:nn #1#2
{ \exp\_args:NV \_keys\_cmd\_set\_direct:nn
\_l_keys\_path\_str { \_keys\_set:n {#1} {#2} } }

(End of definition for \_keys\_meta\_make:n and \_keys\_meta\_make:nn.)
Much the same as other variables, but needs a dedicated auxiliary.

\cs_new_protected:Nm \__keys_prop_put:Nn #1 \#2
\prop_if_exist:NF #1 { \prop_new:N #1 }
\exp_after:wN \__keys_find_key_module:NN \l__keys_path_str \s__keys_stop
\__keys_cmd_set:ne \l__keys_tmpa_tl \l__keys_tmpb_tl
\__keys_cmd_set:ne \l__keys_path_str
\exp_not:c { \prop_#2 put:Nnn }
\exp_not:N #1
\l__keys_tmpb_tl}
\exp_not:n { {##1}}
\exp_not:wN \__keys_find_key_module:N \l__keys_path_str
\exp_not:wN \__keys_find_key_module:N \l__keys_path_str
\cs_generate_variant:Nn \__keys_prop_put:Nn { c }

(End of definition for \__keys_prop_put:Nn.)

\__keys_undefine:
Undefining a key has to be done without \cs_undefine: as that function acts globally.

\cs_new_protected:Nm \__keys_undefine:
\clist_map_inline:nn
{ code , default , groups , inherit , type , check }
\__keys_cs_undefine:c
\tl_use:c { c__keys_##1_root_str } \l__keys_path_str
\cs_generate_variant:Nn \__keys_undefine:c { c }

(End of definition for \__keys_undefine.)

\__keys_value_requirement:nn
Validating key input is done using a second function which runs before the main key
code. Setting that up means setting it equal to a generic stub which does the check. This
approach makes the lookup very fast at the cost of one additional csname per key that
needs it. The cleanup here has to know the structure of the following code.

\cs_new_protected:Nm \__keys_value_requirement:nn #1 #2
\str_case:nnF { #2 }{ true }{ false }
\cs_set_eq:cc
\l__keys_check_root_str \l__keys_path_str
\__keys_check_: #1 :
\cs_if_eq:ccT
\l__keys_check_root_str \l__keys_path_str
\__keys_check_: #1 :
\__keys_cs_undefine: c
\l__keys_check_root_str \l__keys_path_str
\cs_generate_variant:Nn \__keys_check_: ccT { c }

982
\msg_error:nne { keys } { boolean-values-only } { .value_ #1 :n }

\cs_new_protected:Npn \__keys_check_forbidden:
  { \bool_if:NF \l__keys_no_value_bool
    { \msg_error:nnee { keys } { value-forbidden }
      \l_keys_path_str \l_keys_value_tl
      \use_none:nnn }
    }
\cs_new_protected:Npn \__keys_check_required:
  { \bool_if:NT \l__keys_no_value_bool
    { \msg_error:nne { keys } { value-required }
      \l_keys_path_str
      \use_none:nnn }
    }

(End of definition for \__keys_value_requirement:nn, \__keys_check_forbidden:, and \__keys_check_required:.)

\__keys_usage:n \__keys_usage:NN \__keys_usage:w

Save the relevant data.
\cs_new_protected:Npn \__keys_usage:n #1
  { \str_case:nnF {#1}
    { general }
      { \__keys_usage:NN \l_keys_usage_load_prop
        \c_false_bool
        \__keys_usage:NN \l_keys_usage_preamble_prop
        \c_false_bool
      }
    { load }
      { \__keys_usage:NN \l_keys_usage_load_prop
        \c_true_bool
        \__keys_usage:NN \l_keys_usage_preamble_prop
        \c_false_bool
      }
    { preamble }
      { \__keys_usage:NN \l_keys_usage_load_prop
        \c_false_bool
      }
  }

983
65.5 Creating key properties

The key property functions are all wrappers for internal functions, meaning that things stay readable and can also be altered later on.

Importantly, while key properties have “normal” argument specs, the underlying code always supplies one braced argument to these. As such, argument expansion is
handled by hand rather than using the standard tools. This shows up particularly for the two-argument properties, where things would otherwise go badly wrong.

One function for this.

```
\cs_new_protected:cpn { \c__keys_props_root_str .bool_set:N } #1
{ \__keys_bool_set:Nn #1 { } }
```

(End of definition for .bool_set:N and .bool_gset:N. These functions are documented on page 241.)

One function for this.

```
\cs_new_protected:cpn { \c__keys_props_root_str .bool_set_inverse:N } #1
{ \__keys_bool_set_inverse:Nn #1 { } }
```

(End of definition for .bool_set_inverse:N and .bool_gset_inverse:N. These functions are documented on page 241.)

Making a choice is handled internally, as it is also needed by .generate_choices:n.

```
\cs_new_protected:cpn { \c__keys_props_root_str .choice: } 
{ \__keys_choice_make: }
```

(End of definition for .choice:. This function is documented on page 241.)

For auto-generation of a series of mutually-exclusive choices. Here, #1 consists of two separate arguments, hence the slightly odd-looking implementation.

```
\cs_new_protected:cpn { \c__keys_props_root_str .choices:nn } #1
{ \__keys_choices_make:nn #1 }
```

(End of definition for .choices:nn. This function is documented on page 241.)

Creating code is simply a case of passing through to the underlying set function.

```
\cs_new_protected:cpn { \c__keys_props_root_str .code:n } #1
{ \__keys_cmd_set:nn \l_keys_path_str {#1} }
```

(End of definition for .code:n. This function is documented on page 242.)
The functions are documented on page 241.

Expansion is left to the internal functions.

Setting a variable is very easy: just pass the data along.

(End of definition for \clist_set:N and \clist_gset:N. These functions are documented on page 241.)

(End of definition for \clist_set:c and \clist_gset:c. These functions are documented on page 241.)

(End of definition for \clist_set:Np and others. These functions are documented on page 242.)

(End of definition for \clist_gset:N. This function is documented on page 242.)

(End of definition for \dim_set:N. This function is documented on page 242.)

986
Setting a variable is very easy: just pass the data along.

\begin{verbatim}
c_new_protected:cpn { \c__keys_props_root_str .fp_set:N } #1
c_new_protected:cpn { \c__keys_props_root_str .fp_set:c } #1
c_new_protected:cpn { \c__keys_props_root_str .fp_gset:N } #1
c_new_protected:cpn { \c__keys_props_root_str .fp_gset:c } #1
\end{verbatim}

A single property to create groups of keys.

\begin{verbatim}
c_new_protected:cpn { \c__keys_props_root_str .groups:n } #1
c_new_protected:cpn { \c__keys_props_root_str .groups:N } #1
\end{verbatim}

Nothing complex: only one variant at the moment!

\begin{verbatim}
c_new_protected:cpn { \c__keys_props_root_str .inherit:n } #1
c_new_protected:cpn { \c__keys_props_root_str .inherit:N } #1
\end{verbatim}

The standard hand-off approach.

\begin{verbatim}
c_new_protected:cpn { \c__keys_props_root_str .initial:n } #1
c_new_protected:cpn { \c__keys_props_root_str .initial:V } #1
c_new_protected:cpn { \c__keys_props_root_str .initial:e } #1
c_new_protected:cpn { \c__keys_props_root_str .initial:o } #1
c_new_protected:cpn { \c__keys_props_root_str .initial:x } #1
\end{verbatim}

Setting a variable is very easy: just pass the data along.

\begin{verbatim}
c_new_protected:cpn { \c__keys_props_root_str .int_set:N } #1
c_new_protected:cpn { \c__keys_props_root_str .int_set:c } #1
c_new_protected:cpn { \c__keys_props_root_str .int_gset:N } #1
c_new_protected:cpn { \c__keys_props_root_str .int_gset:c } #1
\end{verbatim}
(.legacy_if_set:n)
(.legacy_if_gset:n)
(.legacy_if_set_inverse:n)
(.legacy_if_gset_inverse:n)

\cs_new_protected:cpn{\c__keys_props_root_str .legacy_if_set:n } #1
{\__keys_legacy_if_set:nn {#1} { } }
\cs_new_protected:cpn{\c__keys_props_root_str .legacy_if_gset:n } #1
{\__keys_legacy_if_set:nn {#1} { g } }
\cs_new_protected:cpn{\c__keys_props_root_str .legacy_if_set_inverse:n } #1
{\__keys_legacy_if_set_inverse:nn {#1} { } }
\cs_new_protected:cpn{\c__keys_props_root_str .legacy_if_gset_inverse:n } #1
{\__keys_legacy_if_set_inverse:nn {#1} { g } }

(End of definition for .legacy_if_set:n and others. These functions are documented on page 243.)

.meta:n
Making a meta is handled internally.
\cs_new_protected:cpn{\c__keys_props_root_str .meta:n } #1
{\__keys_meta_make:n {#1} }
(End of definition for .meta:n. This function is documented on page 243.)

.meta:nn
Meta with path: potentially lots of variants, but for the moment no so many defined.
\cs_new_protected:cpn{\c__keys_props_root_str .meta:nn } #1
{\__keys_meta_make:nn #1 }
(End of definition for .meta:nn. This function is documented on page 244.)

.multichoice:
.multichoices:nn
.multichoices:Vn
.multichoices:en
.multichoices:on
.multichoices:xn
The same idea as .choice: and .choices:nn, but where more than one choice is allowed.
\cs_new_protected:cpn{\c__keys_props_root_str .multichoice: }
{\__keys_multichoice_make: }
\cs_new_protected:cpn{\c__keys_props_root_str .multichoices:nn } #1
{\__keys_multichoices_make:nn #1 }
\cs_new_protected:cpn{\c__keys_props_root_str .multichoices:Vn } #1
{\exp_args:NV \__keys_multichoices_make:nn #1 }
\cs_new_protected:cpn{\c__keys_props_root_str .multichoices:en } #1
{\exp_args:Ne \__keys_multichoices_make:nn #1 }
\cs_new_protected:cpn{\c__keys_props_root_str .multichoices:on } #1
{\exp_args:No \__keys_multichoices_make:nn #1 }
\cs_new_protected:cpn{\c__keys_props_root_str .multichoices:xn } #1
{\exp_args:Nx \__keys_multichoices_make:nn #1 }
(End of definition for .multichoice: and .multichoices:nn. These functions are documented on page 244.)

.muskip_set:N
.muskip_set:c
.muskip_gset:N
.muskip_gset:c
Setting a variable is very easy: just pass the data along.
\cs_new_protected:cpn{\c__keys_props_root_str .muskip_set:N } #1
{\__keys_variable_set_required:NnnN #1 { muskip } { } n }
\cs_new_protected:cpn{\c__keys_props_root_str .muskip_set:c } #1
{\__keys_variable_set_required:cnnN {#1} { muskip } { } n }
\cs_new_protected:cpn{\c__keys_props_root_str .muskip_gset:N } #1
{\__keys_variable_set_required:NnnN #1 { muskip } { g } n }
\cs_new_protected:cpn{\c__keys_props_root_str .muskip_gset:c } #1
{\__keys_variable_set_required:cnnN {#1} { muskip } { g } n }
(End of definition for .muskip_set:N and .muskip_gset:N. These functions are documented on page 244.)
Setting a variable is very easy: just pass the data along.

\begin{verbatim}
\cs_new_protected:cpn { \c__keys_props_root_str .prop_put:N } #1
\{ \__keys_prop_put:Nn #1 { } \}
\cs_new_protected:cpn { \c__keys_props_root_str .prop_put:c } #1
\{ \__keys_prop_put:cn {#1} { } \}
\cs_new_protected:cpn { \c__keys_props_root_str .prop_gput:N } #1
\{ \__keys_prop_put:Nn #1 { g } \}
\cs_new_protected:cpn { \c__keys_props_root_str .prop_gput:c } #1
\{ \__keys_prop_put:cn {#1} { g } \}
\end{verbatim}

(End of definition for \texttt{.prop_put:N} and \texttt{.prop_gput:N}. These functions are documented on page 244.)

Setting a variable is very easy: just pass the data along.

\begin{verbatim}
\cs_new_protected:cpn { \c__keys_props_root_str .skip_set:N } #1
\{ \__keys_variable_set_required:NnnN #1 { skip } { } n \}
\cs_new_protected:cpn { \c__keys_props_root_str .skip_set:c } #1
\{ \__keys_variable_set:cnnN {#1} { skip } { } n \}
\cs_new_protected:cpn { \c__keys_props_root_str .skip_gset:N } #1
\{ \__keys_variable_set_required:NnnN #1 { skip } { g } n \}
\cs_new_protected:cpn { \c__keys_props_root_str .skip_gset:c } #1
\{ \__keys_variable_set:cnnN {#1} { skip } { g } n \}
\end{verbatim}

(End of definition for \texttt{.skip_set:N} and \texttt{.skip_gset:N}. These functions are documented on page 244.)

Setting a variable is very easy: just pass the data along.

\begin{verbatim}
\cs_new_protected:cpn { \c__keys_props_root_str .str_set:N } #1
\{ \__keys_variable_set:NnnN #1 { str } { } n \}
\cs_new_protected:cpn { \c__keys_props_root_str .str_set:c } #1
\{ \__keys_variable_set:cnnN {#1} { str } { } n \}
\cs_new_protected:cpn { \c__keys_props_root_str .str_set_e:N } #1
\{ \__keys_variable_set:NnnN #1 { str } { } e \}
\cs_new_protected:cpn { \c__keys_props_root_str .str_set_e:c } #1
\{ \__keys_variable_set:cnnN {#1} { str } { } e \}
\cs_new_protected:cpn { \c__keys_props_root_str .str_gset:N } #1
\{ \__keys_variable_set:NnnN #1 { str } { g } n \}
\cs_new_protected:cpn { \c__keys_props_root_str .str_gset:c } #1
\{ \__keys_variable_set:cnnN {#1} { str } { g } n \}
\cs_new_protected:cpn { \c__keys_props_root_str .str_gset_e:N } #1
\{ \__keys_variable_set:NnnN #1 { str } { g } e \}
\cs_new_protected:cpn { \c__keys_props_root_str .str_gset_e:c } #1
\{ \__keys_variable_set:cnnN {#1} { str } { g } e \}
\end{verbatim}

(End of definition for \texttt{.str_set:N} and others. These functions are documented on page 244.)

Setting a variable is very easy: just pass the data along.

\begin{verbatim}
\cs_new_protected:cpn { \c__keys_props_root_str .tl_set:N } #1
\{ \__keys_variable_set:NnnN #1 { tl } { } n \}
\cs_new_protected:cpn { \c__keys_props_root_str .tl_set:c } #1
\{ \__keys_variable_set:cnnN {#1} { tl } { } n \}
\cs_new_protected:cpn { \c__keys_props_root_str .tl_set_e:N } #1
\{ \__keys_variable_set:NnnN #1 { tl } { } e \}
\cs_new_protected:cpn { \c__keys_props_root_str .tl_set_e:c } #1
\{ \__keys_variable_set:cnnN {#1} { tl } { } e \}
\cs_new_protected:cpn { \c__keys_props_root_str .tl_gset:N } #1
\{ \__keys_variable_set:NnnN #1 { tl } { g } n \}
\cs_new_protected:cpn { \c__keys_props_root_str .tl_gset:c } #1
\{ \__keys_variable_set:cnnN {#1} { tl } { g } n \}
\cs_new_protected:cpn { \c__keys_props_root_str .tl_gset_e:N } #1
\{ \__keys_variable_set:NnnN #1 { tl } { g } e \}
\cs_new_protected:cpn { \c__keys_props_root_str .tl_gset_e:c } #1
\{ \__keys_variable_set:cnnN {#1} { tl } { g } e \}
\end{verbatim}

989
65.6 Setting keys

The aim here is to allow nesting of key setting without needing lots of tracking. That is done by expanding the appropriate tokens “around” the core keyval parsing. As there are several different sub-paths, this needs a few steps and some generic auxiliaries. The arguments here are

1. The root for keys
2. The key groups
3. The keys themselves
4. The relative root for return of unset keys
5. The clist var for returning unset keys
6. The code to set up the correct selection approach
\cs_new_protected:Npn \__keys_set:nnnnNn
{
  \exp_args:Nooo \__keys_set:nnnnnnnNn
  \l__keys_unused_clist
  \l__keys_selective_clist
  \l__keys_relative_tl
}
\cs_new_protected:Npn \__keys_set:nnnnnnnNn #1#2#3#4#5#6#7#8#9
{
  \clist_clear:N \l__keys_unused_clist
  \clist_set:Ne \l__keys_selective_clist { \tl_to_str:n {#5} }
  \tl_set:Nn \l__keys_relative_tl {#7}
  \use:e
  {\exp_not:n}{#9}
  \__keys_set:nn {#4} {#6}
  \__keys_reset_bool:N \l__keys_only_known_bool
  \__keys_reset_bool:N \l__keys_exclude_bool
  \__keys_reset_bool:N \l__keys_selective_bool
}
\clist_set_eq:NN #8 \l__keys_unused_clist
\__kernel_tl_set:Nx \l__keys_unused_clist { \exp_not:n {#1} }
\__kernel_tl_set:Nx \l__keys_selective_clist {\exp_not:n {#3} }
\__kernel_tl_set:Nx \l__keys_relative_tl {\exp_not:n {#3} }
\cs_new:Npn \__keys_reset_bool:N #1
{
  \exp_not:c{ bool_set_ \bool_if:NTF #1 { true } { false } :N }
  \exp_not:N #1
}
\cs_new_protected:Npn \__keys_set:nn #1#2
{
  \__keys_set:nnnnNn
  {#1} { } {#2} { \q__keys_no_value } \l__keys_tmp_clist
  {\bool_set_false:N \l__keys_only_known_bool}
  {\bool_set_false:N \l__keys_exclude_bool}
  {\bool_set_false:N \l__keys_selective_bool}
}
\cs_new_protected:Npn \__keys_set:nn #1#2
{ \exp_args:Noo \__keys_set:nnnnNn \l__keys_module_str
  \__keys_set_keyval:nn {#1}{#2}
  \str_set:Nn \l__keys_module_str {#1}
  \str_set:Nn \l__keys_module_str {#1}
}
\keys_set:nn
A simple wrapper allowing for nesting.
\keys_set:nV
\keys_set:nv
\keys_set:ne
\keys_set:no
\keys_set:nx
(End of definition for \__keys_set:nnnnNn and others.)
Simply set the right variables.

```latex
\keys_set_known:nnnN
\keys_set_known:nVnN
\keys_set_known:nvnN
\keys_set_known:nenN
\keys_set_known:nonN
\keys_set_known:nnN
\keys_set_known:nVn
\keys_set_known:nv
\keys_set_known:ne
\keys_setKnown:n
\keys_set_exclude_groups:nnnnN
\keys_set_exclude_groups:nnVn
\keys_set_exclude_groups:nnvn
\keys_set_exclude_groups:nnon
\keys_set_exclude_groups:nnn
\keys_set_exclude_groups:nV
\keys_set_exclude_groups:nv
\keys_set_exclude_groups:ne
\keys_set_exclude_groups:no
```

The same for (exclusion) groups.

```latex
\keys_set_exclude_groups:nnnnN
\keys_set_exclude_groups:nnV
\keys_set_exclude_groups:nnv
\keys_set_exclude_groups:nno
\keys_set_exclude_groups:nnn
\keys_set_groups:nnnnN
\keys_set_groups:nnVn
\keys_set_groups:nnvn
\keys_set_groups:nnon
\keys_set_groups:nnn
\keys_set_groups:nV
\keys_set_groups:nv
\keys_set_groups:ne
\keys_set_groups:no
```

(End of definition for \keys_set:nn. This function is documented on page 248.)

(End of definition for \keys_set_known:nnnN, \keys_set_known:nvnN, and \keys_set_known:nn. These functions are documented on page 249.)
\bool_set_false:N \l__keys_exclude_bool
\bool_set_true:N \l__keys_selective_bool
}

\cs_generate_variant:Nn \keys_set_groups:nnnnN { nnV , nnv , nno }
\cs_new_protected:Npn \keys_set_groups:nnnN #1#2#3#4

\keys_set_groups:nnnnN {#1} {#2} {#3} { \q__keys_no_value } #4
\keys_set_groups:nnnN #1#2#3

\keys_set_groups:nnnnN {#1} {#2} {#3}
\q__keys_no_value \l__keys_tmp_clist
}
\cs_generate_variant:Nn \keys_set_groups:nnn { nnV , nnv , nno }

\cs_new_protected:Npn \keys_set_groups:nnn #1#2#3

\keys_set_groups:nnnnN {#1} {#2} {\l__keys_tmp_clist}
\cs_generate_variant:Nn \keys_set_groups:nnn { nnV , nnv , nno }

(End of definition for \keys_set_exclude_groups:nnnn and others. These functions are documented on page \pageref{page:keys_set_exclude_groups:nnnn}.)

\keys_precompile:nnN A simple wrapper.
\cs_new_protected:Npn \keys_precompile:nnN #1#2#3

\bool_set_true:N \l__keys_precompile_bool
\tl_clear:N \l__keys_precompile_tl
\keys_set:nn {#1} {#2}
\bool_set_false:N \l__keys_precompile_bool
\tl_set_eq:NN #3 \l__keys_precompile_tl

(End of definition for \keys_precompile:nnN. This function is documented on page \pageref{page:keys_precompile:nnN}.)
\__keys_set_keyval:n
\__keys_set_keyval:nn
\__keys_set_keyval:nnn
\__keys_set_keyval:onn
\__keys_find_key_module:wNN
\__keys_find_key_module_auxi:Nw
\__keys_find_key_module_auxii:Nw
\__keys_find_key_module_auxiii:Nw
\__keys_find_key_module_auxiv:Nw
\__keys_set_selective:

\cs_new_protected:Npn \__keys_set_keyval:n #1
\bool_set_true:N \l__keys_no_value_bool
\__keys_set_keyval:onn \l__keys_module_str {#1} { }

\cs_new_protected:Npn \__keys_set_keyval:nn #1#2
\bool_set_false:N \l__keys_no_value_bool
\__keys_set_keyval:onn \l__keys_module_str {#1} {#2}

\cs_new_protected:Npn \__keys_set_keyval:nnn #1#2#3

\__kernel_tl_set:Nx \l_keys_path_str
\tl_if_blank:nF {#1}
\{ \l__keys_module_str {#1} \}
\\l__keys_module_str {#1} \}

\cs_new_protected:Npn \__keys_set_keyval:onn #1
\bool_set_false:N \l__keys_no_value_bool
\__keys_set_keyval:onn \l__keys_module_str {#1} {#2}

The key path here can be fully defined, after which there is a search for the key and module names: the user may have passed them with part of what is actually the module (for our purposes) in the key name. As that happens on a per-key basis, we use the stack approach to restore the module name without a group.
This function uses \texttt{cs_set_nopar:Npe} internally for performance reasons, the argument \texttt{#1} is already a string in every usage, so turning it into a string again seems unnecessary.

If selective setting is active, there are a number of possible sub-cases to consider. The key name may not be known at all or if it is, it may not have any groups assigned. There is then the question of whether the selection is opt-in or opt-out.
In the case where selective setting requires a comparison of the list of groups which apply to a key with the list of those which have been set active. That requires two mappings, and again a different outcome depending on whether opt-in or opt-out is set. It is safe to use \clist_if_in:NnTF because both \l__keys_selective_clist and \l__keys_groups_clist contain the groups as strings, without leading/trailing spaces in any item, since the l3clist functions were applied to the result of applying \tl_to_str:n.

\cs_new_protected:Npn \__keys_check_groups: 
\{ 
\bool_set_false:N \l__keys_tmp_bool 
\clist_map_inline:Nn \l__keys_selective_clist 
\{ 
\clist_if_in:NnT \l__keys_groups_clist {##1} 
\{ 
\bool_set_true:N \l__keys_tmp_bool 
\clist_map_break: 
\}
\}
\bool_if:NTF \l__keys_tmp_bool 
\{ 
\bool_if:NTF \l__keys_exclude_bool 
\__keys_store_unused: 
\__keys_execute: 
\}
\}
\cs_new_protected:Npn \__keys_value_or_default:n #1 
\{ 
\bool_if:NTF \l__keys_no_value_bool 
\{ 
\cs_if_exist:cTF { \c__keys_default_root_str \l_keys_path_str } 
\{ 
\tl_set_eq:Nc \l_keys_value_tl \c__keys_default_root_str \l_keys_path_str 
\}
\}
\{ 
\tl_clear:N \l_keys_value_tl 
\cs_if_exist:cT 
\{ \c__keys_inherit_root_str \__keys_parent:o \l_keys_path_str \}
\{ \__keys_default_inherit: \}
\}

(End of definition for \__keys_set_keyval:n and others.)

\__keys_value_or_default:n \__keys_default_inherit: If a value is given, return it as #1, otherwise send a default if available.
Actually executing a key is done in two parts. First, look for the key itself, then look for the unknown key with the same path. If both of these fail, complain. What exactly happens if a key is unknown depends on whether unknown keys are being skipped or if an error should be raised.

To deal with the case where there is no hit, we leave \_\_keys_execute_unknown: in the input stream and clean it up using the break function: that avoids needing a boolean.
A key’s code is in the control sequence with csname \c__keys_code_root_str #1. We expand it once to get the replacement text (with argument #2) and call \use:n with this replacement as its argument. This ensures that any undefined control sequence error in the key’s code will lead to an error message of the form <argument>...{control sequence} in which one can read the (undefined) ⟨control sequence⟩ in full, rather than an error message that starts with the potentially very long key name, which would make the (undefined) ⟨control sequence⟩ be truncated or sometimes completely hidden. See https://github.com/latex3/latex2e/issues/351.

When there is no relative path, things here are easy: just save the key name and value. When we are working with a relative path, first we need to turn it into a string: that can’t happen earlier as we need to store \q__keys_no_value. Then, use a standard delimited
approach to fish out the partial path.

\cs_new_protected:Npn \_\_keys_store_unused:
    {\_\_keys_quark_if_no_value:NTF \_\_keys_relative_tl
        {\clist_put_right:Ne \_\_keys_unused_clist
            \_\_keys_key_str
            \bool_if:FN \_\_keys_no_value_bool
                {= { \exp_not:o \_\_keys_value_tl }}
        }
    }
\tl_if_blank:nF {\_\_keys_relative_tl}
    {\clist_put_right:Ne \_\_keys_unused_clist
        \_\_keys_path_str
        \bool_if:FN \_\_keys_no_value_bool
            {= { \exp_not:o \_\_keys_value_tl }}
    }
\_\_keys_store_unused_aux: }
\cs_new_protected:Npn \_\_keys_store_unused_aux:
    {\_kernel_tl_set:Nx \_\_keys_relative_tl
        {\exp_args:No \_\_keys_trim_spaces:n \_\_keys_relative_tl }
        \use:e
            \cs_set_protected:Npn \_\_keys_store_unused:w
                {\tl_if_blank:nF {#1}
                    \msg_error:nnee { keys } { bad-relative-key-path }
                    \_\_keys_path_str
                    \_\_keys_relative_tl
                }
            {\clist_put_right:Ne \_\_keys_unused_clist
                \_\_keys_key_str
                \bool_if:FN \_\_keys_no_value_bool
                    {= { \exp_not:o \_\_keys_value_tl }}
                }
        }
\use:e
    {\_\_keys_store_unused:w \_\_keys_path_str
        \_\_keys_relative_tl / \_\_keys_relative_tl /
Executing a choice has two parts. First, try the choice given, then if that fails call the unknown key. That always exists, as it is created when a choice is first made. So there is no need for any escape code. For multiple choices, the same code ends up used in a mapping.

65.7 Utilities

Used to strip off the ending part of the key path after the last /.

\begin{verbatim}
\cs_new:Npn \__keys_choice_find:n #1
\exp_after:wN \__keys_parent_auxi:w #1 \q_nil \__keys_parent_auxii:w
\__keys_parent_auxiii:n \__keys_parent_auxiv:w
\end{verbatim}
Space stripping has to allow for the fact that the key here might have several parts, and spaces need to be stripped from each part. Since the key name is turned into a string groups can’t be stripped accidentally and the precautions of \_tl_trim_spaces:n aren’t necessary, in this case it is much faster to just directly strip spaces around /.

\keys_if_exist_p:n \keys_if_exist:nnTF

A utility for others to see if a key exists.
Just an alternative view on \keys_if_exist:nnTF.

\begin{Verbatim}
\prg_new_conditional:Npnn \keys_if_choice_exist:nnn #1#2#3
\{ p , T , F , TF \}
\{
\cs_if_exist:cTF
\{ \c__keys_code_root_str \__keys_trim_spaces:n { #1 / #2 / #3 } \}
\{ \prg_return_true: \}
\{ \prg_return_false: \}
\}
\end{Verbatim}

To show a key, show its code using a message.

\begin{Verbatim}
\cs_new_protected:Npn \keys_show:nn
{ \__keys_show:Nnn \msg_show:nneeee }
\cs_new_protected:Npn \keys_log:nn
{ \__keys_show:Nnn \msg_log:nneeee }
\cs_new_protected:Npn \__keys_show:Nnn #1#2#3
{ \keys_if_exist:nnT {#2} {#3}
\{ \exp_args:Nnf \msg_show_item_unbraced:nn { code }
\{ \exp_args:Ne \__keys_show:n
\{ \exp_args:Nc \cs_replacement_spec:N
\{ \c__keys_code_root_str
\__keys_trim_spaces:n { #2 / #3 } \}
\}
\}
\{ \} \{ \}
\}
\cs_new:Npe \__keys_show:n #1
\{
\exp_not:N \__keys_show:w #1
\tl_to_str:n \__keys_precompile:n \}
\exp_not:N \s__keys_stop
\use:e
\{
\cs_new:Npn \exp_not:N \__keys_show:w
\#1 \tl_to_str:n \__keys_precompile:n \}
\exp_not:N \s__keys_stop
\}
\end{Verbatim}
\msg_new:nnnn { keys } { bad-relative-key-path }
\msg_new:nnnn { keys } { boolean-values-only }
\msg_new:nnnn { keys } { choice-unknown }
\msg_new:nnnn { keys } { nested-choice-key }
\msg_new:nnnn { keys } { value-forbidden }
\msg_new:nnnn { keys } { value-required }

(End of definition for \keys_show:n and others. These functions are documented on page 251.)

65.8 Messages

For when there is a need to complain.
\msg_new:nnnn { keys } { unknown }
\msg_new:nnnn { keys } { nested-choice-key }
\msg_new:nnnn { keys } { value-forbidden }
\msg_new:nnnn { keys } { value-required }
\msg_new:nnnn { keys } { bad-relative-key-path }
\msg_new:nnnn { keys } { boolean-values-only }
\msg_new:nnnn { keys } { choice-unknown }
\msg_new:nnnn { keys } { nested-choice-key }
\msg_new:nnnn { keys } { value-forbidden }
\msg_new:nnnn { keys } { value-required }
\msg_new:nnnn { keys } { bad-relative-key-path }
\msg_new:nnnn { keys } { boolean-values-only }
\msg_new:nnnn { keys } { choice-unknown }
\msg_new:nnnn { keys } { nested-choice-key }
\msg_new:nnnn { keys } { value-forbidden }
\msg_new:nnnn { keys } { value-required }
\msg_new:nnnn { keys } { bad-relative-key-path }
\msg_new:nnnn { keys } { boolean-values-only }
\msg_new:nnnn { keys } { choice-unknown }
\msg_new:nnnn { keys } { nested-choice-key }
\msg_new:nnnn { keys } { value-forbidden }
\msg_new:nnnn { keys } { value-required }

1002
The key ‘#1’ must have a value. \\
No value was present: the key will be ignored.
}\n\msg_new:nnn { keys } { show-key }\n\{\n\tl_if_empty:nTF {#2} \n\{ is undefined. \n\{ has-the-properties: #2 . \n\}\n\prop_gput:Nnn \g_msg_module_name_prop { keys } { LaTeX }\n\prop_gput:Nnn \g_msg_module_type_prop { keys } { }\n\}[/package]
Chapter 66

\l3intarray implementation

There are two implementations for this module: One \fontdimen based one for more traditional \TeX engines and a Lua based one for engines with Lua support. Both versions do not allow negative array sizes.

\l__intarray_loop_int A loop index.

66.1 Lua implementation

First, let's look at the Lua variant:

We select the Lua version if the Lua helpers were defined. This can be detected by the presence of \__intarray_gset_count:Nw.

\cs_if_exist:NTF \__intarray_gset_count:Nw

66.1.1 Allocating arrays

\g__intarray_table_int Used to differentiate intarrays in Lua and to record an invalid index.
\l__intarray_bad_index_int

\__intarray:w Used as marker for intarrays in Lua. Followed by an unbraced number identifying the array and a single space. This format is used to make it easy to scan from Lua.

\luacmd{\__intarray:w, function() \scan_int()}

1004
\intarray_new:Nn \intarray_new:cn \__intarray_new:N \__intarray_new:Nn

\cs_new_protected:Npn \__intarray_new:N #1
\__kernel_chk_if_free_cs:N #1 \int_gincr:N \g__intarray_table_int
\cs_gset_nopar:Npe #1 { \__intarray:w \int_use:N \g__intarray_table_int \c_space_tl}
\cs_new_protected:Npn \intarray_new:Nn #1#2
\__intarray_new:N #1 \__intarray_gset_count:Nw #1 \int_eval:n {#2} \scan_stop:
\int_compare:nNnT { \intarray_count:N #1 } < 0
\msg_error:nne { kernel } { negative-array-size } { \intarray_count:N #1 }
\cs_generate_variant:Nn \intarray_new:Nn { c }
\cs_new_protected:Npn \intarray_new:Nn \intarray_new:cn \__intarray_new:N

(End of definition for \intarray_new:Nn and \__intarray_new:N. This function is documented on page 255.)

Before we get to the first command implemented in Lua, we first need some definitions. Since \token.create only works correctly if \TeX has seen the tokens before, we first run a short \TeX sequence to ensure that all relevant control sequences are known.

\__intarray_table

Internal helper to scan an intarray token, extract the associated Lua table and return an error if the input is invalid.

1005
Since in \LaTeX\ this is loaded in the format, we want to preserve any intarrays which are created while format building for the actual run.

To do this, we use the \texttt{register\_luadata} mechanism from \texttt{l3luatex}: Directly before the format get dumped, the following function gets invoked and serializes all existing tables into a string. This string gets compiled and dumped into the format and is made available at the beginning of regular runs as \texttt{get\_luadata'@@'}. 

\begin{verbatim}
if register_luadata then
  register_luadata('_intarray', function()
    local t = '{[0]={},'
    for i=1, #tables do
      t = string.format('%s{%s},', t, table.concat(tables[i], ','))
    end
    return t .. '}'
  end)
end
\end{verbatim}

\texttt{\_intarray\_count:N} Set and get the size of an array. “Setting the size” means in this context that we add zeros until we reach the desired size.

\begin{verbatim}
local sprint = tex.sprint
luacmd('_\_intarray_gset_count:Nw', function()
  local t = _\_intarray_table()
  local n = scan_int()
  for i=#t+1, n do t[i] = 0 end
  end, 'protected', 'global')
luacmd('intarray\_count:N', function()
  sprint(-2, #_\_intarray_table())
  end, 'global')
</lua>
<tex>
cs\_generate\_variant:Nn \intarray\_count:N { c }
</tex>
\end{verbatim}

(End of definition for \texttt{\_\_intarray\_count:N} and \texttt{\_\_intarray\_gset\_count:Nw}. This function is documented on page 256.)
66.1.2 Array items

\__intarray_gset:wF
\__intarray_gset:w

The setter provided by Lua. The argument order somewhat emulates the \fontdimen:
First the array index, followed by the intarray and then the new value. This has been
chosen over a more conventional order to provide a delimiter for the numbers.

```
local i = scan_int()
local t = __intarray_table()
if t[i] then
    t[i] = scan_int()
    put_next(use_none)
else
    tex.count.l__intarray_bad_index_int = i
    scan_int()
    put_next(use_i)
end
```

(End of definition for \__intarray_gset:wF and \__intarray_gset:w.)

\intarray_gzero:N
\intarray_gzero:c
\__kernel_intarray_gset:Nnn
\__kernel_intarray_gset:cnn

The \__kernel_intarray_gset:Nnn function does not use \int_eval:n, namely its
arguments must be suitable for \int_value:w. The user version checks the position and
value are within bounds.

```
\cs_new_protected:Npn \__kernel_intarray_gset:Nnn #1#2#3
    { \__intarray_gset:w #2 #1 #3 \scan_stop: }
\cs_new_protected:Npn \intarray_gset:Nnn #1#2#3
    { \__intarray_gset:wF \int_eval:n {#2} #1 \int_eval:n{#3} }
\cs_generate_variant:Nn \intarray_gset:Nnn { c }
```

(End of definition for \intarray_gset:Nnn and \__kernel_intarray_gset:Nnn. This function is docu-
mented on page 256.)

\intarray_gzero:N
\intarray_gzero:c

Set the appropriate array entry to zero. No bound checking needed.

```
local t = __intarray_table()
for i=1, #t do
t[i] = 0
```

1007
Get the appropriate entry and perform bound checks. The \_\_kernel_intarray\_item:Nn function omits bound checks and omits \inteval:n, namely its argument must be a \TeX{} integer suitable for \intvalue:w.

\newcommand{\__kernel_intarray_item:Nn}{\__intarray_item:w #2 #1}
\newcommand{\intarray_item:Nn}{\__intarray_item:wF \inteval:n {#2} #1}
\generate_variant{\intarray_item:Nn}{c}

(End of definition for \intarray_gzero:N. This function is documented on page 255.)

\newcommand{\intarray_rand_item:N}{\__kernel_intarray_rand_item:NN}
\newcommand{\intarray_rand_item:c}{\__kernel_intarray_rand_item:NC}
\generate_variant{\intarray_item:Nn}{c}

(End of definition for \intarray_item:Nn and others. This function is documented on page 256.)

Importantly, \intarray_item:Nn only evaluates its argument once.
66.1.3 Working with contents of integer arrays

We use the \texttt{\_kernel\_intarray\_gset:Nnn} which does not do bounds checking and instead automatically resizes the array. This is not implemented in Lua to ensure that the clist parsing is consistent with the clist module.

\begin{verbatim}
\cs_new_protected:Npn \intarray_const_from_clist:Nn #1#2
\{ \__intarray_new:N #1 \int_zero:N \l__intarray_loop_int \clist_map_inline:nn {#2}
\{ \__kernel_intarray_gset:Nnn \l__intarray_loop_int { \int_eval:n {##1} } \}
\}\cs_generate_variant:Nn \intarray_const_from_clist:Nn { c }
\end{verbatim}

(End of definition for \texttt{\intarray\_const\_from\_clist:Nn}. This function is documented on page 255.)

\begin{verbatim}
\__intarray_to_clist:Nn \__intarray_to_clist:w
The \texttt{\_intarray\_to\_clist:Nn} auxiliary allows to choose the delimiter and is also used in \texttt{\intarray\_show:N}. Here we just pass the information to Lua and let \texttt{table.concat} do the actual work. We discard the category codes of the passed delimiter but this is not an issue since the delimiter is always just a comma or a comma and a space. In both cases \texttt{sprint(2, \ldots)} provides the right catcodes.

\begin{verbatim}
\langle/\text\rangle \langle/*\text*/\rangle
local concat = table.concat
\protectluacmd{'\_intarray\_to\_clist:Nn', function()
local t = \_intarray\_table()
local sep = \texttt{token.scan\_string()}\texttt{sprint(-2, \texttt{concat(t, sep)})}
\texttt{end, 'global')}
\end{verbatim}

(End of definition for \texttt{\_intarray\_const\_from\_clist:Nn} and \texttt{\_intarray\_to\_clist:w}.)

\begin{verbatim}
\__kernel\_intarray\_range\_to\_clist:Nnn \__intarray\_range\_to\_clist:w
Loop through part of the array.
\begin{verbatim}
\langle/\text\rangle \langle/*\text*/\rangle
\protectluacmd{'\_intarray\_range\_to\_clist:w', function()
local t = \_intarray\_table()
local from = \texttt{scan\_int()}\texttt{local to = \texttt{scan\_int()}}\texttt{sprint(-2, \texttt{concat(t, ', ', from, to))}
\end{verbatim}

1009
Loop through part of the array. We allow additional commas at the end.

\begin{verbatim}
\cs_new_protected:Npn \__kernel_intarray_gset_range_from_clist:Nnn #1#2#3
{\__intarray_gset_range:w \int_eval:w #2 #1 #3 , , \scan_stop: }
\end{verbatim}

In order to allow some code sharing later we provide the \_\_intarray_gset_overflow_test:nw name here. It doesn’t actually test anything since the Lua implementation accepts all integers which could be tested with \tex_ifabsnum:D.

\begin{verbatim}
\cs_new_protected:Npn \__kernel_intarray_gset_overflow_test:nw #1
{ }
\end{verbatim}

\section{Font dimension based implementation}

Go to the false branch of the conditional above.

\begin{verbatim}
\{ }
\end{verbatim}
66.2.1 Allocating arrays

\_\texttt{\_\_intarray\_entry:w} We use these primitives quite a lot in this module.
\_\texttt{\_\_intarray\_count:w}
\cs_new_eq:NN \_\texttt{\_\_intarray\_entry:w} \tex_fontdimen:D
\cs_new_eq:NN \_\texttt{\_\_intarray\_count:w} \tex_hyphenchar:D

(End of definition for \_\texttt{\_\_intarray\_entry:w} and \_\texttt{\_\_intarray\_count:w}.)
\c\texttt{\_\_intarray\_sp\_dim} Used to convert integers to dimensions fast.
\dim_const:Nn \c\texttt{\_\_intarray\_sp\_dim} { 1 sp }

(End of definition for \c\texttt{\_\_intarray\_sp\_dim}.)
\g\texttt{\_\_intarray\_font\_int} Used to assign one font per array.
\int_new:N \g\texttt{\_\_intarray\_font\_int}

(End of definition for \g\texttt{\_\_intarray\_font\_int}.)
\texttt{\_\_intarray\_new:Nn} Declare \#1 to be a font (arbitrarily \texttt{cmr10} at a never-used size). Store the array’s size as the \texttt{hyphenchar} of that font and make sure enough \texttt{fontdimen} are allocated, by setting the last one. Then clear any \texttt{fontdimen} that \texttt{cmr10} starts with. It seems LuaTeX’s \texttt{cmr10} has an extra \texttt{fontdimen} parameter number 8 compared to other engines (for a math font we would replace 8 by 22 or some such). Every \texttt{intarray} must be global; it’s enough to run this check in \texttt{\_\_intarray\_new:Nn}.
\cs_new_protected:Npn \_\_intarray\_new:N #1

\__kernel\_chk\_if\_free\_cs:N #1
\_\texttt{\_\_intarray\_font\_int}
\tex_global:D \tex_font:D #1
= \texttt{cmr10-at-} \_\texttt{\_\_intarray\_font\_int} \_\texttt{\_\_intarray\_sp\_dim} \scan_stop:
\int_step_inline:nn { 8 }
\{ \_\texttt{\_\_kernel\_intarray\_gset:Nnn} \#1 {##1} \_\texttt{c_zero_int} \}
\}
\cs_new_protected:Npn \_\_intarray\_new:Nn \_\_intarray\_new:cn \_\_intarray\_new:N

\{ \cs_new_protected:Npn \_\_intarray\_new:Nn \_\_intarray\_new:cn \_\_intarray\_new:N
\int_gincr:N \g\texttt{\_\_intarray\_font\_int}
\tex_global:D \tex_font:D \_\_intarray\_count:w #1 = \int_eval:n {#2} \scan_stop:
\int_compare:nNnT { \_\_intarray\_count:N #1 } < 0
\{ \msg_error:nne { kernel } { negative-array-size } { \_\_intarray\_count:N #1 } \}
\int_compare:nNnT { \_\_intarray\_count:N #1 } > 0
\{ \_\texttt{\_\_kernel\_intarray\_gset:Nnn} \#1 \{ 0 \} \}
\}
\cs_generate_variant:Nn \_\_intarray\_new:Nn \_\_intarray\_new:cn \_\_intarray\_new:N

(End of definition for \_\_\_intarray\_new:Nn and \_\_\_intarray\_new:N. This function is documented on page 255.)
\texttt{\_\_intarray\_count:N} Size of an array.
\texttt{\_\_intarray\_count:c}
\cs_new:Nn \_\_\_intarray\_count:N \#1 { \int_value:w \_\_\_intarray\_count:w \#1 }
\cs_generate_variant:Nn \_\_\_intarray\_count:N { c }

(End of definition for \_\_\_intarray\_count:N. This function is documented on page 256.)
66.2.2 Array items

Used when an item to be stored is larger than \c{c_max_dim} in absolute value; it is replaced by ±\c{c_max_dim}.

\begin{verbatim}
\cs_new:Npn \__intarray_signed_max_dim:n \#1
\{ \int_value:w \int_compare:nNnT \#1 < 0 \{ - \} \c{c_max_dim} \}
\end{verbatim}

(End of definition for \__intarray_signed_max_dim:n.)

The functions \intarray_gset:Nnn and \intarray_item:Nn share bounds checking. The T branch is used if \#3 is within bounds of the array \#2.

\begin{verbatim}
\cs_new:Npn \__intarray_bounds:NNnTF \#1\#2\#3
\{ \if_int_compare:w 1 > \#3 \exp_stop_f:
\__intarray_bounds_error:NNnw \#1 \#2 \{\#3\}
\else:
\if_int_compare:w \#3 > \intarray_count:N \#2 \exp_stop_f:
\__intarray_bounds_error:NNnw \#1 \#2 \{\#3\}
\fi:
\fi:
\use_i:nn \}
\cs_new:Npn \__intarray_bounds_error:NNnw \#1\#2\#3\#4 \use_i:nn \#5\#6
\{ \#4
#1 \{ kernel \} \{ out-of-bounds \}
\{ \token_to_str:N \} \{ \#3 \} \{ \intarray_count:N \#2 \}
\#6
\}
\end{verbatim}

(End of definition for \__intarray_bounds:NNnTF and \__intarray_bounds_error:NNnw.)

Set the appropriate \fontdimen. The \__kernel_intarray_gset:Nnn function does not use \int_eval:w, namely its arguments must be suitable for \int_value:w. The user version checks the position and value are within bounds.

\begin{verbatim}
\cs_new_protected:Npn \__kernel_intarray_gset:Nnn \#1\#2\#3
\{ \__intarray_entry:w \#2 \#1 \#3 \c{\_intarray_sp_dim} \}
\cs_new_protected:Npn \intarray_gset:Nnn \#1\#2\#3
\{ \exp_after:wN \__intarray_gset:Nww \exp_after:wN \#1 \exp_after:wN \int_value:w \int_eval:n \#2 \exp_after:wN \int_value:w \int_eval:n \#3 \}
\cs_generate_variant:Nn \intarray_gset:Nnn { c }
\cs_new_protected:Npn \__intarray_gset:NNw \#1\#2 ; \#3
\{ \__intarray_bounds:NNnTF \msg_error:nneee \#1 \{ \#2\}
\__intarray_gset_overflow:nw \#3 \}
\__kernel_intarray_gset:Nnn \#1 \#2 \{ \#3\}
\}
\}
\end{verbatim}

1012
\cs_if_exist:NTF \tex_ifabsnum:D 
{ 
\cs_new_protected:Npn \__intarray_gset_overflow_test:nw #1 
{ 
  \tex_ifabsnum:D #1 > \c_max_dim 
  \exp_after:wN \__intarray_gset_overflow:NNnn 
  \fi: 
} 
\cs_new_protected:Npn \__intarray_gset_overflow_test:nw #1 
{ 
  \if_int_compare:w \int_abs:n {#1} > \c_max_dim 
  \exp_after:wN \__intarray_gset_overflow:NNnn 
  \fi: 
} 
\cs_new_protected:Npn \__intarray_gset_overflow:NNnn #1#2#3#4 
{ 
  \msg_error:nneeee { kernel } { overflow } { \token_to_str:N #2 } {#3} {#4} { \__intarray_signed_max_dim:n {#4} } 
#1 #2 {#3} { \__intarray_signed_max_dim:n {#4} } 
} 
\cs_generate_variant:Nn \intarray_gset:Nnn { c } 
\intarray_gzero:N \intarray_gzero:c 
Set the appropriate \fontdimen to zero. No bound checking needed. The \prg-_ 
replicate:nn possibly uses quite a lot of memory, but this is somewhat comparable 
to the size of the array, and it is much faster than an \int_step_inline:nn loop. 
\cs_new_protected:Npn \intarray_gzero:N #1 
{ 
  \int_zero:N \l__intarray_loop_int 
  \prg_replicate:nn { \intarray_count:N #1 } 
  { 
    \int_incr:N \l__intarray_loop_int 
    \__intarray_entry:w \l__intarray_loop_int #1 \c_zero_dim 
  } 
} 
\cs_generate_variant:Nn \intarray_gzero:N { c } 
\intarray_item:Nn \intarray_item:cn \__kernel_intarray_item:Nn \__intarray_item:Nw 
Get the appropriate \fontdimen and perform bound checks. The \__kernel-_ 
\intarray_item:Nn function omits bound checks and omits \int_eval:n, namely its 
argument must be a \TeX integer suitable for \int_value:w. 
\cs_new:Npn \__kernel_intarray_item:Nn {#1#2} 
{ \int_value:w \__intarray_entry:w {#1} #2 } 
\cs_new:Npn \intarray_item:Nn #1#2 
{ \exp_after:wN \__intarray_item:Nw #1 \exp_after:wN \intarray_item:Nw #2 } 
\cs_generate_variant:Nn \intarray_item:Nn { c } 

(End of definition for \intarray_gset:Nnn and others. This function is documented on page 256.) 

(End of definition for \intarray_gzero:N. This function is documented on page 255.) 

1013
66.2.3 Working with contents of integer arrays

Similar to `\intarray_new:N` (which we don’t use because when debugging is enabled that function checks the variable name starts with `g_`). We make use of the fact that \TeX allows allocation of successive `\fontdimen` as long as no other font has been declared: no need to count the comma list items first. We need the code in `\intarray_gset:Nnn` that checks the item value is not too big, namely `\__intarray_gset_overflow_test:nw`, but not the code that checks bounds. At the end, set the size of the intarray.

Loop through the array, putting a comma before each item. Remove the leading comma with \texttt{f}-expansion. We also use the auxiliary in `\intarray_show:N` with argument comma, space.

(End of definition for `\intarray_const_from_clist:N` and `\__intarray_const_from_clist:nW`. This function is documented on page 255.)
\cs_new:Npn \__intarray_to_clist:w #1 ; #2#3 
\{ 
\if_int_compare:w #1 > \__intarray_count:w #2 
\prg_break:n 
\fi:  
\#3 \__kernel_intarray_item:Nn #2 {#1}  
\exp_after:wN \__intarray_to_clist:w  
\int_value:w \int_eval:w #1 + \c_one_int ; #2 {#3} 
\}

(End of definition for \__intarray_to_clist:Nn and \__intarray_to_clist:w)

\cs_new:Npn \__kernel_intarray_range_to_clist:Nnn #1#2#3 
\{  
\exp_last_unbraced:Nf \use_none:n 
\{  
\exp_after:wN \__intarray_range_to_clist:ww  
\int_value:w \int_eval:w #2 \exp_after:wN ; 
\int_value:w \int_eval:w #3 ; 
#1 \prg_break_point:  
\}  
\}
\cs_new:Npn \__intarray_range_to_clist:ww #1 ; #2 ; #3 
\{  
\if_int_compare:w #1 > #2 \exp_stop_f:  
\prg_break:n  
\fi:  
, \__kernel_intarray_item:Nn #3 {#1}  
\exp_after:wN \__intarray_range_to_clist:ww  
\int_value:w \int_eval:w #1 + \c_one_int ; #2 ; #3  
\}

(End of definition for \__kernel_intarray_range_to_clist:Nnn and \__intarray_range_to_clist:ww)

\cs_new:Npn \__kernel_intarray_gset_range_from_clist:Nnn #1#2#3 
\{  
\int_set:Nn \l__intarray_loop_int {#2}  
\__intarray_gset_range:Nw #1 #3 , , \prg_break_point:  
\}
\cs_new_protected:Npn \__intarray_gset_range:Nw #1 #2 ,  
\{  
\if_catcode:w \scan_stop: \tl_to_str:n {#2} \scan_stop:  
\prg_break:n  
\fi:  
\__kernel_intarray_gset:Nnn #1 \l__intarray_loop_int {#2}  
\int_incr:N \l__intarray_loop_int  
\__intarray_gset_range:Nw #1  
\}

(End of definition for \__kernel_intarray_gset_range_from_clist:Nnn and \__intarray_gset_range:Nw)
66.3 Common parts

Copies of the cs functions defined in l3basics.

Convert the list to a comma list (with spaces after each comma)

(EXIT of definition for \intarray_if_exist:NTF. This function is documented on page 256.)

(EXIT of definition for \intarray_show:N and \intarray_log:N. These functions are documented on page 256.)
Chapter 67

l3fp implementation

Nothing to see here: everything is in the subfiles!
Chapter 68

l3fp-aux implementation

Largely for performance reasons, we need to directly access primitives rather than use \int_eval:n. This happens a lot, so we use private names. The same is true for \romannumeral, although it is used much less widely.

Internally, a floating point number \(X\) is a token list containing \s__fp \__fp_chk:w \langle case \rangle \langle sign \rangle \langle body \rangle ; \nl_s__fp \__fp_chk:w \langle case \rangle \langle sign \rangle \langle body \rangle ; \n
Let us explain each piece separately.

Internal floating point numbers are used in expressions, and in this context are subject to f-expansion. They must leave a recognizable mark after f-expansion, to prevent the floating point number from being re-parsed. Thus, \s__fp is simply another name for \relax.

When used directly without an accessor function, floating points should produce an error: this is the role of \__fp_chk:w. We could make floating point variables be protected to prevent them from expanding under \e/x-expansion, but it seems more convenient to treat them as a subcase of token list variables.

The (decimal part of the) IEEE-754-2008 standard requires the format to be able to represent special floating point numbers besides the usual positive and negative cases. We distinguish the various possibilities by their \langle case \rangle, which is a single digit:

0 zeros: +0 and -0,
1 “normal” numbers (positive and negative),
Table 3: Internal representation of floating point numbers.

<table>
<thead>
<tr>
<th>Representation</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 0 \s__fp_... ;</td>
<td>Positive zero.</td>
</tr>
<tr>
<td>0 2 \s__fp_... ;</td>
<td>Negative zero.</td>
</tr>
<tr>
<td>1 0 {\langle \text{exponent} \rangle } {\langle X_1 \rangle } {\langle X_3 \rangle } {\langle X_4 \rangle } ;</td>
<td>Positive floating point.</td>
</tr>
<tr>
<td>1 2 {\langle \text{exponent} \rangle } {\langle X_1 \rangle } {\langle X_3 \rangle } {\langle X_4 \rangle } ;</td>
<td>Negative floating point.</td>
</tr>
<tr>
<td>2 0 \s__fp_... ;</td>
<td>Positive infinity.</td>
</tr>
<tr>
<td>2 2 \s__fp_... ;</td>
<td>Negative infinity.</td>
</tr>
<tr>
<td>3 1 \s__fp_... ;</td>
<td>Quiet \text{n}an.</td>
</tr>
<tr>
<td>3 1 \s__fp_... ;</td>
<td>Signalling \text{n}an.</td>
</tr>
</tbody>
</table>

2 infinities: +\text{inf} and −\text{inf},
3 quiet and signalling \text{n}an.

The \text{\langle sign \rangle} is 0 (positive) or 2 (negative), except in the case of \text{n}an, which have \text{\langle sign \rangle} = 1. This ensures that changing the \text{\langle sign \rangle} digit to 2 − \text{\langle sign \rangle} is exactly equivalent to changing the sign of the number.

Special floating point numbers have the form
\s__fp_... where \s__fp_... is a scan mark carrying information about how the number was formed (useful for debugging).

Normal floating point numbers (\text{\langle case \rangle} = 1) have the form
\s__fp_... \langle \text{exponent} \rangle \{\langle X_1 \rangle \} \{\langle X_2 \rangle \} \{\langle X_3 \rangle \} \{\langle X_4 \rangle \} ;

Here, the \text{\langle exponent \rangle} is an integer, between −10000 and 10000. The body consists in four blocks of exactly 4 digits, 0000 ≤ \text{\langle X_i \rangle} ≤ 9999, and the floating point is
\((-1)^{\langle sign \rangle/2}\langle X_1 \rangle\langle X_2 \rangle\langle X_3 \rangle\langle X_4 \rangle \cdot 10^{\langle exponent \rangle-16}

where we have concatenated the 16 digits. Currently, floating point numbers are normalized such that the \text{\langle exponent \rangle} is minimal, in other words, 1000 ≤ \text{\langle X_1 \rangle} ≤ 9999.

Calculations are done in base 10000, \text{i.e.} one myriad.

68.3 Using arguments and semicolons

\_\_fp_use_none_stop_f:n This function removes an argument (typically a digit) and replaces it by \exp_stop_f:, a marker which stops f-type expansion.
\_\_fp_use_s:n \_\_fp_use_s:nn Those functions place a semicolon after one or two arguments (typically digits).

\_\_fp_use_s:nn \_\_fp_use_s:nn #1 \#2 \_\_fp_use_s:nn #1 \#2 \_\_fp_use_s:nn #1\#2 \_\_fp_use_s:nn #1\#2

(End of definition for \_\_fp_use_s:nn and \_\_fp_use_s:nn.)
Those functions select specific arguments among a set of arguments delimited by a semicolon.

\cs_new:Npn \_fp_use_none_until_s:w #1; { }
\cs_new:Npn \_fp_use_i_until_s:nw #1#2; {#1}
\cs_new:Npn \_fp_use_ii_until_s:nnw #1#2#3; {#2}

(End of definition for \_fp_use_none_until_s:w, \_fp_use_i_until_s:nw, and \_fp_use_ii_until_s:nnw.)

\_fp_reverse_args:Nww

Many internal functions take arguments delimited by semicolons, and it is occasionally useful to swap two such arguments.

\cs_new:Npn \_fp_reverse_args:Nww #1 #2; #3; { #1 #3; #2; }

(End of definition for \_fp_reverse_args:Nww.)

\_fp_rrot:www

Rotate three arguments delimited by semicolons. This is the inverse (or the square) of the Forth primitive \texttt{ROT}, hence the name.

\cs_new:Npn \_fp_rrot:www #1; #2; #3; { #2; #3; #1; }

(End of definition for \_fp_rrot:www.)

\_fp_use_i:ww \_fp_use_i:www

Many internal functions take arguments delimited by semicolons, and it is occasionally useful to remove one or two such arguments.

\cs_new:Npn \_fp_use_i:ww #1; #2; #3; { #1; }
\cs_new:Npn \_fp_use_i:www #1; #2; #3; { #1; }

(End of definition for \_fp_use_i:ww and \_fp_use_i:www.)

\section{Constants, and structure of floating points}

\_fp_misused:n

This receives a floating point object (floating point number or tuple) and generates an error stating that it was misused. This is called when for instance an \texttt{fp} variable is left in the input stream and its contents reach \TeX's stomach.

\cs_new_protected:Npn \_fp_misused:n #1
\{ \msg_error:nne { fp } { misused } { \fp_to_tl:n {#1} } \}

(End of definition for \_fp_misused:n.)

\_s_fp \_fp_chk:w

Floating points numbers all start with \_s_fp \_fp_chk:w, where \_s_fp is equal to the \TeX primitive \texttt{relax}, and \_fp_chk:w is protected. The rest of the floating point number is made of characters (or \texttt{relax}). This ensures that nothing expands under f-expansion, nor under e/x-expansion. However, when typeset, \_s_fp does nothing, and \_fp_chk:w is expanded. We define \_fp_chk:w to produce an error.

\scan_new:N \_s_fp
\cs_new_protected:Npn \_fp_misused:n #1
\{ \msg_error:nne { fp } { misused } { \fp_to_tl:n {#1} } \}

(End of definition for \_s_fp and \_fp_chk:w.)

\_s_fp_expr_mark \_s_fp_expr_stop

Aliases of \texttt{\textbackslash tex\_relax:D}, used to terminate expressions.

\scan_new:N \_s_fp_expr_mark
\scan_new:N \_s_fp_expr_stop

(End of definition for \_s_fp_expr_mark and \_s_fp_expr_stop.)
Generic scan marks used throughout the module.

Functions to gobble up to a scan mark.

A couple of scan marks used to indicate where special floating point numbers come from.

The special floating points. We define the floating points here as “exact”.

The number of digits of floating points.

Blocks have 4 digits so this integer is useful.

Normal floating point numbers have an exponent between $-\text{minus} \text{_min} \text{_exponent}$ and $\text{max} \text{_exponent}$ inclusive. Larger numbers are rounded to $\pm \infty$. Smaller numbers are rounded to $\pm 0$. It would be more natural to define a $\text{min} \text{_exponent}$ with the opposite sign but that would waste one \TeX{} count.

If a number’s exponent is larger than that, its exponential overflows/underflows.

(End of definition for $\backslash \text{c} \_\text{fp} \_\text{prec} \_\text{int}$, $\backslash \text{c} \_\text{fp} \_\text{half} \_\text{prec} \_\text{int}$, and $\backslash \text{c} \_\text{fp} \_\text{block} \_\text{int}$.)

(End of definition for $\backslash \text{c} \_\text{fp} \_\text{myriad} \_\text{int}$.)

(End of definition for $\backslash \text{c} \_\text{fp} \_\text{min} \_\text{us} \text{minus} \text{_min} \text{_exponent} \_\text{int}$ and $\backslash \text{c} \_\text{fp} \_\text{max} \_\text{exponent} \_\text{int}$.)

(End of definition for $\backslash \text{c} \_\text{fp} \_\text{max} \_\text{exp} \_\text{exponent} \_\text{int}$.)
\c__fp_overflowing_fp A floating point number that is bigger than all normal floating point numbers. This replaces infinities when converting to formats that do not support infinities.

\tl_const:Ne \c__fp_overflowing_fp
{\n \s__fp \__fp_chk:w 1 0
{\int_eval:n { \c__fp_max_exponent_int + 1 } }
{1000} {0000} {0000} {0000} ;
}

(End of definition for \c__fp_overflowing_fp.)

\___fp_zero_fp:N In case of overflow or underflow, we have to output a zero or infinity with a given sign.
\cs_new:Npn \__fp_zero_fp:N #1
{\s__fp \__fp_chk:w 0 #1 \s__fp_underflow ;}
\cs_new:Npn \__fp_inf_fp:N #1
{\s__fp \__fp_chk:w 2 #1 \s__fp_overflow ;}

(End of definition for \__fp_zero_fp:N and \__fp_inf_fp:N.)

\___fp_exponent:w For normal numbers, the function expands to the exponent, otherwise to 0. This is used in l3str-format.
\cs_new:Npn \__fp_exponent:w \s__fp \__fp_chk:w #1
{\if_meaning:w 1 #1 \exp_after:wN \__fp_use_ii_until_s:nnw \else: \exp_after:wN \__fp_use_i_until_s:nw \exp_after:wN 0 \fi:}

(End of definition for \__fp_exponent:w.)

\___fp_neg_sign:N When appearing in an integer expression or after \int_value:w, this expands to the sign opposite to #1, namely 0 (positive) is turned to 2 (negative), 1 (nan) to 1, and 2 to 0.
\cs_new:Npn \__fp_neg_sign:N #1
{\__fp_int_eval:w 2 - #1 \__fp_int_eval_end:}

(End of definition for \__fp_neg_sign:N.)

\___fp_kind:w Expands to 0 for zeros, 1 for normal floating point numbers, 2 for infinities, 3 for nan, 4 for tuples.
\cs_new:Npn \__fp_kind:w #1
{\__fp_if_type_fp:NTwFw
 #1 \__fp_use_ii_until_s:nnw
 \s__fp \__fp_use_i_until_s:nw 4 }
\s__fp_stop

(End of definition for \__fp_kind:w.)
68.5 Overflow, underflow, and exact zero

Expects the sign and the exponent in some order, then the significand (which we don’t touch). Outputs the corresponding floating point number, possibly underflowed to ±0 or overflowed to ±∞. The functions \_fp_underflow:w and \_fp_overflow:w are defined in l3fp-traps.

\texttt{\cs_new:Npn \_fp_sanitize:Nw #1 #2;}
\texttt{\if_case:w}
\texttt{\if_int_compare:w #2 > \c__fp_max_exponent_int 1 \else:}
\texttt{\if_int_compare:w #2 < - \c__fp_minus_min_exponent_int 2 \else:}
\texttt{\if_meaning:w 1 #1 3 \fi: \fi: \fi: 0 -}
\texttt{\or: \exp_after:wN \_fp_overflow:w}
\texttt{\or: \exp_after:wN \_fp_underflow:w}
\texttt{\or: \exp_after:wN \_fp_sanitize_zero:w}
\texttt{\fi:}
\texttt{\_fp \_fp_chk:w 1 #1 {#2}}
\texttt{}}
\texttt{\cs_new:Npn \_fp_sanitize:wN #1; #2 { \_fp_sanitize:Nw #2 #1; }}
\texttt{\cs_new:Npn \_fp_sanitize_zero:w \s__fp \_fp_chk:w #1 #2 { \c_zero_fp }}

(End of definition for \_fp_sanitize:Nw, \_fp_sanitize:wN, and \_fp_sanitize_zero:w.)

68.6 Expanding after a floating point number

\_fp_exp_after_o:w \_fp_exp_after_f:nw
\_fp_exp_after_o:w \texttt{(floating point)}
\_fp_exp_after_f:nw \texttt{(tokens)} \texttt{(floating point)}
Places \texttt{(tokens)} (empty in the case of \_fp_exp_after_o:w) between the \texttt{(floating point)} and the following tokens, then hits those tokens with o or f-expansion, and leaves the floating point number unchanged.

We first distinguish normal floating points, which have a significand, from the much simpler special floating points.

\texttt{\cs_new:Npn \_fp_exp_after_o:w \_fp_exp_after_f:nw \_fp_chk:w #1}
\texttt{\{}
\texttt{\if_meaning:w 1 #1}
\texttt{\exp_after:wN \_fp_exp_after_normal:nNw}
\texttt{\else:}
\texttt{\exp_after:wN \_fp_exp_after_special:nNw}
\texttt{\fi:}
\texttt{\}}
\texttt{\cs_new:Npn \_fp_exp_after_f:nw #1 \_fp_chk:w #2}
\texttt{\{}
\texttt{\if_meaning:w 1 #2}
\texttt{\exp_after:wN \_fp_exp_after_normal:nNw}
\texttt{\else:}
\texttt{\exp_after:wN \_fp_exp_after_special:nNw}
\texttt{\fi:}
\texttt{\}}
\texttt{\exp:w \exp_end_continue_f:w #1 \}}
Special floating point numbers are easy to jump over since they contain few tokens.\[
\texttt{\texttt{\__fp\_exp\_after\_special:nNNw}} \{\langle \text{after} \rangle \} \langle \text{case} \rangle \langle \text{sign} \rangle \langle \text{scan mark} \rangle ;
\]

For normal floating point numbers, life is slightly harder, since we have many tokens to jump over. Here it would be slightly better if the digits were not braced but instead were delimited arguments (for instance delimited by \texttt{,}). That may be changed some day.

Floating point tuples take the form \texttt{\__fp\_tuple \__fp\_tuple\_chk:w \{(fp 1) (fp 2) ...\}}; where each \texttt{(fp)} is a floating point number or tuple, hence ends with \texttt{;} itself. When a tuple is typeset, \texttt{\__fp\_tuple\_chk:w} produces an error, just like usual floating point numbers. Tuples may have zero or one element.
Count the number of items in a tuple of floating points by counting semicolons. The technique is very similar to `\tl_count:n`, but with the loop built-in. Checking for the end of the loop is done with the `\use_none:n #1` construction.

```latex
\cs_new:Npn \__fp_tuple_count:w \s__fp_tuple \__fp_tuple_chk:w #1 ;
{ \int_value:w \__fp_int_eval:w 0 \__fp_tuple_count_loop:Nw #1 { ? \prg_break: } ; \prg_break_point: \__fp_int_eval_end: }
\cs_new:Npn \__fp_tuple_count_loop:Nw #1#2;
{ \use_none:n #1 + 1 \__fp_tuple_count_loop:Nw }
```

`\__fp_tuple_count:w`, `\__fp_array_count:n`, and `\__fp_tuple_count_loop:Nw`.

Used as `\__fp_if_type_fp:NTwFw ⟨marker⟩ {⟨true code⟩} \s__fp {⟨false code⟩} \s__fp_stop`, this test whether the ⟨marker⟩ is `\s__fp` or not and runs the appropriate ⟨code⟩. The very unusual syntax is for optimization purposes as that function is used for all floating point operations.

```latex
\cs_new:Npn \__fp_if_type_fp:NTwFw #1 \s__fp #2 #3 \s__fp_stop {#2}
\cs_new:Npe \__fp_if_type_fp:NTwFw #1
{ \__fp_if_type_fp:NTwFw #1 \__fp_if_type_fp:NTwFw \s__fp { \prg_break:n \use_iii:nnn } \s__fp_stop }
```

`\__fp_if_type_fp:NTwFw`.

True if all items are floating point numbers. Used for `min`.

```latex
\cs_new:Npn \__fp_array_if_all_fp:nTF \__fp_array_if_all_fp_loop:w \__fp_array_if_all_fp_loop_other:w \__fp_type_from_scan:w \__fp_type_from_scan_other:w \__fp_type_from_scan: N
\cs_new:Npn \__fp_if_type_fp:NTwFw #1 \s__fp #2 \s__fp_stop {#2}
\cs_new:Npe \__fp_if_type_fp:NTwFw #1
{ \__fp_if_type_fp:NTwFw \s__fp { \__fp_type_from_scan_other:N #1 } \s__fp_stop }
\cs_new:Npe \__fp_if_type_fp:NTwFw #1
{ \__fp_if_type_fp:NTwFw \s__fp { \__fp_type_from_scan: N #1 } \s__fp_stop }
\cs_new:Npn \__fp_if_type_fp:NTwFw #1 \s__fp #2 #3 \s__fp_stop {#2}
\cs_new:Npe \__fp_if_type_fp:NTwFw #1
{ \__fp_if_type_fp:NTwFw \s__fp { \__fp_type_from_scan_other:N #1 } \s__fp_stop }
\cs_new:Npe \__fp_if_type_fp:NTwFw #1
{ \__fp_if_type_fp:NTwFw \s__fp { \__fp_type_from_scan: N #1 } \s__fp_stop }
\cs_new:Npe \__fp_if_type_fp:NTwFw #1
{ \__fp_if_type_fp:NTwFw \s__fp { \__fp_type_from_scan_other:N #1 } \s__fp_stop }
```

`\__fp_if_type_fp:NTwFw`.

Used as `\__fp_type_from_scan:N ⟨token⟩`. Grabs the pieces of the stringified ⟨token⟩ which lies after the first s__fp. If the ⟨token⟩ does not contain that string, the result is _?_.

```latex
\cs_new:Npn \__fp_type_from_scan:N #1
{ \__fp_if_type_fp:NTwFw #1 \s__fp \__fp_type_from_scan_other:N #1 \s__fp_stop }
\cs_new:Npn \__fp_type_from_scan_other:N #1
{ \__fp_if_type_fp:NTwFw \s__fp \__fp_type_from_scan: N #1 \s__fp_stop }
\cs_new:Npn \__fp_type_from_scan:N #1
{ \__fp_if_type_fp:NTwFw \s__fp \__fp_type_from_scan_other:N #1 \s__fp_stop }
\cs_new:Npn \__fp_type_from_scan:w
{ \__fp_if_type_fp:NTwFw \s__fp \__fp_type_from_scan: N \s__fp_stop }
```

`\__fp_type_from_scan:N`, `\__fp_type_from_scan_other:N`, and `\__fp_type_from_scan:w`.

(End of definition for `\__fp_tuple_count:w`, `\__fp_array_count:n`, and `\__fp_tuple_count_loop:Nw`.)
\exp_not:N \token_to_str:N \s__fp_mark
\tl_to_str:n { \s__fp } \s__fp_mark \s__fp_stop
}
\exp_last_unbraced:NNNNo
\cs_new:Npn \__fp_type_from_scan:w #1
{ \tl_to_str:n { s__fp } } #2 \s__fp_mark #3 \s__fp_stop {#2}
(End of definition for \__fp_type_from_scan:N, \__fp_type_from_scan_other:N, and \__fp_type_from_scan:w.)
\__fp_change_func_type:NNN \__fp_change_func_type_aux:w \__fp_change_func_type_chk:NNN
Arguments are ⟨type marker⟩ ⟨function⟩ ⟨recovery⟩. This gives the function obtained by placing the type after @@. If the function is not defined then ⟨recovery⟩ ⟨function⟩ is used instead; however that test is not run when the ⟨type marker⟩ is \s__fp.
\cs_new:Npn \__fp_change_func_type:NNN #1#2#3
{ \__fp_if_type_fp:NTwFw #1 #2 \s__fp
{ \exp_after:wN \__fp_change_func_type_chk:NNN \cs:w __fp \__fp_type_from_scan_other:N #1
\exp_after:wN \__fp_change_func_type_aux:w \token_to_str:N #2 \cs_end:
#2 #3
}\s__fp_stop
}\exp_last_unbraced:NNNNo
\cs_new:Npn \__fp_change_func_type_aux:w #1 { \tl_to_str:n { __fp } } { }
\cs_new:Npn \__fp_change_func_type_chk:NNN #1#2#3
{ \if_meaning:w \scan_stop: #1 \exp_after:wN #3 \exp_after:wN #2 \else:
\exp_after:wN #1 \fi:
}
(End of definition for \__fp_change_func_type:NNN, \__fp_change_func_type_aux:w, and \__fp_change_func_type_chk:NNN.)
\__fp_exp_after_any_f:Nnw \__fp_exp_after_any_f:nw \__fp_exp_after_expr_stop_f:nw
The \Nnw function simply dispatches to the appropriate \_\_fp_exp_after..._f:nw with “...” (either empty or ⟨type⟩) extracted from #1, which should start with ⟨s__fp⟩. If it doesn’t start with ⟨s__fp⟩ the function \_\_fp_exp_after_?_f:nw defined in l3fp-parse gives an error; another special ⟨type⟩ is stop, useful for loops, see below. The \Nnw function has an important optimization for floating points numbers; it also fetches its type marker #2 from the floating point.
\cs_new:Npn \__fp_exp_after_any_f:Nnw #1\__fp_exp_after_any_f::nw #1\__fp_exp_after_expr_stop_f:nw
{ \cs:w \_\_fp_exp_after \__fp_type_from_scan_other:N \_\_fp_exp_after\_\_fp_type_from_scan_other:N \__fp_exp_after\_\_fp_type_from_scan_other:N \_\_fp_exp_after\_\_fp_type_from_scan_other:N \token_to_str:N \_\_fp_exp_after\_\_fp_type_from_scan_other:N \token_to_str:N \_\_fp_exp_after\_\_fp_type_from_scan_other:N \token_to_str:N \_\_fp_exp_after\_\_fp_type_from_scan_other:N \token_to_str:N \_\_fp_exp_after\_\_fp_type_from_scan_other:N \token_to_str:N \_\_fp_exp_after\_\_fp_type_from_scan_other:N \token_to_str:N \_\_fp_exp_after\_\_fp_type_from_scan_other:N \token_to_str:N \_\_fp_exp_after\_\_fp_type_from_scan_other:N \token_to_str:N \_\_fp_exp_after\_\_fp_type_from_scan_other:N \token_to_str:N \_\_fp_exp_after\_\_fp_type_from_scan_other:N \token_to_str:N \_\_fp_exp_after\_\_fp_type_from_scan_other:N \token_to_str:N \_\_fp_exp_after\_\_fp_type_from_scan_other:N \token_to_str:N \_\_fp_exp_after\_\_fp_type_from_scan_other:N \token_to_str:N \_\_fp_exp_after\_\_fp_type_from_scan_other:N \token_to_str:N \_\_fp_exp_after\_\_fp_type_from_scan_other:N \token_to_str:N \_\_fp_exp_after\_\_fp_type_from_scan_other:N \token_to_str:N \_\_fp_exp_after\_\_fp_type_from_scan_other:N \token_to_str:N \_\_fp_exp_after\_\_fp_type_from_scan_other:N \token_to_str:N \_\_fp_exp_after\_\_fp_type_from_scan_other:N \token_to_str:N \_\_fp_exp_after\_\_fp_type_from_scan_other:N \token_to_str:N \_\_fp_exp_after\_\_fp_type_from_scan_other:N \token_to_str:N \_\_fp_exp_after\_\_fp_type_from_scan_other:N \token_to_str:N \_\_fp_exp_after\_\_fp_type_from_scan_other:N \token_to_str:N \_\_fp_exp_after\_\_fp_type_from_scan_other:N \token_to_str:N \_\_fp_exp_after\_\_fp_type_from_scan_other:N \token_to_str:N \_\_fp_exp_after\_\_fp_type_from_scan_other:N \token_to_str:N \_\_fp_exp_after\_\_fp_type_from_scan_other:N \token_to_str:N \_\_fp_exp_after\_\_fp_type_from_scan_other:N \token_to_str:N \_\_fp_exp_after\_\_fp_type_from_scan_other:N \token_to_str:N \_\_fp_exp_after\_\_fp_type_from_scan_other:N \token_to_str:N \_\_fp_exp_after\_\_fp_type_from_scan_other:N \token_to_str:N \_\_fp_exp_after\_\_fp_type_from_scan_other:N \token_to_str:N \_\_fp_exp_after\_\_fp_type_from_scan_other:N \token_to_str:N \_\_fp_exp_after\_\_fp_type_from_scan_other:N \token_to_str:N \_\_fp_exp_after\_\_fp_type_from_scan_other:N \token_to_str:N \_\_fp_exp_after\_\_fp_type_from_scan_other:N \token_to_str:N \_\_fp_exp_after\_\_fp_type_from_scan_other:N \token_to_str:N \_\_fp_exp_after\_\_fp_type_from_scan_other:N \token_to_str:N \_\_fp_exp_after\_\_fp_type_from_scan_other:N \token_to_str:N \_\_fp_exp_after\_\_fp_type_from_scan_other:N \token_to_str:N \_\_fp_exp_after\_\_fp_type_from_scan_other:N \token_to_str:N \_\_fp_exp_after\_\_fp_type_from_scan_other:N \token_to_str:N \_\_fp_exp_after\_\_fp_type_from_scan_other:N \token_to_str:N \_\_fp_exp_after\_\_fp_type_from_scan_other:N \token_to_str:N \_\_fp_exp_after\_\_fp_type_from_scan_other:N \token_to_str:N \_\_fp_exp_after\_\_fp_type_from_scan_other:N \token_to_str:N \_\_fp_exp_after\_\_fp_type_from_scan_other:N \token_to_str:N \_\_fp_exp_after\_\_fp_type_from_scan_other:N \token_to_str:N \_\_fp_exp_after\_\_fp_type_from_scan_other:N \token_to_str:N \_\_fp_exp_after\_\_fp_type_from_scan_other:N \token_to_str:N \_\_fp_exp_after\_\_fp_type_from_scan_other:N \token_to_str:N \_\_fp_exp_after\_\_fp_type_from_scan_other:N \token_to_str:N \_\_fp_exp_after\_\_fp_type_from_scan_other:N \token_to_str:N \_\_fp_exp_after\_\_fp_type_from_scan_other:N \token_to_str:N \_\_fp_exp_after\_\_fp_type_from_scan_other:N \token_to_str:N \_\_fp_exp_after\_\_fp_type_from_scan_other:N \token_to_str:N \_\_fp_exp_after\_\_fp_type_from_scan_other:N \token_to_str:N \_\_fp_exp_after\_\_fp_type_from_scan_other:N \token_to_str:N \_\_fp_exp_after\_\_fp_type_from scanned
The loop works by using the \textit{n} argument of \texttt{\_\_fp_exp_after_any_f:nw} to place the loop macro after the next item in the tuple and expand it.

\begin{verbatim}
\__fp_exp_after_array_f:w
\__fp_exp_after_tuple_f:nw
\__fp_exp_after_tuple_o:w
\__fp_exp_after_expr_stop_f:nn
\use_none:nn
\end{verbatim}

The trick for packing digits is to add $10^8$ to the number, ensuring it has exactly 9 digits, and can then easily find which digits correspond to what position in the number. Of course, this can be modified for any number of digits less or equal to 9 (we are limited by \TeX's integers). This method is very heavily relied upon in \texttt{l3fp-basics}.

More specifically, the auxiliary inserts \texttt{+ #1#2#3#4#5 ; \{#6\}}, which allows us to compute several blocks of 4 digits in a nested manner, performing carries on the fly. Say we want to compute $12345 \times 66778899$. With simplified names, we would do

\begin{verbatim}
\cs_new:Npn \pack:NNNNNw #1 #2#3#4#5 #6; { \{#2#3#4#5\} \{#6\} }
\exp_after:wN \pack:NNNNNw
\__fp_int_value:w \__fp_int_eval:w 1 0000 0000 + #1 ;
\end{verbatim}
The \exp_after:wN \int_value:w \__fp_int_eval:w, which starts a first computation, whose initial value is $-5,0000$ (the “leading shift”). In that computation appears an \exp_after:wN \int_value:w \__fp_int_eval:w with starting value $4,9995\,0000$ (the “middle shift”). That, in turn, expands \exp_after:wN \int_value:w \__fp_int_eval:w which triggers the third computation. The third computation’s value is $5,0000\,0000 + 12345 \times 8899$, which has 9 digits. Adding $5 \cdot 10^8$ to the product allowed us to know how many digits to expect as long as the numbers to multiply are not too big; it also works to some extent with negative results. The \pack function puts the last 4 of those 9 digits into a brace group, moves the semi-colon delimiter, and inserts a +, which combines the carry with the previous computation. The shifts nicely combine into $5,0000\,0000 + 12345 \times 8899$, which has 9 digits. The “leading shift” cancels the combination of the other shifts, and the \post_processing:w takes care of packing the last few digits.

Admittedly, this is quite intricate. It is probably the key in making \lpf as fast as other pure \TeX floating point units despite its increased precision. In fact, this is used so much that we provide different sets of packing functions and shifts, depending on ranges of input.

This set of shifts allows for computations involving results in the range $[-4 \cdot 10^8, 5 \cdot 10^8 - 1]$. Shifted values all have exactly 9 digits.

This set of shifts allows for computations involving results in the range $[-5 \cdot 10^8, 6 \cdot 10^8 - 1]$ (actually a bit more). Shifted values all have exactly 10 digits. Note that the upper bound is due to \TeX’s limit of $2^{31} - 1$ on integers. The shifts are chosen to be roughly the mid-point of $10^9$ and $2^{31}$, the two bounds on 10-digit integers in \TeX.
This set of shifts allows for computations with results in the range \([-1 \cdot 10^9, 147483647]\); the end-point is \(2^{31} - 1 - 2 \cdot 10^9 \approx 1.47 \cdot 10^8\). Shifted values all have exactly 10 digits.

Grabs two sets of 4 digits and places them before the semi-colon delimiter. Putting several copies of this function before a semi-colon packs more digits since each takes the digits packed by the others in its first argument.

Grabs one set of 8 digits and places them before the semi-colon delimiter as a single group. Putting several copies of this function before a semi-colon packs more digits since each takes the digits packed by the others in its first argument.

Addition and multiplication of significands are done in two steps: first compute a (more or less) exact result, then round and pack digits in the final (braced) form. These functions take care of the packing, with special attention given to the case where rounding has caused a carry. Since rounding can only shift the final digit by 1, a carry always produces an exact power of 10. Thus, \(__fp_basics_pack_high_carry:w\) is always followed by four times \(\{0000\}\).

This is used in l3fp-basics and l3fp-extended.

(End of definition for \(__fp_pack_Bigg:NNNNNNw\) and others.)

(End of definition for \(__fp_pack_twice_four:wNNNNNNNN\).)

(End of definition for \(__fp_pack_eight:wNNNNNNNN\).)

(End of definition for \(__fp_basics_pack_low:NNNNNw\), \(__fp_basics_pack_high:NNNNNw\), and \(__fp的基本包装高左移:w\).)
This is used in \texttt{l3fp-basics} for additions and divisions. Their syntax is confusing, hence the name.

\begin{verbatim}
\cs_new:Npn \__fp_basics_pack_weird_low:NNNNw #1 #2#3#4 #5; 
{ \if_meaning:w 2 #1 + 1 \fi: \__fp_int_eval_end: \__fp_int_eval_end: #2#3#4; {#5} ; }
\cs_new:Npn \__fp_basics_pack_weird_high:NNNNNNNNw 1 #1#2#3#4 #5#6#7#8 #9; { ; {#1#2#3#4} {#5#6#7#8} {#9} }
\end{verbatim}

(End of definition for \__fp_basics_pack_weird_low:NNNNw and \__fp_basics_pack_weird_high:NNNNNNNNw.)

68.9 Decimate (dividing by a power of 10)

\begin{verbatim}
\__fp_decimate:nNnnnn {\langle shift\rangle} {\langle f_1\rangle}
\{\langle x_1\rangle\} \{\langle x_2\rangle\} \{\langle x_3\rangle\} \{\langle x_4\rangle\}
Each \langle x_i\rangle consists in 4 digits exactly, and 1000 \leq \langle x_1\rangle < 9999. The first argument determines by how much we shift the digits. \langle f_1\rangle is called as follows:
\langle f_1\rangle \langle\text{rounding}\rangle \{\langle x_1'\rangle\} \{\langle x_2'\rangle\} \{\text{extra-digits}\};
where 0 \leq \langle x_1'\rangle < 10^8 - 1 are 8 digit integers, forming the truncation of our number. In other words,
\[
\left(\sum_{i=1}^{4} \langle x_i\rangle \cdot 10^{-4i} \cdot 10^{-\langle shift\rangle}\right) - (\langle x_1'\rangle \cdot 10^{-8} + \langle x_2'\rangle \cdot 10^{-16}) = 0, (\text{extra-digits}) \cdot 10^{-16} \in [0, 10^{-16}).
\]
To round properly later, we need to remember some information about the difference. The \langle\text{rounding}\rangle digit is 0 if and only if the difference is exactly 0, and 5 if and only if the difference is exactly 0.5 \cdot 10^{-16}. Otherwise, it is the (non-0, non-5) digit closest to 10^{17} times the difference. In particular, if the shift is 17 or more, all the digits are dropped, \langle\text{rounding}\rangle is 1 (not 0), and \langle x_1'\rangle and \langle x_2'\rangle are both zero.

If the shift is 1, the \langle\text{rounding}\rangle digit is simply the only digit that was pushed out of the brace groups (this is important for subtraction). It would be more natural for the \langle\text{rounding}\rangle digit to be placed after the \langle x_1'\rangle, but the choice we make involves less reshuffling.

Note that this function treats negative \langle shift\rangle as 0.
\end{verbatim}
Each of the auxiliaries see the function \( f_1 \), followed by 4 blocks of 4 digits.

(End of definition for \( \_\_fp_{\text{decimate}}:Nnnnnn \))

If the \( \langle \text{shift} \rangle \) is zero, or too big, life is very easy.

\[
\cs_new:Npn \_\_fp_{\text{decimate}}:Nnnnn \ #1 \ #2#3#4#5
\{ \ #1 \ \{#2#3\} \ \{#4#5\} \ ; \ }
\cs_new:Npn \_\_fp_{\text{decimate}}_{\text{tiny}}:Nnnnn \ #1 \ #2#3#4#5
\{ \ #1 \ \{0000 0000\} \ \{0000 0000\} \ 0 \ #2#3#4#5 \ ; \ }
\]

(End of definition for \( \_\_fp_{\text{decimate}}:Nnnnn \) and \( \_\_fp_{\text{decimate}}_{\text{tiny}}:Nnnnnn \))

\[\_\_fp_{\text{decimate}}_{\text{auxi}}:Nnnnn \ \langle f_1 \rangle \ \{\langle X_1 \rangle\} \ \{\langle X_2 \rangle\} \ \{\langle X_3 \rangle\} \ \{\langle X_4 \rangle\}\]

Shifting happens in two steps: compute the \( \langle \text{rounding} \rangle \) digit, and repack digits into two blocks of 8. The sixteen functions are very similar, and defined through \( \_\_fp_{\text{tmp}}:w \). The arguments are as follows: \#1 indicates which function is being defined; after one step of expansion, \#2 yields the “extra digits” which are then converted by \( \_\_fp_{\text{round_digit}}:Nw \) to the \( \langle \text{rounding} \rangle \) digit (note the + separating blocks of digits to avoid overflowing \( \text{T}_{\text{eX}} \text{'s integers} \)). This triggers the \( f \)-expansion of \( \_\_fp_{\text{decimate}}_{\text{pack}}:nnnnnwnn \)\(^\text{10}\) responsible for building two blocks of 8 digits, and removing the rest. For this to work, \#3 alternates between braced and unbraced blocks of 4 digits, in such a way that the 5 first and 5 next token groups yield the correct blocks of 8 digits.

\[
\cs_new:Npn \_\_fp_{\text{tmp}}:w \ #1 \ #2 \ #3
\{ \cs_new:cpn { \_\_fp_{\text{decimate}}_#1 :Nnnnn } ##1 ##2##3##4##5
\{ \exp_after:wN \#1 ##1 \ \int_value:w \exp_after:wN \_\_fp_{\text{round_digit}}:Nw \#2 ; \ \_\_fp_{\text{decimate}}_{\text{pack}}:nnnnnwnn \#3 ; \}
\}
\]

\[
\_\_fp_{\text{tmp}}:w \ {i} \ \{\use_none:nnn \ \#5\} \ \{0(#2)#3(#4)#5 \}
\]

\[
\_\_fp_{\text{tmp}}:w \ {ii} \ \{\use_none:n \ \#5\} \ \{00(#2)#3(#4)#5 \}
\]

\[
\_\_fp_{\text{tmp}}:w \ {iii} \ \{\use_none:n \ \#5\} \ \{000(#2)#3(#4)#5 \}
\]

\[
\_\_fp_{\text{tmp}}:w \ {iv} \ \{\ \#5\} \ \{(0000)\#2(#3)#4 \#5 \}
\]

\[
\_\_fp_{\text{tmp}}:w \ {v} \ \{\use_none:n \ \#4#5\} \ \{0(0000)\#2(#3)#4 \#5 \}
\]

\[
\_\_fp_{\text{tmp}}:w \ {vi} \ \{\use_none:n \ \#4#5\} \ \{00(0000)\#2(#3)#4 \#5 \}
\]

\[
\_\_fp_{\text{tmp}}:w \ {vii} \ \{\use_none:n \ \#4#5\} \ \{000(0000)\#2(#3)#4 \#5 \}
\]

\[
\_\_fp_{\text{tmp}}:w \ {viii}\{\ \#4#5\} \ \{(0000)000(#2)#3 \#4 \#5 \}
\]

\[
\_\_fp_{\text{tmp}}:w \ {ix} \ \{\use_none:n \ \#3#4#5\} \ \{0(0000)000(#2)#3 \#4 \#5 \}
\]

\[
\_\_fp_{\text{tmp}}:w \ {x} \ \{\use_none:n \ \#3#4#5\} \ \{00(0000)000(#2)#3 \#4 \#5 \}
\]

\[
\_\_fp_{\text{tmp}}:w \ {xi} \ \{\use_none:n \ \#3#4#5\} \ \{000(0000)000(#2)#3 \#4 \#5 \}
\]

\[
\_\_fp_{\text{tmp}}:w \ {xii} \ \{\ \#3#4#5\} \ \{(000000000(#2)#3 \#4 \#5 \}
\]

\[
\_\_fp_{\text{tmp}}:w \ {xiii}\{\use_none:n \ \#2#3#4#5\} \ \{0(000000000(#2)#3 \#4 \#5 \}
\]

\[
\_\_fp_{\text{tmp}}:w \ {xiv} \ \{\use_none:n \ \#2#3#4#5\} \ \{00(000000000(#2)#3 \#4 \#5 \}
\]

\[
\_\_fp_{\text{tmp}}:w \ {xv} \ \{\use_none:n \ \#2#3#4#5\} \ \{000(000000000(#2)#3 \#4 \#5 \}
\]

\[
\_\_fp_{\text{tmp}}:w \ {xvi} \ \{\ \#2#3#4#5\} \ \{(000000000(#2)#3 \#4 \#5 \}
\]

(End of definition for \( \_\_fp_{\text{decimate}}_{\text{auxi}}:Nnnnnn \) and others.)

\(^{10}\)No, the argument spec is not a mistake: the function calls an auxiliary to do half of the job.
The computation of the \textit{rounding} digit leaves an unfinished \verb|\int_value:w|, which expands the following functions. This allows us to repack nicely the digits we keep. Those digits come as an alternation of unbraced and braced blocks of 4 digits, such that the first 5 groups of token consist in 4 single digits, and one brace group (in some order), and the next 5 have the same structure. This is followed by some digits and a semicolon.

\begin{verbatim}
\cs_new:Npn \__fp_decimate_pack:nnnnnnnnnnw #1#2#3#4#5
\{ \__fp_decimate_pack:nnnnnw \{ #1#2#3#4#5 \} \}
\cs_new:Npn \__fp_decimate_pack:nnnnnnw #1 #2#3#4#5#6
\{ {#1} {#2#3#4#5#6} \}
\end{verbatim}

(End of definition for \__fp_decimate_pack:nnnnnnnnnnw.)

\section*{68.10 Functions for use within primitive conditional branches}

The functions described in this section are not pretty and can easily be misused. When correctly used, each of them removes one \texttt{\exp_stop_f}: as part of its parameter text, and puts one back as part of its replacement text.

Many computation functions in \texttt{l3fp} must perform tests on the type of floating points that they receive. This is often done in an \texttt{\if_case:w} statement or another conditional statement, and only a few cases lead to actual computations: most of the special cases are treated using a few standard functions which we define now. A typical use context for those functions would be

\begin{verbatim}
\if_case:w \integer \exp_stop_f:
 \_\_fp_case_return_o:Nw \fp \var
 \or: \_\_fp_case_use:nw \{\textit{some computation}\}
 \or: \_\_fp_case_return_same_o:w
 \or: \_\_fp_case_return:nw \{\textit{something}\}
\fi:
\{\textit{junk}\}
\{\textit{floating point}\}
\end{verbatim}

In this example, the case 0 returns the floating point \texttt{\fp \var}, expanding once after that floating point. Case 1 does \textit{some computation} using the \texttt{\floting \point \point} (presumably compute the operation requested by the user in that non-trivial case). Case 2 returns the \texttt{\floting \point \point} without modifying it, removing the \texttt{\textit{junk}} and expanding once after. Case 3 closes the conditional, removes the \texttt{\textit{junk}} and the \texttt{\floting \point \point}, and expands \texttt{\textit{something}} next. In other cases, the ”\textit{junk}” is expanded, performing some other operation on the \texttt{\floting \point \point \point}. We provide similar functions with two trailing \texttt{\floting \point \point \point}.

\begin{verbatim}
\_\_fp_case_use:nw
\end{verbatim}

This function ends a \texttt{\TeX} conditional, removes junk until the next floating point, and places its first argument before that floating point, to perform some operation on the floating point.

\begin{verbatim}
\cs_new:Npn \_\_fp_case_use:nw \#1\#2 \fi: \#3 \\_\_fp \{ \fi: \#1 \_\_fp \}
\end{verbatim}

(End of definition for \_\_fp_case_use:nw.)
This function ends a TeX conditional, removes junk and a floating point, and places its first argument in the input stream. A quirk is that we don’t define this function requiring a floating point to follow, simply anything ending in a semicolon. This, in turn, means that the \langle junk\rangle may not contain semicolons.

\cs_new:Npn \__fp_case_return:nw #1\#2 \fi: #3 ; { \fi: #1 }

(End of definition for \__fp_case_return:nw.)

This function ends a TeX conditional, removes junk and a floating point, and returns its first argument (an \langle fp var\rangle) then expands once after it.

\cs_new:Npn \__fp_case_return_o:Nw #1#2 \fi: #3 #4 ; { \fi: \exp_after:wN #1 }

(End of definition for \__fp_case_return_o:Nw.)

This function ends a TeX conditional, removes junk, and returns the following floating point, expanding once after it.

\cs_new:Npn \__fp_case_return_same_o:w #1 \fi: #2 #3 ; \s__fp #4 ; { \fi: \__fp_exp_after_o:w \s__fp #3 ; }

(End of definition for \__fp_case_return_same_o:w.)

Same as \__fp_case_return_o:Nw but with two trailing floating points.

\cs_new:Npn \__fp_case_return_o:Nww #1#2 \fi: #3 #4 ; #5 ; { \fi: \exp_after:wN #1 }

(End of definition for \__fp_case_return_o:Nww.)

Similar to \__fp_case_return_same_o:w, but this returns the first or second of two trailing floating point numbers, expanding once after the result.

\cs_new:Npn \__fp_case_return_i_o:ww \__fp_case_return_ii_o:ww #1 \fi: #2 \s__fp #3 ; \s__fp #4 ; { \fi: \__fp_exp_after_o:w #3 ; }

(End of definition for \__fp_case_return_i_o:ww and \__fp_case_return_ii_o:ww.)

### 68.11 Integer floating points

\__fp_int_p:w \__fp_int:w TF

Tests if the floating point argument is an integer. For normal floating point numbers, this holds if the rounding digit resulting from \__fp_decimate:nNNNNN is 0.

\prg_new_conditional:Npnn \__fp_int:w \__fp_chk:w #1 #2 #3 ; { #4 ; } { \if_case:w #1 \exp_stop_f: \prg_return_true: \or: \if_charcode:w 0 \__fp_decimate:nNNNNN \c__fp_prec_int - #3 \__fp_use_i_until_s:nw #4 \prg_return_true: \else: \prg_return_false: \end_case:w: }
68.12 Small integer floating points

Tests if the floating point argument is an integer or ±∞. If so, it is clipped to an integer in the range \([-10^8, 10^8]\) and fed as a braced argument to the \(\text{true code}\). Otherwise, the \(\text{false code}\) is performed.

First filter special cases: zeros and infinities are integers, \texttt{nan} is not. For normal numbers, decimate. If the rounding digit is not 0 run the \(\text{false code}\). If it is, then the integer is \#2 \#3; use \#3 if \#2 vanishes and otherwise \(10^8\).

\[
\begin{align*}
\text{\texttt{\_fp_small_int:wTF}} & \equiv \text{\texttt{\_fp_chk:w \_fp_case_return:nw}} \#1 \#2 \\
\text{\texttt{\_fp_small_int_true:wTF}} & \equiv \text{\texttt{\_fp_case_return:nw \_fp_small_int_true:wTF \_fp_small_int_normal:NnwTF}} \\
\text{\texttt{\_fp_small_int_normal:NnwTF}} & \equiv \text{\texttt{\_fp_small_int_test:NnnwNTF}} \\
\text{\texttt{\_fp_small_int_test:NnnwNTF}} & \equiv \text{\texttt{\_fp_small_int_test:NnnwNTF}}(\text{\texttt{\_fp_small_int_normal:NnwTF}}(\text{\texttt{\_fp_small_int_true:wTF}}(\text{\texttt{\_fp_small_int_normal:NnwTF}}(\text{\texttt{\_fp_small_int_test:NnnwNTF}})))})
\end{align*}
\]

(End of definition for \texttt{\_fp_small_int:wTF} and others.)
68.13 Fast string comparison

\__fp_str_if_eq:nn
A private version of the low-level string comparison function.
\cs_new_eq:NN \__fp_str_if_eq:nn \tex_strcmp:D
(End of definition for \__fp_str_if_eq:nn.)

68.14 Name of a function from its l3fp-parse name

\__fp_func_to_name:N
\__fp_func_to_name_aux:w
The goal is to convert for instance \__fp_sin_o:w to sin. This is used in error messages hence does not need to be fast.
\cs_new:Npn \__fp_func_to_name:N #1
{\exp_last_unbraced:Nf \__fp_func_to_name_aux:w { \cs_to_str:N #1 } X}
\cs_set_protected:Npn \__fp_tmp:w #1 #2
{ \cs_new:Npn \__fp_func_to_name_aux:w ##1 #1 ##2 #2 ##3 X {##2} }
\exp_args:Nff \__fp_tmp:w { \tl_to_str:n { __fp_ } }
{ \tl_to_str:n { _o: } }
(End of definition for \__fp_func_to_name:N and \__fp_func_to_name_aux:w.)

68.15 Messages

Using a floating point directly is an error.
\msg_new:nnnn { fp } { misused } { A-floating-point-with-value-'#1'-was-misused. } { To-obtain-the-value-of-a-floating-point-variable,-use- \token_to_str:N \fp_to_decimal:N',-, \token_to_str:N \fp_to_tl:N',-or-other- conversion-functions. }\prop_gput:Nnn \g_msg_module_name_prop { fp } { LaTeX } \prop_gput:Nnn \g_msg_module_type_prop { fp } { }"
Chapter 69

l3fp-traps implementation

Exceptions should be accessed by an n-type argument, among
• invalid_operation
• division_by_zero
• overflow
• underflow
• inexact (actually never used).

69.1 Flags

Flags to denote exceptions.

\l_fp_invalid_operation_flag
\l_fp_division_by_zero_flag
\l_fp_overflow_flag
\l_fp_underflow_flag

(End of definition for \l_fp_invalid_operation_flag and others. These variables are documented on page 270.)

69.2 Traps

Exceptions can be trapped to obtain custom behaviour. When an invalid operation or a
division by zero is trapped, the trap receives as arguments the result as an N-type floating
point number, the function name (multiple letters for prefix operations, or a single symbol
for infix operations), and the operand(s). When an overflow or underflow is trapped, the
trap receives the resulting overly large or small floating point number if it is not too big,
otherwise it receives +∞. Currently, the inexact exception is entirely ignored.

The behaviour when an exception occurs is controlled by the definitions of the func-
tions

• \_fp_invalid_operation

1036
Rather than changing them directly, we provide a user interface as \texttt{\_fp\_trap:nn (\langle exception\rangle) \langle way of trapping\rangle)}, where the \langle way of trapping\rangle is one of error, flag, or none.

We also provide \texttt{\_fp\_invalid\_operation:o:nw}, defined in terms of \texttt{\_fp\_invalid\_operation:mmw}.

\texttt{\_fp\_trap:nn}

\begin{verbatim}
\cs_new_protected:Npn \fp_trap:nn #1#2
{\cs_if_exist_use:cF { __fp_trap_#1_set_#2: }
 {clist_if_in:nnTF { invalid_operation , division_by_zero , overflow , underflow } #1}
 { \msg_error:nnee { fp } unknown-fpu-trap-type } {#1} {#2}
 { \msg_error:nne { fp } unknown-fpu-exception } {#1}
}
\end{verbatim}

(End of definition for \texttt{\_fp\_trap:nn}. This function is documented on page 270.)

We provide three types of trapping for invalid operations: either produce an error and raise the relevant flag; or only raise the flag; or don’t even raise the flag. In most cases, the function produces as a result its first argument, possibly with post-expansion.
We provide three types of trapping for invalid operations and division by zero: either produce an error and raise the relevant flag; or only raise the flag; or don’t even raise the flag. In all cases, the function must produce a result, namely its first argument, ±∞ or nan.
Just as for invalid operations and division by zero, the three different behaviours are obtained by feeding \prg_do_nothing:, \use_none:nnnn or \use_none:nnnnnn to an auxiliary, with a further auxiliary common to overflow and underflow functions. In most cases, the argument of the \__fp_overflow:w and \__fp_underflow:w functions will be an (almost) normal number (with an exponent outside the allowed range), and the error message thus displays that number together with the result to which it overflowed or underflowed. For extreme cases such as 10 ** 1e9999, the exponent would be too large for \TeX{}, and \__fp_overflow:w receives $\pm \infty$ (\__fp_underflow:w would receive $\pm 0$); then we cannot do better than simply say an overflow or underflow occurred.

(End of definition for \__fp_trap_division_by_zero_set_error: and others.)

\__fp_invalid_operation:nnw
\__fp_invalid_operation_o:Nww
\__fp_invalid_operation_tl_o:ff
\__fp_division_by_zero_o:Nww
\__fp_division_by_zero_o:ff
\__fp_overflow:w
\__fp_underflow:w

Initialize the control sequences (to log properly their existence). Then set invalid operations to trigger an error, and division by zero, overflow, and underflow to act silently on their flag.

(End of definition for \__fp_trap_division_by_zero_set_error: and others.)
Convenient short-hands for returning \c_nan_fp for a unary or binary operation, and expanding after.

```
cs_new:Npn \__fp_invalid_operation_o:nw { \__fp_invalid_operation:nnw { \exp_after:wN \c_nan_fp } }
cs_generate_variant:Nn \__fp_invalid_operation_o:nw { f }
```

(End of definition for \__fp_invalid_operation_o:nw.)

### 69.3 Errors

Some messages.

```
\msg_new:nnnn { fp } { unknown-fpu-exception }
{ }
\msg_new:nnnn { fp } { unknown-fpu-trap-type }
{ The-FPU-trap-type-'\#2'-is-not-known. }
\msg_new:nnnn { fp } { unknown-fpu-exception }
{ The-FPU-exception-'\#1'-is-not-known:- that-trap-will-never-be-triggered. }
\msg_new:nnnn { fp } { unknown-fpu-trap-type }
{ The-trap-type-must-be-one-of }\ioverline{\textindent:n}
```

(End of definition for \__fp_error:nnnn.)

### 69.4 Messages
* - none

\msg_new:nnn { fp } { flow } { An - #3 - occurred. }
\msg_new:nnn { fp } { flow-to } { #1 - #3 ed - to - #2 . }
\msg_new:nnn { fp } { zero-div } { Division-by-zero-in- #1 (#2) }
\msg_new:nnn { fp } { zero-div-ii } { Division-by-zero-in- (#1) #3 (#2) }
\msg_new:nnn { fp } { invalid } { Invalid-operation- #1 (#2) }
\msg_new:nnn { fp } { invalid-ii } { Invalid-operation- (#1) #3 (#2) }
\msg_new:nnn { fp } { unknown-type } { Unknown-type-for-`#1` }

(/package)
Chapter 70

l3fp-round implementation

\__fp_parse_word_trunc:N \__fp_parse_word_floor:N \__fp_parse_word_ceil:N
\__fp_parse_word_round:N \__fp_parse_round:Nw

\cs_new:Npn \__fp_parse_word_trunc:N { \__fp_parse_function:NNN \__fp_round_o:Nw \__fp_round_to_zero:NNN }
\cs_new:Npn \__fp_parse_word_floor:N { \__fp_parse_function:NNN \__fp_round_o:Nw \__fp_round_to_ninf:NNN }
\cs_new:Npn \__fp_parse_word_ceil:N { \__fp_parse_function:NNN \__fp_round_o:Nw \__fp_round_to_pinf:NNN }

(End of definition for \__fp_parse_word_trunc:N, \__fp_parse_word_floor:N, and \__fp_parse_word_ceil:N.)

\cs_new:Npn \__fp_parse_word_round:N #1#2
{ \__fp_parse_function:NNN \__fp_round_o:Nw \__fp_round_to_nearest:NNN #1 #2 }
\cs_new:Npn \__fp_parse_round:Nw #1 #2 \__fp_round_to_nearest:NNN #3#4
{ #2 #1 #3 }

(End of definition for \__fp_parse_word_round:N and \__fp_parse_round:Nw.)

70.1 Rounding tools

\c__fp_five_int This is used as the half-point for which numbers are rounded up/down.
\int_const:Nn \c__fp_five_int { 5 }

(End of definition for \c__fp_five_int.)

Floating point operations often yield a result that cannot be exactly represented in a significand with 16 digits. In that case, we need to round the exact result to a representable number. The IEEE standard defines four rounding modes:
• Round to nearest: round to the representable floating point number whose absolute difference with the exact result is the smallest. If the exact result lies exactly at the mid-point between two consecutive representable floating point numbers, round to the floating point number whose last digit is even.

• Round towards negative infinity: round to the greatest floating point number not larger than the exact result.

• Round towards zero: round to a floating point number with the same sign as the exact result, with the largest absolute value not larger than the absolute value of the exact result.

• Round towards positive infinity: round to the least floating point number not smaller than the exact result.

This is not fully implemented in l3fp yet, and transcendental functions fall back on the “round to nearest” mode. All rounding for basic algebra is done through the functions defined in this module, which can be redefined to change their rounding behaviour (but there is not interface for that yet).

The rounding tools available in this module are many variations on a base function \texttt{\_\_fp\_round:NNN}, which expands to \texttt{0\exp_stop_f:} or \texttt{1\exp_stop_f:} depending on whether the final result should be rounded up or down.

\[ \texttt{\_\_fp\_round:NNN \langle sign \rangle \langle digit_1 \rangle \langle digit_2 \rangle} \]

\[ \texttt{\_\_fp\_round\_to\_nearest:NNN} \]

\[ \texttt{\_\_fp\_round\_to\_near\_ninf:NNN} \]

\[ \texttt{\_\_fp\_round\_to\_near\_zero:NNN} \]

\[ \texttt{\_\_fp\_round\_to\_pinf:NNN} \]

\[ \texttt{\_\_fp\_round\_to\_ninf:NNN} \]

\[ \texttt{\_\_fp\_round\_to\_zero:NNN} \]

\[ \texttt{\_\_fp\_round\_to\_pinf:NNN} \]

\[ \texttt{\_\_fp\_round\_s:NNNw \langle sign \rangle \langle digit_1 \rangle \langle digit_2 \rangle \langle more\ digits \rangle} \]

\[ \texttt{\_\_fp\_round\_neg:NNN \langle sign \rangle \langle digit_1 \rangle \langle digit_2 \rangle} \]

If rounding the number \texttt{\langle final\ sign \rangle \langle digit_1 \rangle \langle digit_2 \rangle} to an integer rounds it towards zero (truncates it), this function expands to \texttt{0\exp_stop_f:}, and otherwise to \texttt{1\exp_stop_f:}. Typically used within the scope of an \texttt{\_\_fp\_int\_eval:w}, to add 1 if needed, and thereby round correctly. The result depends on the rounding mode.

It is very important that \texttt{\langle final\ sign \rangle} be the final sign of the result. Otherwise, the result would be incorrect in the case of rounding towards \(-\infty\) or towards \(+\infty\). Also recall that \texttt{\langle final\ sign \rangle} is 0 for positive, and 2 for negative.

By default, the functions below return \texttt{0\exp_stop_f:}, but this is superseded by \texttt{\_\_fp\_round\_return\_one:, which instead returns 1\exp_stop_f:}, expanding everything and removing \texttt{0\exp_stop_f:} in the process. In the case of rounding towards \pm\infty or towards 0, this is not really useful, but it prepares us for the “round to nearest, ties to even” mode.

The “round to nearest” mode is the default. If the \texttt{\langle digit_2 \rangle} is larger than 5, then round up. If it is less than 5, round down. If it is exactly 5, then round such that \texttt{\langle digit_1 \rangle} plus the result is even. In other words, round up if \texttt{\langle digit_1 \rangle} is odd.

The “round to nearest” mode has three variants, which differ in how ties are rounded: down towards \(-\infty\), truncated towards 0, or up towards \(+\infty\).
\cs_new:Npn \__fp_round_return_one: [2 #3]
  { \exp_after:wN 1 \exp_after:wN \exp_stop_f: \exp:w }
\cs_new:Npn \__fp_round_to_ninf:NNN #1 #2 #3
  {
    \if_meaning:w 2 #1
      \if_int_compare:w #3 > \c_zero_int
        \__fp_round_return_one:
      \fi:
    \fi:
    \c_zero_int
  }
\cs_new:Npn \__fp_round_to_zero:NNN #1 #2 #3 { \c_zero_int }
\cs_new:Npn \__fp_round_to_pinf:NNN #1 #2 #3
  {
    \if_meaning:w 0 #1
      \if_int_compare:w #3 > \c_zero_int
        \__fp_round_return_one:
      \fi:
    \fi:
    \c_zero_int
  }
\cs_new:Npn \__fp_round_to_nearest:NNN #1 #2 #3
  {
    \if_int_compare:w #3 > \c__fp_five_int
      \__fp_round_return_one:
    \else:
      \if_meaning:w 5 #3
        \if_int_odd:w #2 \exp_stop_f:
          \__fp_round_return_one:
        \fi:
      \fi:
    \fi:
    \c_zero_int
  }
\cs_new:Npn \__fp_round_to_nearest_ninf:NNN #1 #2 #3
  {
    \if_int_compare:w #3 > \c__fp_five_int
      \__fp_round_return_one:
    \else:
      \if_meaning:w 5 #3
        \if_meaning:w 2 #1
          \__fp_round_return_one:
        \fi:
      \fi:
    \fi:
    \c_zero_int
  }
\cs_new:Npn \__fp_round_to_nearest_zero:NNN #1 #2 #3
  {
    \if_int_compare:w #3 > \c__fp_five_int
      \__fp_round_return_one:
    \fi:
    \c_zero_int
  }

1044
\cs_new:Npn \__fp_round_to_nearest_pinf:NNN #1 #2 #3
{\if_int_compare:w #3 > \c__fp_five_int
\__fp_round_return_one:\else:\if_meaning:w 5 #3
\if_meaning:w 0 #1
\__fp_round_return_one:\fi:\fi:\c_zero_int\fi:\fi:\exp_stop_f:}
\cs_new_eq:NN \__fp_round:NNN \__fp_round_to_nearest:NNN
\__fp_round_s:NNNw \__fp_round_s:NNNw \langle final sign \rangle \langle digit \rangle \langle more digits \rangle;
Similar to \__fp_round:NNN, but with an extra semicolon, this function expands to \texttt{0\exp_stop_f;} if rounding \langle final sign\rangle\langle digit\rangle\langle more digits \rangle to an integer truncates, and to \texttt{1\exp_stop_f;} otherwise. The \langle more digits \rangle part must be a digit, followed by something that does not overflow a \texttt{\int_use:N \__fp_int_eval:w} construction. The only relevant information about this piece is whether it is zero or not.
\cs_new:Npn \__fp_round_digit:Nw #1 #2;
{\if_int_odd:w \if_meaning:w 0 #1 1 \else:\if_meaning:w 5 #1 1 \fi:\exp_stop_f:\fi:\if_int_compare:w \__fp_int_eval:w #2 > \c_zero_int
1 + \fi:\fi:\c_zero_int #3;\exp_stop_f:}
\cs_new_eq:NN \__fp_round_digit:Nw \__fp_round_digit:Nw \langle digit \rangle \langle int expr \rangle;
This function should always be called within an \texttt{\int_value:w} or \texttt{\__fp_int_eval:w} expansion; it may add an extra \texttt{\__fp_int_eval:w}, which means that the integer or integer expression should not be ended with a synonym of \texttt{\relax}, but with a semi-colon for instance.
\cs_new:Npn \__fp_round_digit:Nw \__fp_round_digit:Nw \#1 \#2;
This expands to \(\exp_stop_f:\) or \(1\exp_stop_f:\) after doing the following test. Starting from a number of the form \((final\ sign)0.(15\ digits)(\textit{digit}_1)\) with exactly 15 (non-all-zero) digits before \(\langle\textit{digit}_1\rangle\), subtract from it \((final\ sign)0.0...0(\textit{digit}_2)\), where there are 16 zeros. If in the current rounding mode the result should be rounded down, then this function returns \(1\exp_stop_f:\). Otherwise, \textit{i.e.}, if the result is rounded back to the first operand, then this function returns \(0\exp_stop_f:\). It turns out that this negative “round to nearest” is identical to the positive one. And this is the default mode.

\(70.2\) The \textit{round} function

First check that all arguments are floating point numbers. The \textit{trunc}, \textit{ceil} and \textit{floor} functions expect one or two arguments (the second is 0 by default), and the \textit{round} function also accepts a third argument (\textit{nan} by default), which changes \#1 from \(\_\_\textit{fp\_round\_to\_nearest}\) to one of its analogues.
\begin{verbatim}
\cs_new:Npn \__fp_round_aux_o:Nw #1#2 \@ 
\{ 
  \if_case:w \__fp_int_eval:w \__fp_array_count:n \{#2\} \__fp_int_eval_end:
  \__fp_round_no_arg_o:Nw #1 \exp:w
 \or: \__fp_round:Nwn #1 #2 \{0\} \exp:w
 \or: \__fp_round:Nww #1 #2 \exp:w
 \else: \__fp_round:Nwww #1 #2 \exp:w
  \fi:
  \exp_after:wN \exp_end:
\}

(End of definition for \__fp_round_o:Nw and \__fp_round_aux_o:Nw.)
\__fp_round_no_arg_o:Nw
\cs_new:Npn \__fp_round_no_arg_o:Nw #1
\{ 
  \cs_if_eq:NNTF #1 \__fp_round_to_nearest:NNN
  { \__fp_error:nnnn \{ num-args \} \{ round () \} \{ 1 \} \{ 3 \} }
  { \__fp_error:nffn \{ num-args \} \{ \__fp_round_name_from_cs:N #1 () \} \{ 1 \} \{ 2 \} }
  \exp_after:wN \c_nan_fp
\}

(End of definition for \__fp_round_no_arg_o:Nw.)
\__fp_round:Nwww
Having three arguments is only allowed for round, not trunc, ceil, floor, so check for that case. If all is well, construct one of \__fp_round_to_nearest:NNN, \__fp_round_to_nearest_zero:NNN, \__fp_round_to_nearest_ninf:NNN, \__fp_round_to_nearest_-pinf:NNN and act accordingly. 
\cs_new:Npn \__fp_round:Nwww #1\#2 ; #3 ; \s__fp \__fp_chk:w \#4\#5\#6 ; \#7 @
\{ 
  \cs_if_eq:NNTF #1 \__fp_round_to_nearest:NNN
  { \tl_if_empty:nTF \{\#7\}
  \exp_after:wN \c_nan_fp
  \}
  { \__fp_error:nnnn \{ num-args \} \{ round () \} \{ 1 \} \{ 3 \} }
\}
\end{verbatim}
\__fp_error:nffn { num-args }
\{ \__fp_round_name_from_cs:N \#1 \} \{ 1 \} \{ 2 \}
\exp_after:wN \c_nan_fp
\}

(End of definition for \__fp_round:Nwww.)

\__fp_round_name_from_cs:N
\cs_new:Npn \__fp_round_name_from_cs:N \#1
\{ \cs_if_eq:NNTF \#1 \__fp_round_to_zero:NNN \{ trunc \}
\{ \cs_if_eq:NNTF \#1 \__fp_round_to_ninf:NNN \{ floor \}
\{ \cs_if_eq:NNTF \#1 \__fp_round_to_pinf:NNN \{ ceil \}
\{ round \}
\}
\}

(End of definition for \__fp_round_name_from_cs:N.)

\__fp_round:Nww \__fp_round:Nwn \__fp_round_normal:NwNNnw \__fp_round_normal:NnnwNNnn \__fp_round_normal_end:wwNnn \__fp_round_special:NwwNnn \__fp_round_special_aux:Nw

If the number of digits to round to is an integer or infinity all is good; if it is nan then just produce a nan; otherwise invalid as we have something like round(1,3.14) where the number of digits is not an integer.

\cs_new:Npn \__fp_round:Nww \#1\#2 ; \#3 ;
\{ \__fp_small_int:wTF \#3; \{ \__fp_round:Nwn \#1\#2; \}
\{ \if:w 3 \__fp_kind:w \#3 ;
\exp_after:wN \use_i:nn
\else:
\exp_after:wN \use_ii:nn
\fi:
\{ \exp_after:wN \c_nan_fp \}
\{ \exp_after:wN \c_nan_fp \}
\{ \exp_after:wN \__fp_exp_after_o:w \}
\\__fp_round_name_from_cs:N \#1 \}
\{ \__fp_array_to_clist:n \{ \#2; \#3; \} \}
\}

(End of definition for \__fp_round_name_from_cs:N.)
(End of definition for \_\_fp_round:Nww and others.)
Chapter 71

l3fp-parse implementation

71.1 Work plan

The task at hand is non-trivial, and some previous failed attempts show that the code leads to unreadable logs, so we had better get it (almost) right the first time. Let us first describe our goal, then discuss the design precisely before writing any code.

In this file at least, a (floating point object) is a floating point number or tuple. This can be extended to anything that starts with \s__fp or \s__fp_{type} and ends with ; with some internal structure that depends on the \(type\).

\_\_fp\_parse\_n \{\(fp\ expr\}\}

Evaluates the \(fp\ expr\) and leaves the result in the input stream as a floating point object. This function forms the basis of almost all public l3fp functions. During evaluation, each token is fully f-expanded.

\_\_fp\_parse\_o\_n does the same but expands once after its result.

\TeXhackers note: Registers (integers, toks, etc.) are automatically unpacked, without requiring a function such as \int\_use\_N. Invalid tokens remaining after f-expansion lead to unrecoverable low-level \TeX errors.

Floating point expressions are composed of numbers, given in various forms, infix operators, such as +, **, or , (which joins two numbers into a list), and prefix operators, such as the unary -, functions, or opening parentheses. Here is a list of precedences which control the order of evaluation (some distinctions are irrelevant for the order of evaluation, but serve as signals), from the tightest binding to the loosest binding.

16 Function calls.

13/14 Binary ** and ~ (right to left).

12 Unary +, -, ! (right to left).

11 Juxtaposition (implicit *) with no parenthesis.
10 Binary * and /.
9 Binary + and −.
7 Comparisons.
6 Logical and, denoted by &&.
5 Logical or, denoted by ||.
4 Ternary operator ?:, piece ?.
3 Ternary operator ?:, piece :.
2 Commas.
1 Place where a comma is allowed and generates a tuple.

0 Start and end of the expression.

71.1.1 Storing results

The main question in parsing expressions expandably is to decide where to put the intermediate results computed for various subexpressions.

One option is to store the values at the start of the expression, and carry them together as the first argument of each macro. However, we want \texttt{f}-expand tokens one by one in the expression (as \texttt{\int_eval:n} does), and with this approach, expanding the next unread token forces us to jump with \texttt{\exp_after:wN} over every value computed earlier in the expression. With this approach, the run-time grows at least quadratically in the length of the expression, if not as its cube (inserting the \texttt{\exp_after:wN} is tricky and slow).

A second option is to place those values at the end of the expression. Then expanding the next unread token is straightforward, but this still hits a performance issue: for long expressions we would be reaching all the way to the end of the expression at every step of the calculation. The run-time is again quadratic.

A variation of the above attempts to place the intermediate results which appear when computing a parenthesized expression near the closing parenthesis. This still lets
us expand tokens as we go, and avoids performance problems as long as there are enough parentheses. However, it would be better to avoid requiring the closing parenthesis to be present as soon as the corresponding opening parenthesis is read: the closing parenthesis may still be hidden in a macro yet to be expanded.

Hence, we need to go for some fine expansion control: the result is stored before the start!

Let us illustrate this idea in a simple model: adding positive integers which may be resulting from the expansion of macros, or may be values of registers. Assume that one number, say, 12345, has already been found, and that we want to parse the next number. The current status of the code may look as follows.

\exp_after:wN \add:ww \int_value:w 12345 \exp_after:wN ;
\exp:w \operand:w ⟨stuff⟩

One step of expansion expands \exp_after:wN, which triggers the primitive \int_value:w, which reads the five digits we have already found, 12345. This integer is unfinished, causing the second \exp_after:wN to expand, and to trigger the construction \exp:w, which expands \operand:w, defined to read what follows and make a number out of it, then leave \exp_end:, the number, and a semicolon in the input stream. Once \operand:w is done expanding, we obtain essentially

\exp_after:wN \add:ww \int_value:w 12345 ;
\exp:w \exp_end: 333444 ;

where in fact \exp_after:wN has already been expanded, \int_value:w has already seen 12345, and \exp:w is still looking for a number. It finds \exp_end:, hence expands to nothing. Now, \int_value:w sees the ;, which cannot be part of a number. The expansion stops, and we are left with

\add:ww 12345 ; 333444 ;

which can safely perform the addition by grabbing two arguments delimited by ;.

If we were to continue parsing the expression, then the following number should also be cleaned up before the next use of a binary operation such as \add:ww. Just like \int_value:w 12345 \exp_after:wN ; expanded what follows once, we need \add:ww to do the calculation, and in the process to expand the following once. This is also true in our real application: all the functions of the form \_\_fp..._o:ww expand what follows once. This comes at the cost of leaving tokens in the input stack, and we need to be careful not to waste this memory. All of our discussion above is nice but simplistic, as operations should not simply be performed in the order they appear.

71.1.2 Precedence and infix operators

The various operators we will encounter have different precedences, which influence the order of calculations: \(1 + 2 \times 3 = 1 + (2 \times 3)\) because \(\times\) has a higher precedence than \(+\).

The true analog of our macro \operand:w must thus take care of that. When looking for an operand, it needs to perform calculations until reaching an operator which has lower precedence than the one which called \operand:w. This means that \operand:w must know what the previous binary operator is, or rather, its precedence: we thus rename it \operand:Nw. Let us describe as an example how we plan to do the calculation \(41–2^3+4+5\). More precisely we describe how to perform the first operation in this expression. Here, we abuse notations: the first argument of \operand:Nw should be an integer.
constant (`\_\_fp\_prec\_plus\_int, \ldots`) equal to the precedence of the given operator, not directly the operator itself.

- Clean up 41 and find `-`. We call `\operator:Nw -` to find the second operand.
- Clean up 2 and find `^`.
- Compare the precedences of `-` and `^`. Since the latter is higher, we need to compute the exponentiation. For this, find the second operand with a nested call to `\operator:Nw ^`.
- Clean up 3 and find `*`.
- Compare the precedences of `^` and `*`. Since the former is higher, `\operator:Nw ^` has found the second operand of the exponentiation, which is computed: $2^3 = 8$.
- We now have 41-8*4+5, and `\operator:Nw -` is still looking for a second operand for the subtraction. Is it 8?
- Compare the precedences of `-` and `*`. Since the latter is higher, we are not done with 8. Call `\operator:Nw *` to find the second operand of the multiplication.
- Clean up 4, and find `+`.
- Compare the precedences of `*` and `+`. Since the former is higher, `\operator:Nw *` has found the second operand of the multiplication, which is computed: $8 * 4 = 32$.
- We now have 41-32+5, and `\operator:Nw -` is still looking for a second operand for the subtraction. Is it 32?
- Compare the precedences of `-` and `+`. Since they are equal, `\operator:Nw -` has found the second operand for the subtraction, which is computed: $41 - 32 = 9$.
- We now have 9+5.

The procedure above stops short of performing all computations, but adding a surrounding call to `\operator:Nw` with a very low precedence ensures that all computations are performed before `\operator:Nw` is done. Adding a trailing marker with the same very low precedence prevents the surrounding `\operator:Nw` from going beyond the marker.

The pattern above to find an operand for a given operator, is to find one number and the next operator, then compare precedences to know if the next computation should be done. If it should, then perform it after finding its second operand, and look at the next operator, then compare precedences to know if the next computation should be done. This continues until we find that the next computation should not be done. Then, we stop.

We are now ready to get a bit more technical and describe which of the `\l3fp\_parse` functions correspond to each step above.

First, `\_\_fp\_parse\_operand:Nw` is the `\operator:Nw` function above, with small modifications due to expansion issues discussed later. We denote by `<precedence>` the argument of `\_\_fp\_parse\_operand:Nw`, that is, the precedence of the binary operator whose operand we are trying to find. The basic action is to read numbers from the input stream. This is done by `\_\_fp\_parse\_one:Nw`. A first approximation of this function is that it reads one `<number>`, performing no computation, and finds the following binary `<operator>`. Then it expands to
We now explain the \texttt{infix} auxiliaries. We need some flexibility in how we treat the case of equal precedences: most often, the first operation encountered should be performed, such as \(1-2-3\) being computed as \((1-2)-3\), but \(2\cdot 3^4\) should be evaluated as \(2\cdot (3^4)\) instead. For this reason, and to support the equivalence between \(\ast\ast\) and \(^\) more easily, each binary operator is converted to a control sequence \_	exttt{fp\_parse\_infix\_}\langle\texttt{operator}\rangle:N when it is encountered for the first time. Instead of passing both precedences to a test function to do the comparison steps above, we pass the \langle\texttt{precedence}\rangle of the earlier operator to the \texttt{infix} auxiliary for the following \langle\texttt{operator}\rangle, to know whether to perform the computation of the \langle\texttt{operator}\rangle. If it should not be performed, the \texttt{infix} auxiliary expands to
\begin{verbatim}
  \_\_fp\_parse\_infix\_\langle\texttt{operator}\rangle:N
\end{verbatim}
and otherwise it calls \_	exttt{fp\_parse\_operand}:N\langle\texttt{precedence}\rangle with the precedence of the \langle\texttt{operator}\rangle to find its second operand \langle\texttt{number}\rangle and the next \langle\texttt{operator}\rangle, and expands to
\begin{verbatim}
  \_\_fp\_parse\_apply\_binary:N\langle\texttt{number}\rangle\langle\texttt{operator}\rangle:Na\_\_fp\_parse\_infix\_\langle\texttt{operator}\rangle:N
\end{verbatim}
The \texttt{infix} function is responsible for comparing precedences, but cannot directly call the computation functions, because the first operand \langle\texttt{number}\rangle is before the \texttt{infix} function in the input stream. This is why we stop the expansion here and give control to another function to close the loop.

A definition of \_	exttt{fp\_parse\_operand}:Na\langle\texttt{precedence}\rangle with some of the expansion control removed is
\begin{verbatim}
\exp_after:wN \_\_fp\_parse\_continue:NwN\langle\texttt{precedence}\rangle
\exp_after:wN\_\_fp\_parse\_continue:wN\langle\texttt{number}\rangle\_\_fp\_parse\_one:N\langle\texttt{precedence}\rangle
\end{verbatim}
This expands \_	exttt{fp\_parse\_one}:Na\langle\texttt{precedence}\rangle completely, which finds a number, wraps the next \langle\texttt{operator}\rangle into an \texttt{infix} function, feeds this function the \langle\texttt{precedence}\rangle, and expands it, yielding either
\begin{verbatim}
\_\_fp\_parse\_continue:NwN\langle\texttt{precedence}\rangle\langle\texttt{number}\rangle\_\_fp\_parse\_infix\_\langle\texttt{operator}\rangle:N
\end{verbatim}
or
\begin{verbatim}
\_\_fp\_parse\_continue:NwN\langle\texttt{precedence}\rangle\langle\texttt{number}\rangle\_\_fp\_parse\_apply\_binary:N\langle\texttt{number}\rangle\langle\texttt{operator}\rangle:Na\_\_fp\_parse\_infix\_\langle\texttt{operator}\rangle:N
\end{verbatim}
The definition of \_	exttt{fp\_parse\_continue}:Na\langle\texttt{precedence}\rangle is then very simple:
\begin{verbatim}
\cs_new:Npn \_\_fp\_parse\_continue:NwN #1\#2\#3 \{ #3 \#1 \#2 \_\_fp\_parse\_continue:NwN \}
\end{verbatim}
In the first case, \#3 is \texttt{\use:none:n}, yielding
\begin{verbatim}
\use:none:n \langle precedence \rangle \langle number \rangle @
\end{verbatim}
then \langle number \rangle @ \texttt{\_fp_parse_infix_\langle operator \rangle:N}. In the second case, \#3 is \texttt{\_fp_parse_infix_binary:NwNwN}, whose role is to compute \langle number \rangle \langle operator \rangle \langle number_2 \rangle and to prepare for the next comparison of precedences: first we get
\begin{verbatim}
\_fp_parse_infix_binary:NwNwN
\langle precedence \rangle \langle number \rangle @
\langle operator \rangle \langle number_2 \rangle @ \texttt{\_fp_parse_infix_\langle operator \rangle2:N}
\end{verbatim}
then
\begin{verbatim}
\exp_after:wN \_fp_parse_continue:NwN
\exp_after:wN \_fp_parse_one:Nw
\exp:w \_fp_end_continue_f:w
\_fp_\langle operator \rangle_o:ww \langle number \rangle \langle number_2 \rangle
\exp:w \_fp_end_continue_f:w
\_fp_parse_infix_\langle operator \rangle2:N \langle precedence \rangle
\end{verbatim}
where \texttt{\_fp_\langle operator \rangle_o:ww} computes \langle number \rangle \langle operator \rangle \langle number_2 \rangle and expands after the result, thus triggers the comparison of the precedence of the \langle operator_2 \rangle and the \langle precedence \rangle, continuing the loop.

We have introduced the most important functions here, and the next few paragraphs we describe various subtleties.

### 71.1.3 Prefix operators, parentheses, and functions

Prefix operators (unary \texttt{-}, \texttt{+}, \texttt{!}) and parentheses are taken care of by the same mechanism, and functions (\texttt{sin}, \texttt{exp}, etc.) as well. Finding the argument of the unary \texttt{-}, for instance, is very similar to grabbing the second operand of a binary infix operator, with a subtle precedence explained below. Once that operand is found, the operator can be applied to it (for the unary \texttt{-}, this simply flips the sign). A left parenthesis is just a prefix operator with a very low precedence equal to that of the closing parenthesis (which is treated as an infix operator, since it normally appears just after numbers), so that all computations are performed until the closing parenthesis. The prefix operator associated to the left parenthesis does not alter its argument, but it removes the closing parenthesis (with some checks).

Prefix operators are the reason why we only summarily described the function \texttt{\_fp_parse_one:Nw} earlier. This function is responsible for reading in the input stream the first possible \langle number \rangle and the next infix \langle operator \rangle. If what follows \texttt{\_fp_\langle precedence \rangle} is a prefix operator, then we must find the operand of this prefix operator through a nested call to \texttt{\_fp_parse_operand:Nw} with the appropriate precedence, then apply the operator to the operand found to yield the result of \texttt{\_fp_\langle precedence \rangle}. So far, all is simple.

The unary operators \texttt{+, -}, \texttt{!} complicate things a little bit: \texttt{-3**2} should be \texttt{-3}^2 = \texttt{-9}, and not \texttt{(-(3)^2)} = \texttt{9}. This would easily be done by giving \texttt{-} a lower precedence, equal to that of the infix \texttt{+} and \texttt{-}. Unfortunately, this fails in cases such as \texttt{3**-2*4}, yielding \texttt{3^{-2} \times 4} instead of the correct \texttt{3^{-2} \times 4}. A second attempt would be to call \texttt{\_fp_parse_operand:Nw} with the \langle precedence \rangle of the previous operator, but \texttt{0>-2+3} is then
parsed as 0>(2+3): the addition is performed because it binds more tightly than the comparison which precedes -. The correct approach is for a unary - to perform operations whose precedence is greater than both that of the previous operation, and that of the unary - itself. The unary - is given a precedence higher than multiplication and division. This does not lead to any surprising result, since -(x/y) = (-x)/y and similarly for multiplication, and it reduces the number of nested calls to \_fp_parse_operand:Nw.

Functions are implemented as prefix operators with very high precedence, so that their argument is the first number that can possibly be built.

Note that contrarily to the \textit{infix} functions discussed earlier, the \textit{prefix} functions do perform tests on the previous \textit{precedence} to decide whether to find an argument or not, since we know that we need a number, and must never stop there.

\subsection{Numbers and reading tokens one by one}

So far, we have glossed over one important point: what is a “number”? A number is typically given in the form \textit{significand}\textit{e}\textit{exponent}, where the \textit{significand} is any non-empty string composed of decimal digits and at most one decimal separator (a period), the exponent “\textit{e}\textit{exponent}” is optional and is composed of an exponent mark \texttt{e} followed by a possibly empty string of signs + or - and a non-empty string of decimal digits. The \textit{significand} can also be an integer, dimension, skip, or muskip variable, in which case dimensions are converted from points (or mu units) to floating points, and the \textit{exponent} can also be an integer variable. Numbers can also be given as floating point variables, or as named constants such as \texttt{nan}, \texttt{inf} or \texttt{pi}. We may add more types in the future.

When \_fp_parse_one:Nw is looking for a “number”, here is what happens.

- If the next token is a control sequence with the meaning of \texttt{scan_stop};, it can be: \_fp, in which case our job is done, as what follows is an internal floating point number, or \_fp_expr_mark, in which case the expression has come to an early end, as we are still looking for a number here, or something else, in which case we consider the control sequence to be a bad variable resulting from c-expansion.

- If the next token is a control sequence with a different meaning, we assume that it is a register, unpack it with \texttt{tex_the:D}, and use its value (in pt for dimensions and skips, mu for muskips) as the \textit{significand} of a number: we look for an exponent.

- If the next token is a digit, we remove any leading zeros, then read a significand larger than 1 if the next character is a digit, read a significand smaller than 1 if the next character is a period, or we have found a significand equal to 0 otherwise, and look for an exponent.

- If the next token is a letter, we collect more letters until the first non-letter: the resulting word may denote a function such as \texttt{asin}, a constant such as \texttt{pi} or be unknown. In the first case, we call \_fp_parse_operand:Nw to find the argument of the function, then apply the function, before declaring that we are done. Otherwise, we are done, either with the value of the constant, or with the value \texttt{nan} for unknown words.

- If the next token is anything else, we check whether it is a known prefix operator, in which case \_fp_parse_operand:Nw finds its operand. If it is not known, then either a number is missing (if the token is a known infix operator) or the token is simply invalid in floating point expressions.
Once a number is found, \_fp_parse_one:Nw also finds an infix operator. This goes as follows.

• If the next token is a control sequence, it could be the special marker \s__-fp_expr_mark, and otherwise it is a case of juxtaposing numbers, such as 2\c_zero_int, with an implied multiplication.

• If the next token is a letter, it is also a case of juxtaposition, as letters cannot be proper infix operators.

• Otherwise (including in the case of digits), if the token is a known infix operator, the appropriate \_fp_infix_(operator):N function is built, and if it does not exist, we complain. In particular, the juxtaposition \c_zero_int 2 is disallowed.

In the above, we need to test whether a character token #1 is a digit:

\if_int_compare:w 9 < 1 \token_to_str:N \#1 \exp_stop_f:
  is a digit
\else:
  not a digit
\fi:

To exclude 0, replace 9 by 10. The use of \token_to_str:N ensures that a digit with any catcode is detected. To test if a character token is a letter, we need to work with its character code, testing if ‘#1 lies in [65, 90] (uppercase letters) or [97, 112] (lowercase letters)

\if_int_compare:w \__fp_int_eval:w
  ( ‘\#1 \if_int_compare:w ‘\#1 > ‘Z - 32 \fi: ) / 26 = 3 \exp_stop_f:
  is a letter
\else:
  not a letter
\fi:

At all steps, we try to accept all category codes: when #1 is kept to be used later, it is almost always converted to category code other through \token_to_str:N. More precisely, catcodes \{3, 6, 7, 8, 11, 12\} should work without trouble, but not \{1, 2, 4, 10, 13\}, and of course \{0, 5, 9\} cannot become tokens.

Floating point expressions should behave as much as possible like \e-T\m\e-X-based integer expressions and dimension expressions. In particular, f-expansion should be performed as the expression is read, token by token, forcing the expansion of protected macros, and ignoring spaces. One advantage of expanding at every step is that restricted expandable functions can then be used in floating point expressions just as they can be in other kinds of expressions. Problematically, spaces stop f-expansion: for instance, the macro \X below would not be expanded if we simply performed f-expansion.

\DeclareDocumentCommand \{\test\} {m} \{ \fp_eval:n \{\#1\} }
\ExplSyntaxOff
\test \{ 1 + \X \}

Of course, spaces typically do not appear in a code setting, but may very easily come in document-level input, from which some expressions may come. To avoid this problem, at every step, we do essentially what \use:f would do: take an argument, put it back
in the input stream, then f-expand it. This is not a complete solution, since a macro's expansion could contain leading spaces which would stop the f-expansion before further macro calls are performed. However, in practice it should be enough: in particular, floating point numbers are correctly expanded to the underlying __fp... structure. The f-expansion is performed by __fp_parse_expand:w.

71.2 Main auxiliary functions

\_\_fp_parse_operand:Nw \exp:w \_\_fp_parse_operand:Nw \langle \text{precedence} \rangle \_\_fp_parse_expand:w

Reads the "...", performing every computation with a precedence higher than \langle \text{precedence} \rangle, then expands to

\langle \text{result} \rangle \@ \_\_fp_parse_infix_\langle \text{operation} \rangle:N ...

where the \langle \text{operation} \rangle is the first operation with a lower precedence, possibly end, and the "..." start just after the \langle \text{operation} \rangle.

(End of definition for \_\_fp_parse_operand:Nw.)

\_\_fp_parse_infix_+:N \_\_fp_parse_infix_+:N \langle \text{precedence} \rangle ...

If + has a precedence higher than the \langle \text{precedence} \rangle, cleans up a second \langle \text{operand} \rangle and finds the \langle \text{operation} \rangle which follows, and expands to

\@ \_\_fp_parse_apply_binary:NwNwN + \langle \text{operand} \rangle \@ \_\_fp_parse_infix_\langle \text{operation} \rangle:N ...

Otherwise expands to

\@ \use_none:n \_\_fp_parse_infix_+:N ...

A similar function exists for each infix operator.

(End of definition for \_\_fp_parse_infix_+:N.)

\_\_fp_parse_one:Nw \_\_fp_parse_one:Nw \langle \text{precedence} \rangle ...

Cleans up one or two operands depending on how the precedence of the next operation compares to the \langle \text{precedence} \rangle. If the following \langle \text{operation} \rangle has a precedence higher than \langle \text{precedence} \rangle, expands to

\langle \text{operand}_1 \rangle \@ \_\_fp_parse_apply_binary:NwNwN \langle \text{operation} \rangle \langle \text{operand}_2 \rangle \@ \_\_fp_parse_infix_\langle \text{operation} \rangle:N ...

and otherwise expands to

\langle \text{operand} \rangle \@ \use_none:n \_\_fp_parse_infix_\langle \text{operation} \rangle:N ...

(End of definition for \_\_fp_parse_one:Nw.)
71.3 Helpers

\_\_fp\_parse\_expand:w
\exp:w \__fp_parse_expand:w \langle \text{tokens} \rangle
This function must always come within a \exp:w expansion. The \langle \text{tokens} \rangle should be the part of the expression that we have not yet read. This requires in particular closing all conditionals properly before expanding.
\cs_new:Npn \__fp_parse_expand:w \#1 { \exp_end_continue_f:w \#1 }
(End of definition for \_\_fp\_parse\_expand:w.)

\_fp\_parse\_return\_semicolon:w
This very odd function swaps its position with the following \fi: and removes \__fp\_parse\_expand:w normally responsible for expansion. That turns out to be useful.
\cs_new:Npn \__fp_parse_return_semicolon:w \#1 \fi: \__fp_parse_expand:w \{ \fi: ; \#1 }
(End of definition for \_\_fp\_parse\_return\_semicolon:w.)

\__fp\_parse\_digits\_vii:N
\__fp\_parse\_digits\_vi:N
\__fp\_parse\_digits\_v:N
\__fp\_parse\_digits\_iv:N
\__fp\_parse\_digits\_iii:N
\__fp\_parse\_digits\_ii:N
\__fp\_parse\_digits\_i:N
\__fp\_parse\_digits_:N
These functions must be called within an \int_value:w or \__fp_int_eval:w construction. The first token which follows must be f-expanded prior to calling those functions.
The functions read tokens one by one, and output digits into the input stream, until meeting a non-digit, or up to a number of digits equal to their index. The full expansion is
\langle \text{digits} \rangle ; \langle \text{filling 0} \rangle ; \langle \text{length} \rangle
where \langle \text{filling 0} \rangle is a string of zeros such that \langle \text{digits} \rangle \langle \text{filling 0} \rangle has the length given by the index of the function, and \langle \text{length} \rangle is the number of zeros in the \langle \text{filling 0} \rangle string. Each function puts a digit into the input stream and calls the next function, until we find a non-digit. We are careful to pass the tested tokens through \token_to_str:N to normalize their category code.
\cs_set_protected:Npn \__fp_tmp:w \#1 \#2 \#3
\cs_new:cpn { __fp_parse_digits_ \#1 :N } ##1
\if_int_compare:w 9 < 1 \token_to_str:N ##1 \exp_stop_f:
  \token_to_str:N ##1 \exp_after:wN #2 \exp:w
\else:
  \_\_fp_parse_return_semicolon:w \#3 \#1 \fi:
  \_\_fp_parse_expand:w
\end{definition}
(End of definition for \_\_fp\_parse\_digits\_vii:N and others.)
### 71.4 Parsing one number

This function finds one number, and packs the symbol which follows in an `\_fp_parse_infix...` csname. \#1 is the previous ⟨precedence⟩, and \#2 the first token of the operand. We distinguish four cases: \#2 is equal to `\scan_stop:` in meaning, \#2 is a different control sequence, \#2 is a digit, and \#2 is something else (this last case is split further later). Despite the earlier f-expansion, \#2 may still be expandable if it was protected by `\exp_not:N`, as may happen with the \LaTeXe \texttt{protect} command. Using a well placed `\reverse_if:N`, this case is sent to `\_fp_parse_one_fp:NN` which deals with it robustly.

```latex
\cs_new:Npn \_fp_parse_one:Nw \#1 \#2
\begin{verbatim}
\if_catcode:w \scan_stop: \exp_not:N \#2
\exp_after:wN \if_meaning:w \exp_not:N \#2 \#2 \else:
\exp_after:wN \reverse_if:N
\fi:
\if_meaning:w \scan_stop: \#2
\exp_after:wN \exp_after:wN
\exp_after:wN \_fp_parse_one_fp:NN
\else:
\exp_after:wN \exp_after:wN
\exp_after:wN \_fp_parse_one_register:NN
\fi:
\else:
\if_int_compare:w 9 < 1 \token_to_str:N \#2 \exp_stop_f:
\exp_after:wN \exp_after:wN
\exp_after:wN \_fp_parse_one_digit:NN
\else:
\exp_after:wN \exp_after:wN
\exp_after:wN \_fp_parse_one_other:NN
\fi:
\fi:
\end{verbatim}
```

(End of definition for `\_fp_parse_one:Nw`.)

This function receives a ⟨precedence⟩ and a control sequence equal to `\scan_stop:` in meaning. There are three cases.

- `\s__fp` starts a floating point number, and we call `\_fp_exp_after_f:nw`, which f-expands after the floating point.
- `\s__fp_expr_mark` is a premature end, we call `\_fp_exp_after_expr_mark_end:f:nw`, which triggers an fp-early-end error.
- For a control sequence not containing `\s__fp`, we call `\_fp_exp_after_assoc:f:nw`, causing a bad-variable error.

This scheme is extensible: additional types can be added by starting the variables with a scan mark of the form `\s__fp⟨type⟩` and defining `\_fp_exp_after_assoc⟨type⟩:f:nw`. In all cases, we make sure that the second argument of `\_fp_parse_infix:NN` is correctly expanded. A special case only enabled in \LaTeXe \texttt{2e} is that if \texttt{protect} is encountered then
the error message mentions the control sequence which follows it rather than \protect itself. The test for \LaTeXe uses \unexpandable@protect rather than \protect because \protect is often \scan_stop: hence “does not exist”.

\begin{verbatim}
24122 \cs_new:Npn \__fp_parse_one_fp:NN #1
24123 {  \__fp_exp_after_any_f:nw
24124  \__fp_exp_after_any_f:nw
24125  \exp_after:wN \__fp_parse_infix:NN
24126  \exp_after:wN #1 \exp:w \__fp_parse_expand:w
24127 } \__fp_parse_one_fp:NN #1
24128 }
24129 \cs_new:Npn \__fp_exp_after_expr_mark_f:nw #1
24130 {  \int_case:nnF { \exp_after:wN \use_i:nnn \use_none:nnn #1 } { #1 }
24131     \c__fp_prec_comma_int { } { early-end }
24132     \exp_after:wN \c__fp_empty_tuple_fp
24133     \exp:w \exp_end_continue_f:w
24134 }
24135 { #1 }
24136 \cs_new:cpn { __fp_exp_after_?_f:nw } #1#2
24137 {  \msg_expandable_error:nnn { kernel } { bad-variable } {#2}
24138  \exp_after:wN \c_nan_fp \exp:w \exp_end_continue_f:w #1
24139 }
24140 \cs_set_protected:Npn \__fp_tmp:w #1
24141 {  \cs_if_exist:NT #1
24142     { \cs_gset:cpn { __fp_exp_after_?_f:nw } ##1##2
24143     { \msg_expandable_error:nnn { fp } {##2} { robust-cmd } }
24144     { \msg_expandable_error:nnn { kernel } { bad-variable } {##2}
24145     }
24146     }
24147 }
24148 \cs_new:cpn { __fp_exp_after_?_f:nw } #1#2
24149 {  \msg_expandable_error:nnn { kernel } { bad-variable } {#2}
24150     \exp_after:wN \c_nan_fp \exp:w \exp_end_continue_f:w #1
24151 }
24152 \cs_set_protected:Npn \__fp_tmp:w #1
24153 {  \cs_if_exist:NT #1
24154     { \cs_gset:cpn { __fp_exp_after_?_f:nw } ##1##2
24155     { \exp_after:wN \c_nan_fp \exp:w \exp_end_cont
This is called whenever \#2 is a control sequence other than \texttt{\scan_stop}: in meaning. We special-case \texttt{\wd}, \texttt{\ht}, \texttt{\dp} (see later) and otherwise assume that it is a register, but carefully unpack it with \texttt{\tex_the:D} within braces. First, we find the exponent following \#2. Then we unpack \#2 with \texttt{\tex_the:D}, and the auxii auxiliary distinguishes integer registers from dimensions/skips from muskips, according to the presence of a period and/or of pt. For integers, simply convert \langle value \rangle e \langle exponent \rangle to a floating point number with \texttt{\__fp_parse:n} (this is somewhat wasteful). For other registers, the decimal rounding provided by \TeX{} does not accurately represent the binary value that it manipulates, so we extract this binary value as a number of scaled points with \texttt{\int_value:w \dim_to_decimal_in_sp:n} \{ \texttt{(decimal value)} pt \}, and use an auxiliary of \texttt{\dim_to_fp:n}, which performs the multiplication by \texttt{\texttt{2^{10}}}, correctly rounded.
The \texttt{\textbackslash dim}, \texttt{\ht}, \texttt{\dp} primitives expect an integer argument. We abuse the exponent parser to find the integer argument: simply include the exponent marker \texttt{e}. Once that “exponent” is found, use \texttt{\textbackslash tex_the:D} to find the box dimension and then copy what we did for dimensions.

A digit marks the beginning of an explicit floating point number. Once the number is found, we catch the case of overflow and underflow with \texttt{\_fp_sanitize:Nw},
then \texttt{\_fp_parse_infix_after_operand:NwN} expands \texttt{\_fp_parse_infix:NN} after the number we find, to wrap the following infix operator as required. Finding the number itself begins by removing leading zeros: further steps are described later.

\begin{verbatim}
\cs_new:Npn \__fp_parse_one_digit:NN #1 { \exp_after:wN \__fp_parse_infix_after_operand:NwN \exp_after:wN #1 \exp:w \exp_end_continue_f:w \exp_after:wN \__fp_sanitize:wN \int_value:w \__fp_int_eval:w 0 \__fp_parse_trim_zeros:N }
\end{verbatim}

(End of definition for \texttt{\_fp_parse_one_digit:NN}.)

\texttt{\_fp_parse_one_other:NN} For this function, \texttt{\_fp_parse_one_other:NN} is a character token which is not a digit. If it is an ASCII letter, \texttt{\_fp_parse_letters:N} beyond this one and give the result to \texttt{\_fp_parse_prefix:Nw}. Otherwise, the character is assumed to be a prefix operator, and we build \texttt{\_fp_parse_prefix:(operator):Nw}.

\begin{verbatim}
\cs_new:Npn \__fp_parse_one_other:NN #1 #2 { \if_int_compare:w \__fp_int_eval:w (#2 #2 > 'Z - 32 \fi:) / 26 = 3 \exp_stop_f: \exp_after:wN \__fp_parse_word:Nw \exp_after:wN #1 \exp_after:wN #2 \exp:w \exp_after:wN \__fp_parse_letters:N \exp:w \else: \exp_after:wN \__fp_parse_prefix:NwN \exp_after:wN \__fp_parse_prefix:Nw \exp_after:wN \__fp_parse_prefix:Nw \cs:w \__fp_parse_prefix:\token_to_str:N #2 :Nw \exp_after:wN \cs_end: \exp:w \fi: \__fp_parse_expand:w }
\end{verbatim}

(End of definition for \texttt{\_fp_parse_one_other:NN}.)

\texttt{\_fp_parse_word:Nw} \texttt{\_fp_parse_letters:N} Finding letters is a simple recursion. Once \texttt{\_fp_parse_letters:N} has done its job, we try to build a control sequence from the word \texttt{\_fp_parse_letters:N} has done its job, and look for the following infix operator. Note that the unknown word could be a mistyped function as well as a mistyped constant, so there is no way to tell whether to look for arguments; we do not. The standard requires “\texttt{inf}” and “\texttt{infinity}” and “\texttt{nan}” to be recognized regardless of case, but we probably don’t want to allow every \texttt{l3fp} word to have an arbitrary mixture of lower and upper case, so we test and use a differently-named control sequence.

1065
\cs_new:Npn \__fp_parse_word:Nw #1#2;
{
\cs_if_exist_use:cF \{ \__fp_parse_word_#2:N } 
{
\cs_if_exist_use:cF
\{ \__fp_parse_caseless_\str_casefold:n \{#2\} :N \}
{
\msg_expandable_error:nnn
\{ fp \} \{ unknown-fp-word \} \{#2\}
\exp_after:wN \c_nan_fp \exp:w \exp_end_continue_f:w
\__fp_parse_infix:NN
}
#1
}
\cs_new:Npn \__fp_parse_letters:N #1
{
\exp_end_continue_f:w
\if_int_compare:w
\if_catcode:w \scan_stop: \exp_not:N #1 0
\else:
\__fp_int_eval:w ( '#1 \if_int_compare:w '#1 > 'Z - 32 \fi: ) / 26
\fi:
= 3 \exp_stop_f:
\exp:w \exp_after:wN \__fp_parse_letters:N
\exp:w
\else:
\__fp_parse_return_semicolon:w #1
\fi:
\__fp_parse_expand:w
}

(End of definition for \__fp_parse_word:Nw and \__fp_parse_letters:N.)

\__fp_parse_prefix:NNN
\__fp_parse_prefix_unknown:NNN

For this function, #1 is the previous \langle precedence \rangle, #2 is the operator just seen, and #3 is a control sequence which implements the operator if it is a known operator. If this control sequence is \scan_stop:, then the operator is in fact unknown. Either the expression is missing a number there (if the operator is valid as an infix operator), and we put \texttt{nan}, wrapping the infix operator in a csname as appropriate, or the character is simply invalid in floating point expressions, and we continue looking for a number, starting again from \__fp_parse_one:Nw.
\cs_new:Npn \__fp_parse_prefix:NNN #1#2#3
{
\if_meaning:w \scan_stop: #3
\exp_after:wN \__fp_parse_prefix_unknown:NNN
\exp_after:wN #2
\fi:
#3 #1
}
\cs_new:Npn \__fp_parse_prefix_unknown:NNN #1#2#3
{
}
71.4.1 Numbers: trimming leading zeros

Numbers are parsed as follows: first we trim leading zeros, then if the next character is a digit, start reading a significand \( \geq 1 \) with the set of functions \( \__fp_parse_large \)...; if it is a period, the significand is \(< 1 \); and otherwise it is zero. In the second case, trim additional zeros after the period, counting them for an exponent shift \( \langle \text{exp} \rangle < 0 \), then read the significand with the set of functions \( \__fp_parse_small \)... Once the significand is read, read the exponent if \( e \) is present.

This function expects an already expanded token. It removes any leading zero, then distinguishes three cases: if the first non-zero token is a digit, then call \( \__fp_parse_large \); if it is \( . \), then continue trimming zeros with \( \__fp_parse_strim_zeros \); otherwise, our number is exactly zero, and we call \( \__fp_parse_zero \) to take care of that case.
\_fp\_parse\_trim\_zeros:N \\
\_fp\_parse\_trim\_end:w

If we have removed all digits until a period (or if the body started with a period), then enter the "small trim" loop which outputs \texttt{-1} for each removed 0. Those \texttt{-1} are added to an integer expression waiting for the exponent. If the first non-zero token is a digit, call \texttt{\_fp\_parse\_small:N} (our significand is smaller than \texttt{1}), and otherwise, the number is an exact zero. The name \texttt{strim} stands for "small trim".

\cs{new:Npn}{\_fp\_parse\_strim\_zeros:N}{#1}
\exp{w}{\_fp\_parse\_strim\_zeros:N}{\exp:w}{0}{\exp_not:N}{#1}{-1}
\exp{w}{\_fp\_parse\_strim\_end:w}{#1}{\_fp\_parse\_expand:w}

\cs{new:Npn}{\_fp\_parse\_strim\_end:w}{#1}{\_fp\_parse\_expand:w}{\_fp\_parse\_strim\_zeros:N}{#1}

(End of definition for \texttt{\_fp\_parse\_strim\_zeros:N} and \texttt{\_fp\_parse\_strim\_end:w}).

\_fp\_parse\_small:N

This function is called after we have passed the decimal separator and removed all leading zeros from the significand. It is followed by a non-zero digit (with any catcode). The goal is to read up to 16 digits. But we can’t do that all at once, because \texttt{\_fp\_sanitize:wN}, which removes everything and leaves an exact zero.

\cs{new:Npn}{\_fp\_parse\_zero:}{\_fp\_parse\_small:N}{\_fp\_parse\_zeros:wN}{\_fp\_parse\_small:N}{\_fp\_parse\_zero:}{1}{\exp_stop_f:}{\_fp\_parse\_small:N}{\_fp\_parse\_zero:}{1}{\_fp\_parse\_zero:}{\_fp\_parse\_zeros:wN}{\_fp\_parse\_small:N}{\_fp\_parse\_small:N}{\_fp\_parse\_small:N}{\_fp\_parse\_small:N}

(End of definition for \texttt{\_fp\_parse\_strim\_zeros:N} and \texttt{\_fp\_parse\_strim\_end:w}).

71.4.2 Number: small significand

This function is called after we have passed the decimal separator and removed all leading zeros from the significand. It is followed by a non-zero digit (with any catcode). The goal is to read up to 16 digits. But we can’t do that all at once, because \texttt{\_fp\_sanitize:wN}, which removes everything and leaves an exact zero.

\cs{new:Npn}{\_fp\_parse\_zero:}{\_fp\_parse\_small:N}{\_fp\_parse\_zeros:wN}{\_fp\_parse\_small:N}{\_fp\_parse\_zero:}{1}{\exp_stop_f:}{\_fp\_parse\_small:N}{\_fp\_parse\_zero:}{1}{\_fp\_parse\_zero:}{\_fp\_parse\_zeros:wN}{\_fp\_parse\_small:N}{\_fp\_parse\_small:N}{\_fp\_parse\_small:N}{\_fp\_parse\_small:N}

(End of definition for \texttt{\_fp\_parse\_strim\_zeros:N} and \texttt{\_fp\_parse\_strim\_end:w}).
pack_leading auxiliary puts the various parts in the appropriate order for the processing further up.

```latex
\texttt{\textbackslash cs\_new:Npn \_\_fp\_parse\_small:N \#1}
\begin{verbatim}
  { \exp_after:wN \_\_fp\_parse\_pack\_leading:NNNNNw
  \int_value:w \_\_fp\_int\_eval:w 1 \token_to_str:N \#1
  \exp_after:wN \_\_fp\_parse\_small\_leading:wwNN
  \int_value:w 1
  \exp_after:wN \_\_fp\_parse\_digits\_vii:N
  \exp:w \_\_fp\_parse\_expand:w
}
\end{verbatim}

(End of definition for \_\_fp\_parse\_small:N)

\begin{verbatim}
\texttt{\_\_fp\_parse\_small\_leading:wwNN 1 \{digits\}; \{zeros\}; \{number of zeros\}
\langle next token \rangle}
\end{verbatim}

We leave \langle digits \rangle \langle zeros \rangle in the input stream: the functions used to grab digits are such that this constitutes digits 1 through 8 of the significand. Then prepare to pack 8 more digits, with an exponent shift of zero (this shift is used in the case of a large significand). If \#4 is a digit, leave it behind for the packing function, and read 6 more digits to reach a total of 15 digits: further digits are involved in the rounding. Otherwise put 8 zeros in to complete the significand, then look for an exponent.

```latex
\texttt{\textbackslash cs\_new:Npn \_\_fp\_parse\_small\_leading:wwNN 1 \#1 ; \#2; \#3 \#4}
\begin{verbatim}
  { \#1 \#2
  \exp_after:wN \_\_fp\_parse\_pack\_trailing:NNNNNNw
  \exp_after:wN 0
  \int_value:w \_\_fp\_int\_eval:w 1
  \if_int_compare:w 9 < 1 \token_to_str:N \#4 \exp_stop_f:
    \token_to_str:N \#4
  \exp_after:wN \_\_fp\_parse\_small\_trailing:wwNN
  \int_value:w 1
  \exp_after:wN \_\_fp\_parse\_digits\_vi:N
  \exp:w
  \else:
    0000 0000 \_\_fp\_parse\_exponent:Nw \#4
  \fi:
  \_\_fp\_parse\_expand:w
}
\end{verbatim}

(End of definition for \_\_fp\_parse\_small\_leading:wwNN.)

\begin{verbatim}
\texttt{\_\_fp\_parse\_small\_trailing:wwNN 1 \{digits\}; \{zeros\}; \{number of zeros\}
\langle next token \rangle}
\end{verbatim}

Leave digits 10 to 15 (arguments \#1 and \#2) in the input stream. If the \langle next token \rangle is a digit, it is the 16th digit, we keep it, then the small_round auxiliary considers this digit and all further digits to perform the rounding: the function expands to nothing, to +0 or to +1. Otherwise, there is no 16-th digit, so we put a 0, and look for an exponent.

```latex
\texttt{\textbackslash cs\_new:Npn \_\_fp\_parse\_small\_trailing:wwNN 1 \#1 ; \#2; \#3 \#4}
\begin{verbatim}
  { \#1 \#2
  \if_int_compare:w 9 < 1 \token_to_str:N \#4 \exp_stop_f:
    \token_to_str:N \#4
  \exp_after:wN \_\_fp\_parse\_small\_round:NN
\end{verbatim}

1069
Those functions are expanded after all the digits are found, we took care of the rounding, as well as the exponent. The last argument is the exponent. The previous five arguments are 8 digits which we pack in groups of 4, and the argument before that is 1, except in the rare case where rounding lead to a carry, in which case the argument is 2. The trailing function has an exponent shift as its first argument, which we add to the exponent found in the e... syntax. If the trailing digits cause a carry, the integer expression for the leading digits is incremented (+1 in the code below). If the leading digits propagate this carry all the way up, the function \__fp_parse_pack_carry:w increments the exponent, and changes the significand from 0000... to 1000...: this is simple because such a carry can only occur to give rise to a power of 10.

(End of definition for \__fp_parse_pack_trailing:NNNNw, \__fp_parse_pack_leading:NNNNw, and \__fp_parse_pack_carry:w.)

2.4.3 Number: large significand

Parsing a significand larger than 1 is a little bit more difficult than parsing small significands. We need to count the number of digits before the decimal separator, and add that to the final exponent. We also need to test for the presence of a dot each time we run out of digits, and branch to the appropriate parse_small function in those cases.

This function is followed by the first non-zero digit of a “large” significand (≥ 1). It is called within an integer expression for the exponent. Grab up to 7 more digits, for a total of 8 digits.
\__fp_parse_large_leading:wwNN

We shift the exponent by the number of digits in \#1, namely the target number, 8, minus the \langle number of zeros \rangle (number of digits missing). Then prepare to pack the 8 first digits. If the \langle next token \rangle is a digit, read up to 6 more digits (digits 10 to 15). If it is a period, try to grab the end of our 8 first digits, branching to the small functions since the number of digit does not affect the exponent anymore. Finally, if this is the end of the significand, insert the \langle zeros \rangle to complete the 8 first digits, insert 8 more, and look for an exponent.

\__fp_parse_large_leading:wwNN 1 \langle digits \rangle ; \langle zeros \rangle ; \langle number of zeros \rangle

\__fp_parse_large_leading:wwNN 1 \langle digits \rangle ; \langle zeros \rangle ; \langle number of zeros \rangle

\__fp_parse_large_leading:wwNN 1 \langle digits \rangle ; \langle zeros \rangle ; \langle number of zeros \rangle

\__fp_parse_large_trailing:wwNN

We have just read 15 digits. If the \langle next token \rangle is a digit, then the exponent shift caused by this block of 8 digits is 8, first argument to the pack trailing function. We keep the \langle digits \rangle and this 16-th digit, and find how this should be rounded using \__fp_parse_large_round:NN. Otherwise, the exponent shift is the number of \langle digits \rangle, 7 minus the \langle number of zeros \rangle, and we test for a decimal point. This case happens in 123451234512345.67 with exactly 15 digits before the decimal separator.
Then branch to the appropriate \texttt{small} auxiliary, grabbing a few more digits to complement the digits we already grabbed. Finally, if this is truly the end of the significand, look for an exponent after using the \texttt{⟨zeros⟩} and providing a 16-th digit of 0.

\begin{verbatim}
\cs_new:Npn \_fp_parse_large_trailing:wwNN 1 #1 ; #2; #3 #4
\{
  \if_int_compare:w 9 < 1 \token_to_str:N #4 \exp_stop_f:
    \exp_after:wN \_fp_parse_pack_trailing:NNNNNNww
    \exp_after:wN \c__fp_half_prec_int
    \int_value:w \__fp_int_eval:w 1 #1 \token_to_str:N #4
    \exp_after:wN \_fp_parse_large_round:NN
    \exp_after:wN \int_value:w #4
    \exp:w
  \else:
    \exp_after:wN \_fp_parse_pack_trailing:NNNNNNww
    \int_value:w 7 - #3 \exp_stop_f:
    \int_value:w 1 #1 \if:w . \exp_not:N #4
      \exp_after:wN \_fp_parse_small_trailing:wwNN
      \exp:w
      \else:
        \#2 0 \_fp_parse_exponent:Nw #4
      \fi:
      \exp:w
      \else:
        \_fp_parse_digits_
        \_fp_int_to_roman:w #3
        \_exp_after:wN \int_after:wN N \exp_after:wN \_fp_parse_digits_
        \cs:w__fp_parse_digits_
        \_fp_int_to_roman:w #3
        \cs_end:w
        \exp:w
      \else:
        \_fp_parse_digits_
        \_fp_int_to_roman:w #3
        \cs:w__fp_parse_digits_
        \_fp_int_to_roman:w #3
        \cs_end:w
        \exp:w
      \fi:
  \fi:
  \_fp_parse_expand:w
\}
\end{verbatim}

(End of definition for \_fp_parse_large_trailing:wwNN.)

\textbf{71.4.4 Number: beyond 16 digits, rounding}

This loop is called when rounding a number (whether the mantissa is small or large). It should appear in an integer expression. This function reads digits one by one, until reaching a non-digit, and adds 1 to the integer expression for each digit. If all digits found are 0, the function ends the expression by ;0, otherwise by ;1. This is done by switching the loop to \texttt{round_up} at the first non-zero digit, thus we avoid to test whether digits are 0 or not once we see a first non-zero digit.

\begin{verbatim}
\cs_new:Npn \_fp_parse_round_loop:N \_fp_parse_round_up:N
\{
  \if_int_compare:w 9 < 1 \token_to_str:N \_fp_expr_int \exp_stop_f:
    + 1
  \else:
    0 \token_to_str:N \_fp_expr_int
    \exp_after:wN \_fp_parse_round_loop:N
    \exp:w
  \fi:
\}
\end{verbatim}

1072
\__fp_parse_round_after:wN

After the loop \__fp_parse_round_loop:N, this function fetches an exponent with \__fp_parse_exponent:N, and combines it with the number of digits counted by \__fp_parse_round_loop:N. At the same time, the result 0 or 1 is added to the surrounding integer expression.

\__fp_parse_small_round:NN
\__fp_parse_round_after:wN

Here, \#1 is the digit that we are currently rounding (we only care whether it is even or odd). If \#2 is not a digit, then fetch an exponent and expand to ;⟨exponent⟩ only. Otherwise, we expand to +0 or +1, then ;⟨exponent⟩. To decide which, call \__fp_-
round_s:NNNw to know whether to round up, giving it as arguments a sign 0 (all explicit numbers are positive), the digit \#1 to round, the first following digit \#2, and either +0 or +1 depending on whether the following digits are all zero or not. This last argument is obtained by \__fp_parse_round_after:wN, whose number of digits we discard by multiplying it by 0. The exponent which follows the number is also fetched by \__fp_parse_round_after:wN.
Large numbers are harder to round, as there may be a period in the way. Again, \#1 is the digit that we are currently rounding (we only care whether it is even or odd). If there are no more digits (\#2 is not a digit), then we must test for a period: if there is one, then switch to the rounding function for small significands, otherwise fetch an exponent. If there are more digits (\#2 is a digit), then round, checking with \_\_\_fp\_parse\_round\_\_loop:N if all further digits vanish, or some are non-zero. This loop is not enough, as it is stopped by a period. After the loop, the aux function tests for a period: if it is present, then we must continue looking for digits, this time discarding the number of digits we find.
71.4.5 Number: finding the exponent

Expansion is a little bit tricky here, in part because we accept input where multiplication is implicit.

\__fp_parse:n \{ 3.2 \text{erf}(0.1) \}
\__fp_parse:n \{ 3.2 e\_my\_int \}
\__fp_parse:n \{ 3.2 \c_{\pi\_fp} \}

The first case indicates that just looking one character ahead for an “e” is not enough, since we would mistake the function \text{erf} for an exponent of “rf”. An alternative would be to look two tokens ahead and check if what follows is a sign or a digit, considering in that case that we must be finding an exponent. But taking care of the second case requires that we unpack registers after \text{e}. However, blindly expanding the two tokens ahead completely would break the third example (unpacking is even worse). Indeed, in the course of reading \texttt{3.2}, \c_{\pi\_fp} is expanded to \texttt{\s_{\_fp \_fp\_chk:w} 1 0 {-1} \{3141\} \cdots \;}; and \s_{\_fp\_chk:w} stops the expansion. Expanding two tokens ahead would then force the expansion of \texttt{\s_{\_fp\_chk:w}} (despite it being protected), and that function tries to produce an error.

What can we do? Really, the reason why this last case breaks is that just as \TeX{} does, we should read ahead as little as possible. Here, the only case where there may be an exponent is if the first token ahead is \texttt{e}. Then we expand (and possibly unpack) the second token.

\__fp_parse_exponent:Nw

This auxiliary is convenient to smuggle some material through \texttt{\fi:} ending conditional processing. We place those \texttt{\fi:} (argument \#2) at a very odd place because this allows us to insert \texttt{\_fp\_int\_eval:w ...} there if needed.

\__fp_parse_exponent:N

This function should be called within an \texttt{\int\_value:w} expansion (or within an integer expression). It leaves digits of the exponent behind it in the input stream, and terminates the expansion with a semicolon. If there is no \texttt{e} (or \texttt{E}), leave an exponent of 0. If there is an \texttt{e} or \texttt{E}, expand the next token to run some tests on it. The first rough test is that if the character code of \texttt{\#1} is greater than that of 9 (largest code valid for an exponent, less than any code valid for an identifier), there was in fact no exponent; otherwise, we search for the sign of the exponent.
\__fp_parse_exponent:N Read signs one by one (if there is any).
\cs_new:Npn \__fp_parse_exponent_sign:N #1
\cs_new:Npn \__fp_parse_exponent_sign:N #1
{ \if:w + \if:w - \exp_not:N #1 + \fi: \token_to_str:N #1 \exp_after:wN \__fp_parse_exponent_sign:N \exp:w \exp_after:wN \__fp_parse_exponent_body:N \exp_after:wN \__fp_parse_expand:w \else: \exp_after:wN \__fp_parse_exponent_keep:NTF #1 \{ \__fp_parse_return_semicolon:w #1 \} \fi: \exp_after:wN } \fi: \exp_after:wN \__fp_parse_exponent_body:N \exp_after:wN \__fp_parse_expand:w (End of definition for \__fp_parse_exponent:N and \__fp_parse_exponent_aux:NN.)

\__fp_parse_exponent_body:N An exponent can be an explicit integer (most common case), or various other things (most of which are invalid).
\cs_new:Npn \__fp_parse_exponent_body:N #1
\cs_new:Npn \__fp_parse_exponent_body:N #1
{ \if_int_compare:w 9 < 1 \token_to_str:N #1 \exp_stop_f: \token_to_str:N #1 \exp_after:wN \__fp_parse_exponent_digits:N \exp:w \exp_after:wN \__fp_parse_exponent_keep:NTF \exp_after:wN \__fp_parse_return_semicolon:w #1 \} \{ \__fp_parse_return_semicolon:w \exp_after:wN \__fp_parse_expand:w \exp:w \exp:w \exp:w \exp:w \exp:w \exp:w \exp:w \exp:w \exp:w \exp:w \exp:w \exp:w (End of definition for \__fp_parse_exponent_sign:N.)

1076
\_fp\_parse\_exponent\_digits:N

Read digits one by one, and leave them behind in the input stream. When finding a
non-digit, stop, and insert a semicolon. Note that we do not check for overflow of the
exponent, hence there can be a \TeX{} error. It is mostly harmless, except when parsing
0e9876543210, which should be a valid representation of 0, but is not.

\cs_new:Npn \__fp\_parse\_exponent\_digits:N \exp_after:wN \__fp\_parse\_exponent\_digits:N
\exp:w
\else:
\__fp\_parse\_return\_semicolon:w \exp_after:wN \__fp\_parse\_exponent\_digits:N
\fi:
\__fp\_parse\_expand:w
\end{verbatim}

\_fp\_parse\_exponent\_keep:NTF

This is the last building block for parsing exponents. The argument \#1 is already fully
expanded, and neither + nor - nor a digit. It can be:

- \s__fp, marking the start of an internal floating point, invalid here;
- another control sequence equal to \relax, probably a bad variable;
- a register: in this case we make sure that it is an integer register, not a dimension;
- a character other than +, - or digits, again, an error.

\prg_new_conditional:Npnn \__fp\_parse\_exponent\_keep:N \exp_after:wN \__fp\_parse\_exponent\_keep:N
\exp:w
\else:
\__fp\_parse\_return\_semicolon:w \exp_after:wN \__fp\_parse\_exponent\_digits:N
\fi:
\__fp\_parse\_expand:w
\end{verbatim}
\fi:
\else:
  0
\msg_expandable_error:nnn
  \{ \fp \} \{ missing \} \{ exponent \}
\prg_return_true:
\fi:
}

(End of definition for \__fp_parse_exponent_keep:NTF.)

71.5 Constants, functions and prefix operators

71.5.1 Prefix operators

\__fp_parse_prefix_+:Nw

A unary + does nothing: we should continue looking for a number.
\cs_new_eq:cN { \__fp_parse_prefix_+:Nw } \__fp_parse_one:Nw

(End of definition for \__fp_parse_prefix_+:Nw.)

\__fp_parse_apply_function:NNNw

Here, #1 is a precedence, #2 is some extra data used by some functions, #3 is e.g., \__fp_sin:o:w, and expands once after the calculation, #4 is the operand, and #5 is a \__fp_parse_infix_...:N function. We feed the data #2, and the argument #4, to the function #3, which expands \exp:w thus the infix function #5.
\cs_new:Npn \__fp_parse_apply_function:NNNw #1#2#3#4@#5
  { #3 #2 #4 @ \exp:w \exp_end_continue_f:w #5 #1 }

(End of definition for \__fp_parse_apply_function:NNNw.)

\__fp_parse_apply_unary:NNNw
\__fp_parse_apply_unary_chk:NwNw
\__fp_parse_apply_unary_type:NNN
\__fp_parse_apply_unary_error:NNw

In contrast to \__fp_parse_apply_function:NNNw, this checks that the operand #4 is a single argument (namely there is a single ;). We use the fact that any floating point starts with a "safe" token like \s__fp. If there is no argument produce the fp-no-arg error; if there are at least two produce fp-multi-arg. For the error message extract the mathematical function name (such as sin) from the expl3 function that computes it, such as \__fp_sin:o:w.

In addition, since there is a single argument we can dispatch on type and check that the resulting function exists. This catches things like sin((1,2)) where it does not make sense to take the sine of a tuple.
\cs_new:Npn \__fp_parse_apply_unary:NNNw #1#2#3#4@#5
  { \__fp_parse_apply_unary_chk:NwNw #4 @ ; \s__fp_stop
    \__fp_parse_apply_unary_type:NNN
    #3 #2 #4 @ \exp:w \exp_end_continue_f:w #5 #1 }

\cs_new:Npn \__fp_parse_apply_unary_chk:NwNw #1#2 ; #3#4 \s__fp_stop
  { \if_meaning:w @ #3 \else:
      \token_if_eq_meaning:NNTF . #3
      \{ \__fp_passage_apply_unary_chk:nNNNw \{ no \} } }

1078
\_\_fp\_parse\_prefix\_\-:Nw
\_\_fp\_parse\_prefix\_!:Nw

The unary - and boolean not are harder: we parse the operand using a precedence equal to the maximum of the previous precedence #1 and the precedence \c__fp\_prec\_not\_int of the unary operator, then call the appropriate \_\_fp\_\{operation\}_o:w function, where the \{operation\} is set_sign or not.

\_\_fp\_tmp:w - \c__fp\_prec\_not\_int \_\_fp\_set\_sign:o:w 2
\_\_fp\_tmp:w ! \c__fp\_prec\_not\_int \_\_fp\_not:o:w ?

(End of definition for \_\_fp\_parse\_prefix\_\-:Nw and \_\_fp\_parse\_prefix\_!:Nw.)

\_\_fp\_parse\_prefix\_:Nw

Numbers which start with a decimal separator (a period) end up here. Of course, we do not look for an operand, but for the rest of the number. This function is very similar to \_\_fp\_parse\_one\_digit:N and \_\_fp\_parse\_strim\_zeros:N but calls \_\_fp\_parse\_strim\_zeros:N to trim zeros after the decimal point, rather than the \texttt{trim\_zeros} function for zeros before the decimal point.

\_\_fp\_tmp:w - \c__fp\_prec\_not\_int \_\_fp\_set\_sign:o:w 2
\_\_fp\_tmp:w ! \c__fp\_prec\_not\_int \_\_fp\_not:o:w ?

(End of definition for \_\_fp\_parse\_prefix\_\-:Nw and \_\_fp\_parse\_prefix\_!:Nw.)
The left parenthesis is treated as a unary prefix operator because it appears in exactly the same settings. If the previous precedence is \c__fp_prec_func_int we are parsing arguments of a function and commas should not build tuples; otherwise commas should build tuples. We distinguish these cases by precedence: \c__fp_prec_comma_int for the case of arguments, \c__fp_prec_tuple_int for the case of tuples. Once the operand is found, the lparen_after auxiliary makes sure that there was a closing parenthesis (otherwise it complains), and leaves in the input stream an operand, fetching the following infix operator.

The right parenthesis can appear as a prefix in two similar cases: in an empty tuple or tuple ending with a comma, or in an empty argument list or argument list ending with a comma, such as in \texttt{max(1,2,)} or in \texttt{rand().}
71.5.2 Constants

Some words correspond to constant floating points. The floating point constant is left as a result of \_\_fp_parse_one: NVIDIA after expanding \_\_fp_parse_infix: NN.

\cs_set_protected:Npn \_\_fp_tmp:w #1 #2 {
\cs_new:cpn { __fp_parse_word_#1:N } { \exp_after:wN #2 \exp:w \exp_end_continue_f:w \_\_fp_parse_infix:NN }
}
\_\_fp_tmp:w { inf } \c_inf_fp
\_\_fp_tmp:w { nan } \c_nan_fp
\_\_fp_tmp:w { pi } \c_pi_fp
\_\_fp_tmp:w { deg } \c_one_degree_fp
\_\_fp_tmp:w { true } \c_one_fp
\_\_fp_tmp:w { false } \c_zero_fp

Dimension units are also floating point constants but their value is not stored as a floating point constant. We give the values explicitly here.

\cs_set_protected:Npn \_\_fp_tmp:w #1 #2 {
\cs_new:cpn { __fp_parse_word_#1:N } { \exp_after:wN \_\_fp_exp_after_f:nw { \_\_fp_parse_infix:NN } \s__fp \_\_fp_chk:w 10 #2 ; }
\_\_fp_tmp:w { pt } \c_point_fp
\_\_fp_tmp:w { in } \c_inch_fp
\_\_fp_tmp:w { cm } \c_cm_fp
\_\_fp_tmp:w { mm } \c_mm_fp
\_\_fp_tmp:w { dp } \c_dp_fp
\_\_fp_tmp:w { cc } \c_cc_fp
\_\_fp_tmp:w { nd } \c_nd_fp
\_\_fp_tmp:w { nc } \c_nc_fp
\_\_fp_tmp:w { bp } \c_bp_fp
\_\_fp_tmp:w { sp } \c_sp_fp

(End of definition for \_\_fp_parse_word_in:N and others.)

(End of definition for \_\_fp_parse_word_pi:N, \_\_fp_parse_word_deg:N, \_\_fp_parse_word_infinity:N, \_\_fp_parse_word_nan:N)

(End of definition for \_\_fp_parse_caseless_inf:N, \_\_fp_parse_caseless_infinity:N, and \_\_fp_parse_caseless_nan:N)
The font-dependent units \texttt{em} and \texttt{ex} must be evaluated on the fly. We reuse an auxiliary of \texttt{dim_to_fp:n}.

\begin{verbatim}
cs_new:cpn { __fp_parse_word_#1:N }
{ \exp_after:wN \__fp_parse_apply_unary:NNNwN \exp_after:wN #3 \exp_after:wN #2 \exp_after:wN #1 \exp:w \__fp_parse_operand:Nw \c__fp_prec_func_int \__fp_parse_expand:w }
\end{verbatim}

(End of definition for \texttt{__fp_parse_word_em:N} and \texttt{__fp_parse_word_ex:N}.)

### 71.5.3 Functions

\begin{verbatim}
cs_new:Np \__fp_parse_unary_function:NNN \__fp_parse_function:NNN #1#2#3
{ \exp_after:wN \__fp_parse_apply_unary:NNNwN \exp_after:wN #3 \exp_after:wN #2 \exp_after:wN #1 \exp:w \__fp_parse_operand:Nw \c__fp_prec_func_int \__fp_parse_expand:w }
\end{verbatim}

(End of definition for \texttt{__fp_parse_unary_function:NNN} and \texttt{__fp_parse_function:NNN}.)
### 71.6 Main functions

Start an \exp:w expansion so that \__fp_parse:n expands in two steps. The \__fp_parse_operand:Nw function performs computations until reaching an operation with precedence \c__fp_prec_end_int or less, namely, the end of the expression. The marker \s__fp_expr_mark indicates that the next token is an already parsed version of an infix operator, and \__fp_parse_infix_end:N has infinitely negative precedence. Finally, clean up a (well-defined) set of extra tokens and stop the initial expansion with \exp_end:

\begin{verbatim}
cs_new:Npn \__fp_parse:n #1
c { \exp:w \exp_after:wN \__fp_parse_after:ww \exp:w \__fp_parse_operand:Nw \c__fp_prec_end_int \__fp_parse_expand:w #1 \s__fp_expr_mark \__fp_parse_infix_end:N \s__fp_expr_stop \exp_end: }
cs_new:Npn \__fp_parse_after:ww #1@ \__fp_parse_infix_end:N \s__fp_expr_stop #2 { #2 #1 }
cs_new:Npn \__fp_parse_o:n #1
c { \exp:w \exp_after:wN \__fp_parse_after:ww \exp:w \__fp_parse_operand:Nw \c__fp_prec_end_int \__fp_parse_expand:w #1 \s__fp_expr_mark \__fp_parse_infix_end:N \s__fp_expr_stop \exp_end: }
\end{verbatim}

(End of definition for \__fp_parse:n, \__fp_parse_o:n, and \__fp_parse_after:ww.)

\begin{verbatim}
cs_new:Npn \__fp_parse_operand:Nw \__fp_parse_continue:Nw \__fp_parse_one:Nw
\end{verbatim}

This is just a shorthand which sets up both \__fp_parse_continue:Nw and \__fp_parse_one:Nw with the same precedence. Note the trailing \exp:w.

\begin{verbatim}
cs_new:Npn \__fp_parse_operand:Nw \__fp_parse_one:Nw #1
c { \exp_end_continue_f:w \exp_after:wN \__fp_parse_continue:Nw \exp:w \__fp_parse_one:Nw \exp_after:wN \__fp_parse_continue:Nw \exp_after:wN \__fp_parse_one:Nw \exp:w \exp_end_continue_f:w \exp_after:wN \__fp_parse_one:Nw \exp:w #1 \exp_after:wN \__fp_parse_one:Nw \exp:w \exp_end_continue_f:w \exp_after:wN \__fp_parse_one:Nw \exp:w #1 \exp_after:wN \__fp_parse_one:Nw \exp:w }\end{verbatim}

(End of definition for \__fp_parse_operand:Nw and \__fp_parse_continue:Nw.)
Receives \langle \text{precedence} \rangle \langle \text{operand}_1 \rangle @ \langle \text{operation} \rangle \langle \text{operand}_2 \rangle @ \langle \text{infix command} \rangle. Builds the appropriate call to the \langle \text{operation} \rangle #3, dispatching on both types. If the resulting control sequence does not exist, the operation is not allowed.

This is redefined in 13fp-extras.

\begin{verbatim}
\cs_new:Npn \__fp_parse_apply_binary:NwNwN #1 #2#3@ #4 #5#6@ #7
\{\exp_after:wN \__fp_parse_continue:NwN \exp_after:wN \__fp_parse_apply_binary_chk:NN \__fp_parse_apply_binary_error:NNN \#2 \#4 \#5#6 \#1 \#1 \#2#3 \#4\}
\cs_new:Npn \__fp_binary_type_o:Nww \__fp_binary_rev_type_o:Nww
\{\exp_after:wN \__fp_parse_apply_binary_chk:NN \cs:w \__fp \__fp_type_from_scan:N #2 \#1 \__fp_type_from_scan:N #4 \#5 \__fp_type_from_scan:N #5 \#6 \cs_end:\#1 \#1 \#2 \#2 #3 \; #4 \\cs_end:\#1 \#1 \#2 #3 \; #4 \\cs_end:\#1 \#1 \#2 #3 \; #4 \}
\end{verbatim}

Applies the operator \#1 to its two arguments, dispatching according to their types, and expands once after the result. The rev version swaps its arguments before doing this.
(End of definition for \__fp_binary_type_o:Nww and \__fp_binary_rev_type_o:Nww.)

71.7 Infix operators

\exp_after:wN \__fp_parse_infix_after_operand:NwW
\cs_new:Npn \__fp_parse_infix_after_operand:NwW #1 #2;
\__fp_exp_after_f:nw { \__fp_parse_infix:NN #1 }
\__fp_exp_after:NN #2;
\cs_new:Npn \__fp_parse_infix:NN #1 #2
{\if_catcode:w \scan_stop: \exp_not:N #2\if:w 0 \__fp_str_if_eq:nn { \s__fp_expr_mark } { \exp_not:N #2 }\exp_after:wN \exp_after:wN \exp_after:wN \__fp_parse_infix_juxt:N \else: \exp_after:wN \exp_after:wN \exp_after:wN \__fp_parse_infix_check:NNN \cs:w \__fp_parse_infix_ \token_to_str:N #2 :N \exp_after:wN \exp_after:wN \exp_after:wN \cs_end: \fi: \fi: #1 #2 }\cs_new:Npn \__fp_parse_infix_check:NNN #1#2#3
{\if_meaning:w \scan_stop: \exp_not:N #1 \__fp_int_compare:w \__fp_int_eval:w ('#2 \if_int_compare:w ' #2 > 'Z - 32 \fi: ) / 26 = 3 \exp_stop_f: \exp_after:wN \exp_after:wN \exp_after:wN \__fp_parse_infix_juxt:N \else: \exp_after:wN \exp_after:wN \exp_after:wN \__fp_parse_infix_check:NNN \cs:w \__fp_parse_infix_ \token_to_str:N #2 :N \exp_after:wN \exp_after:wN \exp_after:wN \cs_end: \fi: \fi: #1 #2 }\cs_new:Npn \__fp_parse_infix_check:NNN #1#2#3
{\if_meaning:w \scan_stop: #1
As an infix operator, \s__fp_expr_mark means that the next token (#3) has already
gone through \__fp_parse_infix:NN and should be provided the precedence #1. The
scan mark #2 is discarded.
This one is a little bit odd: force every previous operator to end, regardless of the
precedence.
\begin{verbatim}
\cs_new:Npn \__fp_parse_infix_end:N #1
{ @ \use_none:n \__fp_parse_infix_end:N }
\end{verbatim}

This is very similar to \__fp_parse_infix_end:N, complaining about an extra closing
parenthesis if the previous operator was the beginning of the expression, with precedence
\c__fp_prec_end_int.
\begin{verbatim}
\cs_set_protected:Npn \__fp_tmp:w #1
\cs_new:Npn #1 ##1
\if_int_compare:w ##1 > \c__fp_prec_end_int
\exp_after:wN @ \exp_after:wN \use_none:n \exp_after:wN #1
\else:
\msg_expandable_error:nnn { fp } { extra } { ) }
\exp_after:wN \__fp_parse_infix:NN \exp_after:wN ##1 \exp:w \exp_after:wN \__fp_parse_expand:w \fi:
\exp_args:Nc \__fp_tmp:w { \__fp_parse_infix_):N }
\end{verbatim}

As for other infix operations, if the previous operations has higher precedence the comma
waits. Otherwise we call \__fp_parse_operand:Nw to read more comma-delimited argu-
ments that \__fp_parse_infix_comma:w simply concatenates into a @-delimited array.
The first comma in a tuple that is not a function argument is distinguished: in that case
call \__fp_parse_apply_comma:NwNwN whose job is to convert the first item of the tuple
and an array of the remaining items into a tuple. In contrast to \__fp_parse_apply_-
binary:NwNwN this function’s operands are not single-object arrays.
\begin{verbatim}
\cs_set_protected:Npn \__fp_tmp:w #1
\cs_new:Npn #1 ##1
\if_int_compare:w ##1 > \c__fp_prec_comma_int
\exp_after:wN @ \exp_after:wN \use_none:n \exp_after:wN #1
\else:
\if_int_compare:w ##1 < \c__fp_prec_comma_int
\exp_after:wN @ \exp_after:wN \__fp_parse_apply_comma:NwNwN \exp_after:wN , \exp:w \else:
\exp:w
\end{verbatim}
71.7.2 Usual infix operators

As described in the “work plan”, each infix operator has an associated \..._infix... function, a computing function, and precedence, given as arguments to \__fp_tmp:w. Using the general mechanism for arithmetic operations. The power operation must be associative in the opposite order from all others. For this, we use two distinct precedences.
71.7.3 Juxtaposition

When an opening parenthesis appears where we expect an infix operator, we compute the product of the previous operand and the contents of the parentheses using \_\_fp\_parse\_infix\_mul:N.

\cs_set_protected:Npm \_\_fp_tmp:w #1
{ \cs_new:cpn \_\_fp_parse_infix\_*:N }\#1\#2
{ \if:w * \exp_not:N\#2 
  \exp_after:wN #1
  \exp_after:wN\#1
  \else:
  \exp_after:wN \_\_fp_parse_infix\_mul:N
  \exp_after:wN\#1
  \exp_after:wN\#2
  \fi:
\}
\exp_args:Nc \_\_fp_tmp:w \{ \_\_fp_parse_infix\_:N \} #1
(End of definition for \_\_fp_parse_infix\_:N)

71.7.4 Multi-character cases

\_\_fp_parse_infix\_*:N
\_\_fp_parse_infix\_&:N
\_\_fp_parse_infix\_|:N
(End of definition for \_\_fp_parse_infix\_*:N and others.)
\exp_after:wN \__fp_parse_infix_\&:N \& \__fp_parse_infix_and:N

\__fp parse_infix_or:N
\__fp parse_infix_and:N

(End of definition for \__fp_parse_infix_\|:Nw and \__fp_parse_infix_\&:Nw.)

71.7.5 Ternary operator

\__fp_parse_infix_?:N
\__fp_parse_infix::N

\cs_set_protected:Npn \__fp_tmp:w #1#2#3#4
\cs_new:Npn #1 ##1
\if_int_compare:w ##1 < \c__fp_prec_quest_int
#4
\exp_after:wN @
\exp_after:wN #2
\exp:w
\__fp_parse_operand:Nw #3
\exp_after:wN \__fp_parse_expand:w
\else:
\exp_after:wN @
\exp_after:wN \use_none:n
\exp_after:wN #1
\fi:

\exp_args:Nc \__fp_tmp:w { \__fp_parse_infix_?:Nw }
\__fp_ternary:NwwN \c__fp_prec_quest_int { }
\exp_args:Nc \__fp_tmp:w { \__fp_parse_infix::Nw }
\__fp_ternary_auxii:NwwN \c__fp_prec_colon_int { }
\msg_expandable_error:nnnn
\__fp_compare_end:NNNNw
{ fp } { missing } { ? } { -for-?: }

(End of definition for \__fp_parse_infix_?:Nw and \__fp_parse_infix::Nw.)

71.7.6 Comparisons

\exp_after:wN \__fp_parse_infix_>:N \__fp_parse_infix_!:N
\__fp_parse_excl_error:
\__fp_compare:wNNNNw
\__fp_compare_end:wNNNNw
\__fp_compare:wnNNNNw

\cs_new:cpn { \__fp_parse_infix_<:N } #1
\cs_new:cpn { \__fp_parse_infix=:N } #1
\cs_new:cpn { \__fp_parse_infix=:N } #1
\cs_new:cpn { \__fp_parse_infix=:N } #1
\cs_new:cpn { \__fp_parse_infix=:N } #1

1090
\exp_after:wN \__fp_parse_compare:NNNNNNN
\exp_after:wN \__fp_parse_compare:NNNNNNN \#1
\exp_after:wN \__fp_parse_compare:NNNNNNN 0
\exp_after:wN \__fp_parse_compare:NNNNNNN 1
\exp_after:wN \__fp_parse_compare:NNNNNNN 1
\exp_after:wN \__fp_parse_compare:NNNNNNN 1
\exp_after:wN \__fp_parse_compare:NNNNNNN 1
\cs_new:Npn \__fp_parse_excl_error:
\msg_expandable_error:nnnn { fp } { missing } { = } { ~after~!. }
\cs_new:Npn \__fp_parse_compare:NNNNNNN \#1
\if_int_compare:w \#1 < \c__fp_prec_comp_int
\exp_after:wN \__fp_parse_compare_auxi:NNNNNNN \exp_after:wN \__fp_parse_excl_error:
\else:
\exp_after:wN @ \exp_after:wN \use_none:n \exp_after:wN \__fp_parse_compare:NNNNNNN \fi:
\cs_new:Npn \__fp_parse_compare:NNNNNNN \#1\#2\#3\#4\#5\#6\#7
\if_case:w \__fp_int_eval:w \exp_after:wN \token_to_str:N \#7 - \'<
\__fp_int_eval_end:
\__fp_parse_compare_auxii:NNNNN \#2\#4\#5\#6
\or: \__fp_parse_compare_auxii:NNNNN \#2\#3\#5\#6
\or: \__fp_parse_compare_auxii:NNNNN \#2\#3\#4\#6
\or: \__fp_parse_compare_auxii:NNNNN \#2\#3\#4\#5\#2
\else: \#1 \__fp_parse_compare_end:NNNNw \#3\#4\#5\#6\#7 \fi:
\cs_new:Npn \__fp_parse_compare:NNNNNNN \#1\#2\#3\#4\#5
\if_case:w \__fp_int_eval:w \exp_after:wN \token_to_str:N \#5 - \'<
\__fp_int_eval_end:
\__fp_parse_compare_auxii:NNNNN \#1\#2\#3\#4\#5
\exp_after:wN \__fp_parse_compare_auxii:NNNNN \#1\#2\#3\#4\#5
\exp_after:wN \__fp_parse_compare_auxii:NNNNN \#1\#2\#3\#4\#5 \fi:
\cs_new:Npn \__fp_parse_compare_end:NNNNw \#1\#2\#3\#4\#5 \fi:
\cs_new:Npn \__fp_parse_compare_end:NNNNw \#1\#2\#3\#4\#5
\exp_after:wN \__fp_parse_compare_end:NNNNw \#1
\exp_after:wN \__fp_parse_apply_compare:NwNNN\c_one_fp
\cs_new:Npn \__fp_parse_excl_error:
\msg_expandable_error:nnnn { fp } { missing } { = } { ~after~!. }
\cs_new:Npn \__fp_parse_compare:NNNNNNN \#1
\if_int_compare:w \#1 < \c__fp_prec_comp_int
\exp_after:wN \__fp_parse_compare_auxi:NNNNNNN \exp_after:wN \__fp_parse_excl_error:
\else:
\exp_after:wN @ \exp_after:wN \use_none:n \exp_after:wN \__fp_parse_compare:NNNNNNN \fi:
\cs_new:Npn \__fp_parse_compare:NNNNNNN \#1\#2\#3\#4\#5\#6\#7
\if_case:w \__fp_int_eval:w \exp_after:wN \token_to_str:N \#7 - \'<
\__fp_int_eval_end:
\__fp_parse_compare_auxii:NNNNN \#2\#4\#5\#6
\or: \__fp_parse_compare_auxii:NNNNN \#2\#3\#5\#6
\or: \__fp_parse_compare_auxii:NNNNN \#2\#3\#4\#6
\or: \__fp_parse_compare_auxii:NNNNN \#2\#3\#4\#5\#2
\else: \#1 \__fp_parse_compare_end:NNNNw \#3\#4\#5\#6\#7 \fi:
\cs_new:Npn \__fp_parse_compare:NNNNNNN \#1\#2\#3\#4\#5
\if_case:w \__fp_int_eval:w \exp_after:wN \token_to_str:N \#5 - \'<
\__fp_int_eval_end:
\__fp_parse_compare_auxii:NNNNN \#1\#2\#3\#4\#5
\exp_after:wN \__fp_parse_compare_auxii:NNNNN \#1\#2\#3\#4\#5 \fi:
\cs_new:Npn \__fp_parse_compare_end:NNNNw \#1\#2\#3\#4\#5 \fi:
\cs_new:Npn \__fp_parse_compare_end:NNNNw \#1\#2\#3\#4\#5
\exp_after:wN \__fp_parse_compare_end:NNNNw \#1
\exp_after:wN \__fp_parse_apply_compare:NwNNN\c_one_fp
\cs_new:Npn \__fp_parse_excl_error:
\msg_expandable_error:nnnn { fp } { missing } { = } { ~after~!. }
\cs_new:Npn \__fp_parse_compare:NNNNNNN \#1
\if_int_compare:w \#1 < \c__fp_prec_comp_int
\exp_after:wN \__fp_parse_compare_auxi:NNNNNNN \exp_after:wN \__fp_parse_excl_error:
\else:
\exp_after:wN @ \exp_after:wN \use_none:n \exp_after:wN \__fp_parse_compare:NNNNNNN \fi:
\cs_new:Npn \__fp_parse_compare:NNNNNNN \#1\#2\#3\#4\#5\#6\#7
\if_case:w \__fp_int_eval:w \exp_after:wN \token_to_str:N \#7 - \'<
\__fp_int_eval_end:
\__fp_parse_compare_auxii:NNNNN \#2\#4\#5\#6
\or: \__fp_parse_compare_auxii:NNNNN \#2\#3\#5\#6
\or: \__fp_parse_compare_auxii:NNNNN \#2\#3\#4\#6
\or: \__fp_parse_compare_auxii:NNNNN \#2\#3\#4\#5\#2
\else: \#1 \__fp_parse_compare_end:NNNNw \#3\#4\#5\#6\#7 \fi:
\cs_new:Npn \__fp_parse_compare:NNNNNNN \#1\#2\#3\#4\#5
\if_case:w \__fp_int_eval:w \exp_after:wN \token_to_str:N \#5 - \'<
\__fp_int_eval_end:
\__fp_parse_compare_auxii:NNNNN \#1\#2\#3\#4\#5
\exp_after:wN \__fp_parse_compare_auxii:NNNNN \#1\#2\#3\#4\#5 \fi:
\cs_new:Npn \__fp_parse_compare_end:NNNNw \#1\#2\#3\#4\#5 \fi:
\cs_new:Npn \__fp_parse_compare_end:NNNNw \#1\#2\#3\#4\#5
\exp_after:wN \__fp_parse_compare_end:NNNNw \#1
\exp_after:wN \__fp_parse_apply_compare:NwNNN\c_one_fp
\cs_new:Npn \__fp_parse_excl_error:
\msg_expandable_error:nnnn { fp } { missing } { = } { ~after~!. }
\cs_new:Npn \__fp_parse_compare:NNNNNNN \#1
\if_int_compare:w \#1 < \c__fp_prec_comp_int
\exp_after:wN \__fp_parse_compare_auxi:NNNNNNN \exp_after:wN \__fp_parse_excl_error:
\else:
\exp_after:wN @ \exp_after:wN \use_none:n \exp_after:wN \__fp_parse_compare:NNNNNNN \fi:
\cs_new:Npn \__fp_parse_compare:NNNNNNN \#1\#2\#3\#4\#5\#6\#7
\if_case:w \__fp_int_eval:w \exp_after:wN \token_to_str:N \#7 - \'<
\__fp_int_eval_end:
\__fp_parse_compare_auxii:NNNNN \#2\#4\#5\#6
\or: \__fp_parse_compare_auxii:NNNNN \#2\#3\#5\#6
\or: \__fp_parse_compare_auxii:NNNNN \#2\#3\#4\#6
\or: \__fp_parse_compare_auxii:NNNNN \#2\#3\#4\#5\#2
\else: \#1 \__fp_parse_compare_end:NNNNw \#3\#4\#5\#6\#7 \fi:
\cs_new:Npn \__fp_parse_compare:NNNNNNN \#1\#2\#3\#4\#5
\if_case:w \__fp_int_eval:w \exp_after:wN \token_to_str:N \#5 - \'<
\__fp_int_eval_end:
\__fp_parse_compare_auxii:NNNNN \#1\#2\#3\#4\#5
\exp_after:wN \__fp_parse_compare_auxii:NNNNN \#1\#2\#3\#4\#5 \fi:
\_\_fp\_parse\_apply\_compare:NNNNNNN\\
#1 \#2 \#3 \#4\#5\#6 \#7 \#8 \#9
}\n\cs\new\Npn \_\_fp\_parse\_apply\_compare\_aux:NN\w#1#2\w#3\w#4#5#6#7#8#9
\cs\new\Npn \_\_fp\_parse\_continue\_compare:NN\w#1#2#3#4#5#6#7\w#8#9
\cs\new\Npn \_\_fp\_parse\_infix_<:N\\
\cs\new\Npn \_\_fp\_parse\_function\_all\_fp\_o:fnw

(End of definition for \_\_fp\_parse\_infix_<:N and others.)

### 71.8 Tools for functions

\_\_fp\_parse\_function\_all\_fp\_o:fnw

Followed by \{\textit{(function name)}\} \{\textit{(code)}\} \{\textit{float array}\} @ this checks all floats are floating point numbers (no tuples).
This is followed by \texttt{⟨function name⟩} \texttt{⟨code⟩} \texttt{⟨float array⟩} \@. It checks that the \texttt{⟨float array⟩} consists of one or two floating point numbers (not tuples), then leaves the \texttt{⟨code⟩} (if there is one float) or its tail (if there are two floats) followed by the \texttt{⟨float array⟩}. The \texttt{⟨code⟩} should start with a single token such as \texttt{\__fp_atan_default:w} that deals with the single-float case.

The first \texttt{\__fp_if_type_fp:NTwFw} test catches the case of no argument and the case of a tuple argument. The next one distinguishes the case of a single argument (no error, just add \texttt{\c_one_fp}) from a tuple second argument. Finally check there is no further argument.
Apply #1 to all items in the following tuple and expand once afterwards. The code #1 should itself expand once after its result.

```
cs_new:Npn \_fp_tuple_map_o:nw \_fp_tuple_map_loop_o:nw \_fp_tuple_mapthread_o:nw

\_fp_tuple_mapthread_loop_o:nw
```

Apply #1 to pairs of items in the two following tuples and expand once afterwards.

```
cs_new:Npn \_fp_tuple_mapthread_o:nw \_fp_tuple_mapthread_loop_o:nw
```

(End of definition for \_fp_tuple_map_o:nw, \_fp_tuple_map_loop_o:nw and others.)

Apply #1 to all items in the following tuple and expand once afterwards. The code #1 should itself expand once after its result.
71.9 Messages

\msg_new:nnn { fp } { deprecated }\closemessage{ deprecated }
\msg_new:nnn { fp } { use-#2 }\closemessage{ #1-deprecated; use-#2 }
\msg_new:nnn { fp } { unknown-fp-word }\closemessage{ Unknown-fp-word-#1. }
\msg_new:nnn { fp } { missing }\closemessage{ Missing-#1-inserted #2. }
\msg_new:nnn { fp } { extra }\closemessage{ Extra-#1-ignored. }
\msg_new:nnn { fp } { early-end }\closemessage{ Premature-end-in-fp-expression. }
\msg_new:nnn { fp } { after-e }\closemessage{ Cannot-use-#1 after-`e`. }
\msg_new:nnn { fp } { missing-number }\closemessage{ Missing-number-before-`#1`. }
\msg_new:nnn { fp } { unknown-symbol }\closemessage{ Unknown-symbol-#1-ignored. }
\msg_new:nnn { fp } { extra-comma }\closemessage{ Unexpected-comma-turned-to-nan-result. }
\msg_new:nnn { fp } { no-arg }\closemessage{ #1-got-no-argument; used-nan. }
\msg_new:nnn { fp } { multi-arg }\closemessage{ #1-got-more-than-one-argument; used-nan. }
\msg_new:nnn { fp } { num-args }\closemessage{ #1-expects-between-#2-and-#3-arguments. }
\msg_new:nnn { fp } { bad-args }\closemessage{ Arguments-in-#1#2-are-invalid. }
\msg_new:nnn { fp } { infy-pi }\closemessage{ Math-command-#1 is-not-an-fp }
\cs_if_exist:cT { @unexpandable@protect }\closemessage{ \unexpandable@protect }
\msg_new:nnn { fp } { robust-cmd }\closemessage{ Robust-command-#1 invalid-in-fp-expression! }
\msg_new:nnn { fp } \{ /*package*/ }
Chapter 72

l3fp-assign implementation

72.1 Assigning values

Floating point variables are initialized to be +0.

```latex
\fp_new:N
\cs_new_protected:Npn \fp_new:N #1
\cs_new_eq:NN #1 \c_zero_fp
\cs_generate_variant:Nn \fp_new:N {c}
```

(End of definition for \fp_new:N. This function is documented on page 260.)

```latex
\fp_set:Nn
\fp_set:cn
\fp_gset:Nn
\fp_gset:cn
\fp_const:Nn
\fp_const:cn
\cs_new_protected:Npn \fp_set:Nn #1#2
\cs_new_protected:Npn \fp_gset:Nn #1#2
\cs_new_protected:Npn \fp_const:Nn #1#2
\cs_generate_variant:Nn \fp_set:Nn {c}
\cs_generate_variant:Nn \fp_gset:Nn {c}
\cs_generate_variant:Nn \fp_const:Nn {c}
```

(End of definition for \fp_set:Nn, \fp_gset:Nn, and \fp_const:Nn. These functions are documented on page 260.)

```latex
\fp_set_eq:NN
\fp_set_eq:cN
\fp_set_eq:Nc
\fp_gset_eq:NN
\fp_gset_eq:cc
\fp_gset_eq:cN
\fp_gset_eq:Nc
\fp_gset_eq:cc
\cs_new_eq:NN \fp_set_eq:NN \tl_set_eq:NN
\cs_new_eq:NN \fp_gset_eq:NN \tl_gset_eq:NN
\cs_generate_variant:Nn \fp_set_eq:NN { c , Nc , cc }
\cs_generate_variant:Nn \fp_gset_eq:NN { c , Nc , cc }
```

(End of definition for \fp_set_eq:NN and \fp_gset_eq:NN. These functions are documented on page 260.)
Setting a floating point to zero: copy \texttt{c_zero_fp}.

\begin{verbatim}
\cs_new_protected:Npn \fp_zero:N #1 { \fp_set_eq:NN #1 \c_zero_fp }
\cs_new_protected:Npn \fp_gzero:N #1 { \fp_gset_eq:NN #1 \c_zero_fp }
\cs_generate_variant:Nn \fp_zero:N { c }
\cs_generate_variant:Nn \fp_gzero:N { c }
\end{verbatim}

(End of definition for \texttt{fp_zero:N} and \texttt{fp_gzero:N}. These functions are documented on page 260.)

Set the floating point to zero, or define it if needed.

\begin{verbatim}
\cs_new_protected:Npn \fp_zero_new:N #1 { \fp_if_exist:NTF #1 { \fp_zero:N #1 } { \fp_new:N #1 } }
\cs_new_protected:Npn \fp_gzero_new:N #1 { \fp_if_exist:NTF #1 { \fp_gzero:N #1 } { \fp_new:N #1 } }
\cs_generate_variant:Nn \fp_zero_new:N { c }
\cs_generate_variant:Nn \fp_gzero_new:N { c }
\end{verbatim}

(End of definition for \texttt{fp_zero_new:N} and \texttt{fp_gzero_new:N}. These functions are documented on page 260.)

### 72.2 Updating values

These match the equivalent functions in \texttt{l3int} and \texttt{l3skip}.

\begin{verbatim}
\cs_new_protected:Npn \fp_add:Nn { \__fp_add:NNNn \fp_set:Nn + }
\cs_generate_variant:Nn \fp_add:Nn { c }
\cs_new_protected:Npn \fp_gadd:Nn { \__fp_add:NNNn \fp_gset:Nn + }
\cs_generate_variant:Nn \fp_gadd:Nn { c }
\cs_new_protected:Npn \fp_sub:Nn { \__fp_add:NNNn \fp_set:Nn - }
\cs_generate_variant:Nn \fp_sub:Nn { c }
\cs_new_protected:Npn \fp_gsub:Nn { \__fp_add:NNNn \fp_gset:Nn - }
\cs_generate_variant:Nn \fp_gsub:Nn { c }
\end{verbatim}

(End of definition for \texttt{fp_add:Nn} and others. These functions are documented on page 260.)

### 72.3 Showing values

This shows the result of computing its argument by passing the right data to \texttt{\tl_show:n} or \texttt{\tl_log:n}.

\begin{verbatim}
\cs_new_protected:Npn \fp_show:N { \__fp_show:NN \tl_show:n }
\cs_generate_variant:Nn \fp_show:N { c }
\cs_new_protected:Npn \fp_log:N { \__fp_show:NN \tl_log:n }
\cs_generate_variant:Nn \fp_log:N { c }
\end{verbatim}

\begin{verbatim}
\cs_new_protected:Npn \__kernel_chk_tl_type:NnnT #1 #2 #3 { \fp }
\end{verbatim}
To support symbolic expression, validation has to be done recursively. Two \@@_show_validate:nn wrappers are used to distinguish between initial and recursive calls, in which the former provides a demo of possible forms a fp variable would have.
\begin{verbatim}
\cs_new:Npn \__fp_tuple_show_validate:w #1 \s__fp_tuple \__fp_tuple_chk:w #2 #3 #4 \s__fp_stop {
  \str_if_eq:nnF { #2 } {?}
  { \s__fp_tuple \__fp_tuple_chk:w { \__fp_show_validate_aux:n { #2 } } ; }
\}
\cs_new:Npn \__fp_symbolic_show_validate:w #1 \s__fp_symbolic \__fp_symbolic_chk:w #2 , #3 #4 \s__fp_stop {
  \str_if_eq:nnF { #2 } {?}
  { \s__fp_symbolic \__fp_symbolic_chk:w \exp_not:n { #2 } ,
    \{ \__fp_show_validate_aux:n { #3 } \} ;
    \__fp_show_validate_aux:n { #5 } }
\}
\end{verbatim}

(End of definition for \__fp_show_validate:n and others.)

\fp_show:n
\fp_log:n
Use general tools.

\begin{verbatim}
\cs_new_protected:Npn \fp_show:n { \__kernel_msg_show_eval:Nn \fp_to_tl:n }
\cs_new_protected:Npn \fp_log:n { \__kernel_msg_log_eval:Nn \fp_to_tl:n }
\end{verbatim}

(End of definition for \fp_show:n and \fp_log:n. These functions are documented on page 271.)

\subsection{Some useful constants and scratch variables}

Some constants.

\begin{verbatim}
\c_one_fp
\c_e_fp
\fp_const:Nn \c_e_fp { 2.718 2818 2845 9045 }
\fp_const:Nn \c_one_fp { 1 }
\end{verbatim}

(End of definition for \c_one_fp and \c_e_fp. These variables are documented on page 269.)

\begin{verbatim}
\c_pi_fp
\c_one_degree_fp
\fp_const:Nn \c_pi_fp { 3.141 5926 5358 9793 }
\fp_const:Nn \c_one_degree_fp { 0.0 1745 3292 5199 4330 }
\end{verbatim}

(End of definition for \c_pi_fp and \c_one_degree_fp. These variables are documented on page 269.)

Scratch variables are simply initialized there.

\begin{verbatim}
\l_tmpa_fp
\l_tmpb_fp
\g_tmpa_fp
\g_tmpb_fp
\fp_new:N \l_tmpa_fp
\fp_new:N \l_tmpb_fp
\fp_new:N \g_tmpa_fp
\fp_new:N \g_tmpb_fp
\end{verbatim}

(End of definition for \l_tmpa_fp and others. These variables are documented on page 269.)

\end{verbatim}

72.4 Some useful constants and scratch variables
Chapter 73

\textbf{l3fp-logic implementation}

\begin{verbatim}
25541 \{+package\}
25542 \{@@=fp\}
\__fp_parse_word_max:N \__fp_parse_word_min:N

Those functions may receive a variable number of arguments.

\begin{verbatim}
25543 \cs_new:Npn \__fp_parse_word_max:N { \__fp_parse_function:NNN \__fp_minmax_o:Nw 2 }
25544 \cs_new:Npn \__fp_parse_word_min:N { \__fp_parse_function:NNN \__fp_minmax_o:Nw 0 }
\end{verbatim}

(End of definition for \__fp_parse_word_max:N and \__fp_parse_word_min:N.)

73.1 Syntax of internal functions

- \texttt{\_fp_compare_npos:nwnw \{\langle expon1\rangle \} \{\langle body1\rangle \}; \{\langle expon2\rangle \} \{\langle body2\rangle \};}
- \texttt{\_fp_minmax_o:Nw \langle sign\rangle \langle floating point array\rangle}
- \texttt{\_fp_not_o:w \? \langle floating point array\rangle} (with one floating point number only)
- \texttt{\_fp_&_o:ww \langle floating point\rangle \langle floating point\rangle}
- \texttt{\_fp_|_o:ww \langle floating point\rangle \langle floating point\rangle}
- \texttt{\_fp_ternary:NwwN, \_fp_ternary_auxi:NwwN, \_fp_ternary_auxii:NwwN}

have to be understood.

73.2 Tests

\texttt{\fp_if_exist_p:N} \texttt{\fp_if_exist_p:c} \texttt{\fp_if_exist:NTF} \texttt{\fp_if_exist:cTF}

Copies of the cs functions defined in \texttt{l3basics}.

\begin{verbatim}
25547 \prg_new_eq_conditional:NNn \fp_if_exist:N \cs_if_exist:N \{ TF , T , F , p \}
25548 \prg_new_eq_conditional:NNn \fp_if_exist:c \cs_if_exist:c \{ TF , T , F , p \}
\end{verbatim}

(End of definition for \texttt{\fp_if_exist:NTF}. This function is documented on page 262.)

1100
Evaluate and check if the result is a floating point of the same kind as \texttt{nan}.

\begin{verbatim}
\if:3 \exp_last_unbraced:Nf \__fp_kind:w { \__fp_parse:n (#1) }
\prg_return_true:
\else:
\prg_return_false:
\fi:
\end{verbatim}

(End of definition for \texttt{\fp_if_nan:nTF}. This function is documented on page 264.)

### 73.3 Comparison

Within floating point expressions, comparison operators are treated as operations, so we evaluate \#1, then compare with $\pm 0$. Tuples are \texttt{true}.

\begin{verbatim}
\exp_after:wN \__fp_compare_return:w
\exp:w \exp_end_continue_f:w \__fp_parse:n {#1}
\end{verbatim}

\begin{verbatim}
\if_charcode:w 0 \__fp_if_type_fp:NTwFw #1 { \__fp_use_i_delimit_by_s_stop:nw #3 \s__fp_stop }
\s__fp 1 \s__fp_stop
\prg_return_false:
\else:
\prg_return_true:
\fi:
\end{verbatim}

(End of definition for \texttt{\fp_compare:nTF} and \texttt{\__fp_compare_return:w}. This function is documented on page 264.)

Evaluate \#1 and \#3, using an auxiliary to expand both, and feed the two floating point numbers swapped to \texttt{\__fp_compare_back_any:ww}, defined below. Compare the result with `'#2'='`, which is $-1$ for $\lt$, $0$ for $=$, $1$ for $\gt$ and $2$ for $\equiv$.

\begin{verbatim}
\if_int_compare:w \exp_after:wN \__fp_compare_back_any:ww \exp:w \exp_end_continue_f:w \__fp_parse:n {#1} {#3} = \__fp_int_eval:w '#2 - '=' \__fp_int_eval_end:
\prg_return_true:
\else:
\prg_return_false:
\fi:
\end{verbatim}

Evaluate #1 and #3, using an auxiliary to expand both, and feed the two floating point numbers swapped to \texttt{\__fp_compare_back_any:ww}, defined below. Compare the result with `'#2'='`, which is $-1$ for $\lt$, $0$ for $=$, $1$ for $\gt$ and $2$ for $\equiv$.
\__fp_compare_back:ww \__fp_bcmp:ww \__fp_compare_back_any:ww \__fp_compare_nan:ww

\__fp_compare_back:ww \langle y \rangle ; \langle x \rangle ;
Expands (in the same way as \int_eval:n) to −1 if \( x < y \), 0 if \( x = y \), 1 if \( x > y \), and 2 otherwise (denoted as \( x' y \)). If either operand is \texttt{nan}, stop the comparison with \__fp_compare_nan:w returning 2. If \( x \) is negative, swap the outputs 1 and −1 (i.e., \( > \) and \( < \)); we can henceforth assume that \( x \geq 0 \). If \( y \geq 0 \), and they have the same type, either they are normal and we compare them with \__fp_compare_npos:nww, or they are equal. If \( y \geq 0 \), but of a different type, the highest type is a larger number. Finally, if \( y \leq 0 \), then \( x > y \), unless both are zero.

\cs_new:Npn \__fp_compare_back:ww #1#2; #3#4;
{\cs:w \__fp \__fp_type_from_scan:N #1 _bcmp \__fp_type_from_scan:N #3 :ww \cs_end: #1#2; #3#4; }
\cs_new:Npn \__fp_compare_back_any:ww #1#2; #3
{\__fp_if_type_fp:NTwFw #1 { \__fp_if_type_fp:NTwFw #3 \use_i:nn \s__fp \use_ii:nn \s__fp_stop } \s__fp \use_ii:nn \s__fp_stop \__fp_compare_back:ww {\cs:w \__fp \__fp_type_from_scan:N #1 _compare_back \__fp_type_from_scan:N #3 :ww \cs_end: } \#1#2; #3 }
\cs_new:Npn \__fp_bcmp:ww \s__fp \__fp_chk:w #1 #2 #3; \s__fp \__fp_chk:w #4 #5 #6;
{\int_value:w \if_meaning:w 3 #1 \exp_after:wN \__fp_compare_nan:w \fi: \if_meaning:w 3 #4 \exp_after:wN \__fp_compare_nan:w \fi: \if_meaning:w 2 #5 \fi: \if_meaning:w 2 #2 \#5 \if_meaning:w 2 #1 \#4 \if_meaning:w 1 \#1 \if_meaning:w 1 \#1 \__fp_compare_npos:nww #6; #3; }

(End of definition for \fp_compare:nNnTF and \__fp_compare_aux:wn. This function is documented on page 263.)
Tuple and floating point numbers are not comparable so return 2 in mixed cases or when tuples have a different number of items. Otherwise compare pairs of items with \texttt{\_\_fp\_compare\_back\_any:ww} and if any don’t match return 2 (as \texttt{\int\_value:w \_\_fp\_compare\_back\_tuple:ww \_\_fp\_compare\_back\_loop:ww \_\_fp\_compare\_back\_tuple:ww \_\_fp\_compare\_back\_loop:ww}).
Within an \int_value:w \exp_stop_f: construction, this expands to 0 if the
two numbers are equal, \(-1\) if the first is smaller, and \(1\) if the first is bigger. First
compare the exponents: the larger one denotes the larger number. If they are equal, we
must compare significands. If both the first 8 digits and the next 8 digits coincide, the
numbers are equal. If only the first 8 digits coincide, the next 8 decide. Otherwise, the
first 8 digits are compared.

\cs_new:Npn \__fp_compare_npos:nwnw #1#2; #3#4;
{\if_int_compare:w #1 = #3 \exp_stop_f:
  \__fp_compare_significand:nnnnnnnn #2 #4
\else:
  \if_int_compare:w #1 < #3 - \fi: 1
\fi:}
\cs_new:Npn \__fp_compare_significand:nnnnnnnn #1#2#3#4#5#6#7#8
{\if_int_compare:w #1#2 = #5#6 \exp_stop_f:
  \if_int_compare:w #3#4 = #7#8 \exp_stop_f:
    0
  \else:
    \if_int_compare:w #3#4 < #7#8 - \fi: 1
  \fi:
\else:
  \if_int_compare:w #1#2 < #5#6 - \fi: 1
\fi:}

(End of definition for \__fp_compare_npos:nwnw and \__fp_compare_significand:nnnnnnnn.)

### 73.4 Floating point expression loops

These are quite easy given the above functions. The do_until and do_while versions
execute the body, then test. The until_do and while_do do it the other way round.

\cs_new:Npn \fp_do_until:nn #1#2
{\fp_compare:nF {#1} {\fp_do_until:nn {#1} {#2} }}
\cs_new:Npn \fp_do_while:nn #1#2
{\fp_compare:nT {#1} {\fp_do_while:nn {#1} {#2} }}
\cs_new:Npn \fp_until_do:nn #1#2
{\fp_compare:nF {#1} {\fp_until_do:nn {#1} {#2} }}
\cs_new:Npn \fp_while_do:nn #1#2
{\fp_compare:nT {#1} {\fp_while_do:nn {#1} {#2} }}
\cs_new:Npn \fp_while_do:nn {#1}{#2} \fp_while_do:nn {#1} {#2} \fp_while_do:nn {#1}{#2}

(End of definition for \fp_do_until:nNnn and others. These functions are documented on page 265.)

\cs_new:Npn \fp_do_until:nNnn #1#2#3#4
\cs_new:Npn \fp_do_while:nNnn #1#2#3#4
\cs_new:Npn \fp_until_do:nNnn #1#2#3#4
\cs_new:Npn \fp_while_do:nNnn #1#2#3#4

As above but not using the nNn syntax.

\cs_new:Npn \fp_do_until:nNnn #1#2#3#4
\cs_new:Npn \fp_do_while:nNnn #1#2#3#4
\cs_new:Npn \fp_until_do:nNnn #1#2#3#4
\cs_new:Npn \fp_while_do:nNnn #1#2#3#4

The approach here is somewhat similar to \int_step_function:nnnN. There are two
subtleties: we use the internal parser \__fp_parse:n to avoid converting back and forth
from the internal representation; and (due to rounding) even a non-zero step does not
guarantee that the loop counter increases.

\cs_new:Npn \fp_step_function:nnnN \fp_step_function:nnnc
\cs_new:Npn \fp_step_function:nnnN #1#2#3
\exp_after:wN \__fp_step:wwwN \__fp_parse_o:n {#1} \__fp_parse_o:n {#2} 

(End of definition for \fp_do_until:nNnn and others. These functions are documented on page 264.)
Only floating point numbers (not tuples) are allowed arguments. Only “normal” floating points (not ±0, ±\text{inf}, nan) can be used as step; if positive, call \_\_fp_step:Nnnnn with argument > otherwise <. This function has one more argument than its integral counterpart, namely the previous value, to catch the case where the loop has made no progress. Conversion to decimal is done just before calling the user’s function.

\begin{verbatim}
\cs_new:Npn \_\_fp_step:wwwN #1#2; #3#4; #5#6; #7
{ \_\_fp_if_type_fp:NTwFw #1 { } \s__fp \prg_break: \s__fp_stop
\_\_fp_if_type_fp:NTwFw #3 { } \s__fp \prg_break: \_\_fp_step
\_\_fp_if_type_fp:NTwFw #5 { } \s__fp \prg_break: \_\_fp_step
\use_i:nnnn { \_\_fp_step_fp:wwwN #1#2; #3#4; #5#6; #7 }
\prg_break_point:
\use:n
{ \_\_fp_error:nfff { step-tuple } { \fp_to_tl:n { #1#2; } }
{ \fp_to_tl:n { #3#4; } } { \fp_to_tl:n { #5#6; } }
}
\cs_new:Npn \_\_fp_step_fp:wwwN #1; \s__fp \_\_fp_chk:w #2#3#4; #5; #6
{ \token_if_eq_meaning:NNTF #2 1
{ \token_if_eq_meaning:NNTF #3 0
{ \_\_fp_step:Nnnnn #1#2; #3; #5; #6
{ \_\_fp_error:nfff { step-tuple } { \fp_to_tl:n { #1#2; } }
{ \fp_to_tl:n { #3#4; } } { \fp_to_tl:n { #5#6; } }
}
\_\_fp_error:nfff { zero-step } { \_\_fp_chk:w #2#3#4; #5; #6
{ \_\_fp_error:nfff { bad-step } { }
{ \fp_to_tl:n { \s__fp \_\_fp_chk:w #2#3#4; } } {#6}
\use_none:nnnn
{ #1; } \{ \c_nan_fp \} \{ \_\_fp_chk:w #2#3#4; \} \{ #5; \} #6
}
\cs_new:Npn \_\_fp_step:Nnnnn #1#2#3#4#5#6
{ \fp_compare:nNnTF {#2} = {#3}
{ \_\_fp_error:nffn { tiny-step }
{ \fp_to_tl:n {#3} } { \fp_to_tl:n {#4} } {#6}
}
{ \fp_compare:nNnF {#2} #1 {#5}
{}

1106
\end{verbatim}
\exp_args:Nf #6 \{ \_\_fp_to_decimal_dispatch:w \#2 \}
\_\_fp_step:NfnnnN
\#1 \{ \_\_fp_parse:n \{ \#2 + \#4 \} \{\#2\} \{\#4\} \{\#5\} \#6 \}
}
\cs_generate_variant:Nn \_\_fp_step:NnnnnN { Nf }

(End of definition for \fp_step_function:nnnn and others. This function is documented on page 266.)

As for \int_step_inline:nnnn, create a global function and apply it, following up with a break point.

\cs_new_protected:Npm \fp_step_inline:nnnn
\{ \int_gincr:N \g__kernel_prg_map_int
\exp_args:NNc \_\_fp_step:NNnnnn
\cs_gset_protected:Npn
\{ \_\_fp_map_\int_use:N \g__kernel_prg_map_int :w \}
\}
\cs_new_protected:Npm \fp_step_variable:nnnNn #1#2#3#4#5
\{ \int_gincr:N \g__kernel_prg_map_int
\exp_args:NNc \_\_fp_step:NNnnnn
\cs_gset_protected:Npe
\{ \_\_fp_map_\int_use:N \g__kernel_prg_map_int :w \}
\{ #1 \} \{#2\} \{#3\}
{ \tl_set:Nn \exp_not:N #4 {##1} \exp_not:n \{#5\}
\}
\exp_not:n \{ #5\}
\}
shop catastrophe \int_gdecr:N \g__kernel_prg_map_int
\}

(End of definition for \fp_step_inline:nnnn, \fp_step_variable:nnnNn, and \_\_fp_step:NNnnnn.
These functions are documented on page 266.)

\msg_new:nnn { fp } { step-tuple }
\{ Tuple-argument-in-fp_step....-(#1)#2(#3). \}
\msg_new:nnn { fp } { bad-step }
\{ Invalid-step-size=#2-for-function=#3. \}
\msg_new:nnn { fp } { tiny-step }
\{ Tiny-step-size-(#1+#2=#1)-for-function=#3. \}

73.5 Extrema

First check all operands are floating point numbers. The argument \#1 is 2 to find the maximum of an array \#2 of floating point numbers, and 0 to find the minimum. We read numbers sequentially, keeping track of the largest (smallest) number found so far. If numbers are equal (for instance ±0), the first is kept. We append −∞ (∞), for the case
of an empty array. Since no number is smaller (larger) than that, this additional item only affects the maximum (minimum) in the case of \texttt{max()} and \texttt{min()} with no argument. The weird \texttt{fp-like} trailing marker breaks the loop correctly: see the precise definition of \texttt{\_\_fp_minmax_loop:Nww}.

\begin{verbatim}
\cs_new:Npn \__fp_minmax_o:Nw #1
 { \__fp_parse_function_all_fp_o:fnw
\{ \token_if_eq_meaning:NNTF 0 #1 { min } { max } \}
\{ \__fp_minmax_aux_o:Nw #1 \}
}
\cs_new:Npn \__fp_minmax_aux_o:Nw #1#2 @
 { \if_meaning:w 0 #1
 \exp_after:wN \__fp_minmax_loop:Nww \exp_after:wN +
 \else:
 \exp_after:wN \__fp_minmax_loop:Nww \exp_after:wN -
 \fi:
 #2
 \s__fp \__fp_chk:w 2 #1 \s__fp_exact ;
 \s__fp \__fp_chk:w { 3 \__fp_minmax_break_o:w } ;
}
\end{verbatim}

(End of definition for \texttt{\_\_fp_minmax_o:Nw} and \texttt{\_\_fp_minmax_aux_o:Nw})

\texttt{\_\_fp_minmax_loop:Nww} The first argument is \texttt{-} or \texttt{+} to denote the case where the currently largest (smallest) number found (first floating point argument) should be replaced by the new number (second floating point argument). If the new number is \texttt{nan}, keep that as the extremum, unless that extremum is already a \texttt{nan}. Otherwise, compare the two numbers. If the new number is larger (in the case of \texttt{max}) or smaller (in the case of \texttt{min}), the test yields \texttt{true}, and we keep the second number as a new maximum; otherwise we keep the first number. Then loop.

\begin{verbatim}
\cs_new:Npn \__fp_minmax_loop:Nww
 #1 \s__fp \__fp_chk:w #2#3; \s__fp \__fp_chk:w #4#5;
 { \if_meaning:w 3 #4
 \if_meaning:w 3 #2
 \__fp_minmax_auxi:ww\else:
 \__fp_minmax_auxii:ww
 \fi:
\else:
 \if_int_compare:w
 \__fp_compare_back:ww
 \s__fp \__fp_chk:w #4#5;
 \s__fp \__fp_chk:w #2#3;
 = #1 1 \exp_stop_f:
 \__fp_minmax_auxii:ww\else:
 \__fp_minmax_auxi:ww
 \fi:
 \fi:
 \__fp_minmax_loop:Nww #1
 \s__fp \__fp_chk:w #2#3;
\end{verbatim}

1108
\__fp \__fp_chk:w #4#5;
}

(End of definition for \__fp\textunderscore minmax\textunderscore loop:Nww.)
\__fp\textunderscore minmax\textunderscore auxi:ww
Keep the first/second number, and remove the other.
\__fp\textunderscore minmax\textunderscore auxii:ww
\__fp\textunderscore minmax\textunderscore break\textunderscore o:w
This function is called from within an \texttt{if\textunderscore meaning:w} test. Skip to the end of the tests, close the current test with \texttt{\iffi:}, clean up, and return the appropriate number with one post-expansion.
\__fp\textunderscore not\textunderscore o:w
\__fp\textunderscore tuple\textunderscore not\textunderscore o:w
Return \texttt{true} or \texttt{false}, with two expansions, one to exit the conditional, and one to please l3fp-parse. The first argument is provided by l3fp-parse and is ignored.
\__fp\textunderscore &\textunderscore o:ww
\__fp\textunderscore &\textunderscore tuple\textunderscore o:ww
\__fp\textunderscore &\textunderscore tuple\textunderscore tuple\textunderscore o:ww
\__fp\textunderscore |\textunderscore o:ww
\__fp\textunderscore |\textunderscore tuple\textunderscore o:ww
\__fp\textunderscore |\textunderscore tuple\textunderscore tuple\textunderscore o:ww
\__fp\textunderscore and\textunderscore return:wWw
For \texttt{and}, if the first number is zero, return it (with the same sign). Otherwise, return the second one. For \texttt{or}, the logic is reversed: if the first number is non-zero, return it, otherwise return the second number: we achieve that by hi-jacking \__fp\textunderscore &\textunderscore o:ww, inserting an extra argument, \texttt{\else:}, before \texttt{s\_fp}. In all cases, expand after the floating point number.

73.6 Boolean operations

\__fp\textunderscore not\textunderscore o:w
\__fp\textunderscore tuple\textunderscore not\textunderscore o:w
\__fp\textunderscore &\textunderscore o:ww
\__fp\textunderscore &\textunderscore tuple\textunderscore o:ww
\__fp\textunderscore &\textunderscore tuple\textunderscore tuple\textunderscore o:ww
\__fp\textunderscore |\textunderscore o:ww
\__fp\textunderscore |\textunderscore tuple\textunderscore o:ww
\__fp\textunderscore |\textunderscore tuple\textunderscore tuple\textunderscore o:ww
\__fp\textunderscore and\textunderscore return:wWw
\__fp\textunderscore not\textunderscore o:w
\__fp\textunderscore tuple\textunderscore not\textunderscore o:w
\__fp\textunderscore &\textunderscore o:ww
\__fp\textunderscore &\textunderscore tuple\textunderscore o:ww
\__fp\textunderscore &\textunderscore tuple\textunderscore tuple\textunderscore o:ww
\__fp\textunderscore |\textunderscore o:ww
\__fp\textunderscore |\textunderscore tuple\textunderscore o:ww
\__fp\textunderscore |\textunderscore tuple\textunderscore tuple\textunderscore o:ww
\__fp\textunderscore and\textunderscore return:wWw
\__fp\textunderscore not\textunderscore o:w
\__fp\textunderscore tuple\textunderscore not\textunderscore o:w
\__fp\textunderscore &\textunderscore o:ww
\__fp\textunderscore &\textunderscore tuple\textunderscore o:ww
\__fp\textunderscore &\textunderscore tuple\textunderscore tuple\textunderscore o:ww
\__fp\textunderscore |\textunderscore o:ww
\__fp\textunderscore |\textunderscore tuple\textunderscore o:ww
\__fp\textunderscore |\textunderscore tuple\textunderscore tuple\textunderscore o:ww
\__fp\textunderscore and\textunderscore return:wWw
\__fp\textunderscore not\textunderscore o:w
\__fp\textunderscore tuple\textunderscore not\textunderscore o:w
\__fp\textunderscore &\textunderscore o:ww
\__fp\textunderscore &\textunderscore tuple\textunderscore o:ww
\__fp\textunderscore &\textunderscore tuple\textunderscore tuple\textunderscore o:ww
\__fp\textunderscore |\textunderscore o:ww
\__fp\textunderscore |\textunderscore tuple\textunderscore o:ww
\__fp\textunderscore |\textunderscore tuple\textunderscore tuple\textunderscore o:ww
\__fp\textunderscore and\textunderscore return:wWw

1109
The first function receives the test and the true branch of the `?:` ternary operator. It calls `\__fp_ternary_auxii:NwwN` if the test branch is a floating point number ±0, and otherwise calls `\__fp_ternary_auxi:NwwN`. These functions select one of their two arguments.

```
\cs_new:Npn \__fp_ternary:NwwN #1 #2#3@ #4@ #5
  { \if_meaning:w \__fp_parse_infix_::N #5
    \if_charcode:w 0
      \__fp_if_type_fp:NTwFw
        \__fp_use_i_delimit_by_s_stop:nw #3 \s__fp_stop
      \else:
        \__fp_use_i_delimit_by_s_stop:nw #3 \s__fp_stop
      \fi:
      \__fp_use_i_delimit_by_s_stop:nw #3 \s__fp_stop
    \else:
      \msg_expandable_error:nnnn { fp } { missing } { : } { ~for~?: }
    \fi:
    \__fp_use_i_delimit_by_s_stop:nw #3 \s__fp_stop
  }
\cs_new:Npn \__fp_ternary_auxi:NwwN
  { \__fp_ternary:NwwN #1 \exp:w #4 \__fp_expr_stop
    \exp_after:wN \__fp瘌parse_operand:Nw \c__fp_prec_colon_int
    \__fp爴parse-expand:w
  }
\cs_new:Npn \__fp_ternary_auxii:NwwN
  { \__fp_ternary:NwwN #1 \exp:w #4 \__fp_expr_stop
    \exp_after:wN \__fp爴parse_operand:Nw \c__fp爴prec_colon_int
    \__fp爴parse-expand:w
  }
```

(End of definition for `\__fp_\&_o:ww` and others.)

### 73.7 Ternary operator

The first function receives the test and the true branch of the `?:` ternary operator. It calls `\__fp_ternary_auxii:NwwN` if the test branch is a floating point number ±0, and otherwise calls `\__fp_ternary_auxi:NwwN`. These functions select one of their two arguments.
\cs_new:Npn \__fp_ternary_auxi:NwwN #1\#2\#3\#4
{
\exp_after:wN \__fp_parse_continue:NwN
\exp_after:wN #1
\exp:w \exp_end_continue_f:w
\__fp_exp_after_array_f:w #2 \s__fp_expr_stop
#4 #1
}
\cs_new:Npn \__fp_ternary_auxii:NwwN #1\#2\#3\#4
{
\exp_after:wN \__fp_parse_continue:NwN
\exp_after:wN #1
\exp:w \exp_end_continue_f:w
\__fp_exp_after_array_f:w #3 \s__fp_expr_stop
#4 #1
}

(End of definition for \__fp_ternary:NwwN, \__fp_ternary_auxi:NwwN, and \__fp_ternary_auxii:NwwN.)

(/package)
Chapter 74

l3fp-basics implementation

The l3fp-basics module implements addition, subtraction, multiplication, and division of two floating points, and the absolute value and sign-changing operations on one floating point. All operations implemented in this module yield the outcome of rounding the infinitely precise result of the operation to the nearest floating point.

Some algorithms used below end up being quite similar to some described in “What Every Computer Scientist Should Know About Floating Point Arithmetic”, by David Goldberg, which can be found at http://cr.yp.to/2005-590/goldberg.pdf.

Unary functions.

\__fp_parse_word_abs:N
\__fp_parse_word_logb:N
\__fp_parse_word_sign:N
\__fp_parse_word_sqrt:N

(End of definition for \__fp_parse_word_abs:N and others.)

74.1 Addition and subtraction

We define here two functions, \_\_fp_-_o:ww and \_\_fp_+_o:ww, which perform the subtraction and addition of their two floating point operands, and expand the tokens following the result once.

A more obscure function, \_\_fp_add_big_i_o:ww, is used in l3fp-expo.

The logic goes as follows:

- \_\_fp_-_o:ww calls \_\_fp_+_o:ww to do the work, with the sign of the second operand flipped;
- \_\_fp_+_o:ww dispatches depending on the type of floating point, calling specialized auxiliaries;
• in all cases except summing two normal floating point numbers, we return one or the other operands depending on the signs, or detect an invalid operation in the case of $\infty - \infty$;

• for normal floating point numbers, compare the signs;

• to add two floating point numbers of the same sign or of opposite signs, shift the significand of the smaller one to match the bigger one, perform the addition or subtraction of significands, check for a carry, round, and pack using the `__fp_basics_pack...` functions.

The trickiest part is to round correctly when adding or subtracting normal floating point numbers.

### 74.1.1 Sign, exponent, and special numbers

\_\_fp\_\_o:ww

The `\_\_fp\_\_o:ww` auxiliary has a hook: it takes one argument between the first `\s__fp` and `\_\_fp\_chk:w`, which is applied to the sign of the second operand. Positioning the hook there means that `\_\_fp\_\_o:ww` can still perform the sanity check that it was followed by `\a__fp`.

```latex
\cs_new:c { \__fp_+o:ww } \s__fp \exp_not:c { \__fp_+o:ww } \exp_not:n { \s__fp \__fp_neg_sign:N } \{ \exp_after:wN \__fp_add_zeros_o:Nww \int_value:w \exp_after:wN \__fp_add_normal_o:Nww \int_value:w \}
```

(End of definition for `\_\_fp\_\_o:ww`.)

\_\_fp\_\_t\_o:ww

This function is either called directly with an empty #1 to compute an addition, or it is called by `\_\_fp\_\_o:ww` with `\_\_fp\_neg_sign:N` as #1 to compute a subtraction, in which case the second operand’s sign should be changed. If the ⟨types⟩ #2 and #4 are the same, dispatch to case #2 (0, 1, 2, or 3), where we call specialized functions: thanks to `\int_value:w`, those receive the tweaked ⟨sign⟩ (expansion of #1#5) as an argument. If the ⟨types⟩ are distinct, the result is simply the floating point number with the highest ⟨type⟩. Since case 3 (used for two `nan`) also picks the first operand, we can also use it when ⟨type1⟩ is greater than ⟨type2⟩. Also note that we don’t need to worry about ⟨sign2⟩ in that case since the second operand is discarded.

```latex
\cs_new:cpn { \_\_fp\_t\_o:ww } \s__fp #1 \_\_fp\_chk:w #2 #3 ; \_\_fp \_\_fp\_chk:w #4 #5 \{ \if_case:w #2 \if_meaning:w #2 #4 \exp_stop_f: \else: \if_int_compare:w #2 > #4 \exp_stop_f: 3 \else: 4 \fi: \fi: \exp_stop_f: \exp_after:wN \_\_fp_add_zeros_o:Nww \int_value:w \or: \exp_after:wN \_\_fp_add_normal_o:Nww \int_value:w \}
```

1113
\begin{verbatim}
\__fp_add_return_ii_o:Nww
\cs_new:Npn \__fp_add_return_ii_o:Nww \#1 \s__fp \__fp_chk:w \#2 \#3 ; \s__fp \__fp_chk:w \#4 \#5

(End of definition for \__fp_add_return_ii_o:Nww.)
\end{verbatim}

\begin{verbatim}
\__fp_add_zeros_o:Nww
\cs_new:Npn \__fp_add_zeros_o:Nww \#1 \s__fp \__fp_chk:w 0 \#2

\if_int_compare:w \#2 \#1 = 20 \exp_stop_f:
  \__fp_case_return_i_o:ww
\else:
  \__fp_case_use:nw
  \exp_last_unbraced:Nf \__fp_invalid_operation_o:Nww
  { \token_if_eq_meaning:NNTF \#1 \#4 + - }
\fi:
\s__fp \__fp_chk:w 2 \#2 \#3;
\s__fp \__fp_chk:w 2 \#4

(End of definition for \__fp_add_zeros_o:Nww.)
\end{verbatim}

\begin{verbatim}
\__fp_add_inf_o:Nww
\cs_new:Npn \__fp_add_inf_o:Nww \#1 \s__fp \__fp_chk:w 2 \#2 \#3; \s__fp \__fp_chk:w 2 \#4

\if_meaning:w \#1 \#2
  \__fp_case_return_i_o:ww
\else:
  \__fp_case_use:nw
  \exp_last_unbraced:Nf \__fp_invalid_operation_o:Nww
  { \token_if_eq_meaning:NNTF \#1 \#4 + - }
\fi:
\s__fp \__fp_chk:w 2 \#2 \#3;
\s__fp \__fp_chk:w 2 \#4

(End of definition for \__fp_add_inf_o:Nww.)
\end{verbatim}
We now have two normal numbers to add, and we have to check signs and exponents more carefully before performing the addition.

\begin{verbatim}
\cs_new:Npn \__fp_add_normal_o:Nww #1 \s__fp \__fp_chk:w 1 #2 
\{ \begin{cases} \exp_after:wN \__fp_add_npos_o:NnwNnw \text{if meaning:w #1#2} \\
\exp_after:wN \__fp_sub_npos_o:NnwNnw \text{else:} \end{cases}
\exp_after:wN \__fp_sanitize:Nw \fi:
\int_value:w \__fp_int_eval:w #2 \#1 \#2 #3 \#1 \#3;
\}
\end{verbatim}

(End of definition for \texttt{\_fp_add_normal_o:Nww}.)

### 74.1.2 Absolute addition

In this subsection, we perform the addition of two positive normal numbers.

\begin{verbatim}
\cs_new:Npn \__fp_add_npos_o:NnwNnw #1; #2; #3; #4;
\{ \begin{cases} \__fp_decimate:nNnnnn {#1} \__fp_add_significand_o:NnnwnnnnN \\
\__fp_int_eval:w #5 - #2 \#1 \#3; \end{cases}
\}
\end{verbatim}

(End of definition for \texttt{\_fp_add_npos_o:NnwNnw}.)

Used in \texttt{l3fp-expo}. Shift the significand of the small number, then add with \texttt{\_fp_add_significand_o:NnnwnnnnN}.
To round properly, we must know at which digit the rounding should occur. This requires to know whether the addition produces an overall carry or not. Thus, we do the computation now and check for a carry, then go back and do the rounding. The rounding may cause a carry in very rare cases such as 0.99⋅⋅⋅95 → 1.00⋅⋅⋅0, but this situation always give an exact power of 10, for which it is easy to correct the result at the end.

If there's no carry, grab all the digits again and round. The packing function \_fp_add_significand_no_carry_o:wwNN takes care of the case where rounding brings a carry.
The case where there is a carry is very similar. Rounding can even raise the first digit from 1 to 2, but we don’t care.

Rounding properly in some modes requires to know what the sign of the result will be. Thus, we start by comparing the exponents and significands. If the numbers coincide, return zero. If the second number is larger, swap the numbers and call \(_{fp_sub_npos-}\) with the opposite of \(<sign>_1</sign>_1\).
\__fp_sub_npos_i_o:Nnwnw

After the computation is done, \__fp_sanitize:Nw checks for overflow/underflow. It expects the ⟨final sign⟩ and the ⟨exponent⟩ (delimited by ;). Start an integer expression for the exponent, which starts with the exponent of the largest number, and may be decreased if the two numbers are very close. If the two numbers have the same exponent, call the near auxiliary. Otherwise, decimate y, then call the far auxiliary to evaluate the difference between the two significands. Note that we decimate by 1 less than one could expect.

\__fp_sub_npos_i_o:Nnwnw

\__fp_sub_back_near_o:nnnnnnnN
\__fp_sub_back_near_pack:NNNNNNw
\__fp_sub_back_near_after:wNNNw

\__fp_sub_back_near_o:nnnnnnnN \{(Y_1)\} \{(Y_2)\} \{(Y_3)\} \{(Y_4)\} \{(X_1)\}
\{(X_2)\} \{(X_3)\} \{(X_4)\} ⟨final sign⟩

In this case, the subtraction is exact, so we discard the ⟨final sign⟩ #9. The very large shifts of $10^9$ and $1.1 \cdot 10^9$ are unnecessary here, but allow the auxiliaries to be reused later. Each integer expression produces a 10 digit result. If the resulting 16 digits start with a 0, then we need to shift the group, padding with trailing zeros.

\__fp_sub_back_near_o:nnnnnnnN \#1\#2\#3\#4 \#5\#6\#7\#8 \#9

\__fp_sub_back_near_o:nnnnnnnN \#1\#2\#3\#4 \#5\#6\#7\#8 \#9

\__fp_sub_back_near_o:nnnnnnnN \#1\#2\#3\#4 \#5\#6\#7\#8 \#9

\__fp_sub_back_near_o:nnnnnnnN \#1\#2\#3\#4 \#5\#6\#7\#8 \#9

\__fp_sub_back_near_o:nnnnnnnN \#1\#2\#3\#4 \#5\#6\#7\#8 \#9

\if_meaning:w 0 \#1
\exp_after:wN \__fp_sub_back_shift:wnnnn
\fi:
; {#1#2#3#4} {#5}
}

(End of definition for \__fp_sub_back_near_o:nnnnnnN, \__fp_sub_back_near_pack:NNNNNNw, and \__fp_sub_back_near_after:whNNNw.)

\__fp_sub_back_shift:wnnnn ; {\langle Z_1 \rangle} {\langle Z_2 \rangle} {\langle Z_3 \rangle} {\langle Z_4 \rangle} ;

This function is called with \langle Z_1 \rangle \leq 999. Act with \number to trim leading zeros from \langle Z_1 \rangle \langle Z_2 \rangle (we don't do all four blocks at once, since non-zero blocks would then overflow \TeX{}'s integers). If the first two blocks are zero, the auxiliary receives an empty \#1 and trims \#2\#30 from leading zeros, yielding a total shift between 7 and 16 to the exponent. Otherwise we get the shift from \#1 alone, yielding a result between 1 and 6. Once the exponent is taken care of, trim leading zeros from \#1\#2\#3\#4 (when \#1 is empty, the space before \#2\#3 is ignored), get four blocks of 4 digits and finally clean up. Trailing zeros are added so that digits can be grabbed safely.

\cs_new:Npn \__fp_sub_back_shift:wnnnn ; #1#2
{\exp_after:wN \__fp_sub_back_shift_ii:ww
\int_value:w #1 #2 0 ;}

\cs_new:Npn \__fp_sub_back_shift_ii:ww #1 0 ; #2#3 ;
{\if_meaning:w @ #1 @
-7
-\exp_after:wN \use_i:nnn
\exp_after:wN \__fp_sub_back_shift_iii:NNNNNNNNw
\int_value:w #2#3 0 ~ 123456789;
\else:
-\__fp_sub_back_shift_iii:NNNNNNNNw #1 123456789;
\fi:
\exp_after:wN \__fp_pack_twice_four:wwNNNNNNNN
\exp_after:wN \__fp_pack_twice_four:wwNNNNNNNN
\exp_after:wN \__fp_sub_back_shift_iv:nnnnw
\exp_after:wN ;
\int_value:w
#1 ~ #2#3 0 ~ 0000 0000 0000 000 ;
}

\cs_new:Npn \__fp_sub_back_shift_iii:NNNNNNNNw #1#2#3#4; #5#6#7#8
{\langle rounding \rangle \{Y'_1\} \{Y'_2\} \{X_1\} \{X_2\} \{X_3\} \{X_4\} \{final sign\}

If the difference is greater than 10^{\langle expo \rangle}, call the very_far auxiliary. If the result is less than 10^{\langle expo \rangle}, call the not_far auxiliary. If it is too close a call to know yet, namely if 1\langle Y'_1\rangle\langle Y'_2\rangle = \langle X_1\rangle\langle X_2\rangle\langle X_3\rangle\langle X_4\rangle0, then call the quite_far auxiliary. We use the odd combination of space and semi-colon delimiters to allow the not_far auxiliary to grab each piece individually, the very_far auxiliary to use \__fp_pack_eight:wwNNNNNNNN, and the quite_far to ignore the significands easily (using the ; delimiter).

\cs_new:Npn \__fp_sub_back_far_o:NNnnnnnn {rounding} \{\langle Y'_1\rangle\} \{\langle Y'_2\rangle\}
{extra-digits} ; \{\langle X_1\rangle\} \{\langle X_2\rangle\} \{\langle X_3\rangle\} \{\langle X_4\rangle\} \{final sign\}

(End of definition for \__fp_sub_back_shift:wnnnn and others.)
The easiest case is when \( x - y \) is extremely close to a power of 10, namely the first digit of \( x \) is 1, and all others vanish when subtracting \( y \). Then the \( \langle \text{rounding} \rangle \) #3 and the \( \langle \text{final sign} \rangle \) #4 control whether we get 1 or 0.

In the usual round-to-nearest mode, we get 1 whenever the \( \langle \text{rounding} \rangle \) digit is less than or equal to 5 (remember that the \( \langle \text{rounding} \rangle \) digit is only equal to 5 if there was no further non-zero digit).

In the present case, \( x \) and \( y \) have different exponents, but \( y \) is large enough that \( x - y \) has a smaller exponent than \( x \). Decrement the exponent (with \(-1\)). Then proceed in a way similar to the near auxiliaries seen earlier, but multiplying \( x \) by 10 (#30 and #40 below), and with the added quirk that the \( \langle \text{rounding} \rangle \) digit has to be taken into account. Namely, we may have to decrease the result by one unit if \( \langle \text{rounding} \rangle \) returns 1. This function expects the \( \langle \text{final sign} \rangle \) #6, the last digit of 1100000000+40=#2, and the \( \langle \text{rounding} \rangle \) digit. Instead of redoing the computation for the second argument, we note that \( \langle \text{final sign} \rangle \) only cares about its parity, which is identical to that of the last digit of #2.
The case where \( x - y \) and \( x \) have the same exponent is a bit more tricky, mostly because it cannot reuse the same auxiliaries. Shift the \( y \) significand by adding a leading 0. Then the logic is similar to the not\_far functions above. Rounding is a bit more complicated: we have two (rounding) digits \#3 and \#6 (from the decimation, and from the new shift) to take into account, and getting the parity of the main result requires a computation. The first \int_value:w triggers the second one because the number is unfinished; we can thus not use 0 in place of 2 there.

\[
\begin{align*}
\text{(End of definition for } \_\_\_\_\_\_\_\text{.)}
\end{align*}
\]

### 74.2 Multiplication

#### 74.2.1 Signs, and special numbers

We go through an auxiliary, which is common with \_\_\_\_\_\_\_/o:w. The first argument is the operation, used for the invalid operation exception. The second is inserted in a formula to dispatch cases slightly differently between multiplication and division. The
third is the operation for normal floating points. The fourth is there for extra cases needed in \_fp/_o:ww.

\cs_new:cpn { \_fp_*_o:ww }
\{ \__fp_mul_cases_o:NnNnww
\* \{ - 2 + \}
\__fp_mul_npos_o:Nww \}

(End of definition for \_fp_*_o:ww.)

\__fp_mul_cases_o:nNnnww

\cs_new:Npn \__fp_mul_cases_o:NnNnww #1#2#3#4 \s__fp \__fp_chk:w #5#6#7; \s__fp \__fp_chk:w #8#9
\{ \if_case:w \__fp_int_eval:w \if_int_compare:w #5 #8 = 11 ~
1
\else:
\if_meaning:w 3 #8 3
\else:
\if_meaning:w 3 #5 2
\else:
\if_int_compare:w #5 #8 = 10 ~
9 #2 - 2
\else:
(\#5 \#2 \#8) / 2 * 2 + 7
\fi:
\fi:
\fi:
\fi:
\fi:
\if_meaning:w \#6 \#9 - 1 \fi:
\__fp_int_eval_end:
\__fp_case_use:nw \{ #3 0 \}
\__fp_case_use:nw \{ #3 2 \}
\__fp_case_return_i_o:ww \c_zero_fp
\__fp_case_return_ii_o:ww \c_minus_zero_fp

\_fp_mul_cases_o:nnww

Split into 10 cases (12 for division). If both numbers are normal, go to case 0 (same sign) or case 1 (opposite signs): in both cases, call \_fp_mul_npos_o:Nww to do the work. If the first operand is nan, go to case 2, in which the second operand is discarded; if the second operand is nan, go to case 3, in which the first operand is discarded (note the weird interaction with the final test on signs). Then we separate the case where the first number is normal and the second is zero: this goes to cases 4 and 5 for multiplication, 10 and 11 for division. Otherwise, we do a computation which dispatches the products \(0 \times 0 = 0 \times 1 = 1 \times 0 = 0\) to case 4 or 5 depending on the combined sign, the products \(0 \times \infty\) and \(\infty \times 0\) to case 6 or 7 (invalid operation), and the products \(1 \times \infty = \infty \times 1 = \infty \times \infty = \infty\) to cases 8 and 9. Note that the code for these two cases (which return \(\pm \infty\)) is inserted as argument #4, because it differs in the case of divisions.

\cs_new:Npm \_fp_mul_cases_o:nnww
\#1\#2\#3\#4 \s__fp \_fp_chk:w \#5\#6\#7; \s__fp \_fp_chk:w \#8\#9
\{ \if_case:w \_fp_int_eval:w \if_int_compare:w \#5 \#8 = 11 ~
1
\else:
\if_meaning:w 3 \#8 3
\else:
\if_meaning:w 3 \#5 2
\else:
\if_int_compare:w \#5 \#8 = 10 ~
9 \#2 - 2
\else:
(\#5 \#2 \#8) / 2 * 2 + 7
\fi:
\fi:
\fi:
\fi:
\if_meaning:w \#6 \#9 - 1 \fi:
\_fp_int_eval_end:
\_fp_case_use:nw \{ #3 0 \}
\_fp_case_use:nw \{ #3 2 \}
\_fp_case_return_i_o:ww \c_zero_fp
\_fp_case_return_ii_o:ww \c_zero_fp
\_fp_case_return_o:Nww \c_minus_zero_fp

74.2.2 Absolute multiplication

In this subsection, we perform the multiplication of two positive normal numbers.

\[ \text{\_fp_mul_npos_o:Nww} \text{(final sign)} \text{s\_fp \_fp_chk:w 1 \{sign1\} \{exp1\}} \text{body1}; \text{s\_fp \_fp_chk:w 1 \{sign2\} \{exp2\}} \text{body2}; \]

After the computation, \text{\_fp_sanitize:Nw} checks for overflow or underflow. As we did for addition, \text{\_fp_int_eval:w} computes the exponent, catching any shift coming from the computation in the significand. The \text{(final sign)} is needed to do the rounding properly in the significand computation. We setup the post-expansion here, triggered by \text{\_fp_mul_significand_o:nnnnNnnnn}.

This is also used in \text{l3fp-convert}.

\[ \text{\_fp_mul_significand_o:nnnnNnnnn} \text{\_fp_mul_significand_drop:NNNNNw} \text{\_fp_mul_significand_keep:NNNNNw} \]

\text{\_fp_mul_significand_o:nnnnNnnnn} \{X_1\} \{(X_2)\} \{(X_3)\} \{(X_4)\} \{sign\} \{(Y_1)\} \{(Y_2)\} \{(Y_3)\} \{(Y_4)\} \]

Note the three semicolons at the end of the definition. One is for the last \text{\_fp_mul_significand_drop:NNNNNw}; one is for \text{\_fp_round_digit:Nw} later on; and one, preceded by \text{\exp_after:wN}, which is correctly expanded (within an \text{\_fp_int_eval:w}), is used by \text{\_fp_basics_pack_low:NNNNNw}.

The product of two 16 digit integers has 31 or 32 digits, but it is impossible to know which one before computing. The place where we round depends on that number of digits, and may depend on all digits until the last in some rare cases. The approach is thus to compute the 5 first blocks of 4 digits (the first one is between 100 and 9999 inclusive), and a compact version of the remaining 3 blocks. Afterwards, the number of digits is known, and we can do the rounding within yet another set of \text{\_fp_int_eval:w}.

\text{(End of definition for \_fp_mul_cases_o:nNnnww.)}
\exp_after:wN \__fp_mul_significand_keep:NNNNNw
\int_value:w \__fp_int_eval:w 99990000 + #1*#7 + #2*#6 +
\exp_after:wN \__fp_mul_significand_keep:NNNNNw
\int_value:w \__fp_int_eval:w 99990000 + #1*#8 + #2*#7 + #3*#6 +
\exp_after:wN \__fp_mul_significand_drop:NNNNNw
\int_value:w \__fp_int_eval:w 99990000 + #1*#9 + #2*#8 +
#3*#7 + #4*#6 +
\exp_after:wN \__fp_mul_significand_drop:NNNNNw
\int_value:w \__fp_int_eval:w 99990000 + #2*#9 + #3*#8 +
#4*#7 +
\exp_after:wN \__fp_mul_significand_drop:NNNNNw
\int_value:w \__fp_int_eval:w 100000000 + #4*#9 ;
\exp_after:wN ;
\exp_after:wN {
\cs_new:Npn \__fp_mul_significand_test_f:NNN #1 #2 #3
\begin{tabular}{l}
\if_meaning:w 0 #3
\exp_after:wN \__fp_mul_significand_small_f:NNwwwN
\else:
\exp_after:wN \__fp_mul_significand_large_f:NwwNNNN
\fi:
\#1 #3
\end{tabular}
\}

(End of definition for \_\_fp_mul_significand_o:nnnnNnnnn, \_\_fp_mul_significand_drop:NNNNNw, and
\_\_fp_mul_significand_keep:NNNNNw.)

\_\_fp_mul_significand_test_f:NNN \langle sign \rangle 1 \langle digits 1-8 \rangle ; \langle digits 9-12 \rangle ; \langle digits 13-16 \rangle ; \langle digits 17-20 \rangle + \langle digits 21-24 \rangle + \langle digits 25-28 \rangle + \langle digits 29-32 \rangle ; \exp_after:wN ;

If the \langle digit 1 \rangle is non-zero, then for rounding we only care about the digits 16 and 17, and whether further digits are zero or not (check for exact ties). On the other hand, if \langle digit 1 \rangle is zero, we care about digits 17 and 18, and whether further digits are zero.

\cs_new:Npn \_\_fp_mul_significand_test_f:NNN #1 #2 #3
\begin{tabular}{l}
\if_meaning:w 0 #3
\exp_after:wN \_\_fp_mul_significand_small_f:NNwwwN
\else:
\exp_after:wN \_\_fp_mul_significand_large_f:NwwNNNN
\fi:
\#1 #3
\end{tabular}

(End of definition for \_\_fp_mul_significand_test_f:NNN.)

\_\_fp_mul_significand_large_f:NwwNNNN

In this branch, \langle digit 1 \rangle is non-zero. The result is thus \langle digits 1-16 \rangle, plus some rounding which depends on the digits 16, 17, and whether all subsequent digits are zero or not. Here, \_\_fp_round_digit:Nw takes digits 17 and further (as an integer expression), and replaces it by a \langle rounding digit \rangle, suitable for \_\_fp_round:NNN.

\cs_new:Npn \_\_fp_mul_significand_large_f:NwwNNNN #1 #2 #3 #4#5#6#7 ;
\begin{tabular}{l}
\exp_after:wN \_\_fp_basics_pack_high:NNNNNw
\int_value:w \_\_fp_int_eval:w 1 #2
\exp_after:wN \_\_fp_basics_pack_low:NNNNNw
\int_value:w \_\_fp_int_eval:w 1 #3#4#5#6#7
+ \exp_after:wN \_\_fp_round:NNN
\exp_after:wN \#1
\end{tabular}

1124
In this branch, \textit{digit 1} is zero. Our result is thus \textit{digits 2–17}, plus some rounding which depends on the digits 17, 18, and whether all subsequent digits are zero or not. The 8 digits 1#3 are followed, after expansion of the \texttt{small\_pack} auxiliary, by the next digit, to form a 9 digit number.

\begin{verbatim}
\cs_new:Npn \__fp_mul_significand_small_f:NNwwwN #1 #2#3; #4#5; #6; + #7 {
  - 1
  \exp_after:wN \__fp_basics_pack_high:NNNNNw
  \int_value:w \__fp_int_eval:w 1#3#4
  \exp_after:wN \__fp_basics_pack_low:NNNNNw
  \int_value:w \__fp_int_eval:w 1#5#6#7
  + \exp_after:wN \__fp_round:NNN
    \exp_after:wN #1 \exp_after:wN #7
  \exp_after:wN \__fp_round_digit:Nw
}
\end{verbatim}

\textit{(End of definition for \_\_fp_mul_significand\_large\_f:NwwNNN.)}

## 74.3 Division

### 74.3.1 Signs, and special numbers

Time is now ripe to tackle the hardest of the four elementary operations: division.

Filtering special floating point is very similar to what we did for multiplications, with a few variations. Invalid operation exceptions display \texttt{/} rather than \texttt{*}. In the formula for dispatch, we replace \texttt{- 2 +} by \texttt{-}. The case of normal numbers is treated using \_\_fp\_div\_npos\_o:Nww rather than \_\_fp\_mul\_npos\_o:Nww. There are two additional cases: if the first operand is normal and the second is a zero, then the division by zero exception is raised; cases 10 and 11 of the \texttt{\if_case:w} construction in \_\_fp\_mul\_cases\_o:Nnnww are provided as the fourth argument here.

\begin{verbatim}
\cs_new:cpn { \_\_fp\_/\_o:ww }
  { \_\_fp_mul\_cases\_o:Nnnww / \_\_fp\_div\_npos\_o:Nww
    \_\_fp\_division\_by\_zero\_o:NNww \c_inf_fp / }
  \_\_fp\_division\_by\_zero\_o:NNww \c_minus\_inf_fp / }
\end{verbatim}

1125
We want to compute $A/Z$. As for multiplication, \__fp\_sanitize:Nw checks for overflow or underflow; we provide it with the \langle final sign \rangle, and an integer expression in which we compute the exponent. We set up the arguments of \__fp\_div\_significand_i_o:wnnw, namely an integer $\langle y \rangle$ obtained by adding 1 to the first 5 digits of $Z$ (explanation given soon below), then the four $\{A_i\}$, then the four $\{Z_i\}$, a semi-colon, and the \langle final sign \rangle, used for rounding at the end.

26391 \cs_new:Npn \__fp\_div\_npos_o:Nww
26392 #1 \s__fp \__fp\_chk:w 1 #2 #3 #4 ; \s__fp \__fp\_chk:w 1 #5 #6 #7 #8 #9;
26393 \{ \exp_after:wN \__fp\_sanitize:Nw
26394 \exp_after:wN \#1 \s__fp\_chk:w 1 \#2 \#3 \#4 ; \s__fp\_chk:w 1 \#5 \#6 \#7 \#8 \#9;
26395 \int_value:w \__fp\_int\_eval:w
26396 \#3 - \#6
26397 \exp_after:wN \__fp\_div\_significand_i_o:wnnw
26398 \int_value:w \__fp\_int\_eval:w \#7 \use_i:nnnn \#8 + 1 ;
26399 \#4
26400 \{\#7\}{\#8}\#9 ;
26401 \#1
26402 \}

(End of definition for \__fp\_div\_npos_o:Nww.)

### 74.3.2 Work plan

In this subsection, we explain how to avoid overflowing \TeX's integers when performing the division of two positive normal numbers.

We are given two numbers, $A = 0.A_1A_2A_3A_4$ and $Z = 0.Z_1Z_2Z_3Z_4$, in blocks of 4 digits, and we know that the first digits of $A_1$ and of $Z_1$ are non-zero. To compute $A/Z$, we proceed as follows.

- Find an integer $Q_A \simeq 10^4 A/Z$.
- Replace $A$ by $B = 10^4 A - Q_A Z$.
- Find an integer $Q_B \simeq 10^4 B/Z$.
- Replace $B$ by $C = 10^4 B - Q_B Z$.
- Find an integer $Q_C \simeq 10^4 C/Z$.
- Replace $C$ by $D = 10^4 C - Q_C Z$.
- Find an integer $Q_D \simeq 10^4 D/Z$.
- Consider $E = 10^4 D - Q_D Z$, and ensure correct rounding.
The result is then \( Q = 10^{-4} Q_A + 10^{-8} Q_B + 10^{-12} Q_C + 10^{-16} Q_D \) + rounding. Since the \( Q_i \) are integers, \( B, C, D, \) and \( E \) are all exact multiples of \( 10^{-16} \), in other words, computing with 16 digits after the decimal separator yields exact results. The problem is the risk of overflow: in general \( B, C, D, \) and \( E \) may be greater than 1.

Unfortunately, things are not as easy as they seem. In particular, we want all intermediate steps to be positive, since negative results would require extra calculations at the end. This requires that \( Q_A \leq 10^4 A/Z \) etc. A reasonable attempt would be to define \( Q_A \) as

\[
\texttt{\textbackslash int\_eval:n } \left\{ \frac{A_1 A_2}{Z_1 + 1} - 1 \right\} \leq 10^4 \frac{A}{Z}
\]

Subtracting 1 at the end takes care of the fact that \( \varepsilon\text{-\TeX}\)’s \texttt{\_fp\_int\_eval:w} rounds divisions instead of truncating (really, \( 1/2 \) would be sufficient, but we work with integers). We add 1 to \( Z_1 \) because \( Z_1 \leq 10^4 Z < Z_1 + 1 \) and we need \( Q_A \) to be an underestimate. However, we are now underestimating \( Q_A \) too much: it can be wrong by up to 100, for instance when \( Z = 0.1 \) and \( A \approx 1 \). Then \( B \) could take values up to \( 10 \) (maybe more), and a few steps down the line, we would run into arithmetic overflow, since \( \text{\TeX} \) can only handle integers less than roughly \( 2 \cdot 10^9 \).

A better formula is to take

\[
Q_A = \texttt{\textbackslash int\_eval:n } \left\{ \frac{10 \cdot A_1 A_2}{10^{-3} \cdot Z_1 Z_2} + 1 \right\}.
\]

This is always less than \( 10^9 A/(10^5 Z) \), as we wanted. In words, we take the 5 first digits of \( Z \) into account, and the 8 first digits of \( A \), using 0 as a 9-th digit rather than the true digit for efficiency reasons. We shall prove that using this formula to define all the \( Q_i \) avoids any overflow. For convenience, let us denote

\[
y = \left\lfloor 10^{-3} \cdot Z_1 Z_2 \right\rfloor + 1,
\]

so that, taking into account the fact that \( \varepsilon\text{-\TeX} \) rounds ties away from zero,

\[
Q_A = \left\lfloor \frac{A_1 A_2}{y} - \frac{1}{2} \right\rfloor,
\]

\[
> A_1 A_2 \cdot \frac{1}{y} - \frac{3}{2}.
\]

Note that \( 10^4 < y \leq 10^5 \), and \( 999 < Q_A \leq 999899 \). Also note that this formula does not cause an overflow as long as \( A < (2^{31} - 1)/10^9 \approx 2.147 \cdot \cdot \cdot \), since the numerator involves an integer slightly smaller than \( 10^9 A \).

Let us bound \( B \):

\[
10^9 B = A_1 A_2 0 + 10 \cdot 0. A_3 A_4 - 10 \cdot Z_1 Z_2 Z_3 Z_4 \cdot Q_A
\]

\[
< A_1 A_2 0 \cdot \left( 1 - 10 \cdot \frac{Z_1 Z_2 Z_3 Z_4}{y} \right) + 3 \cdot 10 \cdot Z_1 Z_2 Z_3 Z_4 + 10
\]

\[
\leq A_1 A_2 0 \cdot \left( y - 10 \cdot Z_1 Z_2 Z_3 Z_4 \right) + 3 \cdot 2 y + 10
\]

\[
\leq A_1 A_2 0 \cdot \frac{1}{y} + \frac{3}{2} y + 10 \leq \frac{10^9 A}{y} + 1.6 \cdot y.
\]

1127
At the last step, we hide $10$ into the second term for later convenience. The same reasoning yields

\[\begin{align*}
10^5 B &< 10^9 A/y + 1.6y, \\
10^5 C &< 10^9 B/y + 1.6y, \\
10^5 D &< 10^9 C/y + 1.6y, \\
10^5 E &< 10^9 D/y + 1.6y.
\end{align*}\]

The goal is now to prove that none of $B$, $C$, $D$, and $E$ can go beyond $(2^{31} - 1)/10^9 = 2.147\ldots$.

Combining the various inequalities together with $A < 1$, we get

\[\begin{align*}
10^5 B &< 10^9/y + 1.6y, \\
10^5 C &< 10^{13}/y^2 + 1.6(y + 10^4), \\
10^5 D &< 10^{17}/y^3 + 1.6(y + 10^4 + 10^8/y), \\
10^5 E &< 10^{21}/y^4 + 1.6(y + 10^4 + 10^8/y + 10^{12}/y^2).
\end{align*}\]

All of those bounds are convex functions of $y$ (since every power of $y$ involved is convex, and the coefficients are positive), and thus maximal at one of the end-points of the allowed range $10^4 < y \leq 10^5$. Thus,

\[\begin{align*}
10^5 B &< \max(1.16 \cdot 10^5, 1.7 \cdot 10^5), \\
10^5 C &< \max(1.32 \cdot 10^5, 1.77 \cdot 10^5), \\
10^5 D &< \max(1.48 \cdot 10^5, 1.777 \cdot 10^5), \\
10^5 E &< \max(1.64 \cdot 10^5, 1.7777 \cdot 10^5).
\end{align*}\]

All of those bounds are less than $2.147 \cdot 10^5$, and we are thus within \TeX’s bounds in all cases!

We later need to have a bound on the $Q_i$. Their definitions imply that $Q_A < 10^9 A/y - 1/2 < 10^5 A$ and similarly for the other $Q_i$. Thus, all of them are less than $177770$.

The last step is to ensure correct rounding. We have

\[A/Z = 10\sum_{i=1}^{4} (10^{-4i}Q_i) + 10^{-16}E/Z\]

exactly. Furthermore, we know that the result is in $[0.1, 10)$, hence will be rounded to a multiple of $10^{-16}$ or of $10^{-15}$, so we only need to know the integer part of $E/Z$, and a “rounding” digit encoding the rest. Equivalently, we need to find the integer part of $2E/Z$, and determine whether it was an exact integer or not (this serves to detect ties). Since

\[\frac{2E}{Z} = 2 \frac{10^5 E}{10^9 Z} \leq 2 \frac{10^5 E}{10^9} < 36,\]

1128
this integer part is between 0 and 35 inclusive. We let $\varepsilon$-\TeX{} round
\[ P = \texttt{\textbackslash int\_eval:n} \left\{ \frac{2 \cdot E_1 E_2}{Z_1 Z_2} \right\}, \]
which differs from $2E/Z$ by at most
\[ \frac{1}{2} + 2 \frac{E}{Z} - \frac{E}{10^{-8}Z_1 Z_2} + 2 \frac{10^8 E - E_1 E_2}{Z_1 Z_2} < 1, \]
(1/2 comes from $\varepsilon$-\TeX{}’s rounding) because each absolute value is less than $10^{-7}$. Thus $P$ is either the correct integer part, or is off by $P/2$. If it is zero, then $E/Z \in \{P/2, (P-1)/2, P/2\}$. If it is positive, then $E/Z \in (P/2, (P-1)/2)$. In each case, we know how to round to an integer, depending on the parity of $P$, and the rounding mode.

74.3.3 Implementing the significand division

\[ \texttt{\_fp\_div\_significand\_calc:wwnnnnnn} \]
\begin{verbatim}
\_fp_div_significand_i_o:wnnw \{ \langle A_1 \rangle \langle A_2 \rangle \langle A_3 \rangle \langle A_4 \rangle \}
\{ \langle Z_1 \rangle \langle Z_2 \rangle \langle Z_3 \rangle \langle Z_4 \rangle \}; \langle \text{sign} \rangle

Compute $10^6 + Q_A$ (a 7 digit number thanks to the shift), unbrace $\langle A_1 \rangle$ and $\langle A_2 \rangle$, and prepare the $\langle \text{continuation} \rangle$ arguments for 4 consecutive calls to \_fp\_div\_significand\_calc:wwnnnnnn. Each of these calls needs $\langle y \rangle$ (#1), and it turns out that we need post-expansion there, hence the \texttt{\textbackslash int\_value:w}. Here, #4 is six brace groups, which give the first n-type arguments of the \texttt{calc} function.

(End of definition for \_fp\_div\_significand\_i_o:wnnw.)

\[ \_fp_div_significand_calc:wwnnnnnn \langle 10^6 + Q_A \rangle \langle A_1 \rangle \langle A_2 \rangle ; \langle A_3 \rangle \langle Z_1 \rangle \langle Z_2 \rangle \langle Z_3 \rangle \langle Z_4 \rangle \langle \text{continuation} \rangle \]

expands to
\[ \langle 10^6 + Q_A \rangle \langle \text{continuation} \rangle ; \langle B_1 \rangle \langle B_2 \rangle ; \langle B_3 \rangle \langle B_4 \rangle \langle Z_1 \rangle \langle Z_2 \rangle \langle Z_3 \rangle \langle Z_4 \rangle \]

where $B = 10^4 A - Q_A \cdot Z$. This function is also used to compute $C$, $D$, $E$ (with the input shifted accordingly), and is used in l3fp-expo.

We know that $0 < Q_A < 1.8 \cdot 10^5$, so the product of $Q_A$ with each $Z_i$ is within \TeX{}'s bounds. However, it is a little bit too large for our purposes: we would not be able to

1129
use the usual trick of adding a large power of 10 to ensure that the number of digits is
fixed.

The bound on $Q_A$, implies that $10^6 + Q_A$ starts with the digit 1, followed by 0 or
1. We test, and call different auxiliaries for the two cases. An earlier implementation
did the tests within the computation, but since we added a \langle continuation\rangle, this is not
possible because the macro has 9 parameters.

The result we want is then (the overall power of 10 is arbitrary):

$$
10^{-4}(\#2 - \#1 \cdot \#5 - 10 \cdot \langle i \rangle \cdot \#5\#6) + 10^{-8}(\#3 - \#1 \cdot \#6 - 10 \cdot \langle i \rangle \cdot \#7)
$$

$$
+ 10^{-12}(\#4 - \#1 \cdot \#7 - 10 \cdot \langle i \rangle \cdot \#8) + 10^{-16}(-\#1 \cdot \#8),
$$

where $\langle i \rangle$ stands for the $10^5$ digit of $Q_A$, which is 0 or 1, and $\#1$, $\#2$, etc. are the
parameters of either auxiliary. The factors of 10 come from the fact that $Q_A = 10 \cdot
10^4 \cdot \langle i \rangle + \#1$. As usual, to combine all the terms, we need to choose some shifts which
must ensure that the number of digits of the second, third, and fourth terms are each
fixed. Here, the positive contributions are at most $10^8$ and the negative contributions
can go up to $10^9$. Indeed, for the auxiliary with $\langle i \rangle = 1$, $\#1$ is at most 80000, leading
to contributions of at worse $-8 \cdot 10^4$, while the other negative term is very small < $10^6$
(except in the first expression, where we don’t care about the number of digits); for the
auxiliary with $\langle i \rangle = 0$, $\#1$ can go up to 99999, but there is no other negative term.
Hence, a good choice is $2 \cdot 10^9$, which produces totals in the range $[10^9, 2 \cdot 10^9]$. We are
flirting with \TeX’s limits once more.

\begin{verbatim}
\cs_new:Npn \__fp_div_significand_calc:wwnnnnnnn 1#1
\{  \exp_after:wN \__fp_div_significand_calc_i:wwnnnnnnn \else: \exp_after:wN \__fp_div_significand_calc_ii:wwnnnnnnn \fi: \}
\cs_new:Npn \__fp_div_significand_calc_i:wwnnnnnnn #1; #2;#3#4 #5#6#7#8 #9
\{ 1 1 #1 \exp_after:wN \int_value:w \__fp_int_eval:w \c__fp_Bigg_leading_shift_int
+ \#2 - \#1 * \#5 - \#5\#60 \exp_after:wN \__fp_pack_Bigg:NNNNNNw \int_value:w \__fp_int_eval:w \c__fp_Bigg_middle_shift_int
+ \#3 - \#1 * \#6 - \#70 \exp_after:wN \__fp_pack_Bigg:NNNNNNw \int_value:w \__fp_int_eval:w \c__fp_Bigg_middle_shift_int
+ \#4 - \#1 * \#7 - \#80 \exp_after:wN \__fp_pack_Bigg:NNNNNNw \int_value:w \__fp_int_eval:w \c__fp_Bigg_trailing_shift_int
- \#1 * \#8 ; \}
\cs_new:Npn \__fp_div_significand_calc_ii:wwnnnnnnn #1; #2;#3#4 #5#6#7#8 #9
\{ 1 0 #1 \}
\end{verbatim}
\cs_new:Npn \__fp_div_significand_ii:wwn #1; #2;#3
\exp_after:wN \__fp_div_significand_pack:NNN
\int_value:w \__fp_int_eval:w \exp_after:wN \__fp_div_significand_calc:wwnnnnnnn
\int_value:w \__fp_int_eval:w 999999 + #2 #3 0 / #1 ; #2 #3 ;
\__fp_div_significand_ii:wwn \langle y \rangle; \langle E_1 \rangle; \{ \langle E_2 \rangle \} \{ \langle E_3 \rangle \} \{ \langle E_4 \rangle \}
\{ \langle Z_1 \rangle \} \{ \langle Z_2 \rangle \} \{ \langle Z_3 \rangle \} \{ \langle Z_4 \rangle \} \{ \langle \text{sign} \rangle \}
Compute \( Q_B \) by evaluating \( (B_1)(B_2)/y - 1 \). The result is output to the left, in an \__fp_int_eval:w \ which we start now. Once that is evaluated (and the other \( Q_i \) also, since later expansions are triggered by this one), a packing auxiliary takes care of placing the digits of \( Q_B \) in an appropriate way for the final addition to obtain \( Q \). This auxiliary is also used to compute \( Q_C \) and \( Q_D \) with the inputs \( C \) and \( D \) instead of \( B \).
\cs_new:Npn \__fp_div_significand_iii:wwnnnnn #1; #2;#3#4#5 #6#7
\exp_after:wN \__fp_div_significand_iv:wwnnnnnnn
\int_value:w \__fp_int_eval:w ( 2 * #2 #3) / #6 #7 ; \% \leftarrow P \( #2 ; \langle #3 \rangle \langle #4 \rangle \langle #5 \rangle \langle #6 \rangle \langle #7 \rangle \);
\__fp_div_significand_iii:wwnnnn \langle P \rangle; \langle E_1 \rangle; \{ \langle E_2 \rangle \} \{ \langle E_3 \rangle \} \{ \langle E_4 \rangle \}
\{ \langle Z_1 \rangle \} \{ \langle Z_2 \rangle \} \{ \langle Z_3 \rangle \} \{ \langle Z_4 \rangle \} \{ \langle \text{sign} \rangle \}
We compute \( P \approx 2E/Z \) by rounding \( 2E_1E_2/Z_1Z_2 \). Note the first 0, which multiplies \( Q_D \) by 10: we later add (roughly) \( 5 \cdot P \), which amounts to adding \( P/2 \approx E/Z \) to \( Q_D \), the appropriate correction from a hypothetical \( Q_E \).
\cs_new:Npn \__fp_div_significand_iv:wwnnnnnnn \langle P \rangle; \langle E_1 \rangle; \{ \langle E_2 \rangle \} \{ \langle E_3 \rangle \} \{ \langle E_4 \rangle \}
\{ \langle Z_1 \rangle \} \{ \langle Z_2 \rangle \} \{ \langle Z_3 \rangle \} \{ \langle Z_4 \rangle \} \{ \langle \text{sign} \rangle \}
(End of definition for \__fp_div_significand_i:wwnmmmn, and \__fp_div_significand_calc_i:wwnnnnnn.)
This adds to the current expression \((10^7 + 10 \cdot Q_D)\) a contribution of \(5 \cdot P + \text{sign}(T)\) with \(T = 2E - PZ\). This amounts to adding \(P/2\) to \(Q_D\), with an extra (rounding) digit. This (rounding) digit is 0 or 5 if \(T\) does not contribute, i.e., if \(0 = T = 2E - PZ\), in other words if \(10^{16}A/Z\) is an integer or half-integer. Otherwise it is in the appropriate range, \([1, 4]\) or \([6, 9]\). This is precise enough for rounding purposes (in any mode).

It seems an overkill to compute \(T\) exactly as I do here, but I see no faster way right now.

Once more, we need to be careful and show that the calculation \#1 \#6 \#7 below does not cause an overflow: naively, \(P\) can be up to 35, and \#6 \#7 up to \(10^8\), but both cannot happen simultaneously. To show that things are fine, we split in two (non-disjoint) cases.

- For \(P < 10\), the product obeys \(P \cdot \#6 \#7 < 10^8 \cdot P < 10^9\).
- For large \(P \geq 3\), the rounding error on \(P\), which is at most 1, is less than a factor of 2, hence \(P \leq 4E/Z\). Also, \#6 \#7 \(\leq 10^8 \cdot Z\), hence \(P \cdot \#6 \#7 \leq 4E \cdot 10^8 < 10^9\).

Both inequalities could be made tighter if needed.

Note however that \(P \#8 \#9\) may overflow, since the two factors are now independent, and the result may reach \(3.5 \cdot 10^9\). Thus we compute the two lower levels separately. The rest is standard, except that we use + as a separator (ending integer expressions explicitly). \(T\) is negative if the first character is -; otherwise, \(T\) is positive if the first argument of \_]_fp_div_significand_vi:Nw, a sum of several terms, is also zero. Otherwise, there was an exact agreement: \(T = 0\).

\(\text{End of definition for } \_]_fp_div_significand_iv:wwnnn; \; \text{#1}; \; \#2; \; \#3; \; \#4; \; \#5; \; \#6; \; \#7; \; \#8; \; \#9\)

At this stage, we are in the following situation: \TeX is in the process of expanding several integer expressions, thus functions at the bottom expand before those above.
\texttt{\_fp\_div\_significand\_test\_o:w 10^6 + Q_A \_fp\_div\_significand\_pack:NNN 10^6 + Q_B \_fp\_div\_significand\_pack:NNN 10^6 + Q_C \_fp\_div\_significand\_pack:NNN 10^7 + 10 \cdot Q_D + 5 \cdot P + \varepsilon; \langle \text{sign} \rangle}

Here, $\varepsilon = \text{sign}(T)$ is 0 in case $2E = PZ$, 1 in case $2E > PZ$, which means that $P$ was the correct value, but not with an exact quotient, and $-1$ if $2E < PZ$, $i.e.$, $P$ was an overestimate. The packing function we define now does nothing special: it removes the $10^6$ and carries two digits (for the $10^5$'s and the $10^4$'s).

\texttt{\cs_new:Npn \_fp\_div\_significand\_pack:NNN 1 #1 #2 \{ + #1 #2; \}}

\texttt{(End of definition for \_fp\_div\_significand\_pack:NNN.)}

\texttt{\_fp\_div\_significand\_test\_o:w \_fp\_div\_significand\_test\_o:w 1 0 \langle 5d \rangle; \langle 4d \rangle; \langle 5d \rangle; \langle \text{sign} \rangle}

The reason we know that the first two digits are 1 and 0 is that the final result is known to be between 0.1 (inclusive) and 10, hence $Q_A$ (the tilde denoting the contribution from the other $Q_i$) is at most 99999, and $10^6 + Q_A = 10^{\ldots}$.

It is now time to round. This depends on how many digits the final result will have.

\texttt{\cs_new:Npn \_fp\_div\_significand\_small\_o:wwwNNNNwN \_fp\_div\_significand\_small\_o:wwwNNNNwN 0 \langle 4d \rangle; \langle 4d \rangle; \langle 5d \rangle; \langle \text{final \ sign} \rangle}

Standard use of the functions \_fp\_basics\_pack\_low:NNNNw and \_fp\_basics\_pack\_high:NNNNw. We finally get to use the \texttt{(final sign)} which has been sitting there for a while.

\texttt{\cs_new:Npn \_fp\_div\_significand\_large\_o:wwwNNNNwN \_fp\_div\_significand\_large\_o:wwwNNNNwN \langle 5d \rangle; \langle 4d \rangle; \langle 4d \rangle; \langle 5d \rangle; \langle \text{sign} \rangle}

We know that the final result cannot reach 10, hence 1#1#2, together with contributions from the level below, cannot reach $2 \cdot 10^9$. For rounding, we build the \texttt{(rounding digit)} from the last two of our 18 digits.
74.4 Square root

\_\_fp_sqrt_o:w

Zeros are unchanged: \( \sqrt{-0} = -0 \) and \( \sqrt{+0} = +0 \). Negative numbers (other than \(-0\)) have no real square root. Positive infinity, and \texttt{nan}, are unchanged. Finally, for normal positive numbers, there is some work to do.

\_\_fp_sqrt_npos_o:w

Prepare \_\_fp_sanitize:Nw to receive the final sign 0 (the result is always positive) and the exponent, equal to half of the exponent \#1 of the argument. If the exponent \#1 is even, find a first approximation of the square root of the significand \(10^8a_1 + a_2 = 10^8#2#3 + #4#5\) through Newton's method, starting at \(x = 57234133 \approx 10^{7.75}\). Otherwise, first shift the significand of the argument by one digit, getting \(a'_{1} \in [10^{6}, 10^{7})\) instead of \([10^{7}, 10^{8})\), then use Newton's method starting at \(17782794 \approx 10^{7.25}\).
Newton’s method maps $x \mapsto [(x + [10^8 a_1 / x]) / 2]$ in each iteration, where $[b/c]$ denotes $\varepsilon$-TeX’s division. This division rounds the real number $b/c$ to the closest integer, rounding ties away from zero, hence when $c$ is even, $b/c - 1/2 + 1/c \leq [b/c] \leq b/c + 1/2$ and when $c$ is odd, $b/c - 1/2 + 1/(2c) \leq [b/c] \leq b/c + 1/2 - 1/(2c)$. For all $c$, $b/c - 1/2 + 1/(2c) \leq [b/c] \leq b/c + 1/2$.

Let us prove that the method converges when implemented with $\varepsilon$-TeX integer division, for any $10^6 \leq a_1 < 10^9$ and starting value $10^6 \leq x < 10^9$. Using the inequalities above and the arithmetic-geometric inequality $(x + t) / 2 \geq \sqrt{xt}$ for $t = 10^8 a_1 / x$, we find

$$x' = \frac{x + [10^8 a_1 / x]}{2} \geq \frac{x + 10^8 a_1 / x - 1/2 + 1/(2x)}{2} \geq \sqrt{10^8 a_1} - \frac{1}{4} + \frac{1}{4x}.$$ 

After any step of iteration, we thus have $\delta = x - \sqrt{10^8 a_1} \geq -0.25 + 0.25 \cdot 10^{-8}$. The new difference $\delta' = x' - \sqrt{10^8 a_1}$ after one step is bounded above as

$$x' - \sqrt{10^8 a_1} \leq \frac{x + 10^8 a_1 / x + 1/2}{2} - \sqrt{10^8 a_1} \leq \frac{\delta}{2} + \frac{\delta}{\sqrt{10^8 a_1} + \delta} + \frac{3}{4}.$$ 

For $\delta > 3/2$, this last expression is $\leq \delta / 2 + 3/4 < \delta$, hence $\delta$ decreases at each step: since all $x$ are integers, $\delta$ must reach a value $-1/4 < \delta \leq 3/2$. In this range of values, we get $\delta' \leq \frac{3}{2 \sqrt{10^8 a_1}} + \frac{3}{4} \leq 0.75 + 1.125 \cdot 10^{-7}$. We deduce that the difference $\delta = x - \sqrt{10^8 a_1}$ eventually reaches a value in the interval $[-0.25 + 0.25 \cdot 10^{-8}, 0.75 + 11.25 \cdot 10^{-8}]$, whose width is $1 + 11 \cdot 10^{-8}$. The corresponding interval for $x$ may contain two integers, hence $x$ might oscillate between those two values.

However, the fact that $x \mapsto x - 1$ and $x - 1 \mapsto x$ puts stronger constraints, which are not compatible: the first implies

$$x + [10^8 a_1 / x] \leq 2x - 2$$

hence $10^8 a_1 / x \leq x - 3/2$, while the second implies

$$x - 1 + [10^8 a_1 / (x - 1)] \geq 2x - 1$$

hence $10^8 a_1 / (x - 1) \geq x - 1/2$. Combining the two inequalities yields $x^2 - 3x/2 \geq 10^8 a_1 \geq x - 3x/2 + 1/2$, which cannot hold. Therefore, the iteration always converges to a single integer $x$. To stop the iteration when two consecutive results are equal, the function $\_\_fp_sqrt_Newton_o:wwn$ receives the newly computed result as $#1$, the previous result as $#2$, and $a_1$ as $#3$. Note that $\varepsilon$-TeX combines the computation of a multiplication and a following division, thus avoiding overflow in $#3 * 100000000 / #1$. In any case, the result is within $[10^7, 10^8]$. 

1135
This function is followed by \( 10^8 + x - 1 \), which has 9 digits starting with 1, then \\
\{a_1\} \{a_2\} \{a_3\}. Here, \( x \approx \sqrt{10^{-8}a_1} \) and we want to estimate the square root of \( \alpha = 10^{-8}a_1 + 10^{-16}a_2 + 10^{-17}a_3 \). We set up an initial underestimate
\[
y = (x - 1)10^{-8} + 0.2499998875 \cdot 10^{-8} \approx \sqrt{\alpha}.
\]
From the inequalities shown earlier, we know that \( y \leq \sqrt{10^{-8}a_1} \leq \sqrt{\alpha} \) and that \( \sqrt{10^{-8}a_1} \leq y + 10^{-8} + 11 \cdot 10^{-16} \) hence (using \( 0.1 \leq y \leq \sqrt{\alpha} \leq 1 \))
\[
a - y^2 \leq 10^{-8}a_1 + 10^{-8} - y^2 \leq (y + 10^{-8} + 11 \cdot 10^{-16})^2 - y^2 + 10^{-8} < 3.2 \cdot 10^{-8},
\]
and \( \sqrt{\alpha} - y = (a - y^2)/(\sqrt{\alpha} + y) \leq 16 \cdot 10^{-8} \). Next, \( \llap{\_}\llap{\_}\llap{\_}\llap{\_}\llap{\_}\llap{\_}\llap{\_}\llap{\_}\llap{\_}\llap{\_}\llap{\_}\llap{\_}\text{fp_sqrt_auxii_o:NNNNnnnn} \) is called several times to get closer and closer underestimates of \( \sqrt{\alpha} \). By construction, the underestimates \( y \) are always increasing, \( a - y^2 < 3.2 \cdot 10^{-8} \) for all. Also, \( y < 1 \).

This receives a continuation function \#1, then five blocks of 4 digits for \( y \), then two 8-digit blocks and a single digit for \( \alpha \). A common estimate of \( \sqrt{\alpha} - y = (a - y^2)/(\sqrt{\alpha} + y) \) is \( (a - y^2)/(2y) \), which leads to alternating overestimates and underestimates. We tweak this, to only work with underestimates (no need then to worry about signs in the computation). Each step finds the largest integer \( j \leq 6 \) such that \( 10^{4j}(a - y^2) < 2 \cdot 10^8 \), then computes the integer (with ɛ-T\( \TeX \)’s rounding division)
\[
10^{4j}z = \left\lfloor \left( \left\lfloor 10^{4j}(a - y^2) \right\rfloor - 257 \right\rfloor \cdot (0.5 \cdot 10^8) / \left\lfloor 10^8y + 1 \right\rfloor \right\rfloor.
\]
The choice of \( j \) ensures that \( 10^{4j}z < 2 \cdot 10^8 \cdot 0.5 \cdot 10^8 / 10^7 = 10^9 \), thus \( 10^9 + 10^{4j}z \) has exactly 10 digits, does not overflow T\( \TeX \)’s integer range, and starts with 1. Incidentally, since all \( a - y^2 \leq 3.2 \cdot 10^{-8} \), we know that \( j \geq 3 \).

Let us show that \( z \) is an underestimate of \( \sqrt{\alpha} - y \). On the one hand, \( \sqrt{\alpha} - y \leq 16 \cdot 10^{-8} \) because this holds for the initial \( y \) and values of \( y \) can only increase. On the other hand, the choice of \( j \) implies that \( \sqrt{\alpha} - y \leq 5(\sqrt{\alpha} + y)(\sqrt{\alpha} - y) = 5(a - y^2) < 10^{9 - 4j} \). For \( j = 3 \),
the first bound is better, while for larger $j$, the second bound is better. For all $j \in \{3, 6\}$, we find $\sqrt{a} - y < 16 \cdot 10^{-2j}$. From this, we deduce that

$$10^{4j}(\sqrt{a} - y) = \frac{10^{4j}(a - y^2 - (\sqrt{a} - y)^2)}{2y} \geq \frac{[10^{4j}(a - y^2)] - 257}{2 \cdot 10^{-8} [10^8 y + 1]} + \frac{1}{2}$$

where we have replaced the bound $10^{4j}(16 \cdot 10^{-2j}) = 256$ by $257$ and extracted the corresponding term $1/(2 \cdot 10^{-8} [10^8 y + 1]) \geq 1/2$. Given that $c$-FPX's integer division obeys $b/c \leq b/c + 1/2$, we deduce that $10^{4j} z \leq 10^{4j}(\sqrt{a} - y)$, hence $y + z \leq \sqrt{a}$ is an underestimate of $\sqrt{a}$, as claimed. One implementation detail: because the computation involves $-\#4\#4 - 2\#3\#5 - 2\#2\#6$ which may be as low as $-5 \cdot 10^8$, we need to use the pack_big functions, and the big shifts.

```latex
\cs_new:Npn \__fp_sqrt_auxii_o:Nnnnnnnn #1 #2 #3 #4 #5 #6 #7 #8 #9
\exp_after:wN \__fp_int_eval:w \c__fp_big_leading_shift_int + #7 - #2 \times #2
\exp_after:wN \__fp_pack_big:NNNNNNw
\int_value:w \c__fp_int_eval:w \c__fp_big_middle_shift_int - 2 \times #2 \times #3
\exp_after:wN \__fp_pack_big:NNNNNNw
\int_value:w \c__fp_int_eval:w \c__fp_big_middle_shift_int + #8 \times #3 - 2 \times #2 \times #4
\exp_after:wN \__fp_pack_big:NNNNNNw
\int_value:w \c__fp_int_eval:w \c__fp_big_middle_shift_int - 2 \times #3 \times #4 - 2 \times #2 \times #5
\exp_after:wN \__fp_pack_big:NNNNNNw
\int_value:w \c__fp_int_eval:w \c__fp_big_middle_shift_int + #9 \times 000 \times 000 - #4 \times #4 - 2 \times #3 \times #5 - 2 \times #2 \times #6
\exp_after:wN \__fp_pack_big:NNNNNNw
\int_value:w \c__fp_int_eval:w \c__fp_big_middle_shift_int - 2 \times #4 \times #5 - 2 \times #3 \times #6
\exp_after:wN \__fp_pack_big:NNNNNNw
\int_value:w \c__fp_int_eval:w \c__fp_big_middle_shift_int - #5 \times #5 - 2 \times #4 \times #6
\exp_after:wN \__fp_pack_big:NNNNNNw
\int_value:w \c__fp_int_eval:w \c__fp_big_middle_shift_int - 2 \times #5 \times #6
\exp_after:wN \__fp_pack_big:NNNNNNw
\int_value:w \c__fp_int_eval:w \c__fp_big_middle_shift_int - #6 \times #6
\exp_after:wN \__fp_pack_big:NNNNNNw
\int_value:w \c__fp_int_eval:w \c__fp_big_trailing_shift_int - #6 \times #6
\% 257 \times 5000 \times 0000 \times (#2 \times #3 + 1) + 10 \times 0000 \times 0000$
\{#2\{#3\{#4\{#5\{#6\{#7\{#8\{#9
\}
\}
\}
\}
\}
\}
\}
```

(End of definition for \_\_fp_sqrt_auxii_o:Nnnnnnnn.)

We receive here the difference $a - y^2 = d = \sum_i d_i \cdot 10^{-4i}$, as $(d_4)$; $(d_4)$; ... $(d_{10})$, where each block has 4 digits, except $(d_2)$. This function finds the largest $j \leq 6$ such that $10^{4j}(a - y^2) < 2 \cdot 10^8$, then leaves an open parenthesis and the integer $\lfloor 10^{4j}(a - y^2) \rfloor$
in an integer expression. The closing parenthesis is provided by the caller \_\_fp_sqrt\_auxii_o:\nnnnnnnnN, which completes the expression

\[
10^{ij}z = \left[\left(\lfloor 10^{4j}(a - y^2) \rfloor - 257 \right) \cdot (0.5 \cdot 10^8) / \lfloor 10^8y + 1 \rfloor\right]
\]

for an estimate of \(10^{ij}\sqrt{a - y}\). If \(d_2 \geq 2, j = 3\) and the auxiv auxiliary receives \(10^{j2}z\). If \(d_2 \leq 1\) but \(10^{j3}d_3 + 10^{j4}d_4 + d_5 \geq 1\). In this last case, we detect when \(10^{j2}z < 10^j\), which essentially means \(\sqrt{a - y} \lesssim 10^{-17}\): once this threshold is reached, there is enough information to find the correctly rounded \(\sqrt{a}\) with only one more call to \_\_fp_sqrt_auxii_o:\nnnnnnnnN. Note that the iteration cannot be stuck before reaching \(j = 6\), because for \(j < 6\), one has \(2 \cdot 10^8 \leq 10^{2(j+1)}(a - y^2)\), hence

\[
10^{ij}z \geq \frac{(200000 - 257)(0.5 \cdot 10^8)}{\lfloor 10^y + 1 \rfloor} \geq (200000 - 257) \cdot 0.5 > 0.
\]

\[
\cs_new:Npn \__fp_sqrt_auxiii_o:wnnnnnnnn #1; #2#3#4#5#6#7#8#9 \{ \if_int_compare:w #1 > \c_one_int \exp_after:wN \__fp_sqrt_auxiv_o:NNNNNw \int_value:w \__fp_int_eval:w (#1#2 %) \else: \if_int_compare:w #1#2 > \c_one_int \exp_after:wN \__fp_sqrt_auxv_o:NNNNNw \int_value:w \__fp_int_eval:w (#1#2#3 %) \else: \if_int_compare:w #1#2#3 > \c_one_int \exp_after:wN \__fp_sqrt_auxvi_o:NNNNNw \int_value:w \__fp_int_eval:w (#1#2#3#4 %) \else: \exp_after:wN \__fp_sqrt_auxvii_o:NNNNNw \int_value:w \__fp_int_eval:w (#1#2#3#4#5 %) \fi: \fi: \fi: \fi: \cs_new:Npn \__fp_sqrt_auxiv_o:NNNNNw 1#1#2#3#4#5#6; \cs_new:Npn \__fp_sqrt_auxv_o:NNNNNw 1#1#2#3#4#5#6; \cs_new:Npn \__fp_sqrt_auxvi_o:nnnnnnn {\nnnnnnn \#1#2#3#4#5#6} \cs_new:Npn \__fp_sqrt_auxvii_o:nnnnnnn {\nnnnnnn \#1#2#3#4#5#6}; \cs_new:Npn \__fp_sqrt_auxviii_o:nnnnnnn {\nnnnnnn \#1#2#3#4#5#6}; \if_int_compare:w #1#2 = \c_zero_int \exp_after:wN \__fp_sqrt_auxx_o:NNnnnnnn \fi: \__fp_sqrt_auxviii_o:nnnnnnn \if_int_compare:w \#1#2#3#4#5#6 = \c_zero_int \exp_after:wN \__fp_sqrt_auxx_o:NNnnnnnn \fi: \end{definition}
Simply add the two 8-digit blocks of \( z \), aligned to the last four of the five 4-digit blocks of \( y \), then call the \texttt{auxii} auxiliary to evaluate \( y^2 = (y + z)^2 \).

\[
\begin{align*}
\text{\texttt{\_fp_sqrt_auxviii\_o:nnnnnnn}} & \quad \text{\texttt{\_fp_sqrt_auxix\_o:wnwnw}} \\
\text{\texttt{\_fp_sqrt_auxx\_o:Nnnnnnnn}} & \quad \text{\texttt{\_fp_sqrt_auxxi\_o:wwnnN}}
\end{align*}
\]

At this stage, \( j = 6 \) and \( 10^{24}z < 10^7 \), hence

\[
10^7 + 1/2 > 10^{24}z + 1/2 \geq (10^{24}(a - y^2) - 258) \cdot (0.5 \cdot 10^8) / (10^8 y + 1),
\]

then \( 10^{24}(a - y^2) - 258 < 2(10^7 + 1/2)(y + 10^{-8}) \), and

\[
10^{24}(a - y^2) < (10^7 + 1290.5)(1 + 10^{-8}/y)(2y) < (10^7 + 1290.5)(1 + 10^{-7})(y + \sqrt{a}),
\]

which finally implies \( 0 \leq \sqrt{a} - y < 0.2 \cdot 10^{-16} \). In particular, \( y \) is an underestimate of \( \sqrt{a} \) and \( y + 0.5 \cdot 10^{-16} \) is a (strict) overestimate. There is at exactly one multiple \( m \) of \( 0.5 \cdot 10^{-10} \) in the interval \( [y, y + 0.5 \cdot 10^{-16}] \). If \( m^2 > a \), then the square root is inexact and is obtained by rounding \( m - \epsilon \) to a multiple of \( 10^{-16} \) (the precise shift \( 0 < \epsilon < 0.5 \cdot 10^{-16} \) is irrelevant for rounding). If \( m^2 = a \) then the square root is exactly \( m \), and there is no rounding. If \( m^2 < a \) then we round \( m + \epsilon \). For now, discard a few irrelevant arguments \#1, \#2, \#3, and find the multiple of \( 0.5 \cdot 10^{-16} \) within \( [y, y + 0.5 \cdot 10^{-16}] \); rather, only the last 4 digits \#8 of \( y \) are considered, and we do not perform any carry yet. The \texttt{auxxi} auxiliary sets up \texttt{auxii} with a continuation function \texttt{auxxii} instead of \texttt{auxiii} as before. To prevent \texttt{auxii} from giving a negative results \( a - m^2 \), we compute \( a + 10^{-16} - m^2 \) instead, always positive since \( m < \sqrt{a} + 0.5 \cdot 10^{-16} \) and \( a \leq 1 - 10^{-16} \).
The difference $0 \leq a + 10^{-16} - m^2 \leq 10^{-16} + (\sqrt{a} - m)(\sqrt{a} + m) \leq 2 \cdot 10^{-16}$ was just computed: its first 8 digits vanish, as do the next four, $#1$, and most of the following four, $#2$. The guess $m$ is an underestimate if $a + 10^{-16} - m^2 < 10^{-16}$, that is, $#1#2$ vanishes. Otherwise it is an underestimate, unless $a + 10^{-16} - m^2 = 10^{-16}$ exactly. For an underestimate, call the auxxiv function with argument 9998. For an exact result call it with 9999, and for an overestimate call it with 10000.

This receives 9998, 9999 or 10000 as $#1$ when $m$ is an underestimate, exact, or an overestimate, respectively. Then comes $m$ as five blocks of 4 digits, but where the last block $#6$ may be 0, 5000, or 10000. In the latter case, we need to add a carry, unless $m$ is an underestimate ($#1$ is then 10000). Then comes $a$ as three arguments. Rounding is done by \_fp_round:NNN, whose first argument is the final sign (square roots are positive). We fake its second argument. It should be the last digit kept, but this is only used when ties are “rounded to even”, and only when the result is exactly half-way between two representable numbers rational square roots of numbers with 16 significant digits have: this situation never arises for the square root, as any exact square root of a 16 digit number has at most 8 significant digits. Finally, the last argument is the next digit, possibly shifted by 1 when there are further nonzero digits. This is achieved by \_fp_round_digit:Nw, which receives (after removal of the 10000’s digit) one of 0000, 0001, 4999, 5000, 5001, or 9999, which it converts to 0, 1, 4, 5, 6, and 9, respectively.
The exponent of a normal number is its \( \langle \text{exponent} \rangle \) minus one.

\[
\text{cs_new:Npn } \_\_\_\_\_\_\text{fp_logb_o:w} \ ? \ _\_\_\_\_\_\text{fp_chk:w} \ #1 \ #2 ; @
\{
  \text{if_case:w} \ #1 \ \text{\exp_stop_f:}:
  \_\_\_\_\_\_\_\_\text{fp_case_use:nw}
  \{
    \_\_\_\_\_\_\_\_\_\text{fp_division_by_zero_o:Nnw} \ c_{-\text{inf_fp}} \ c_{\text{logb}} \}
  \text{or: } \_\_\_\_\_\_\text{exp_after:wN} \_\_\_\_\_\_\_\text{fp_logbAux_o:w}
  \text{or: } \_\_\_\_\_\_\text{fp_case_return_o:Nw} \ c_{\text{inf_fp}}
  \text{else: } \_\_\_\_\_\_\_\_\text{fp_case_return_same_o:w}
  \text{fi:}
  \_\_\_\_\_\_\text{fp_chk:w} \ #1 \ #2 ;
\}
\text{cs_new:Npn } \_\_\_\_\_\_\_\_\text{fp_logbAux_o:w} \ _\_\_\_\_\_\_\_\text{fp_chk:w} \ #1 \ #2 \ #3 \ #4 ;
\{
  \_\_\_\_\_\_\text{exp_after:wN} \_\_\_\_\_\_\_\text{fp_parse:n} \ \_\_\_\_\_\_\text{exp_after:wN}
  \{ \_\_\_\_\_\_\text{int_value:w} \_\_\_\_\_\_\text{int_eval:w} \ #3 \ - \ 1 \_\_\_\_\_\_\text{exp_after:wN} \}
\}
\]

(End of definition for \_\_\_\_\_\_\_\_fp_logb_o:w and \_\_\_\_\_\_\_\_fp_logbAux_o:w.)

Find the sign of the floating point: nan, +0, −0, +1 or −1.

\[
\text{cs_new:Npn } \_\_\_\_\_\_\_\_\text{fp_sign_o:w} \ ? \ _\_\_\_\_\_\_\_\text{fp_chk:w} \ #1 \ #2 ; @
\{
  \text{if_case:w} \ #1 \ \text{\exp_stop_f:}:
  \_\_\_\_\_\_\_\_\text{fp_case_return_same_o:w}
  \text{or: } \_\_\_\_\_\_\text{exp_after:wN} \_\_\_\_\_\_\_\text{fp_signAux_o:w}
  \text{or: } \_\_\_\_\_\_\text{exp_after:wN} \_\_\_\_\_\_\_\text{fp_signAux_o:w}
  \text{else: } \_\_\_\_\_\_\_\_\text{fp_case_return_same_o:w}
  \text{fi:}
  \_\_\_\_\_\_\_\_\text{fp_chk:w} \ #1 \ #2 ;
\}
\text{cs_new:Npn } \_\_\_\_\_\_\_\_\_\text{fp_signAux_o:w} \ _\_\_\_\_\_\_\_\text{fp_chk:w} \ #1 \ #2 \ #3 ;
\{
  \_\_\_\_\_\_\text{exp_after:wN} \_\_\_\_\_\_\_\text{fp_setSign_o:w} \_\_\_\_\_\_\text{exp_after:wN} \ c_{\text{one_fp}} \ @
\}
\]

(End of definition for \_\_\_\_\_\_\_\_fp_sign_o:w and \_\_\_\_\_\_\_\_fp_signAux_o:w.)
This function is used for the unary minus and for abs. It leaves the sign of nan invariant,
turns negative numbers (sign 2) to positive numbers (sign 0) and positive numbers (sign 0)
to positive or negative numbers depending on #1. It also expands after itself in the input
stream, just like \_\_fp-_o:ww.

\cs_new:Npn \__fp_set_sign_o:w #1 \s__fp \__fp_chk:w #2#3#4; @
{ \exp_after:wN \__fp_exp_after_o:w \exp_after:wN \s__fp \exp_after:wN \__fp_chk:w \exp_after:wN \int_value:w \if_case:w #3 \exp_stop_f: #1 \or: 1 \or: 0 \fi: \exp_stop_f: #4; }

(End of definition for \__fp_set_sign_o:w.)

### 74.6 Operations on tuples

Two cases: abs(⟨tuple⟩) for which #1 is 0 (invalid for tuples) and ~(tuple) for which
#1 is 2. In that case, map over all items in the tuple an auxiliary that dispatches to the
type-appropriate sign-flipping function.

\cs_new:Npn \__fp_tuple_set_sign_o:w #1#2 @
{ \if_meaning:w 2 #1 \exp_after:wN \__fp_tuple_set_sign_aux_o:Nnw \fi: \__fp_invalid_operation_o:nw { abs } #2 }
\cs_new:Npn \__fp_tuple_set_sign_aux_o:Nnw #1#2
{ \__fp_tuple_map_o:nw \__fp_tuple_set_sign_aux_o:w }
\cs_new:Npn \__fp_change_func_type:NNN #1 \__fp_set_sign_o:w \__fp_parse_apply_unary_error:Nnw 2 #1 #2 ; @

(End of definition for \_\_fp_tuple_set_sign_o:w, \_\_fp_tuple_set_sign_aux_o:Nnw, and \_\_fp_-
tuple_set_sign_aux_o:w)

\_\_fp_*_tuple_o:ww  For (number)*⟨tuple⟩ and ⟨tuple⟩*(number) and ⟨tuple⟩/⟨number⟩, loop through the
\_\_fp_tuple_*_o:ww  ⟨tuple⟩ some code that multiplies or divides by the appropriate ⟨number⟩. Importantly
\_\_fp_tuple_/o:ww  we need to dispatch according to the type, and we make sure to apply the operator in
\_\_fp*_tuple_o:ww the correct order.

\cs_new:cpn { \__fp_*_tuple_o:ww } #1
{ \__fp_tuple_map_o:nw \__fp_binary_type_o:Nww * #1 ; }
\cs_new:cpn { \__fp_tuple_*_o:ww } #1
{ \__fp_tuple_map_o:nw \__fp_binary_rev_type_o:Nww * #1 ; }
\cs_new:cpn { \__fp_tuple_/o:ww } #1
{ \__fp_tuple_map_o:nw \__fp_binary_rev_type_o:Nww / #1 ; }
Check the two tuples have the same number of items and map through these a helper that dispatches appropriately depending on the types. This means \((1,2)+((1,1),2)\) gives \((\text{nan},4)\).

```
\setprotect:Npn __fp_tmp:w #1
{
\cs_new:cpn { __fp_tuple_#1_tuple_o:ww }
\s__fp_tuple __fp_tuple_chk:w ##1 ;
\s__fp_tuple __fp_tuple_chk:w ##2 ;
{ \int_compare:nNnTF
  { \__fp_array_count:n {##1} } = { \__fp_array_count:n {##2} }
  { \__fp_tuple_mapthread_o:nww { \__fp_binary_type_o:Nww #1 } }
  { \__fp_invalid_operation_o:nww #1 }
\s__fp_tuple __fp_tuple_chk:w {##1} ;
\s__fp_tuple __fp_tuple_chk:w {##2} ;
}
\__fp_tmp:w +
\__fp_tmp:w -
```

(End of definition for __fp_*_tuple_o:ww, __fp_tuple_*_o:ww, and __fp_tuple_/o:ww.)
Chapter 75

l3fp-extended implementation

75.1 Description of fixed point numbers

This module provides a few functions to manipulate positive floating point numbers with extended precision (24 digits), but mostly provides functions for fixed-point numbers with this precision (24 digits). Those are used in the computation of Taylor series for the logarithm, exponential, and trigonometric functions. Since we eventually only care about the 16 first digits of the final result, some of the calculations are not performed with the full 24-digit precision. In other words, the last two blocks of each fixed point number may be wrong as long as the error is small enough to be rounded away when converting back to a floating point number. The fixed point numbers are expressed as

\[ \{a_1\} \{a_2\} \{a_3\} \{a_4\} \{a_5\} \{a_6\} ; \]

where each \(a_i\) is exactly 4 digits (ranging from 0000 to 9999), except \(a_1\), which may be any “not-too-large” non-negative integer, with or without leading zeros. Here, “not-too-large” depends on the specific function (see the corresponding comments for details). Checking for overflow is the responsibility of the code calling those functions. The fixed point number \(a\) corresponding to the representation above is \(a = \sum_{i=1}^{6} a_i \cdot 10^{-4i}\).

Most functions we define here have the form

```
\__fp_fixed_(calculation):wwn (operand_1) ; (operand_2) ; { (continuation) }
```

They perform the \(\text{calculation}\) on the two \(\text{operands}\), then feed the result (6 brace groups followed by a semicolon) to the \(\text{continuation}\), responsible for the next step of the calculation. Some functions only accept an \(N\)-type \(\text{continuation}\). This allows constructions such as

```
\__fp_fixed_add:wwn (X_1) ; (X_2) ;
\__fp_fixed_mul:wwn (X_3) ;
\__fp_fixed_add:wwn (X_4) ;
```
to compute \((X_1 + X_2) \cdot X_3 + X_4\). This turns out to be very appropriate for computing continued fractions and Taylor series.

At the end of the calculation, the result is turned back to a floating point number using \texttt{fp\_fixed\_to\_float\_o:}wN. This function has to change the exponent of the floating point number; it must be used after starting an integer expression for the overall exponent of the result.

### 75.2 Helpers for numbers with extended precision

\texttt{\texttt{\_\_fp\_one\_fixed\_tl}}

The fixed-point number 1, used in \texttt{l3fp-expo}.

\begin{verbatim}
\tl_const:Nn \_\_fp\_one\_fixed\_tl
{ {10000} {0000} {0000} {0000} {0000} ; }
\end{verbatim}

(End of definition for \texttt{\_\_fp\_one\_fixed\_tl}.)

\texttt{\_\_fp\_fixed\_continue:kn}

This function simply calls the next function.

\begin{verbatim}
\cs_new:Npn \_\_fp\_fixed\_continue:kn #1; #2 { #2 #1; }
\end{verbatim}

(End of definition for \texttt{\_\_fp\_fixed\_continue:kn}.)

\texttt{\_\_fp\_fixed\_add\_one:kn}

\begin{verbatim}
\_\_fp\_fixed\_add\_one:kn \{ a \}; (continuation)
This function adds 1 to the fixed point \{a\}, by changing \(a_1\) to 10000 + \(a_1\), then calls
the \{continuation\}. This requires \(a_1 + 10000 < 2^{31}\).
\cs_new:Npn \_\_fp\_fixed\_add\_one:kn \{ a \}; #2 { #2 \{ a + 1 \}; }
\end{verbatim}

(End of definition for \texttt{\_\_fp\_fixed\_add\_one:kn}.)

\texttt{\_\_fp\_fixed\_div\_myriad:kn}

Divide a fixed point number by 10000. This is a little bit more subtle than just removing
the last group and adding a leading group of zeros: the first group \#1 may have any
number of digits, and we must split \#1 into the new first group and a second group of
exactly 4 digits. The choice of shifts allows \#1 to be in the range \([0, 5 \cdot 10^8 - 1]\).

\begin{verbatim}
\cs_new:Npn \_\_fp\_fixed\_div\_myriad:kn \#1\#2\#3\#4\#5\#6; \{ \exp_after:wN \_\_fp\_fixed\_mul\_after:wn
\int_value:w \_\_fp\_int\_eval:w \_\_fp\_leading\_shift_int
\exp_after:wN \_\_fp\_pack:NNNNNW
\int_value:w \_\_fp\_int\_eval:w \_\_fp\_trailing\_shift_int
+ \#1 ; \#2\#3\#4\#5\#6;
\}
\end{verbatim}

(End of definition for \texttt{\_\_fp\_fixed\_div\_myriad:kn}.)

\texttt{\_\_fp\_fixed\_mul\_after:wn}

The fixed point operations which involve multiplication end by calling this auxiliary. It
braces the last block of digits, and places the \{continuation\} \#3 in front.

\begin{verbatim}
\cs_new:Npn \_\_fp\_fixed\_mul\_after:wn \#1\#2\#3 { \#3 \{ \#1 \} \#2; }
\end{verbatim}

(End of definition for \texttt{\_\_fp\_fixed\_mul\_after:wn}.)
75.3 Multiplying a fixed point number by a short one

\begin{verbatim}
\__fp_fixed_mul_short:wwn
 \{⟨a⟩1\} \{⟨a⟩2\} \{⟨a⟩3\} \{⟨a⟩4\} \{⟨a⟩5\} \{⟨a⟩6\};
 \{⟨b⟩0\} \{⟨b⟩1\} \{⟨b⟩2\}; \{⟨continuation⟩\}
\end{verbatim}

Computes the product \( c = ab \) of \( a = \sum_i \langle a_i \rangle 10^{-4i} \) and \( b = \sum_i \langle b_i \rangle 10^{-4i} \), rounds it to the closest multiple of \( 10^{-24} \), and leaves \( \{⟨continuation⟩\} \ldots \{⟨c_i⟩\} \) in the input stream, where each of the \( \langle c_i \rangle \) are blocks of 4 digits, except \( ⟨c_1⟩ \), which is any \TeX integer. Note that indices for \( ⟨b⟩ \) start at 0: for instance a second operand of \{0001\}{0000\}{0000\} leaves the first operand unchanged (rather than dividing it by \( 10^4 \), as \( \__fp_fixed_mul:wwn \) would).

\begin{verbatim}
\cs_new:Npn \__fp_fixed_mul_short:wwn #1#2#3#4#5#6; #7#8#9;
 \exp_after:wN \__fp_fixed_mul_after:wwn
 \int_value:w \__fp_int_eval:w \c__fp_leading_shift_int
 + #1*#7
 \exp_after:wN \__fp_pack:NNNNNw
 \int_value:w \__fp_int_eval:w \c__fp_middle_shift_int
 + #1*#8 + #2*#7
 \exp_after:wN \__fp_pack:NNNNNw
 \int_value:w \__fp_int_eval:w \c__fp_middle_shift_int
 + #1*#9 + #2*#8 + #3*#7
 \exp_after:wN \__fp_pack:NNNNNw
 \int_value:w \__fp_int_eval:w \c__fp_middle_shift_int
 + #2*#9 + #3*#8 + #4*#7
 \exp_after:wN \__fp_pack:NNNNNw
 \int_value:w \__fp_int_eval:w \c__fp_middle_shift_int
 + #3*#9 + #4*#8 + #5*#7
 \exp_after:wN \__fp_pack:NNNNNw
 \int_value:w \__fp_int_eval:w \c__fp_trailing_shift_int
 + #4*#9 + #5*#8 + #6*#7
 + ( #5*#9 + #6*#8 + #6*#9 / \c__fp_myriad_int )
 / \c__fp_myriad_int ;
\end{verbatim}

(End of definition for \( \__fp_fixed_mul_short:wwn \).)

75.4 Dividing a fixed point number by a small integer

\begin{verbatim}
\end{verbatim}

\begin{verbatim}
\__fp_fixed_div_int:wwN \{⟨a⟩\} \{⟨n⟩\}; \{⟨continuation⟩\}
\end{verbatim}

Divides the fixed point number \( ⟨a⟩ \) by the (small) integer \( 0 < ⟨n⟩ < 10^4 \) and feeds the result to the \( ⟨continuation⟩ \). There is no bound on \( a_1 \).

The arguments of the \( i \) auxiliary are 1: one of the \( a_i \), 2: \( n \), 3: the \( ii \) or the \( iii \) auxiliary. It computes a (somewhat tight) lower bound \( Q_i \) for the ratio \( a_i/n \).

The \( ii \) auxiliary receives \( Q_i, n, \) and \( a_i \) as arguments. It adds \( Q_i \) to a surrounding integer expression, and starts a new one with the initial value \( 9999 \), which ensures that the result of this expression has 5 digits. The auxiliary also computes \( a_i - n \cdot Q_i \), placing the result in front of the 4 digits of \( a_{i+1} \). The resulting \( a_{i+1} \) = \( 10^4(a_i - n \cdot Q_i) + a_{i+1} \) serves as the first argument for a new call to the \( i \) auxiliary.

When the \( iii \) auxiliary is called, the situation looks like this:
where expansion is happening from the last line up. The \( a \) auxiliary adds \( a_6 + 2 \approx a_6/n + 1 \) to the last 9999, giving the integer closest to 10000 + \( a_6/n \).

Each pack auxiliary receives 5 digits followed by a semicolon. The first digit is added as a carry to the integer expression above, and the 4 other digits are braced. Each call to the pack auxiliary thus produces one brace group. The last brace group is produced by the after auxiliary, which places the \langle \text{continuation} \rangle \text{as appropriate.}

(End of definition for \texttt{\_fp_fixed_div_int:wwN} and others.)

### 75.5 Adding and subtracting fixed points

\texttt{\_fp_fixed_add:wwN} (a) ; (b) \{\langle \text{continuation} \rangle \}

\texttt{\_fp_fixed_sub:wwN}
Computes \(a+b\) (resp. \(a-b\)) and feeds the result to the \(\langle\text{continuation}\rangle\). This function requires \(0 \leq a_1, b_1 \leq 114748\), its result must be positive (this happens automatically for addition) and its first group must have at most 5 digits: \((a \pm b)_1 < 100000\). The two functions only differ by a sign, hence use a common auxiliary. It would be nice to grab the 12 brace groups in one go; only 9 parameters are allowed. Start by grabbing the sign, \(a_1, \ldots, a_4\), the rest of \(a\), and \(b_1\) and \(b_2\). The second auxiliary receives the rest of \(a\), the sign multiplying \(b\), the rest of \(b\), and the \(\langle\text{continuation}\rangle\) as arguments. After going down through the various level, we go back up, packing digits and bringing the \(\langle\text{continuation}\rangle\) (#8, then #7) from the end of the argument list to its start.

\[
\begin{align*}
\cs_new:Npn \_fp_fixed_add:wwn & { \_fp_fixed_add:Nnnnnwnn + } \\
\cs_new:Npn \_fp_fixed_sub:wwn & { \_fp_fixed_add:Nnnnnwnn - } \\
\cs_new:Npn \_fp_fixed_add:Nnnnnwnn & \int_value:w \_fp_int_eval:w 9 9999 9998 + #2#3 #1 #7#8 \\
\cs_new:Npn \_fp_fixed_add:nnNnnnwnn & \int_value:w \_fp_int_eval:w 2 0000 0000 #3 #6#7 + #1#2 ; \{#8\} \\
\cs_new:Npn \_fp_fixed_add_after:NNNNNwn & \int_value:w \_fp_int_eval:w 2 0000 0000 #3 #6#7 + #1#2 ; \{#8\} \\
\cs_new:Npn \_fp_fixed_add:nnnNnnnwnn & \int_value:w \_fp_int_eval:w 2 0000 0000 #3 #6#7 + #1#2 ; \{#8\} \\
\cs_new:Npn \_fp_fixed_add_after:NNNNNwnn & \int_value:w \_fp_int_eval:w 2 0000 0000 #3 #6#7 + #1#2 ; \{#8\} \end{align*}
\]

(End of definition for \_fp_fixed_add:wwn and others.)

### 75.6 Multiplying fixed points

\[
\begin{align*}
\_fp_fixed_mul:wwn & \_fp_fixed_mul:nnnNnnnwnn \end{align*}
\]

Computes \(a \times b\) and feeds the result to \(\langle\text{continuation}\rangle\). This function requires \(0 \leq a_1, b_1 < 10000\). Once more, we need to play around the limit of 9 parameters for \TeX{} macros. Note that we don’t need to obtain an exact rounding, contrarily to the \texttt{*} operator, so things could be harder. We wish to perform carries in

\[
a \times b = a_1 \cdot b_1 \cdot 10^{-8} + (a_1 \cdot b_2 + a_2 \cdot b_1) \cdot 10^{-12} + (a_1 \cdot b_3 + a_2 \cdot b_2 + a_3 \cdot b_1) \cdot 10^{-16} + (a_1 \cdot b_4 + a_2 \cdot b_3 + a_3 \cdot b_2 + a_4 \cdot b_1) \cdot 10^{-20} + \left( a_2 \cdot b_4 + a_3 \cdot b_3 + a_4 \cdot b_2 
\begin{align*}
&+ a_3 \cdot b_4 + a_4 \cdot b_3 + a_1 \cdot b_6 + a_2 \cdot b_5 + a_5 \cdot b_2 + a_6 \cdot b_1 \right) \cdot 10^4 + a_1 \cdot b_5 + a_5 \cdot b_1 \right) \cdot 10^{-24} + O(10^{-24}),
\end{align*}
\]
where the $O(10^{-24})$ stands for terms which are at most $5 \cdot 10^{-24}$; ignoring those leads to an error of at most 5 ulp. Note how the first 15 terms only depend on $a_1, \ldots, a_4$ and $b_1, \ldots, b_4$, while the last 6 terms only depend on $a_1, a_2, a_5, a_6$, and the corresponding parts of $b$. Hence, the first function grabs $a_1, \ldots, a_4$, the rest of $a$, and $b_1, \ldots, b_4$, and writes the 15 first terms of the expression, including a left parenthesis for the fraction. The i auxiliary receives $a_5, a_6, b_1, b_2, a_1, a_2, b_5, b_6$ and finally the (continuation) as arguments. It writes the end of the expression, including the right parenthesis and the denominator of the fraction. The (continuation) is finally placed in front of the 6 brace groups by \_\_fp_fixed_mul_after:wwn.

\begin{verbatim}
\cs_new:Npn \_\_fp_fixed_mul:wwn #1#2#3#4 #5; #6#7#8#9
\exp_after:wN \_\_fp_fixed_mul_after:wwn
\int_value:w \_\_fp_int_eval:w \_c_fp.leading_shift_int
+ #1*#6
\exp_after:wN \_\_fp_pack:NNNNNW
\int_value:w \_\_fp_int_eval:w \_c_fp.middle_shift_int
+ #1*#7 + #2*#6
\exp_after:wN \_\_fp_pack:NNNNNW
\int_value:w \_\_fp_int_eval:w \_c_fp.middle_shift_int
+ #1*#8 + #2*#7 + #3*#6
\exp_after:wN \_\_fp_pack:NNNNNW
\int_value:w \_\_fp_int_eval:w \_c_fp.middle_shift_int
+ #1*#9 + #2*#8 + #3*#7 + #4*#6
\exp_after:wN \_\_fp_pack:NNNNNW
\int_value:w \_\_fp_int_eval:w \_c_fp.trailing_shift_int
+ #2*#9 + #3*#8 + #4*#7
+ ( #3*#9 + #4*#8
+ \_\_fp_fixed_mul:annnnnnn #5 #6 #7 #8 #1 #2)
\cs_new:Npn \_\_fp_fixed_mul:nnnnnnnw #1#2 #3#4 #5#6 #7#8 ;
\exp_after:wN \_\_fp_fixed_mul:nnnnnnn
\exp_after:wN \_\_fp_fixed_mul:nnnnnnn
\exp_after:wN \_\_fp_fixed_mul:nnnnnnn
\int_value:w \_\_fp_int_eval:w \_c_fp.myriad_int
+ #1*#4 + #2*#3 + #5*#8 + #6*#7 / \_\_fp_fixed_mul:nnnnnnnnn
\end{verbatim}

(End of definition for \_\_fp_fixed_mul:wwn and \_\_fp_fixed_mul:nnnnnnnw.)

75.7 Combining product and sum of fixed points

\begin{verbatim}
\_\_fp_fixed_mul_add:wwn (a); (b); (c); {\langle continuation\rangle}
\_\_fp_fixed_mul_sub_back:wwn (a); (b); (c); {\langle continuation\rangle}
\_\_fp_fixed_one_minus_mul:wwn (a); (b); {\langle continuation\rangle}
\end{verbatim}

Sometimes called FMA (fused multiply-add), these functions compute $a \times b + c, c – a \times b,$ and $1 – a \times b$ and feed the result to the (continuation). These functions require $0 \leq a_1, b_1, c_1 \leq 10000$. Since those functions are at the heart of the computation of Taylor expansions, we over-optimize them a bit, and in particular we do not factor out the common parts of the three functions.
For definiteness, consider the task of computing \( a \times b + c \). We perform carries in

\[
a \times b + c = (a_1 \cdot b_1 + c_{12}) \cdot 10^{-8} \\
+ (a_1 \cdot b_2 + a_2 \cdot b_1) \cdot 10^{-12} \\
+ (a_1 \cdot b_3 + a_2 \cdot b_2 + a_3 \cdot b_1 + c_{23}) \cdot 10^{-16} \\
+ (a_1 \cdot b_4 + a_2 \cdot b_3 + a_3 \cdot b_2 + a_4 \cdot b_1) \cdot 10^{-20} \\
+ \left( a_2 \cdot b_4 + a_3 \cdot b_3 + a_4 \cdot b_2 \\
+ \frac{a_1 \cdot b_4 + a_4 \cdot b_1 + a_1 \cdot b_2 + a_2 \cdot b_5 + a_5 \cdot b_2 + a_6 \cdot b_1}{10^4} \\
+ a_1 \cdot b_5 + a_5 \cdot b_1 + c_{56} \right) \cdot 10^{-24} + O(10^{-24}),
\]

where \( c_{12}, c_{23}, c_{56} \) denote the 8-digit number obtained by juxtaposing the two blocks of digits of \( c \), and \( \cdot \) denotes multiplication. The task is obviously tough because we have 18 brace groups in front of us.

Each of the three function starts the first two levels (the first, corresponding to \( 10^{-4} \), is empty), with \( c_{12} \) in the first level, calls the \( i \) auxiliary with arguments described later, and adds a trailing \( + c_{56} \); \{ (continuation) \};. The \( + c_{56} \) piece, which is omitted for \( \_\_fp\_fixed\_one\_minus\_mul:wwn \), is taken in the integer expression for the \( 10^{-24} \) level.

```latex
\begin{verbatim}
\cs_new:Npn \_fp_fixed_mul_add:wwwn #1; #2; #3#4#5#6#7#8; 
#1 \_fp_fixed_mul_add:wwwn \_fp_fixed_mul_after:wwn 
\_fp_int_eval:w \_fp_big_middle_shift_int + #3 #4 
\_fp_fixed_mul_add:Nwnnnwnnn +
+ #5 #6 ; #2 ; #1 ; #2 ; +
+ #7 #8 ; ;
\}
\cs_new:Npn \_fp_fixed_mul_sub_back:wwwn #1; #2; #3#4#5#6#7#8; 
#1 \_fp_fixed_mul_sub_back:wwwn \_fp_fixed_mul_after:wwn 
\_fp_int_eval:w \_fp_big_middle_shift_int + #3 #4 
\_fp_fixed_mul_sub_back:wwwn \_fp_fixed_mul_after:wwn 
\_fp_int_eval:w \_fp_big_middle_shift_int + #3 #4 
\_fp_fixed_mul_sub_back:wwwn -
+ #5 #6 ; #2 ; #1 ; #2 ; -
+ #7 #8 ; ;
\}
\cs_new:Npn \_fp_fixed_one_minus_mul:wwn #1; #2; 
#1 \_fp_one_minus_mul:wwn \_fp_fixed_mul_after:wwn 
\_fp_int_eval:w \_fp_big_middle_shift_int + #3 #4 
\_fp_one_minus_mul:wwn \_fp_fixed_mul_after:wwn 
\_fp_int_eval:w \_fp_big_middle_shift_int + #3 #4 
\_fp_one_minus_mul:wwn -
+ #5 #6 ; #2 ; #1 ; #2 ; -
+ #7 #8 ; ;
\}
\end{verbatim}
```

1150
Expressions \langle\text{all parts of} \__fp_fixed_mul_one_minus_mul:wwn \rangle \text{was inserted by the} \langle a \rangle \cdot \langle b \rangle \text{;} \langle \text{op} \rangle.

To do all this, we keep \langle b \rangle \text{; arguments} #3, #4, #5 \text{are} \langle b_1 \rangle, \langle b_2 \rangle, \langle b_3 \rangle; \text{arguments} #7, #8, #9 \text{are} \langle a_1 \rangle, \langle a_2 \rangle, \langle a_3 \rangle. \text{We can build three levels:} a_1 \cdot b_1 \text{for} 10^{-8}, (a_1 \cdot b_2 + a_2 \cdot b_1) \text{for} 10^{-12}, \text{and} (a_1 \cdot b_3 + a_2 \cdot b_2 + a_3 \cdot b_1 + c_3 c_4) \text{for} 10^{-16}. \text{The} a \cdot b \text{products use the sign} #1. \text{Note that} #2 \text{is empty for} \__fp_fixed_one_minus_mul:wwn. \text{We call the} \langle a \rangle \cdot \langle b \rangle \text{auxiliary for levels} 10^{-20} \text{and} 10^{-24}, \text{keeping the pieces of} \langle a \rangle \text{we've read, but not} \langle b \rangle, \text{since there is another copy later in the input stream.}

Level \text{is} \text{another copy later in the input stream.} \text{for} 10^{-20} \text{is} (a_1 \cdot b_4 + a_2 \cdot b_3 + a_3 \cdot b_2 + a_4 \cdot b_1), \text{multiplied by the sign, which was inserted by the} \begin{array}{c} \text{auxiliary.} \end{array} \text{Then we prepare level} 10^{-24}. \text{We don't have access to all parts of} \langle a \rangle \text{and} \langle b \rangle \text{needed to make all products. Instead, we prepare the partial expressions} \begin{array}{c} b_1 + a_1 \cdot b_4 + a_2 \cdot b_3 + a_3 \cdot b_2 + a_4 \cdot b_1 \end{array} \begin{array}{c} b_2 + a_1 \cdot b_4 + a_2 \cdot b_3 + a_3 \cdot b_2 + a_4 \cdot b_1, \end{array} \text{and of course with the trailing} + c_5 c_6. \text{To do all this, we keep} a_1, a_5, a_6, \text{and the corresponding pieces of} \langle b \rangle.
75.8 Extended-precision floating point numbers

In this section we manipulate floating point numbers with roughly 24 significant figures ("extended-precision" numbers, in short, "ep"), which take the form of an integer exponent, followed by a comma, then six groups of digits, ending with a semicolon. The first group of digits may be any non-negative integer, while other groups of digits have 4 digits. In other words, an extended-precision number is an exponent ending in a comma, then a fixed point number. The corresponding value is \(0.\langle\text{digits}\rangle\cdot10^{\langle\text{exponent}\rangle}\). This convention differs from floating points.

Converting an extended-precision number with an exponent at most 4 and a first block less than \(10^8\) to a fixed point number whose first block has 12 digits, hopefully starting with many zeros.

\[
\text{\texttt{\_fp_ep_to_fixed:wwn}} \quad \text{\texttt{\_fp_ep_to_fixed_auxi:www}} \quad \text{\texttt{\_fp_ep_to_fixed_auxii:nnnnnnnwn}}
\]
Normalize an extended-precision number. More precisely, leading zeros are removed from
the mantissa of the argument, decreasing its exponent as appropriate. Then the digits
are packed into 6 groups of 4 (discarding any remaining digit, not rounding). Finally,
the continuation #8 is placed before the resulting exponent–mantissa pair. The input
exponent may in fact be given as an integer expression. The loop auxiliary grabs a
digit: if it is 0, decrement the exponent and continue looping, and otherwise call the end
auxiliary, which places all digits in the right order (the digit that was not
0, and any
remaining digits), followed by some 0, then packs them up neatly in 3 × 2 = 6 blocks of
four. At the end of the day, remove with ___fp_use_i:ww any digit that did not make it
in the final mantissa (typically only zeros, unless the original first block has more than 4
digits).

\cs_new:Npn \__fp_ep_to_ep:wwN #1,#2#3#4#5#6#7; \#8
\exp_after:wN \#8
\int_value:w \__fp_int_eval:w #1 + 4
\exp_after:wN \use_i:nn
\exp_after:wN \__fp_ep_to_ep_loop:N
\int_value:w \__fp_int_eval:w 1 0000 0000 + \#2 \__fp_int_eval_end:
\#3#4#5#6#7 ; ; !
\__fp_ep_to_ep_loop:N #1
\__fp_ep_to_ep_end:www #1 \fi: \__fp_ep_to_ep_loop:N #2; #3!
\__fp_ep_to_ep_zero:ww \fi: #1; #2; #3;
\__fp_pack_twice_four:wNNNNNNNN \__fp_pack_twice_four:wNNNNNNNN
\__fp_pack_twice_four:wNNNNNNNN \__fp_use_i:ww , ;
\#1 \#2 0000 0000 0000 0000 0000 0000 0000 0000 ;
\__fp_ep_to_ep_zero:ww \fi: \#1; \#2; \#3;
\__fp_ep_to_ep_zero:ww \fi: #1; #2; #3;
\{ \fi: , {10000000000000000000000000000000} ; } }

(End of definition for ___fp_ep_to_ep:wwN and others.)

In l3fp-trig we need to compare two extended-precision numbers. This is based on the
same function for positive floating point numbers, with an extra test if comparing only
16 decimals is not enough to distinguish the numbers. Note that this function only works
if the numbers are normalized so that their first block is in [1000, 9999].
\cs_new:Npn \__fp_ep_compare:wwww #1,#2#3#4#5#6#7;
\__fp_ep_mul:wwwwn
\__fp_ep_mul_raw:wwwwN

Multiply two extended-precision numbers: first normalize them to avoid losing too much precision, then multiply the mantissas #2 and #4 as fixed point numbers, and sum the exponents #1 and #3. The result's first block is in \([100,9999]\).

\cs_new:Npn \__fp_ep_mul:wwwwn #1,#2; #3,#4; #5
\__fp_ep_to_ep:wwN #3,#4;
\__fp_fixed_continue:wn
\__fp_ep_to_ep:wwN #1,#2;
\__fp_ep_mul_raw:wwwwN
\__fp_fixed_continue:wn
\cs_new:Npn \__fp_ep_mul_raw:wwwwN #1,#2; #3,#4; #5
\__fp_fixed_mul:wn #2; #4;
\exp_after:wN #5 \int_value:w \__fp_int_eval:w #1 + #3 ,
\}
\}

(End of definition for \__fp_ep_mul:wwwwn and \__fp_ep_mul_raw:wwwwN.)

### 75.9 Dividing extended-precision numbers

Divisions of extended-precision numbers are difficult to perform with exact rounding: the technique used in l3fp-basics for 16-digit floating point numbers does not generalize easily to 24-digit numbers. Thankfully, there is no need for exact rounding.

Let us call \(n\) the numerator and \(d\) the denominator. After a simple normalization step, we can assume that \(\langle n \rangle \in [0.1,1)\) and \(\langle d \rangle \in [0.1,1)\), and compute \(\langle n \rangle/(10\langle d \rangle)\) \(\in (0.01,1)\). In terms of the 6 blocks of digits \(\langle n_1 \rangle \cdots \langle n_6 \rangle\) and the 6 blocks \(\langle d_1 \rangle \cdots \langle d_6 \rangle\), the condition translates to \(\langle n_1 \rangle, \langle d_1 \rangle \in [1000,9999]\).
We first find an integer estimate $a \simeq 10^8/\langle d \rangle$ by computing

\[
\alpha = \left\lfloor \frac{10^9}{\langle d_1 \rangle + 1} \right\rfloor \\
\beta = \left\lfloor \frac{10^9}{\langle d_1 \rangle} \right\rfloor \\
a = 10^3\alpha + (\beta - \alpha) \cdot \left( 10^3 - \left\lfloor \frac{\langle d_2 \rangle}{10} \right\rfloor \right) - 1250,
\]

where $\left\lfloor \frac{a}{b} \right\rfloor$ denotes $\varepsilon$-TEX’s rounding division, which rounds ties away from zero. The idea is to interpolate between $10^3\alpha$ and $10^3\beta$ with a parameter $\langle d_2 \rangle/10^4$, so that when $\langle d_2 \rangle = 0$ one gets $a = 10^3\beta - 1250 \simeq 10^{12}/\langle d_1 \rangle \simeq 10^8/\langle d \rangle$, while when $\langle d_2 \rangle = 9999$ one gets $a = 10^3\alpha - 1250 \simeq 10^{12}/(\langle d_1 \rangle + 1) \simeq 10^8/\langle d \rangle$. The shift by 1250 helps to ensure that $a$ is an underestimate of the correct value. We shall prove that

\[
1 - 1.755 \cdot 10^{-5} < \frac{\langle d \rangle a}{10^8} < 1.
\]

We can then compute the inverse of $\langle d \rangle a/10^8 = 1 - \epsilon$ using the relation $1/(1 - \epsilon) \simeq (1 + \epsilon)(1 + \epsilon^2) + \epsilon^4$, which is correct up to a relative error of $\epsilon^5 < 1.6 \cdot 10^{-24}$. This allows us to find the desired ratio as

\[
\frac{\langle n \rangle}{\langle d \rangle} = \frac{\langle n \rangle a}{10^8} (1 + \epsilon)(1 + \epsilon^2) + \epsilon^4).
\]

Let us prove the upper bound first (multiplied by $10^{15}$). Note that $10^7\langle d \rangle < 10^3\langle d_1 \rangle + 10^{-1}(\langle d_2 \rangle + 1)$, and that $\varepsilon$-TEX’s division $\left\lfloor \frac{\langle d_2 \rangle}{10} \right\rfloor$ underestimates $10^{-1}(\langle d_2 \rangle + 1)$ by 0.5 at most, as can be checked for each possible last digit of $\langle d_2 \rangle$. Then,

\[
10^7\langle d \rangle a < \left( 10^3\langle d_1 \rangle + \left\lfloor \frac{\langle d_2 \rangle}{10} \right\rfloor + \frac{1}{2} \right) \left( 10^3 - \left\lfloor \frac{\langle d_2 \rangle}{10} \right\rfloor \right) \beta + \left\lfloor \frac{\langle d_2 \rangle}{10} \right\rfloor \alpha - 1250 \tag{1}
\]

\[
< 10^3\langle d_1 \rangle + \left\lfloor \frac{\langle d_2 \rangle}{10} \right\rfloor + \frac{1}{2} \left( 10^3 - \left\lfloor \frac{\langle d_2 \rangle}{10} \right\rfloor \right) \left( 10^3 - \left\lfloor \frac{\langle d_2 \rangle}{10} \right\rfloor + \frac{1}{2} \right) + \left\lfloor \frac{\langle d_2 \rangle}{10} \right\rfloor \langle d_1 \rangle + \left\lfloor \frac{\langle d_2 \rangle}{10} \right\rfloor \langle d_1 \rangle + \frac{1}{2} - 1250 \tag{2}
\]

\[
< 10^3\langle d_1 \rangle + \left\lfloor \frac{\langle d_2 \rangle}{10} \right\rfloor + \frac{1}{2} \left( 10^3 - \left\lfloor \frac{\langle d_2 \rangle}{10} \right\rfloor \right) \left( 10^3 - \left\lfloor \frac{\langle d_2 \rangle}{10} \right\rfloor + \frac{1}{2} \right) - 750 \tag{3}
\]

\[
< 10^3\langle d_1 \rangle + \left\lfloor \frac{\langle d_2 \rangle}{10} \right\rfloor + \frac{1}{2} \left( 10^3 - \langle d_2 \rangle \right) \langle d_1 \rangle - \left\lfloor \frac{\langle d_2 \rangle}{10} \right\rfloor \langle d_1 \rangle \langle d_1 \rangle + \frac{1}{2} \langle d_2 \rangle - \left\lfloor \frac{\langle d_2 \rangle}{10} \right\rfloor \langle d_1 \rangle \langle d_1 \rangle + \frac{1}{2} \langle d_2 \rangle - 1250 \tag{4}
\]

We recognize a quadratic polynomial in $\left\lfloor \langle d_2 \rangle/10 \right\rfloor$ with a negative leading coefficient: this polynomial is bounded above, according to $(\left\lfloor \langle d_2 \rangle/10 \right\rfloor + a)(b - c(\langle d_2 \rangle/10)) \leq (b + ca)^2/(4c)$. Hence,

\[
10^7\langle d \rangle a < \frac{10^{15}}{\langle d_1 \rangle \left( \langle d_1 \rangle + 1 \right)} \left( \langle d_1 \rangle + \frac{1}{2} + \frac{1}{4} \cdot 10^{-3} - \frac{3}{8} \cdot 10^{-9} \langle d_1 \rangle \langle d_1 \rangle + 1 \right)^2
\]

Since $\langle d_1 \rangle$ takes integer values within $[1000, 9999]$, it is a simple programming exercise to check that the squared expression is always less than $\langle d_1 \rangle \langle d_1 \rangle + 1$, hence $10^7\langle d \rangle a < 10^{15}$. The upper bound is proven. We also find that $\frac{3}{8}$ can be replaced by slightly smaller numbers, but nothing less than 0.374563... , and going back through the derivation of
the upper bound, we find that 1250 is as small a shift as we can obtain without breaking
the bound.

Now, the lower bound. The same computation as for the upper bound implies

$$10^7(d)a > \left( 10^3(d_1) + \left\lfloor \frac{(d_2)}{10} \right\rfloor - \frac{1}{2} \right) \left( 10^{12} \frac{(d_2)}{10} - \left\lfloor \frac{(d_1)}{(d_1) + 1} \right\rfloor - 1750 \right)$$

This time, we want to find the minimum of this quadratic polynomial. Since the leading
coefficient is still negative, the minimum is reached for one of the extreme values $[y/10] = 0$ or $[y/10] = 100$, and we easily check the bound for those values.

We have proven that the algorithm gives us a precise enough answer. Incidentally,
the upper bound that we derived tells us that $a < 10^8/\langle d \rangle \leq 10^9$, hence we can compute
$a$ safely as a TeX integer, and even add $10^9$ to it to ease grabbing of all the digits. The lower
bound implies $10^8 - 1755 < a$, which we do not care about.

\begin{verbatim}
\__fp_ep_div:wwwwn

Compute the ratio of two extended-precision numbers. The result is an extended-
precision number whose first block lies in the range $[100,9999]$, and is placed after the
continuation once we are done. First normalize the inputs so that both first block
lie in $[1000,9999]$, then call \__fp_ep_div_esti:wwwwn (denominator) \langle numerator \rangle,
responsible for estimating the inverse of the denominator.

\cs_new:Npn \__fp_ep_div:wwwwn #1,#2; #3,#4;
\__fp_ep_to_ep:wwN #1,#2;
\__fp_fixed_continue:wn
\__fp_ep_to_ep:wwN #3,#4;
\__fp_ep_div_esti:wwwwn
\}

(End of definition for \__fp_ep_div:wwwwn.)
\end{verbatim}

The esti function evaluates $\alpha = 10^9/(\langle d_1 \rangle + 1)$, which is used twice in the expression for
$a$, and combines the exponents #1 and #4 (with a shift by 1 because we later compute
$\langle n \rangle/(10^d)$). Then the estii function evaluates $10^9 + a$, and puts the exponent #2
after the continuation #7: from there on we can forget exponents and focus on the
mantissa. The estiii function multiplies the denominator #7 by $10^{-8}a$ (obtained as a
split into the single digit #1 and two blocks of 4 digits, #2#3#4#5 and #6). The result
$10^{-9}a(d) = (1 - \epsilon)$, and a partially packed $10^{-9}a$ (as a block of four digits, and five
individual digits, not packed by lack of available macro parameters here) are passed to
\__fp_ep_div_estii:wwnnnnn, which computes $10^{-9}a/(1 - \epsilon)$, that is, $1/(10^d)$ and we
finally multiply this by the numerator #8.

\begin{verbatim}
\cs_new:Npn \__fp_ep_div_esti:wwwwn \#1,#2; \#3,#4;
\exp_after:wN \__fp_ep_div_estii:wwnnn\n\int_value:w \__fp_int_eval:w 10 0000 0000 / ( \#2 + 1 )
\exp_after:wN ;
\int_value:w \__fp_int_eval:w #4 - #1 + 1 ,
{\#2} \#3;
\}
\cs_new:Npn \__fp_ep_div_estii:wwnnnnn \#1; \#2,#3#4#5; \#6; \#7
\{
\exp_after:wN \__fp_ep_div_estiii:NNNNNwwwn
\}
\end{verbatim}
The bounds shown above imply that the \texttt{epsi} function's first operand is \((1 - \epsilon)\) with \(\epsilon \in [0, 1.755 \cdot 10^{-5}]\). The \texttt{epsi} function computes \(\epsilon\) as \(1 - (1 - \epsilon)\). Since \(\epsilon < 10^{-4}\), its first block vanishes and there is no need to explicitly use \#1 (which is 9999). Then \texttt{epsii} evaluates \(10^{-a}/(1 - \epsilon)\) as \((1 + \epsilon^2)(1 + \epsilon)(10^{-9a} + 10^{-9a})\). Importantly, we compute \(10^{-9a}\) before multiplying it with the rest, rather than multiplying by \(\epsilon\) and then \(10^{-9a}\), as this second option loses more precision. Also, the combination of \texttt{short_mul} and \texttt{div_myriad} is both faster and more precise than a simple \texttt{mul}.

\begin{verbatim}
\cs_new:Npn \__fp_ep_div_epsii:wwn#1#2#3#4#5#6; #7; 
\{ \__fp_fixed_mul_short:wwn #7; \{#1\}{#2#3#4#5}{#6}; \__fp_ep_div_epsii:wn#1#2#3#4#5#6; \__fp_fixed_mul:wwn \}
\cs_new:Npn \__fp_ep_div_eps_pack:NNNNNw #1#2#3#4#5#6; 
\{ + #1 ; \{#2#3#4#5#6\} \}
\cs_new:Npn \__fp_ep_div_epsii:wwn#1; #2; #3#4#5#6#7#8 
\{ \__fp_fixed_mul:wn \{0000\}\{#1\}\{#2\}; \__fp_fixed_add_one:wn \__fp_fixed_mul:wn \{10000\} \{#1\} \#2 ; \{ \__fp_fixed_mul_short:wwn \{0000\}\{#1\}\{#2\}; \{#3\}{#4#5#6#7}\{#8000\}; \__fp_fixed_div_myriad:wn \__fp_fixed_mul:wn \} \__fp_fixed_add:wn \{#3\}{#4#5#6#7}\{#8000\}\{0000\}\{0000\}; \}
\end{verbatim}

(End of definition for \texttt{\_fp_ep_div_epsii:wwn}, \texttt{\_fp_ep_div_eps_pack:NNNNNw}, and \texttt{\_fp_ep_div_\--estiii:NNNNNwwwn}.)
75.10 Inverse square root of extended precision numbers

The idea here is similar to division. Normalize the input, multiplying by powers of 100 until we have \( x \in [0.01, 1) \). Then find an integer approximation \( r \in [101, 1003] \) of \( 10^2/\sqrt{x} \), as the fixed point of iterations of the Newton method: essentially \( r \mapsto (r + 10^8/(x r))/2 \), starting from a guess that optimizes the number of steps before convergence. In fact, just as there is a slight shift when computing divisions to ensure that some inequalities hold, we replace \( 10^8 \) by a slightly larger number which ensures that \( r^2 x \geq 10^4 \). This also causes \( r \in [101, 1003] \). Another correction to the above is that the input is actually normalized to \([0.1, 1)\], and we use either \( 10^8 \) or \( 10^9 \) in the Newton method, depending on the parity of the exponent. Skipping those technical hurdles, once we have the approximation \( r \), we set \( y = 10^{-4} r^2 x \) (or rather, the correct power of 10 to get \( y \approx 1 \)) and compute \( y^{-1/2} \) through another application of Newton’s method. This time, the starting value is \( z = 1 \), each step maps \( z \mapsto z(1.5 - 0.5 y z^2) \), and we perform a fixed number of steps. Our final result combines \( r \) with \( y^{-1/2} \) as \( x^{-1/2} = 10^{-2} ry^{-1/2} \).

---

\[
\text{\textbackslash cs new\textunderscore Npn \textbackslash _fp\textunderscore ep\textunderscore isqrt:wwn #1,#2;}
\text{\{}
\text{\textbackslash _fp\textunderscore ep\textunderscore to\textunderscore ep:wwN #1,#2;}
\text{\textbackslash _fp\textunderscore ep\textunderscore isqrt\textunderscore auxi:wwn}
\text{\}}
\text{\textbackslash cs new\textunderscore Npn \textbackslash _fp\textunderscore ep\textunderscore isqrt\textunderscore auxi:wwn #1,}
\text{\{}
\text{\exp\textunderscore after:wwN \textbackslash _fp\textunderscore ep\textunderscore isqrt\textunderscore auxii:wwnnwn}
\text{\int\textunderscore value:ww \textbackslash _fp\textunderscore int\textunderscore eval:w}
\text{\int\textunderscore if\textunderscore odd:nTF \{#1\}
\text{\{ (1 - #1) / 2 , 535 , \{ 0 \} \{ \} \}}
\text{\{ 1 - #1 / 2 , 168 , \{ \} \{ 0 \} \}}
\text{\}}
\text{\} \text{\textbackslash cs new\textunderscore Npn \textbackslash _fp\textunderscore ep\textunderscore isqrt\textunderscore auxii:wwnnwn #1, #2, #3#4 #5#6; #7
\text{\{}
\text{\textbackslash _fp\textunderscore ep\textunderscore isqrt\textunderscore esti:wwnnwn #2, 0, #5, \{#3\} \{#4\}
\text{\{#5\} #6 ; \{ #7 #1 , \}
\text{\}}
\text{\}}
\text{\textbackslash _fp\textunderscore ep\textunderscore isqrt\textunderscore estii:wwnnwn}
\text{\_fp\textunderscore ep\textunderscore isqrt\textunderscore estiii:NNNNNwwwn}
\text{(End of definition for \textbackslash _fp\textunderscore ep\textunderscore isqrt:wwn, \textbackslash _fp\textunderscore ep\textunderscore isqrt\textunderscore aux:wwn, and \textbackslash _fp\textunderscore ep\textunderscore isqrt\textunderscore auxii:wwnnwn.)}

If the last two approximations gave the same result, we are done: call the estii function to clean up. Otherwise, evaluate \((\langle \text{prev} \rangle + 1.005 \cdot 10^8 \text{ or } 9/(\langle \text{prev} \rangle \cdot x))/2\), as the next approximation: omitting the 1.005 factor, this would be Newton’s method. We can check
by brute force that if \( #4 \) is empty (the original exponent was even), the process computes an integer slightly larger than \( 100 / \sqrt{x} \), while if \( #4 \) is 0 (the original exponent was odd), the result is an integer slightly larger than \( 100 / \sqrt{x/10} \). Once we are done, we evaluate \( 100r^2/2 \) or \( 10r^2/2 \) (when the exponent is even or odd, respectively) and feed that to \texttt{estiii}. This third auxiliary finds \( y_{\text{even}}/2 = 10^{-4}r^2x/2 \) or \( y_{\text{odd}}/2 = 10^{-5}r^2x/2 \) (again, depending on earlier parity). A simple program shows that \( y \in [1, 1.0201] \). The number \( y/2 \) is fed to \texttt{\_\_fp_ep_isqrt_estii:wwN}, which computes \( 1/\sqrt{y} \), and we finally multiply the result by \( r \).

\begin{verbatim}
cs_new:Npn \_\_fp_ep_isqrt_estii:wwN #1, #2, #3, #4
\{ \exp_after:wN \_\_fp_ep_isqrt_estii:wwN \int_value:w \__fp_int_eval:w (\#1 + 1 0050 0000 \#4 / (#1 * #3)) / 2 , \#1, #3, {#4} \}
\}
\cs_new:Npn \_\_fp_ep_isqrt_estii:wwNNNNNnnn #1#2#3#4#5#6, 1#7#8; #9;
{ \__fp_fixed_mul_short:wwn #9; {#1} {#2#3#4#5} {#600} ; \__fp_ep_isqrt_epsii:wwN \__fp_fixed_mul_short:wwn {#7} {#80} {0000} ; }
\}
\end{verbatim}

(End of definition for \texttt{\_\_fp_ep_isqrt_estii:wwN}, \texttt{\_\_fp_ep_isqrt_estii:wwN}, and \texttt{\_\_fp_ep_isqrt_estii:NNNNnnn}.)

Here, we receive a fixed point number \( y/2 \) with \( y \in [1, 1.0201] \). Starting from \( z = 1 \) we iterate \( z \mapsto z(3/2 - z^2y/2) \). In fact, we start from the first iteration \( z = 3/2 - y/2 \) to avoid useless multiplications. The \texttt{epsii} auxiliary receives \( z \) as \#1 and \( y \) as \#2.

\begin{verbatim}
cs_new:Npn \_\_fp_ep_isqrt_epsii:wwN #1;
{ \__fp_fixed_sub:wwn {15000}{0000}{0000}{0000}{0000}{0000}; #1; \__fp_isqrt_epsii:wwN \__fp_isqrt_epsii:wwN \__fp_isqrt_epsii:wwN \__fp_isqrt_epsii:wwN \__fp_isqrt_epsii:wwN \}
\}
\cs_new:Npn \_\_fp_ep_isqrt_epsii:wwN #1; 1#2;
{ \__fp_fixed_mul:wwn \__fp_fixed_mul:wwn \__fp_fixed_mul_sub_back:wwn \{15000\}{0000}\}
\}
\end{verbatim}

(End of definition for \texttt{\_\_fp_ep_isqrt_epsii:wwN} and \texttt{\_\_fp_ep_isqrt_epsii:wwN}.)
75.11 Converting from fixed point to floating point

After computing Taylor series, we wish to convert the result from extended precision (with or without an exponent) to the public floating point format. The functions here should be called within an integer expression for the overall exponent of the floating point.

An extended-precision number is simply a comma-delimited exponent followed by a fixed point number. Leave the exponent in the current integer expression then convert the fixed point number.

```
cs_new:Npn \__fp_ep_to_float_o:wwN #1, { \__fp_int_eval:w #1 \__fp_fixed_to_float_o:wN }\__fp_ep_inv_to_float_o:wwN #1,#2; { \__fp_ep_div:wwwwn 1,{1000}{0000}{0000}{0000}{0000}{0000}; #1,#2; \__fp_ep_to_float_o:wwN }\__fp_fixed_inv_to_float_o:wN \__fp_fixed_to_float_rad_o:wN \__fp_fixed_to_float_o:wN \__fp_fixed_to_float_o:Nw ...
```

And the to_fixed version gives six brace groups instead of 4, ensuring that 1000 ≤ \langle a’1⟩ ≤ 9999. At this stage, we know that ⟨a1⟩ is positive (otherwise, it is sign of an error before), and we assume that it is less than 10⁸.¹¹

```
cs_new:Npn \__fp_fixed_to_float_o:wn \__fp_fixed_to_float_o:Nw \__fp_int_eval:w (exponent) \__fp_fixed_to_float_o:wn \__fp_fixed_to_float_o:wn \__fp_fixed_to_float_o:Nw \__fp_fixed_to_float_o:Nw \__fp_fixed_to_float_o:Nw \__fp_int_eval:w \c__fp_block_int \exp_after:wN \exp_after:wN \exp_after:wN \exp_after:wN \exp_after:wN \exp_after:wN
```

¹¹Bruno: I must double check this assumption.
\cs_new:Npn \_fp_fixed_to_loop:N #1
\{ 
\if_meaning:w 0 #1 
- 1 
\exp_after:wN \_fp_fixed_to_loop:N
\else: 
\exp_after:wN \_fp_fixed_to_loop_end:w 
\exp_after:wN \_fp_fixed_to_float_zero:w
\fi:
\}
\cs_new:Npn \_fp_fixed_to_loop_end:w #1 #2 ;
\{ 
\if_meaning:w ; #1 
\exp_after:wN \_fp_pack_twice_four:wNNNNNNNN 
\else: 
\exp_after:wN \_fp_pack_twice_four:wNNNNNNNN 
\exp_after:wN \_fp_fixed_to_float_pack:ww
\exp_after:wN ;
\fi:
#1 #2 0000 0000 0000 0000 ;
\}
\cs_new:Npn \_fp_fixed_to_float_zero:w ; 0000 0000 0000 0000 ;
\{ 
- 2 * \c__fp_max_exponent_int ; 
\{0000} \{0000} \{0000} \{0000} ;
\}
\cs_new:Npn \_fp_fixed_to_float_pack:ww #1 ; #2 #3 #4 ;
\{ 
\if_int_compare:w #2 > 4 \exp_stop_f: 
\exp_after:wN \_fp_fixed_to_float_round_up:wnnnw 
\exp_after:wN \_fp_fixed_to_float_round_up:wnnnw 
\exp_after:wN \_fp_fixed_to_float_round_up:wnnnw 
\exp_after:wN \_fp_fixed_to_float_round_up:wnnnw 
\fi:
\}
\cs_new:Npn \_fp_fixed_to_float_round_up:wnnnw
\{ 
\exp_after:wN \_fp_basics_pack_high:NNNNNw 
\int_value:w \_fp_int_eval:w 1 #1 #2
\exp_after:wN \_fp_basics_pack_low:NNNNNw 
\int_value:w \_fp_int_eval:w 1 #3 #4 + 1 ;
\}
\end{definition_for \_fp_fixed_to_float_o:wN and \_fp_fixed_to_float_o:Nw.}
\end{package}
Chapter 76

l3fp-expo implementation

Unary functions.

\_\_fp_parse_word_exp:N
\_\_fp_parse_word_ln:N
\_\_fp_parse_word_fact:N

(End of definition for \_\_fp_parse_word_exp:N, \_\_fp_parse_word_ln:N, and \_\_fp_parse_word_fact:N.)

76.1 Logarithm

76.1.1 Work plan

As for many other functions, we filter out special cases in \_\_fp_ln_o:w. Then \_\_fp_-ln_npos_o:w receives a positive normal number, which we write in the form \(a \cdot 10^b\) with \(a \in [0,1)\).

The rest of this section is actually not in sync with the code. Or is the code not in sync with the section? In the current code, \(c \in [1,10]\) is such that \(0.7 \leq ac < 1.4\).

We are given a positive normal number, of the form \(a \cdot 10^b\) with \(a \in [0.1,1)\). To compute its logarithm, we find a small integer \(5 \leq c < 50\) such that \(0.91 \leq ac/5 < 1.1\), and use the relation

\[
\ln(a \cdot 10^b) = b \cdot \ln(10) - \ln(c/5) + \ln(ac/5).
\]

The logarithms \(\ln(10)\) and \(\ln(c/5)\) are looked up in a table. The last term is computed using the following Taylor series of \(\ln\) near 1:

\[
\ln\left(\frac{ac}{5}\right) = \ln\left(\frac{1 + t}{1 - t}\right) = 2t \left(1 + t^2 \left(\frac{1}{3} + t^2 \left(\frac{1}{5} + t^2 \left(\frac{1}{7} + \ldots\right)\right)\right)\right)
\]
where \( t = 1 - 10/(ac + 5) \). We can now see one reason for the choice of \( ac \sim 5 \): then \( ac + 5 = 10(1 - \epsilon) \) with \(-0.05 < \epsilon \leq 0.045\), hence

\[
t = \frac{\epsilon}{1-\epsilon} = \epsilon(1 + \epsilon)(1 + \epsilon^2)(1 + \epsilon^4),
\]

is not too difficult to compute.

### 76.1.2 Some constants

A few values of the logarithm as extended fixed point numbers. Those are needed in the implementation. It turns out that we don’t need the value of \( \ln(5) \).

```
\tl_const:Nn \c__fp_ln_i_fixed_tl \{ {0000}{0000}{0000}{0000}{0000}{0000};}
\tl_const:Nn \c__fp_ln_ii_fixed_tl \{ \{6931\} \{4718\} \{0559\} \{9453\} \{0941\} \{7232\};
\tl_const:Nn \c__fp_ln_iii_fixed_tl \{ \{10986\} \{1228\} \{8668\} \{1096\} \{9139\} \{5245\};
\tl_const:Nn \c__fp_ln_iv_fixed_tl \{ \{13862\} \{9436\} \{1119\} \{8906\} \{1883\} \{4464\};
\tl_const:Nn \c__fp_ln_vii_fixed_tl \{ \{17917\} \{5946\} \{9228\} \{0550\} \{0081\} \{2477\};
\tl_const:Nn \c__fp_ln_viii_fixed_tl \{ \{19459\} \{1014\} \{9055\} \{3133\} \{0510\} \{5353\};
\tl_const:Nn \c__fp_ln_ix_fixed_tl \{ \{20794\} \{4154\} \{1679\} \{8359\} \{2825\} \{1696\};
\tl_const:Nn \c__fp_ln_x_fixed_tl \{ \{21972\} \{2457\} \{7336\} \{2193\} \{8279\} \{0490\};
\tl_const:Nn \c__fp_ln_x_fixed_tl \{ \{23025\} \{8509\} \{2994\} \{0456\} \{8401\} \{7991\};
```

(End of definition for \c__fp_ln_i_fixed_tl and others.)

### 76.1.3 Sign, exponent, and special numbers

The logarithm of negative numbers (including \(-\infty\) and \(-0\)) raises the “invalid” exception. The logarithm of \(+0\) is \(-\infty\), raising a division by zero exception. The logarithm of \(+\infty\) or a \texttt{nan} is itself. Positive normal numbers call \_\_fp_ln_npos_o:w.

```
\cs_new:Npn \_\_fp_ln_o:w #1 \s__fp \_\_fp_chk:w #2#3#4; @
{ \if_meaning:w 2 #3
\_\_fp_case_use:nw { \_\_fp_invalid_operation_o:nw { ln } }
\fi:
\if_case:w #2 \exp_stop_f:
\_\_fp_case_use:nw { \_\_fp_division_by_zero_o:Nnw \c_minus_inf_fp { ln } }
\or:
\else:
\_\_fp_case_return_same_o:w
\fi:
\_\_fp_ln_npos_o:w \s__fp \_\_fp_chk:w #2#3#4;
}
```

(End of definition for \_\_fp_ln_o:w.)

### 76.1.4 Absolute ln

We catch the case of a significand very close to 0.1 or to 1. In all other cases, the final result is at least \( 10^{-4} \), and then an error of \( 0.5 \cdot 10^{-20} \) is acceptable.

```
\cs_new:Npn \_\_fp_ln_npos_o:w \s__fp \_\_fp_chk:w #10#1#2#3;
{ \_\_fp_case_use:nw { \_\_fp_infinity_o:w } \_\_fp_case_use:nw { \_\_fp_normalize_o:w } \_\_fp_case_use:nw { \_\_fp_exact_zero_o:w } \_\_fp_case_use:nw { \_\_fp_division_by_zero_o:Nnw \c_minus_inf_fp { ln } } \_\_fp_case_use:nw { \_\_fp_underflow_o:nw { ln } } }
```

1163
This function expands to

\[ Y = -\ln(X) \]

as an extended fixed point.

We have thus found \( c \in [1, 10] \) such that \( 0.7 \leq ac < 1.4 \) in all cases. Compute \( 1 + x = 1 + ac \in [1.7, 2.4] \).
The Taylor series to be used is expressed in terms of $t = (x - 1)/(x + 1) = 1 - 2/(x + 1)$. We now compute the quotient with extended precision, reusing some code from \_\_fp_/o:ww. Note that $1 + x$ is known exactly.

To reuse notations from l3fp-basics, we want to compute $A/Z$ with $A = 2$ and $Z = x + 1$. In l3fp-basics, we considered the case where both $A$ and $Z$ are arbitrary, in the range $[0.1, 1)$, and we had to monitor the growth of the sequence of remainders $A, B, C, \ldots$ to ensure that no overflow occurred during the computation of the next quotient. The main source of risk was our choice to define the quotient as roughly $10^{9} \cdot A/10^{5} \cdot Z$: then $A$ was bound to be below $2 \cdot 10^{147} \ldots$, and this limit was never far.

In our case, we can simply work with $10^{8} \cdot A$ and $10^{4} \cdot Z$, because our reason to work with higher powers has gone: we needed the integer $y \approx 10^{5} \cdot Z$ to be at least $10^{4}$, and now, the definition $y \approx 10^{4} \cdot Z$ suffices.

Let us thus define $y = \left\lfloor \frac{10^{4} \cdot Z}{10^{4}} \right\rfloor + 1 \in (1.7 \cdot 10^{4}, 2.4 \cdot 10^{4}]$, and

$$Q_1 = \left\lfloor \frac{10^{8} \cdot A}{y} - \frac{1}{2} \right\rfloor.$$

(The $1/2$ comes from how $\varepsilon$-\TeX{} rounds.) As for division, it is easy to see that $Q_1 \leq 10^{4}A/Z$, i.e., $Q_1$ is an underestimate.

Exactly as we did for division, we set $B = 10^{8}A - Q_1Z$. Then

$$10^{4}B \leq A_1A_2A_3A_4 - \left( \frac{A_1A_2}{y} - \frac{3}{2} \right) 10^{4}Z$$

$$\leq A_1A_2 \left( 1 - \frac{10^{4}Z}{y} \right) + 1 + \frac{3}{2}y$$

$$\leq 10^{8} \frac{A}{y} + 1 + \frac{3}{2}y$$
In the same way, and using $1.7 \cdot 10^4 \leq y \leq 2.4 \cdot 10^4$, and convexity, we get

\[
10^4 A = 2 \cdot 10^4 \\
10^4 B \leq 10^8 \frac{A}{y} + 1.6y \leq 4.7 \cdot 10^4 \\
10^4 C \leq 10^8 \frac{B}{y} + 1.6y \leq 5.8 \cdot 10^4 \\
10^4 D \leq 10^8 \frac{C}{y} + 1.6y \leq 6.3 \cdot 10^4 \\
10^4 E \leq 10^8 \frac{D}{y} + 1.6y \leq 6.5 \cdot 10^4 \\
10^4 F \leq 10^8 \frac{E}{y} + 1.6y \leq 6.6 \cdot 10^4
\]

Note that we compute more steps than for division: since $t$ is not the end result, we need to know it with more accuracy (on the other hand, the ending is much simpler, as we don’t need an exact rounding for transcendental functions, but just a faithful rounding).

\[
\text{The number is } x. \text{ Compute } y \text{ by adding 1 to the five first digits.}
\]

\[
\cs_new:Npn \__fp_ln_x_iv:wnnnnnnnn \{ 1 \text{ or } 2 \} \{ 8d \} ; \{ \langle 4d \rangle \} \{ \langle 4d \rangle \} \{ \text{fixed-tl} \}
\]

\[
\text{\exp_after:wN } \__fp_div_significand_pack:NNN \int_value:w \__fp_int_eval:w
\text{\int_value:w } 999999 + 200000000 \div \text{#1} \; \% \text{ Q1}
\]

\[
\cs_new:Npn \__fp_ln_div_i:w \text{#1; #2} \{ #3 \} \{ #4 \} \{ #5 \}
\]

\[
\cs_new:Npn \__fp_ln_div_ii:wwn \text{#1; #2; #3} \{ \langle 4d \rangle \} \{ \langle 4d \rangle \}
\]

\[
\cs_new:Npn \__fp_ln_div_vi:wwn \text{#1; #2; #3; #4; #5; #6; #7; #8; #9}
\]

\[
\text{\exp_after:wN } \__fp_div_significand_pack:NNN \int_value:w \__fp_int_eval:w
\text{\int_value:w } 999999 + \text{#2} \div \text{#3} \; \% \text{ Q2}
\]

\[
\cs_new:Npn \__fp_ln_div_vii:w \text{#1; #2; #3} \{ \langle 4d \rangle \} \{ \langle 4d \rangle \}
\]

\[
\cs_new:Npn \__fp_div_significand_pack:NNN
\]

1166
We now have essentially

\[ \text{\_\_fp\_ln\_div\_after:Nw \langle fixed\ tl \rangle } \]
\[ \text{\_\_fp\_div\_significand\_pack:NNN \ 10^6 + Q_1 } \]
\[ \text{\_\_fp\_div\_significand\_pack:NNN \ 10^6 + Q_2 } \]
\[ \text{\_\_fp\_div\_significand\_pack:NNN \ 10^6 + Q_3 } \]
\[ \text{\_\_fp\_div\_significand\_pack:NNN \ 10^6 + Q_4 } \]
\[ \text{\_\_fp\_div\_significand\_pack:NNN \ 10^6 + Q_5 } \]
\[ \text{\_\_fp\_div\_significand\_pack:NNN \ 10^6 + Q_6 } ; \]
\[ \langle exponent \rangle ; \langle \text{continuation} \rangle \]

where \( \langle fixed\ tl \rangle \) holds the logarithm of a number in \([1, 10]\), and \( \langle exponent \rangle \) is the exponent. Also, the expansion is done backwards. Then \( \text{\_\_fp\_div\_significand\_pack:NNN} \) puts things in the correct order to add the \( Q_i \) together and put semicolons between each piece. Once those have been expanded, we get

\[ \text{\_\_fp\_ln\_div\_after:Nw \langle fixed-tl \rangle \langle 1d \rangle ; \langle 4d \rangle ; \langle 4d \rangle ; \langle 4d \rangle ; \langle 4d \rangle ; \langle 4d \rangle ; \langle 4d \rangle ; \langle exponent \rangle ; \langle \text{continuation} \rangle } \]

Just as with division, we know that the first two digits are 1 and 0 because of bounds on the final result of the division \( 2/(x+1) \), which is between roughly 0.8 and 1.2. We then compute \( 1 - 2/(x+1) \), after testing whether \( 2/(x+1) \) is greater than or smaller than 1.

Where \( \langle fixed\ tl \rangle \) holds the logarithm of a number in \([1, 10]\), and \( \langle exponent \rangle \) is the exponent. Also, the expansion is done backwards. Then \( \text{\_\_fp\_div\_significand\_pack:NNN} \) puts things in the correct order to add the \( Q_i \) together and put semicolons between each piece. Once those have been expanded, we get

\[ \text{\_\_fp\_ln\_div\_after:Nw \langle fixed-tl \rangle \langle 1d \rangle ; \langle 4d \rangle ; \langle 4d \rangle ; \langle 4d \rangle ; \langle 4d \rangle ; \langle 4d \rangle ; \langle 4d \rangle ; \langle exponent \rangle ; \langle \text{continuation} \rangle } \]

Just as with division, we know that the first two digits are 1 and 0 because of bounds on the final result of the division \( 2/(x+1) \), which is between roughly 0.8 and 1.2. We then compute \( 1 - 2/(x+1) \), after testing whether \( 2/(x+1) \) is greater than or smaller than 1.

Compute the square \( t^2 \), and keep \( t \) at the end with its sign. We know that \( t < 0.1765 \), so every piece has at most 4 digits. However, since we were not careful in \( \text{\_\_fp\_ln\_t\_small:w} \), they can have less than 4 digits.
\texttt{\cs_new:Npn \__fp\_ln\_t\_large:NNw \#1 \#2 \#3; \#4; \#5; \#6; \#7; \#8;}
{\exp_after:wN \__fp\_ln\_square\_t\_after:w}
\texttt{\int_value:w \__fp\_int\_eval:w 9999 0000 + \#3*\#3}
\exp_after:wN \__fp\_ln\_square\_t\_pack:NNNNNw
\texttt{\int_value:w \__fp\_int\_eval:w 9999 0000 + 2*\#3*\#4}
\exp_after:wN \__fp\_ln\_square\_t\_pack:NNNNNw
\texttt{\int_value:w \__fp\_int\_eval:w 9999 0000 + 2*\#3*\#5 + \#4*\#4}
\exp_after:wN \__fp\_ln\_square\_t\_pack:NNNNNw
\texttt{\int_value:w \__fp\_int\_eval:w 9999 0000 + 2*\#3*\#6 + 2*\#4*\#5}
\exp_after:wN \__fp\_ln\_square\_t\_pack:NNNNNw
\texttt{\int_value:w \__fp\_int\_eval:w 9999 0000 + 2*\#3*\#7 + 2*\#4*\#6 + \#5*\#5}
+ (2*\#3*\#8 + 2*\#4*\#7 + \#5*\#6) / 1 0000
% ; ;
\exp_after:wN \__fp\_ln\_twice\_t\_after:w
\int_value:w \__fp\_int\_eval:w -1 + 2*\#3
\exp_after:wN \__fp\_ln\_twice\_t\_pack:Nw
\int_value:w \__fp\_int\_eval:w 9999 + 2*\#4
\exp_after:wN \__fp\_ln\_twice\_t\_pack:Nw
\int_value:w \__fp\_int\_eval:w 9999 + 2*\#5
\exp_after:wN \__fp\_ln\_twice\_t\_pack:Nw
\int_value:w \__fp\_int\_eval:w 9999 + 2*\#6
\exp_after:wN \__fp\_ln\_twice\_t\_pack:Nw
\int_value:w \__fp\_int\_eval:w 9999 + 2*\#7
\exp_after:wN \__fp\_ln\_twice\_t\_pack:Nw
\int_value:w \__fp\_int\_eval:w 10000 + 2*\#8 ; ;
\texttt{\__fp\_ln\_c:NwNw \#1}
\#2
\}

\texttt{(End of definition for \_\_fp\_ln\_x\_ii:wnnn.)}

\_\_fp\_ln\_Taylor:wwNw Denoting $T = t^2$, we get

\begin{verbatim}
\texttt{\_fp\_ln\_Taylor:wwNw}
\texttt{\{\langle T_1\rangle \langle T_3\rangle \langle T_5\rangle \langle T_7\rangle \langle T_9\rangle \}}
\texttt{\};
\texttt{\{\langle (2t_1)\rangle \langle (2t_3)\rangle \langle (2t_5)\rangle \langle (2t_7)\rangle \langle (2t_9)\rangle \}}
\texttt{\};
\texttt{\} \_\_fp\_ln\_c:NwNw (sign)}
\texttt{\(\text{(fixed r1)} \langle \text{exponent} \rangle \}; \langle \text{continuation} \rangle}
\end{verbatim}

And we want to compute

\[
\ln \left( \frac{1 + t}{1 - t} \right) = 2t \left( 1 + T \left( \frac{1}{3} + T \left( \frac{1}{5} + T \left( \frac{1}{7} + T \left( \frac{1}{9} + \cdots \right) \right) \right) \right) \right)
\]

The process looks as follows
\loop 5; A;
\div_int 5; 1.0; \add A; \mul T; \{\loop \eval 5-2;\}
\add 0.2; A; \mul T; \{\loop \eval 5-2;\}
\mul B; T; \{\loop 3;\}
\loop 3; C;

This uses the routine for dividing a number by a small integer ( < 10^4).
\cs_new:Npn \__fp_ln_Taylor:wwNw
\cs_new:Npn \__fp_ln_Taylor_loop:www #1; {0000}{0000}{0000}{0000} ; }
\cs_new:Npn \__fp_ln_Taylor_loop:www #1; #2; #3;
\__fp_fixed_div_int:wwN \c__fp_one_fixed_tl #1;
\__fp_fixed_add:wwn #2;
\__fp_fixed_mul:wwn #3;
\__fp_ln_Taylor_break:w #1 #2 ;
\__fp_ln_Taylor_break:w \fi: #1 \__fp_fixed_add:wwn #2#3 ;
\__fp_ln_Taylor_break:w \exp_after:wN \__fp_fixed_sub:wwn #3 #4 ;
\__fp_ln_Taylor_break:w
\__fp_ln_c:NwNw \__fp_ln_c:NwNw \__fp_ln_c:NwNw
\{ (sign) \} \{ (r1) \} \{ (r2) \} \{ (r3) \} \{ (r4) \} \{ (r5) \} ;
\{ (fixed tl) \} \{ (exponent) \} ; \{ (continuation) \}

We are now reduced to finding $\ln(c)$ and $(\text{exponent}) \ln(10)$ in a table, and adding it to the mixture. The first step is to get $\ln(c) - \ln(x) = -\ln(a)$, then we get $b \ln(10)$ and add or subtract.

For now, $\ln(x)$ is given as $\cdot 10^9$. Unless both the exponent is 1 and $c = 1$, we shift to working in units of $\cdot 10^4$, since the final result is at least $\ln(10/7) \approx 0.35$.
\cs_new:Npn \__fp_ln_c:NwNw #1 #2; #3
\if_meaning:w + #1
\exp_after:wN \__fp_fixed_sub:wwn \exp_after:wN \__fp_fixed_add:wwn
\else:
\exp_after:wN \__fp_fixed_add:wwn \exp_after:wN \__fp_fixed_sub:wwn
\fi:
\__fp_ln_c:NwNw #3 #2 ;
(End of definition for \__fp_ln_c:NwNw.)
Compute \(\text{(exponent)}\) times \(\ln(10)\). Apart from the cases where \(\text{(exponent)}\) is 0 or 1, the result is necessarily at least \(\ln(10) \approx 2.3\) in magnitude. We can thus drop the least significant 4 digits. In the case of a very large (positive or negative) exponent, we can (and we need to) drop 4 additional digits, since the result is of order \(10^4\). Naively, one would think that in both cases we can drop 4 more digits than we do, but that would be slightly too tight for rounding to happen correctly. Besides, we already have addition and subtraction for 24 digits fixed point numbers.

Now we painfully write all the cases.\(^{12}\) No overflow nor underflow can happen, except when computing \(\ln(1)\).

For small exponents, we just drop one block of digits, and set the exponent of the log to 4 (minus any shift coming from leading zeros in the conversion from fixed point to floating point). Note that here the exponent has been made positive.

\(^{12}\)Bruno: do rounding.
76.2 Exponential

76.2.1 Sign, exponent, and special numbers

\_fp_exp_o:w

\cs_new:Npn \_fp_exp_o:w #1 \s__fp \_fp_chk:w #2#3#4; @
{\if_case:w #2 \exp_stop_f:
  \_fp_case_return_o:Nw \c_one_fp
\or:
  \exp_after:wN \_fp_exp_normal_o:w
\or:
  \if_meaning:w 0 #3
    \exp_after:wN \_fp_case_return_o:Nw \c_inf_fp
  \else:
    \exp_after:wN \_fp_case_return_o:Nw \c_zero_fp
  \fi:
\or:
  \_fp_case_return_same_o:w
\fi:
\s__fp \_fp_chk:w #2#3#4; }

(End of definition for \_fp_exp_o:w.)

\_fp_exp_normal_o:w
\_fp_exp_pos_o:Nnwnw
\_fp_exp_overflow:NN

\cs_new:Npn \_fp_exp_normal_o:w \s__fp \_fp_chk:w 1#1 {\if_meaning:w 0 #1
{\if_int_compare:w #4 > \c__fp_max_exp_exponent_int
  \token_if_eq_charcode:NNTF + #1
  \_fp_exp_overflow:NN \_fp_overflow:w \c_inf_fp
\else:
  \_fp_exp_overflow:NN \_fp_underflow:w \c_zero_fp
\exp:w
\fi:
\fi:
\if_int_compare:w #4 > \c__fp_max_exp_exponent_int
\token_if_eq_charcode:NNTF + #1
{ \_fp_exp_overflow:NN \_fp_overflow:w \c_inf_fp }
\else:
\_fp_sanitize:Nw
0 \int_value:w #1 \__fp_int_eval:w
\if_int_compare:w #4 < \c_zero_int
\use_i:nn
\else:

1171
This function is called for numbers in the range \(10^{-9}, 10^{-1}\). We compute 10 terms of the Taylor series. The first argument is irrelevant (rounding digit used by some other functions). The next three arguments, at least 16 digits, delimited by a semicolon, form a fixed point number, so we pack it in blocks of 4 digits.
The integer array has \(6 \times 9 \times 4 = 216\) items encoding the values of \(\exp(j \times 10^i)\) for \(j = 1, \ldots, 9\) and \(i = -1, \ldots, 4\). Each value is expressed as \(\simeq 10^p \times m_1 \times m_2 \times m_3\) with three 8-digit blocks \(m_1, m_2, m_3\) and an integer exponent \(p\) (one more than the scientific exponent), and these are stored in the integer array as four items: \(p, 10^8 + m_1, 10^8 + m_2, 10^8 + m_3\). The various exponentials are stored in increasing order of \(j \times 10^i\).

Storing this data in an integer array makes it slightly harder to access (slower, too), but uses 16 bytes of memory per exponential stored, while storing as tokens used around 40 tokens; tokens have an especially large footprint in Unicode-aware engines.
The first two arguments are irrelevant (a rounding digit, and a brace group with 8 zeros). The third argument is the integer part of our number, then we have the decimal part delimited by a semicolon, and finally the exponent, in the range $[0, 5]$. Remove leading zeros from the integer part: putting #4 in there too ensures that an integer part of 0 is also removed. Then read digits one by one, looking up $\exp(\langle \text{digit} \rangle \cdot 10^{\langle \text{exponent} \rangle})$ in a table, and multiplying that to the current total. The loop is done by \_fp_exp_large:NwN, whose #1 is the $\langle \text{exponent} \rangle$, #2 is the current mantissa, and #3 is the $\langle \text{digit} \rangle$. At the end, \_fp_exp_large_after:wwn moves on to the Taylor series, eventually multiplied with the mantissa that we have just computed.
76.3 Power

Raising a number \( a \) to a power \( b \) leads to many distinct situations.

\[
\begin{array}{cccccccc}
  a^b & -
  \infty & (\infty, -0) & \text{-integer} & \pm 0 & \text{+integer} & (0, \infty) & +\infty & \text{nan} \\
  \infty & +0 & +0 & +1 & +\infty & +\infty & \text{nan} \\
  (1, \infty) & +0 & +|a|^b & +1 & +|a|^b & +\infty & \text{nan} \\
  +1 & +1 & +1 & +1 & +1 & +1 & +1 \\
  (0, 1) & +\infty & +|a|^b & +1 & +|a|^b & +0 & \text{nan} \\
  +0 & +\infty & +\infty & +1 & +0 & +0 & \text{nan} \\
  -0 & +\infty & \text{nan} & (-1)^b\infty & +1 & (-1)^b0 & +0 & +0 & \text{nan} \\
  (-1, 0) & +\infty & \text{nan} & (-1)^b|a|^b & +1 & (-1)^b|a|^b & \text{nan} & +0 & \text{nan} \\
  -1 & +\infty & \text{nan} & (-1)^b & +1 & (-1)^b & \text{nan} & +1 & \text{nan} \\
  (-\infty, -1) & +\infty & \text{nan} & (-1)^b|a|^b & +1 & (-1)^b|a|^b & \text{nan} & +\infty & \text{nan} \\
  -\infty & +\infty & \text{nan} & (-1)^b0 & +1 & (-1)^b & \text{nan} & +\infty & \text{nan} \\
  \text{nan} & \text{nan} & \text{nan} & \text{nan} & \text{nan} & \text{nan} & \text{nan} & \text{nan} & \text{nan}
\end{array}
\]

We distinguished in this table the cases of finite (positive or negative) integer exponents, as \((-1)^b\) is defined in that case. One peculiarity of this operation is that \(\text{nan}^0 = 1\text{nan} = 1\), because this relation is obeyed for any number, even \(\pm \infty\).

We cram most of the tests into a single function to save csnames. First treat the case \(b = 0\): \(a^0 = 1\) for any \(a\), even \text{nan}. Then test the sign of \(a\).
• If it is positive, and \( a \) is a normal number, call \( \texttt{\_fp\_pow\_normal\_o:ww} \) followed by the two \( \texttt{fp a} \) and \( b \). For \( a = +0 \) or \(+\infty\), call \( \texttt{\_fp\_pow\_zero\_or\_inf:ww} \) instead, to return either \(+0\) or \(+\infty\) as appropriate.

• If \( a \) is a \texttt{nan}, then skip to the next semicolon (which happens to be conveniently the end of \( b \)) and return \texttt{nan}.

• Finally, if \( a \) is negative, compute \( a^b \) (\( \texttt{\_fp\_pow\_normal\_o:ww} \) which ignores the sign of its first operand), and keep an extra copy of \( a \) and \( b \) (the second brace group, containing \( \{ b \ a \} \), is inserted between \( a \) and \( b \)). Then do some tests to find the final sign of the result if it exists.

\[
\begin{align*}
\cs_new:cpn \{ \_fp_ \char:N \w ^ _o:ww \\
\_fp_{ } \_fp\_chk:w \#1\#2\#3; \_s_{ } \_fp \_fp\_chk:w \#4\#5\#6; \\
\{ \end{align*}
\]

\( \texttt{\_fp\_pow\_zero\_or\_inf:ww} \) Raising \(-0\) or \(-\infty\) to \texttt{nan} yields \texttt{nan}. For other powers, the result is \(+0\) if \( 0 \) is raised to a positive power or \( +\infty \) to a negative power, and \(+\infty\) otherwise. Thus, if the type of \( a \) and the sign of \( b \) coincide, the result is \( 0 \), since those conveniently take the same possible values, \( 0 \) and \( 2 \). Otherwise, either \( a = \pm\infty \) and \( b > 0 \) and the result is \(+\infty\), or \( a = \pm0 \) with \( b < 0 \) and we have a division by zero unless \( b = -\infty \).

\[
\begin{align*}
\cs_new:cpn \_fp_{ } \_fp\_pow\_zero\_or\_inf:w w \\
\{ \end{align*}
\]

(End of definition for \( \texttt{\_fp\_\_o:ww} \)).
We have in front of us $a$, and $b \neq 0$, we know that $a$ is a normal number, and we wish to compute $|a|^b$. If $|a| = 1$, we return 1, unless $a = -1$ and $b$ is nan. Indeed, returning 1 at this point would wrongly raise "invalid" when the sign is considered. If $|a| \neq 1$, test the type of $b$:

0 Impossible, we already filtered $b = \pm 0$.
1 Call \__fp_pow_npos_o:ww.
2 Return $+\infty$ or $+0$ depending on the sign of $b$ and whether the exponent of $a$ is positive or not.
3 Return $b$.

\cs_new:Npn \__fp_pow_normal_o:ww
\s__fp \__fp_chk:w 1#1#2#3; \s__fp \__fp_chk:w #4#5
{ \if:w 0 \__fp_str_if_eq:nn { #2 #3 } { 1 {1000} {0000} {0000} {0000} } \if_int_compare:w #4 #1 = 32 \exp_stop_f:
\exp_after:wN \__fp_case_return_ii_o:ww
\exp_after:wN #5
\else:
\__fp_case_return_ii_o:ww
\fi:
\fi:
\if_case:w #4 \exp_stop_f:
\exp_after:wN \__fp_pow_npos_o:ww
\exp_after:wN \#5
\else:
\exp_after:wN \__fp_pow_npos_o:ww
\exp_after:wN \c_zero_fp
\fi:
\or:
\exp_after:wN \__fp_pow_npos_o:ww
\exp_after:wN \#5
\or:
\if_meaning:w 2 #5 \exp_after:wN \reverse_if:N \fi:
\if_int_compare:w #2 > \c_zero_int
\exp_after:wN \__fp_case_return_ii_o:ww
\exp_after:wN \c_inf_fp
\else:
\exp_after:wN \__fp_case_return_ii_o:ww
\exp_after:wN \c_zero_fp
\fi:
\or:
\__fp_case_return_ii_o:ww
We now know that $a \neq \pm 1$ is a normal number, and $b$ is a normal number too. We want to compute $|a|^b = (|x| \cdot 10^n)^y \cdot 10^p = \exp((\ln|x| + n \ln(10)) \cdot y \cdot 10^p) = \exp(z)$. To compute the exponential accurately, we need to know the digits of $z$ up to the 16-th position. Since the exponential of $10^5$ is infinite, we only need at most 21 digits, hence the fixed point result of $\_\_fp_ln_o:w$ is precise enough for our needs. Start an integer expression for the decimal exponent of $e^{|z|}$. If $z$ is negative, negate that decimal exponent, and prepare to take the inverse when converting from the fixed point to the floating point result.

The first argument is the conversion function from fixed point to float. Then comes an exponent and the 4 brace groups of $x$, followed by $b$. Compute $-\ln(x)$.

(End of definition for $\_\_fp_pow_normal_o:ww$.)
\cs_new:Npn \__fp_pow_exponent:Nnnnnnw #1#2; #3#4#5#6#7#8;
\begin{verbatim}
{ %^^A todo: use that in ln.
\exp_after:wN \__fp_fixed_mul_after:wnn
\int_value:w \c__fp_leading_shift_int
\exp_after:wN \__fp_pack:NNNNNw
\int_value:w \c__fp_middle_shift_int
#1#2*23025 - #1 #3
\exp_after:wN \__fp_pack:NNNNNw
\int_value:w \c__fp_middle_shift_int
#1 #2*8509 - #1 #4
\exp_after:wN \__fp_pack:NNNNNw
\int_value:w \c__fp_middle_shift_int
#1 #2*2994 - #1 #5
\exp_after:wN \__fp_pack:NNNNNw
\int_value:w \c__fp_middle_shift_int
#1 #2*0456 - #1 #6
\exp_after:wN \__fp_pack:NNNNNw
\int_value:w \c__fp_trailing_shift_int
#1 #2*8401 - #1 #7
#1 ( #2*7991 - #8 ) / 1 0000 ; ;
}\end{verbatim}
\cs_new:Npn \__fp_pow_B:wwN #1#2#3#4#5#6; #7;
\begin{verbatim}
\if_int_compare:w #7 < \c_zero_int
\exp_after:wN \__fp_pow_C_neg:w \int_value:w -
\else:
\if_int_compare:w #7 < 22 \exp_stop_f:
\exp_after:wN \__fp_pow_C_pos:w \int_value:w
\else:
\exp_after:wN \__fp_pow_C_overflow:w \int_value:w
\fi:
#7 \exp_after:wN ;
\int_value:w \c__fp_int_eval:w 10 0000 + #1 \__fp_int_eval_end:
#2#3#4#5#6 0000 0000 0000 0000 0000 0000 ; %^^A todo: how many 0?
\end{verbatim}
\cs_new:Npn \__fp_pow_C_overflow:w #1; #2; #3
\begin{verbatim}
{ 2 * \c__fp_max_exponent_int
\exp_after:wN \__fp_fixed_continue:wn \c__fp_one_fixed_tl
\end{verbatim}
\cs_new:Npn \__fp_pow_C_neg:w #1 ; 1
\begin{verbatim}
{ \__fp_pow_C_pos_loop:wN #1; }
\end{verbatim}
\cs_new:Npn \__fp_pow_C_overw #1; #2; #3
\begin{verbatim}
\exp_after:wN \exp_after:wN \exp_after:wN \__fp_pow_C_pack:w
\prg_replicate:nnn \c#1 \{0\}
\end{verbatim}
\cs_new:Npn \__fp_pow_C_pack:w
\begin{verbatim}
\begin{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}

1179
This function is followed by three floating point numbers: \( a^b, a \in [-\infty, -0], \) and \( b. \) If \( b \) is an even integer (case \( -1 \)), \( a^b = a^b. \) If \( b \) is an odd integer (case \( 0 \)), \( a^b = -a^b, \) obtained by a call to \( \_fp_pow_neg_aux:www. \) Otherwise, the sign is undefined. This is invalid, unless \( a^b \) turns out to be \( +0 \) or \( \text{nan}, \) in which case we return that as \( a^b. \) In particular, since the underflow detection occurs before \( \_fp_pow_neg:www \) is called, \((-0.1)^{12345.67}\) gives \( +0 \) rather than complaining that the sign is not defined.

This function expects a floating point number, and determines its “parity”. It should be used after \( \text{if_case:w} \) or in an integer expression. It gives \(-1\) if the number is an even integer, \( 0 \) if the number is an odd integer, and \( 1 \) otherwise. Zeros and \( \pm \infty \) are even
(because very large finite floating points are even), while \texttt{nan} is a non-integer. The sign of normal numbers is irrelevant to parity. After \texttt{__fp_decimate:nNnnnn} the argument \texttt{#1} of \texttt{__fp_pow_neg_case_aux:Nnnw} is a rounding digit, 0 if and only if the number was an integer, and \texttt{#3} is the 8 least significant digits of that integer.

\begin{verbatim}
\cs_new:Npn \__fp_pow_neg_case:w \s__fp \__fp_chk:w #1#2#3; {
 \if_case:w #1 \exp_stop_f:
 \or: \__fp_pow_neg_case_aux:nnnnn #3
 \or: -1 \else: 1 \fi:
 \exp_stop_f:
}
\cs_new:Npn \__fp_pow_neg_case_aux:nnnnn #1#2#3#4#5 {
 \if_int_compare:w #1 > \c__fp_prec_int
 -1 \else:
 \__fp_decimate:nNnnnn { \c__fp_prec_int - #1 } \__fp_pow_neg_case_aux:Nnnw
 \{#2\} \{#3\} \{#4\} \{#5\} \fi:
}
\cs_new:Npn \__fp_pow_neg_case_aux:Nnnw #1#2#3#4 ; {
 \if_meaning:w 0 #1
 \if_int_odd:w #3 \exp_stop_f:
 0 \else:
 1 \fi:
 \else:
 1 \fi:
}
\end{verbatim}

(End of definition for \texttt{\__fp_pow_neg_case:w}, \texttt{\__fp_pow_neg_case_aux:nnnnn}, and \texttt{\__fp_pow_neg_-

\texttt{\_fp_case_w}}.)

76.4 Factorial

\texttt{\c__fp_fact_max_arg_int} The maximum integer whose factorial fits in the exponent range is 3248, as 3249! \sim 10^{10000}.

\begin{verbatim}
\int_const:Nn \c__fp_fact_max_arg_int { 3248 }
\end{verbatim}

(End of definition for \texttt{\c__fp_fact_max_arg_int}.)

\texttt{\_fp_fact_o:w} First detect \pm 0 and \pm \infty and \texttt{nan}. Then note that factorial of anything with a negative
sign (except \texttt{-0}) is undefined. Then call \texttt{\_fp_small_int:wTF} to get an integer as the argument, and start a loop. This is not the most efficient way of computing the factorial, but it works all right. Of course we work with 24 digits instead of 16. It is easy to check that computing factorials with this precision is enough.
\cs_new:Npn \__fp_fact_o:w #1 \s__fp \__fp_chk:w #2 #3 #4 ; @
{ 
\if_case:w #2 \exp_stop_f:
  \__fp_case_return_o:Nw \c_one_fp
\or:
  \or:
  \if_meaning:w 0 #3
    \exp_after:wN \__fp_case_return_same_o:w
  \fi:
\or:
  \__fp_case_return_same_o:w
\fi:
\if_meaning:w 2 #3
  \__fp_case_use:nw { \__fp_invalid_operation_o:fw { fact } }
\fi:
\__fp_fact_pos_o:w
\s__fp \__fp_chk:w #2 #3 #4 ; }

(End of definition for \__fp_fact_o:w.)

\__fp_fact_pos_o:w Then check the input is an integer, and call \__fp_facorial_int_o:n with that int as an argument. If it’s too big the factorial overflows. Otherwise call \__fp_sanitize:Nw with a positive sign marker 0 and an integer expression that will mop up any exponent in the calculation.

\cs_new:Npn \__fp_fact_pos_o:w #1; 
{ 
\__fp_small_int:wTF #1; 
\{ \__fp_fact_int_o:n }
\} 
\cs_new:Npn \__fp_fact_int_o:n #1 
{ 
\if_int_compare:w #1 > \c__fp_fact_max_arg_int 
  \__fp_case_return:nw
  \exp_after:wN \exp_after:wN \exp_after:wN \__fp_overflow:w
  \exp_after:wN \c_inf_fp
\fi:
\exp_after:wN \__fp_sanitize:Nw
\exp_after:wN 0
\int_value:w \__fp_int_eval:w
\__fp_fact_loop_o:w #1 . 4 , \{ \} { } { } { } { } { } ; }

(End of definition for \__fp_fact_pos_o:w and \__fp_fact_int_o:w.)

\__fp_fact_loop_o:w The loop receives an integer \#1 whose factorial we want to compute, which we progressively decrement, and the result so far as an extended-precision number \#2 in the form \langle exponent\rangle,\langle mantissa\rangle;. The loop goes in steps of two because we compute \#1*\#1-1 as an integer expression (it must fit since \#1 is at most 3248), then multiply with the result so far. We don’t need to fill in most of the mantissa with zeros because \__fp_ep_mul:wwwwn first normalizes the extended precision number to avoid loss of precision.
When reaching a small enough number simply use a table of factorials less than $10^8$. This limit is chosen because the normalization step cannot deal with larger integers.

\begin{verbatim}
cs_new:Npn \_fp_fact_loop_o:w #1 . #2 ;
    {\if_int_compare:w #1 < 12 \exp_stop_f:
        \_fp_fact_small_o:w #1 \fi:
        \exp_after:wN \_fp_ep_mul:wwww
        \exp_after:wN #2 ,
        \exp_after:wN \_fp_fact_loop_o:w
        \int_value:w \_fp_int_eval:w #1 * (#1 - 1) ;}
    \exp_after:wN \_fp_fact_loop_o:w
    \int_value:w \_fp_int_eval:w #1 . #2 ; \__fp_to_float_o:wwN 0}
\end{verbatim}

(End of definition for \_fp_fact_loop_o:)
Chapter 77

l3fp-trig implementation

 Unary functions.

\tl_map_inline:nn

\cs_new:cpe { __fp_parse_word_#1:N }

\exp_not:N \__fp_parse_unary_function:NNN

\exp_not:c { __fp_#1_o:w }\use_i:nn

\cs_new:cpe { __fp_parse_word_#1d:N }

\exp_not:N \__fp_parse_function:NNN \__fp_acot_o:Nw \use_ii:nn

(End of definition for \__fp_parse_word_acos:N and others.)

Those functions may receive a variable number of arguments.

\cs_new:Npn \__fp_parse_word_acot:N

\cs_new:Npn \__fp_parse_word_atan:N

\cs_new:Npn \__fp_parse_word_atand:N

(End of definition for \__fp_parse_word_acot:N and others.)
77.1 Direct trigonometric functions

The approach for all trigonometric functions (sine, cosine, tangent, cotangent, cosecant, and secant), with arguments given in radians or in degrees, is the same.

- Filter out special cases ($\pm 0$, $\pm \infty$ and nan).
- Keep the sign for later, and work with the absolute value $|x|$ of the argument.
- Small numbers ($|x| < 1$ in radians, $|x| < 10$ in degrees) are converted to fixed point numbers (and to radians if $|x|$ is in degrees).
- For larger numbers, we need argument reduction. Subtract a multiple of $\pi/2$ (in degrees, 90) to bring the number to the range to $[0, \pi/2)$ (in degrees, $[0, 90)$).
- Reduce further to $[0, \pi/4)$ (in degrees, $[0, 45)$) using $\sin x = \cos(\pi/2 - x)$, and when working in degrees, convert to radians.
- Use the appropriate power series depending on the octant $\lfloor x \pi/4 \rfloor \mod 8$ (in degrees, the same formula with $\pi/4 \to 45$), the sign, and the function to compute.

77.1.1 Filtering special cases

\_fp_sin_o:w

This function, and its analogs for cos, csc, sec, tan, and cot instead of sin, are followed either by \use_i:nn and a float in radians or by \use_ii:nn and a float in degrees. The sine of $\pm 0$ or nan is the same float. The sine of $\pm \infty$ raises an invalid operation exception with the appropriate function name. Otherwise, call the trig function to perform argument reduction and if necessary convert the reduced argument to radians. Then, \_fp_sin_series_o:NNwwww is called to compute the Taylor series: this function receives a sign #3, an initial octant of 0, and the function \_fp_ep_to_float_o:wwN which converts the result of the series to a floating point directly rather than taking its inverse, since $\sin(x) = #3 \sin|x|$.

\_fp_cos_o:w

The cosine of $\pm 0$ is 1. The cosine of $\pm \infty$ raises an invalid operation exception. The cosine of nan is itself. Otherwise, the trig function reduces the argument to at most half a right-angle and converts if necessary to radians. We then call the same series as
for sine, but using a positive sign \(0\) regardless of the sign of \(x\), and with an initial octant of 2, because \(\cos(x) = + \sin(\pi/2 + |x|)\).

\[
\begin{align*}
\text{\texttt{\_\_fp_cos_o:w}} & \quad \text{\texttt{\_\_fp_case_return_o:Nw \_\_fp_case_use:nuw}} \\
\{ & \quad \{ \text{\_\_fp_trig:NNNNWm} \#1 \_\_fp_sin_series_o:NNwwww \\
& \quad \_\_fp_ep_to_float_o:wwN 0 2 \\
\} & \quad \} \\
\text{\_\_fp_case_use:nuw} & \quad \{ \_\_fp_invalid_operation_o:fw \{ \#1 \{ \cos \} \{ \cosd \} \} \} \\
\text{\texttt{\_\_fp_case_return_same_o:w}} & \quad \} \\
\text{\texttt{\_\_fp_chk:w \#2 \#3;}} & \quad \\
\end{align*}
\]

(End of definition for \texttt{\_\_fp_cos_o:w})

\[
\text{\texttt{\_\_fp_csc_o:w}}
\]

The cosecant of ±0 is ±\(\infty\) with the same sign, with a division by zero exception (see \texttt{\_\_fp_cot_zero_o:Nfw} defined below), which requires the function name. The cosecant of ±\(\infty\) raises an invalid operation exception. The cosecant of \texttt{nan} is itself. Otherwise, the \texttt{trig} function performs the argument reduction, and converts if necessary to radians before calling the same series as for sine, using the sign \#3, a starting octant of 0, and inverting during the conversion from the fixed point sine to the floating point result, because \(\csc(x) = \#3(\sin|x|)^{-1}\).

\[
\begin{align*}
\text{\texttt{\_\_fp_sec_o:w}} & \quad \text{\texttt{\_\_fp_case_return_o:Nw \_\_fp_case_use:nuw}} \\
\{ & \quad \{ \text{\_\_fp_trig:NNNNWm} \#1 \_\_fp_sin_series_o:NNwwww \\
& \quad \_\_fp_ep_inv_to_float_o:wwN \#3 0 \\
\} & \quad \} \\
\text{\_\_fp_case_use:nuw} & \quad \{ \_\_fp_invalid_operation_o:fw \{ \#1 \{ \csc \} \{ \cscd \} \} \} \\
\text{\texttt{\_\_fp_case_return_same_o:w}} & \quad \} \\
\text{\texttt{\_\_fp_chk:w \#2 \#3 \#4;}} & \quad \\
\end{align*}
\]

(End of definition for \texttt{\_\_fp_csc_o:w})

\[
\text{\texttt{\_\_fp_sec_o:w}}
\]

The secant of ±0 is 1. The secant of ±\(\infty\) raises an invalid operation exception. The secant of \texttt{nan} is itself. Otherwise, the \texttt{trig} function reduces the argument and turns it to radians before calling the same series as for sine, using a positive sign \(0\), a starting octant of 2, and inverting upon conversion, because \(\sec(x) = +1/\sin(\pi/2 + |x|)\).

\[
\begin{align*}
\text{\texttt{\_\_fp_case_return_o:Nw \_\_fp_case_use:nuw}} & \quad \} \\
\text{\texttt{\_\_fp_chk:w \#2 \#3;}} & \quad \\
\end{align*}
\]
The tangent of ±0 or nan is the same floating point number. The tangent of ±∞ raises an invalid operation exception. Once more, the trig function does the argument reduction step and conversion to radians before calling \_fp_tan_series_o:NNwwww, with a sign #3 and an initial octant of 1 (this shift is somewhat arbitrary). See \_fp_cot_o:w for an explanation of the 0 argument.

\_fp_tan_o:w The tangent of ±0 or nan is the same floating point number. The tangent of ±∞ raises an invalid operation exception. Once more, the trig function does the argument reduction step and conversion to radians before calling \_fp_tan_series_o:NNwwww, with a sign #3 and an initial octant of 1 (this shift is somewhat arbitrary). See \_fp_cot_o:w for an explanation of the 0 argument.

\_fp_cot_o:w The cotangent of ±0 is ±∞ with the same sign, with a division by zero exception (see \_fp_cot_zero_o:Nfw. The cotangent of ±∞ raises an invalid operation exception. The cotangent of nan is itself. We use cot x = −tan(π/2 + x), and the initial octant for the tangent was chosen to be 1, so the octant here starts at 3. The change in sign is obtained by feeding \_fp_tan_series_o:NNwwww two signs rather than just the sign of the argument: the first of those indicates whether we compute tangent or cotangent. Those signs are eventually combined.
77.1.2 Distinguishing small and large arguments

The first argument is \use_i:nn if the operand is in radians and \use_ii:nn if it is in degrees. Arguments #2 to #5 control what trigonometric function we compute, and #6 to #8 are pieces of a normal floating point number. Call the _series function #2, with arguments #3, either a conversion function (\_fp_to_float:o:wN or \_fp_\text{-}inv_to_float:o:wN) or a sign 0 or 2 when computing tangent or cotangent; #4, a sign 0 or 2; the octant, computed in an integer expression starting with #5 and stopped by a period; and a fixed point number obtained from the floating point number by argument reduction (if necessary) and conversion to radians (if necessary). Any argument reduction adjusts the octant accordingly by leaving a (positive) shift into its integer expression. Let us explain the integer comparison. Two of the four \exp_after:wN are expanded, the expansion hits the test, which is true if the float is at least 1 when working in radians, and at least 10 when working in degrees. Then one of the remaining \exp_after:wN hits #1, which picks the trig or trigd function in whichever branch of the conditional was taken. The final \exp_after:wN closes the conditional. At the end of the day, a number is large if it is $\geq 1$ in radians or $\geq 10$ in degrees, and small otherwise. All four trig/trigd auxiliaries receive the operand as an extended-precision number.
77.1.3 Small arguments

\_\_fp\_trig\_small:ww

This receives a small extended-precision number in radians and converts it to a fixed point number. Some trailing digits may be lost in the conversion, so we keep the original floating point number around: when computing sine or tangent (or their inverses), the last step is to multiply by the floating point number (as an extended-precision number) rather than the fixed point number. The period serves to end the integer expression for the octant.

28122 \cs_new:Npn \__fp_trig_small:ww #1,#2; \{ \__fp_ep_to_fixed:wwn #1,#2; . #1,#2; \}

(End of definition for \__fp_trig_small:ww.)

\__fp\_trigd\_small:ww

Convert the extended-precision number to radians, then call \__fp_trig_small:ww to massage it in the form appropriate for the _series auxiliary.

28124 \cs_new:Npn \__fp_trigd_small:ww #1,#2; \{ \__fp_ep_mul_raw:wwwwN -1,{1745}{3292}{5199}{4329}{5769}{2369}; #1,#2; \__fp_trig_small:ww \}

(End of definition for \__fp_trigd_small:ww.)

77.1.4 Argument reduction in degrees

\__fp\_trigd\_large:ww \__fp\_trigd\_large\_auxi:nnnnwNNNN \__fp\_trigd\_large\_auxii:wNw \__fp\_trigd\_large\_auxiii:www

Note that $25 \times 360 = 9000$, so $10^{k+1} \equiv 10^k \pmod{360}$ for $k \geq 3$. When the exponent #1 is very large, we can thus safely replace it by 22 (or even 19). We turn the floating point number into a fixed point number with two blocks of 8 digits followed by five blocks of 4 digits. The original float is $100 \times \langle \text{block}_1 \rangle \cdots \langle \text{block}_3 \rangle \langle \text{block}_4 \rangle \cdots \langle \text{block}_7 \rangle$, or is equal to it modulo 360 if the exponent #1 is very large. The first auxiliary finds $\langle \text{block}_1 \rangle + \langle \text{block}_2 \rangle \pmod{9}$, a single digit, and prepends it to the 4 digits of $\langle \text{block}_3 \rangle$. It also unpacks $\langle \text{block}_4 \rangle$ and grabs the 4 digits of $\langle \text{block}_7 \rangle$. The second auxiliary grabs the $\langle \text{block}_3 \rangle$ plus any contribution from the first two blocks as #1, the first digit of $\langle \text{block}_4 \rangle$ (just after the decimal point in hundreds of degrees) as #2, and the three other digits as #3. It finds the quotient and remainder of #1#2 modulo 9, adds twice the quotient to the integer expression for the octant, and places the remainder (between 0 and 8) before #3 to form a new $\langle \text{block}_4 \rangle$. The resulting fixed point number is $x \in \mathbb{[0, 0.9]}$. If $x \geq 0.45$, we add 1 to the octant and feed $0.9 - x$ with an exponent of 2 (to compensate the fact that we are working in units of hundreds of degrees rather than degrees) to \_\_fp_trigd_small:ww. Otherwise, we feed it $x$ with an exponent of 2. The third auxiliary also discards digits which were not packed into the various $\langle \text{blocks} \rangle$. Since the original exponent #1 is at least 2, those are all 0 and no precision is lost (#6 and #7 are four 0 each).

28130 \cs_new:Npn \__fp_trigd_large:ww #1, #2#3#4#5#6#7; \{ \exp_after:wN \__fp_pack_eight:wNNNNNNNN \exp_after:wN \__fp_pack_eight:wNNNNNNNN \exp_after:wN \__fp_pack_twice_four:wNNNNNNNN \exp_after:wN \__fp_pack_twice_four:wNNNNNNNN \exp_after:wN \__fp_trigd_large_auxi:mmmmwNNNN \exp_after:wN \__fp_trigd_large:ww \}

(End of definition for \__fp_trigd_large:ww.)
77.1.5 Argument reduction in radians

Arguments greater or equal to 1 need to be reduced to a range where we only need a few terms of the Taylor series. We reduce to the range \([0, 2\pi]\) by subtracting multiples of \(2\pi\), then to the smaller range \([0, \pi/2]\) by subtracting multiples of \(\pi/2\) (keeping track of how many times \(\pi/2\) is subtracted), then to \([0, \pi/4]\) by mapping \(x \rightarrow \pi/2 - x\) if appropriate. When the argument is very large, say, \(10^{100}\), an equally large multiple of \(2\pi\) must be subtracted, hence we must work with a very good approximation of \(2\pi\) in order to get a sensible remainder modulo \(2\pi\).

Specifically, we multiply the argument by an approximation of \(1/(2\pi)\) with 10048 digits, then discard the integer part of the result, keeping 52 digits of the fractional part. From the fractional part of \(x/(2\pi)\) we deduce the octant (quotient of the first three digits by 125). We then multiply by 8 or \(-8\) (the latter when the octant is odd), ignore any integer part (related to the octant), and convert the fractional part to an extended precision number, before multiplying by \(\pi/4\) to convert back to a value in radians in \([0, \pi/4]\).

It is possible to prove that given the precision of floating points and their range of exponents, the 52 digits may start at most with 24 zeros. The 5 last digits are
affected by carries from computations which are not done, hence we are left with at least
52 − 24 − 5 = 23 significant digits, enough to round correctly up to 0.6 · ulp in all cases.

This integer array stores blocks of 8 decimals of $10^{-16}/(2\pi)$. Each entry is $10^8$ plus an
8 digit number storing 8 decimals. In total we store 10112 decimals of $10^{-16}/(2\pi)$. The
number of decimals we really need is the maximum exponent plus the number of digits
we later need, 52, plus 12 (4 − 1 groups of 4 digits). The memory footprint (1/2 byte per
digit) is the same as an earlier method of storing the data as a control sequence name,
but the major advantage is that we can unpack specific subsets of the digits without
unpacking the 10112 decimals.

\[\text{\c__fp_trig_intarray}\]

This integer array stores blocks of 8 decimals of $10^{-16}/(2\pi)$. Each entry is $10^8$ plus an
8 digit number storing 8 decimals. In total we store 10112 decimals of $10^{-16}/(2\pi)$. The
number of decimals we really need is the maximum exponent plus the number of digits
we later need, 52, plus 12 (4 − 1 groups of 4 digits). The memory footprint (1/2 byte per
digit) is the same as an earlier method of storing the data as a control sequence name,
but the major advantage is that we can unpack specific subsets of the digits without
unpacking the 10112 decimals.

\[\text{\c__fp_trig_intarray}\]

This integer array stores blocks of 8 decimals of $10^{-16}/(2\pi)$. Each entry is $10^8$ plus an
8 digit number storing 8 decimals. In total we store 10112 decimals of $10^{-16}/(2\pi)$. The
number of decimals we really need is the maximum exponent plus the number of digits
we later need, 52, plus 12 (4 − 1 groups of 4 digits). The memory footprint (1/2 byte per
digit) is the same as an earlier method of storing the data as a control sequence name,
but the major advantage is that we can unpack specific subsets of the digits without
unpacking the 10112 decimals.

\[\text{\c__fp_trig_intarray}\]

This integer array stores blocks of 8 decimals of $10^{-16}/(2\pi)$. Each entry is $10^8$ plus an
8 digit number storing 8 decimals. In total we store 10112 decimals of $10^{-16}/(2\pi)$. The
number of decimals we really need is the maximum exponent plus the number of digits
we later need, 52, plus 12 (4 − 1 groups of 4 digits). The memory footprint (1/2 byte per
digit) is the same as an earlier method of storing the data as a control sequence name,
but the major advantage is that we can unpack specific subsets of the digits without
unpacking the 10112 decimals.

\[\text{\c__fp_trig_intarray}\]

This integer array stores blocks of 8 decimals of $10^{-16}/(2\pi)$. Each entry is $10^8$ plus an
8 digit number storing 8 decimals. In total we store 10112 decimals of $10^{-16}/(2\pi)$. The
number of decimals we really need is the maximum exponent plus the number of digits
we later need, 52, plus 12 (4 − 1 groups of 4 digits). The memory footprint (1/2 byte per
digit) is the same as an earlier method of storing the data as a control sequence name,
but the major advantage is that we can unpack specific subsets of the digits without
unpacking the 10112 decimals.

\[\text{\c__fp_trig_intarray}\]

This integer array stores blocks of 8 decimals of $10^{-16}/(2\pi)$. Each entry is $10^8$ plus an
8 digit number storing 8 decimals. In total we store 10112 decimals of $10^{-16}/(2\pi)$. The
number of decimals we really need is the maximum exponent plus the number of digits
we later need, 52, plus 12 (4 − 1 groups of 4 digits). The memory footprint (1/2 byte per
digit) is the same as an earlier method of storing the data as a control sequence name,
but the major advantage is that we can unpack specific subsets of the digits without
unpacking the 10112 decimals.

\[\text{\c__fp_trig_intarray}\]

This integer array stores blocks of 8 decimals of $10^{-16}/(2\pi)$. Each entry is $10^8$ plus an
8 digit number storing 8 decimals. In total we store 10112 decimals of $10^{-16}/(2\pi)$. The
number of decimals we really need is the maximum exponent plus the number of digits
we later need, 52, plus 12 (4 − 1 groups of 4 digits). The memory footprint (1/2 byte per
digit) is the same as an earlier method of storing the data as a control sequence name,
but the major advantage is that we can unpack specific subsets of the digits without
unpacking the 10112 decimals.
The exponent #1 is between 1 and 10000. We wish to look up decimals \(10^{#1-16}/(2\pi)\)
starting from the digit \(#1+1\). Since they are stored in batches of 8, compute \([#1/8]\)
and fetch blocks of 8 digits starting there. The numbering of items in \c__fp_trig_intarray\nstarts at 1, so the block \([#1/8]+1\) contains the digit we want, at one of the eight
positions. Each call to \texttt{\textbackslash \int_value:w \_\textbackslash \texttt{kernel_intarray_item:Nn}} expands the
next, until being stopped by \texttt{\_\textbackslash \texttt{fp_trig_large_auxii:w}} using \texttt{\exp_stop_f:}. Once
all these blocks are unpacked, the \texttt{\exp_stop_f:} and 0 to 7 digits are removed by \texttt{\use_\textbackslash \texttt{none:n...n}}. Finally, \texttt{\_\textbackslash \texttt{fp_trig_large_auxii:w}} packs 64 digits (there are between 65
and 72 at this point) into groups of 4 and the auxv auxiliary is called.

\begin{verbatim}
\cs_new:Npn \_\textbackslash \texttt{fp_trig_large:ww} #1, #2#3#4#5#6;
{ \exp_after:wN \_\textbackslash \texttt{fp_trig_large_auxi:w} \int_value:w (#1 - 4) / 8 \exp_after:wN , \int_value:w #1 , \{#2}{#3}{#4}{#5} ; }
\cs_new:Npn \_\textbackslash \texttt{fp_trig_large_auxi:w} #1, #2,
{ \exp_after:wN \exp_after:wN \exp_after:wN \_\textbackslash \texttt{fp_trig_large_auxii:w} \cs:w use_none:n \prg_replicate:nn { #2 - #1 * 8 } { n } \cs_end: \exp_after:wN \int_value:w \_\textbackslash \texttt{kernel_intarray_item:Nn} \c__fp_trig_intarray { \_\textbackslash \texttt{fp_int_eval:w} #1 + 1 \scan_stop: } \exp_after:wN \_\textbackslash \texttt{fp_trig_large_auxii:w} \int_value:w \_\textbackslash \texttt{kernel_intarray_item:Nn} \c__fp_trig_intarray { \_\textbackslash \texttt{fp_int_eval:w} #1 + 2 \scan_stop: } \_\textbackslash \texttt{kernel_intarray_item:Nn} \c__fp_trig_intarray { \_\textbackslash \texttt{fp_int_eval:w} #1 + 3 \scan_stop: } \_\textbackslash \texttt{kernel_intarray_item:Nn} \c__fp_trig_intarray { \_\textbackslash \texttt{fp_int_eval:w} #1 + 4 \scan_stop: } \_\textbackslash \texttt{kernel_intarray_item:Nn} \c__fp_trig_intarray { \_\textbackslash \texttt{fp_int_eval:w} #1 + 5 \scan_stop: } \_\textbackslash \texttt{kernel_intarray_item:Nn} \c__fp_trig_intarray { \_\textbackslash \texttt{fp_int_eval:w} #1 + 6 \scan_stop: } }
\end{verbatim}
First come the first 64 digits of the fractional part of $10^{1-16}/(2\pi)$, arranged in 16 blocks of 4, and ending with a semicolon. Then a few more digits of the same fractional part, ending with a semicolon, then 4 blocks of 4 digits holding the significand of the original argument. Multiply the 16-digit significand with the 64-digit fractional part: the `__fp_trig_large_auxvi` auxiliary receives the significand as \texttt{#2#3#4#5} and 16 digits of the fractional part as \texttt{#6#7#8#9}, and computes one step of the usual ladder of pack functions we use for multiplication (see e.g., \texttt{__fp_fixed_mul:wwn}), then discards one block of the fractional part to set things up for the next step of the ladder. We perform 13 such steps, replacing the last \texttt{middle_shift} by the appropriate trailing shift, then discard the significand and remaining 3 blocks from the fractional part, as there are not enough digits to compute any more step in the ladder. The last semicolon closes the ladder, and we return control to the \texttt{__fp_trig_large_auxvii} auxiliary.
The \texttt{auxvii} auxiliary is followed by 52 digits and a semicolon. We find the octant as the integer part of 8 times what follows, or equivalently as the integer part of \( \frac{\#1\#2\#3}{125} \), and add it to the surrounding integer expression for the octant. We then compute 8 times the 52-digit number, with a minus sign if the octant is odd. Again, the last middle shift is converted to a trailing shift. Any integer part (including negative values which come up when the octant is odd) is discarded by \texttt{\_fp\_use\_i\_until\_s:nw}. The resulting fractional part should then be converted to radians by multiplying by \( 2\pi/8 \), but first, build an extended precision number by abusing \texttt{\_fp\_ep\_to\_ep\_loop:N} with the appropriate trailing markers. Finally, \texttt{\_fp\_trig\_small:ww} sets up the argument for the functions which compute the Taylor series.
Here we receive a conversion function \__fp_to_float_o:wwN or \__fp_inv-_to_float_o:wwN, a \langle sign \rangle (0 or 2), a \langle non-negative \rangle \langle octant \rangle delimited by a dot, a \langle fixed point \rangle number delimited by a semicolon, and an extended-precision number. The auxiliary receives:

- the conversion function \#1;
- the final sign, which depends on the octant \#3 and the sign \#2;
- the octant \#3, which controls the series we use;
- the square \#4 * \#4 of the argument as a fixed point number, computed with \__fp_fixed_mul:wwn;
- the number itself as an extended-precision number.

If the octant is in \{1, 2, 5, 6, ...\}, we are near an extremum of the function and we use the series

\[
\cos(x) = 1 - x^2 \left(\frac{1}{2!} - x^2 \left(\frac{1}{4!} - x^2 \left(\cdots\right)\right)\right).
\]

Otherwise, the series

\[
\sin(x) = x \left(1 - x^2 \left(\frac{1}{3!} - x^2 \left(\frac{1}{5!} - x^2 \left(\cdots\right)\right)\right)\right)
\]

is used. Finally, the extended-precision number is converted to a floating point number with the given sign, and \__fp_sanitize:Nw checks for overflow and underflow.
Contrarily to \_\_fp_sin_series_o:NNwww which received a conversion auxiliary as \#1, here, \#1 is 0 for tangent and 2 for cotangent. Consider first the case of the tangent. The octant \#3 starts at 1, which means that it is 1 or 2 for $|x| \in [0, \pi/2]$, it is 3 or 4 for $|x| \in [\pi/2, \pi]$, and so on: the intervals on which $\tan|x| \geq 0$ coincide with those for which \((\#3 + 1)/2\) is odd. We also have to take into account the original sign of $x$ to get the sign of the final result; it is straightforward to check that the first \texttt{int_value:w} expansion produces 0 for a positive final result, and 2 otherwise. A similar story holds for \texttt{cot(x)}.

The auxiliary receives the sign, the octant, the square of the (reduced) input, and the (reduced) input (an extended-precision number) as arguments. It then computes the numerator and denominator of

$$
\tan(x) \simeq x\left(1 - x^2(a_1 - x^2(a_2 - x^2(a_3 - x^2(a_4 - x^2(a_5))))))
- 1 - x^2(b_1 - x^2(b_2 - x^2(b_3 - x^2(b_4 - x^2b_5))))
\right).
$$

The ratio is computed by \_\_fp_ep_div:wwww, then converted to a floating point number. For octants \#3 (really, quadrants) next to a pole of the functions, the fixed point
numerator and denominator are exchanged before computing the ratio. Note that this
\textbackslash if\_int\_odd:w test relies on the fact that the octant is at least 1.

77.2 Inverse trigonometric functions

All inverse trigonometric functions (arcsine, arccosine, arctangent, arccotangent, arccscant, and arccosecant) are based on a function often denoted \texttt{atan2}. This func-
Arctangent and arccotangent

Rather than introducing a new function, \texttt{atan2}, the arctangent function \texttt{atan} is overloaded: it can take one or two arguments. In the comments below, following many texts, we call the first argument \(y\) and the second \(x\), because \(\text{atan}(y, x) = \text{atan}(y/x)\) is the angular coordinate of the point \((x, y)\).

As for direct trigonometric functions, the first step in computing \(\text{atan}(y, x)\) is argument reduction. The sign of \(y\) gives that of the result. We distinguish eight regions where the point \((x, |y|)\) can lie, of angular size roughly \(\pi/8\), characterized by their “octant”, between 0 and 7 included. In each region, we compute an arctangent as a Taylor series, then shift this arctangent by the appropriate multiple of \(\pi/4\) and sign to get the result. Here is a list of octants, and how we compute the arctangent (we assume \(y > 0\): otherwise replace \(y\) by \(-y\) below):

1. \(0 < y < 0.41421 x\), then \(\text{atan} \frac{|y|}{x}\) is given by a nicely convergent Taylor series;
2. \(0 < 0.41421 x < |y| < x\), then \(\text{atan} \frac{|y|}{x} = \frac{\pi}{4} - \text{atan} \frac{x-|y|}{x+|y|}\);
3. \(0 < x < 0.41421 |y|\), then \(\text{atan} \frac{|y|}{x} = \frac{\pi}{4} + \text{atan} \frac{x+|y|}{x-|y|}\);
4. \(0 < -x < 0.41421 |y|\), then \(\text{atan} \frac{|y|}{x} = \frac{\pi}{4} + \text{atan} \frac{x-|y|}{x+|y|}\);
5. \(0 < 0.41421 |y| < -x < |y|\), then \(\text{atan} \frac{|y|}{x} = \frac{3\pi}{4} - \text{atan} \frac{x+|y|}{x-|y|}\);
6. \(0 < -0.41421 x < |y| < -x\), then \(\text{atan} \frac{|y|}{x} = \frac{3\pi}{4} + \text{atan} \frac{x-|y|}{x+|y|}\);
7. \(0 < |y| < -0.41421 x\), then \(\text{atan} \frac{|y|}{x} = \pi - \text{atan} \frac{|y|}{x}\).

In the following, we denote by \(z\) the ratio among \(|\frac{x}{y}|\), \(\frac{x}{y}\), \(|\frac{x+y}{y-x}|\), \(|\frac{x-y}{y+x}|\) which appears in the right-hand side above.

77.2.1 Arctangent and arccotangent

The parsing step manipulates \texttt{atan} and \texttt{acot} like \texttt{min} and \texttt{max}, reading in an array of operands, but also leaves \texttt{use_i:nn} or \texttt{use_i:nn} depending on whether the result
should be given in radians or in degrees. The helper \_\_fp\_parse\_function\_one\_two:nnw checks that the operand is one or two floating point numbers (not tuples) and leaves its second argument or its tail accordingly (its first argument is used for error messages). More precisely if we are given a single floating point number \_\_fp\_atan\_default:w places \c\_one\_fp (expanded) after it; otherwise \_\_fp\_atan\_default:w is omitted by \_\_fp\_parse\_function\_one\_two:nnw.

\cs_new:Npn \__fp\_atan\_o:Nw #1
\__fp\_parse\_function\_one\_two:nnw
\if #1
\__fp\_atan\_default:w \__fp\_atanii\_o:Nww #1
\else:
\__fp\_acot\_o:Nw #1
\fi:
\cs_new:Npn \__fp\_acot\_o:Nw #1
\__fp\_parse\_function\_one\_two:nnw
\if #1
\__fp\_acot\_default:w \__fp\_acotii\_o:Nww #1
\else:
\__fp\_atan\_default:w \__fp\_atanii\_o:Nww #1
\fi:
\cs_new:Npe \__fp\_atan\_default:w #1#2#3 @ { #1 #2 #3 \c\_one\_fp @ }

\__fp\_atanii\_o:Nww \__fp\_acotii\_o:Nww
If either operand is nan, we return it. If both are normal, we call \_\_fp\_atan\_normal:\o:NN\NNw. If both are zero or both infinity, we call \_\_fp\_atan\_inf:o:NN\NNw with argument 2, leading to a result among \{±\pi/4,±3\pi/4\} (in degrees, \{±45,±135\}). Otherwise, one is much bigger than the other, and we call \_\_fp\_atan\_inf:o:NN\NNw with either an argument of 4, leading to the values \{±\pi/2\} (in degrees, ±90), or 0, leading to \{±0,±\pi\} (in degrees, ±180). Since acot(x,y) = atan(y,x), \_\_fp\_acotii\_o:ww simply reverses its two arguments.

\cs_new:Npn \__fp\_atannii\_o:Nw #1
\s\__fp \__fp\_chk:w #2#3#4; \s\__fp \__fp\_chk:w #5 #6 @
\if_meaning:w 3 #2 \__fp\_case\_return\_i\_o:ww \fi:
\if_meaning:w 3 #5 \__fp\_case\_return\_ii\_o:ww \fi:
\if_case:w
\if_meaning:w #2 #5
\if_meaning:w 1 #2 10 \else: 0 \fi:
\else:
\if_int_compare:w #2 > #5 \exp_stop_f: 1 \else: 2 \fi:
\fi:
\exp_stop_f:
\__fp\_case\_return:ww \{ \__fp\_atan\_inf:o:NN\NNw #1 #3 2 \}
\or:\__fp\_case\_return:ww \{ \__fp\_atan\_inf:o:NN\NNw #1 #3 4 \}
\or:\__fp\_case\_return:ww \{ \__fp\_atan\_inf:o:NN\NNw #1 #3 0 \}
\fi:
\_\_fp\_atan\_normal:o:NN\NNw\NNw #1
\s\__fp \_\_fp\_chk:w #2#3#4;
\s\__fp \_\_fp\_chk:w #5 #6
\}
\cs_new:Npn \__fp\_acotii\_o:Nww #1#2; #3;
\__fp\_atannii\_o:Nw #1#2 #3;
\cs_new:Npn \__fp\_acotii\_o:Nww #1#2; #3;
(End of definition for \_\_fp\_atanii:o:Nww and \_\_fp\_acotii:o:Nww)
This auxiliary is called whenever one number is ±0 or ±∞ (and neither is nan). Then the result only depends on the signs, and its value is a multiple of π/4. We use the same auxiliary as for normal numbers, \_fp_atan_combine_o:NwwwN, with arguments the final sign #2; the octant #3; atan \( z/z = 1 \) as a fixed point number; \( z = 0 \) as a fixed point number; and \( z = 0 \) as an extended-precision number. Given the values we provide, atan \( z \) is computed to be 0, and the result is \([3/2] \cdot \pi/4\) if the sign #5 of \( x \) is positive, and \([(7 - #3)/2] \cdot \pi/4\) for negative \( x \), where the divisions are rounded up.

\(28630\)
\begin{verbatim}
\cs_new:Npn \__fp_atan_inf_o:NNNw #1#2#3 \s__fp \__fp_chk:w #4#5#6; 
\end{verbatim}

(End of definition for \_fp_atan_inf_o:NNNw.)

\_fp_atan_normal_o:NNnwN

Here we simply reorder the floating point data into a pair of signed extended-precision numbers, that is, a sign, an exponent ending with a comma, and a six-block mantissa ending with a semi-colon. This extended precision is required by other inverse trigonometric functions, to compute things like \( \text{atan}(x, \sqrt{1 - x^2}) \) without intermediate rounding errors.

\(28640\)
\begin{verbatim}
\cs_new_protected:Npn \__fp_atan_normal_o:NNnwNnw #1 \s__fp \__fp_chk:w 1#2#3#4; \s__fp \__fp_chk:w 1#5#6#7; 
\end{verbatim}

(End of definition for \_fp_atan_normal_o:NNnwN.)

\_fp_atan_test_o:NwwNwwN

This receives: the sign #1 of \( y \), its exponent #2, its 24 digits #3 in groups of 4, and similarly for \( x \). We prepare to call \_fp_atan_combine_o:NwwwN which expects the sign #1, the octant, the ratio \( (\text{atan} z)/z = 1 - \cdots \), and the value of \( z \), both as a fixed point number and as an extended-precision floating point number with a mantissa in [0.01,1). For now, we place #1 as a first argument, and start an integer expression for the octant. The sign of \( x \) does not affect \( z \), so we simply leave a contribution to the octant: \( \langle \text{octant} \rangle \to 7 - \langle \text{octant} \rangle \) for negative \( x \). Then we order \(|y|\) and \(|x|\) in a non-decreasing order: if \(|y| > |x|\), insert 3 in the expression for the octant, and swap the two numbers. The finer test with 0.41421 is done by \_fp_atan_div:wnwwNw after the operands have been ordered.

\(28640\)
\begin{verbatim}
\exp_after:wN \_fp_atan_test_o:NwwNwwN \exp_after:wN \_fp_atan_combi
\end{verbatim}
This receives two positive numbers \( a \) and \( b \) (equal to \( |x| \) and \( |y| \) in some order), each as an exponent and 6 blocks of 4 digits, such that \( 0 < a < b \). If \( 0.41421b < a \), the two numbers are “near”, hence the point \((y, x)\) that we started with is closer to the diagonals \( \{|y| = |x|\} \) than to the axes \( \{xy = 0\}\). In that case, the octant is 1 (possibly combined with the \(7\)− and \(3\)− inserted earlier) and we wish to compute \( \arctan \frac{b}{a} + \frac{\pi}{4} \). Otherwise, the octant is 0 (again, combined with earlier terms) and we wish to compute \( \arctan \frac{a}{b} \). In any case, call \( \__fp_atan_auxi:ww \) followed by \( z \), as a comma-delimited exponent and a fixed point number.

\[ \text{if} \int \text{compare:w} \]
\[ \__fp_atan_div:wnwwnw \#2,#3; #5,#6; > \c_zero_int \]
\[ 3 - \]
\[ \exp_after:wN \__fp_reverse_args:Nww \]
\[ \__fp_atan_div:wnwwnw \#2,#3; #5,#6; \]
\]

(End of definition for \( \__fp_atan_test_o:NwwNwwN \).)

\[ \__fp_atan_div:wnwwnw \]
\[ \__fp_atan_near:wwwn \]
\[ \__fp_atan_near_aux:wwn \]

This converts \( z \) from a representation as an exponent and a fixed point number in \([0.01, 1)\) to a fixed point number only, then set up the call to \( \__fp_atan_Taylor_loop:www \), followed by the fixed point representation of \( z \) and the old representation.

\[ \text{cs_new:Npn \__fp_atan_div:wnwwnw \#1,#2#3; #4,#5#6;} \]
\[ \{ \]
\[ \text{if} \int \text{compare:w} \]
\[ \__fp_int_eval:w 41421 * \#5 < #2 000 \]
\[ \text{if case:w \__fp_int_eval:w \#4 - \#1 \__fp_int_eval_end:} \]
\[ 00 \text{ or: 0} \fi: \]
\[ \exp_stop_f: \]
\[ \exp_after:wN \__fp_atan_near:wwwn \]
\[ \__fp_ep_div:wwwwn #1,#2; #3, \]
\[ \__fp_atan_near_aux:ww \]
\]

(End of definition for \( \__fp_atan_div:wnwwnw, \__fp_atan_near:wwwn, \text{and} \__fp_atan_near_aux:wwn.)

\[ \__fp_atan_auxi:ww \]
\[ \__fp_atan_auxii:w \]
We compute the series of \((\text{atan } z)/z\). A typical intermediate stage has \(\#1 = 2k - 1\), \(\#2 = \frac{1}{2k+1} - z^2 (\frac{1}{2k+3} - z^2 (\cdots - z^2 (\frac{1}{2k+39})))\), and \(\#3 = z^2\). To go to the next step \(k \rightarrow k - 1\), we compute \(\frac{1}{2k-1}\), then subtract from it \(z^2\) times \(\#2\). The loop stops when \(k = 0\): then \(\#2\) is \((\text{atan } z)/z\), and there is a need to clean up all the unnecessary data, end the integer expression computing the octant with a semicolon, and leave the result \(\#2\) afterwards.

\begin{verbatim}
\cs_new:Npn \__fp_atan_Taylor_loop:www #1; #2; #3; 
{ \if_int_compare:w #1 = - \c_one_int \__fp_atan_Taylor_break:w \fi: \exp_after:wN \__fp_fixed_div_int:wwN \c__fp_one_fixed_tl #1; \__fp_rrot:www \__fp_fixed_mul_sub_back:wwwn #2; #3; \} \exp_after:wN \__fp_atan_Taylor_loop:www \int_value:w \__fp_int_eval:w #1 - 2 ; \__fp fixed_mul_sub_back:wwwn \int_value:w #2 #3; 
\cs_new:Npn \__fp_atan_Taylor_break:w \fi: #1 \__fp_fixed_mul_sub_back:wwwn #2; \fi: ; #2; }
\end{verbatim}

(End of definition for \__fp_atan_Taylor_loop:www and \__fp_atan_Taylor_break:w.)

This receives a \langle\text{sign}\rangle, an \langle\text{octant}\rangle, a fixed point value of \((\text{atan } z)/z\), a fixed point number \(z\), and another representation of \(z\), as an \langle\text{exponent}\rangle and the fixed point number \(10^{-\langle\text{exponent}\rangle} z\), followed by either \texttt{\textbackslash use\_i:nn} (when working in radians) or \texttt{\textbackslash use\_ii:nn} (when working in degrees). The function computes the floating point result

\begin{equation}
\langle\text{sign}\rangle \left( \left( \frac{\langle\text{octant}\rangle}{2} \right) \frac{\pi}{4} + (-1)^{\langle\text{octant}\rangle} \frac{\text{atan} z}{z} \cdot z \right),
\end{equation}

multiplied by 180/\(\pi\) if working in degrees, and using in any case the most appropriate representation of \(z\). The floating point result is passed to \texttt{\textbackslash __fp\textunderscore sanitize:Nw}, which checks for overflow or underflow. If the octant is 0, leave the exponent \#5 for \texttt{\textbackslash __fp\textunderscore sanitize:Nw}, and multiply \#3 = \text{atan} \(z\) with \#6, the adjusted \(z\). Otherwise, multiply \#3 = \text{atan} \(z\) with \#4 = \(z\), then compute the appropriate multiple of \(\frac{\pi}{4}\) and add or subtract the product \#3 \cdot \#4. In both cases, convert to a floating point with \texttt{\textbackslash __fp\textunderscore fixed\textunderscore to\textunderscore float\textunderscore o:wN}.

\begin{verbatim}
\cs_new:Npn \__fp_atan_combine_o:NwwwwN \__fp_atan_combine_aux:ww 
\{ \exp_after:wN \__fp_sanitize:Nw \exp_after:wN \__fp\textunderscore sanitize:Nw \exp_after:wN \int_value:w \__fp_int_eval:w #1 #2; #3; #4; #5; #6; #7 
\cs_new:Npn \__fp_atan_combine_o:NwwwwN \__fp_atan_combine_aux:ww 
\{ \exp_after:wN \__fp\textunderscore sanitize:Nw \exp_after:wN \int_value:w \__fp_int_eval:w #1 #2; #3; #4; #5; #6; #7 
\cs_new:Npn \__fp\textunderscore sanitize:Nw 
\end{verbatim}
77.2.2 Arcsine and arccosine

\__fp_asin_o:w Again, the first argument provided by l3fp-parse is \use_i:nn if we are to work in radians and \use_i:nn for degrees. Then comes a floating point number. The arcsine of $\pm 0$ or nan is the same floating point number. The arcsine of $\pm \infty$ raises an invalid operation exception. Otherwise, call an auxiliary common with \__fp_acos_o:w, feeding it information about what function is being performed (for “invalid operation” exceptions).

(End of definition for \__fp_atan_combine_o:NwwwwwN and \__fp_atan_combine_aux:ww.)
The arccosine of ±0 is π/2 (in degrees, 90). The arccosine of ±∞ raises an invalid operation exception. The arccosine of nan is itself. Otherwise, call an auxiliary common with \_fp_sin_o:w, informing it that it was called by acos or acosd, and preparing to swap some arguments down the line.

\_fp_acos_o:w

If the exponent #5 is at most 0, the operand lies within (−1, 1) and the operation is permitted: call \_fp_asin_auxi_o:NnNww with the appropriate arguments. If the number is exactly ±1 (the test works because we know that #5 ≥ 1, #6#7 ≥ 1000000, #8#9 ≥ 0, with equality only for ±1), we also call \_fp_asin_auxi_o:NnNww. Otherwise, \_fp_use_i:ww gets rid of the asin auxiliary, and raises instead an invalid operation, because the operand is outside the domain of arcsine or arccosine.

\_fp_asin_normal_o:NfwNnnnnw

We compute \( x/\sqrt{1-x^2} \). This function is used by asin and acos, but also by acsc and asec after inverting the operand, thus it must manipulate extended-precision numbers. First evaluate \( 1-x^2 \) as \( (1+x)(1-x) \): this behaves better near \( x = 1 \). We do the...
addition/subtraction with fixed point numbers (they are not implemented for extended-precision floats), but go back to extended-precision floats to multiply and compute the inverse square root $1/\sqrt{1-x^2}$. Finally, multiply by the (positive) extended-precision float $|x|$, and feed the (signed) result, and the number $+1$, as arguments to the arctangent function. When computing the arccosine, the arguments $x/\sqrt{1-x^2}$ and $+1$ are swapped by \_\_fp_reverse_args:Nww in that case before \_\_fp_atan_test_o:NwwNwwN is evaluated. Note that the arctangent function requires normalized arguments, hence the need for ep_to_ep and continue after ep_mul.

\begin{verbatim}
cs_new:Npn \_\_fp_asin_auxi_o:NnNww #1#2#3#4,#5;
{ \_\_fp_ep_to_fixed:wwn #4,#5; \_\_fp_asin_isqrt:wn \_\_fp_ep_mul:wwwwn #4,#5; \_\_fp_ep_to_ep:wwN \_\_fp_fixed_continue:wn { #2 \_\_fp_atan_test_o:NwwNwwN #3 } 0 1,\{0000\}{0000\}{0000\}{0000\}{0000\}; #1}
cs_new:Npn \_\_fp_asin_isqrt:wn #1;
{ \exp_after:wN \_\_fp_fixed_sub:wwn \c__fp_one_fixed_tl #1; \_\_fp_fixed_add_one:wN #1; \_\_fp_fixed_continue:wn \{ \_\_fp_ep_mul:wwwwn 0, } 0, \_\_fp_ep_isqrt:wwn}
\end{verbatim}

(End of definition for \_\_fp_asin_auxi_o:Nww and \_\_fp_asin_isqrt:wn.)

77.2.3 Arc cosecant and arccosecant

\_\_fp_acsc_o:w

Cases are mostly labelled by \#2, except when \#2 is 2: then we use \#3\#2, which is 02 = 2 when the number is $+\infty$ and 22 when the number is $-\infty$. The arccosecant of $\pm0$ raises an invalid operation exception. The arccosecant of $\pm\infty$ is $\pm0$ with the same sign. The arccosecant of nan is itself. Otherwise, \_\_fp_acsc_normal_o:NfwNnw does some more tests, keeping the function name (acsc or acscd) as an argument for invalid operation exceptions.

\begin{verbatim}
cs_new:Npn \_\_fp_acsc_o:w #1 \s__fp \_\_fp_chk:w #2#3#4; @
{ \if_case:w \if_meaning:w 2 #2 #3 \fi: #2 \exp_stop_f:
\_\_fp_case_use:nw
\} \_\_fp_invalid_operation_o:fw \{ #1 \{ acsc \} \{ acscd \} \} \_\_fp_case_use:nw
\else: \_\_fp_acsc_normal_o:NfwNnw \#1 \{ #1 \{ acsc \} \{ acscd \} \}
\or: \_\_fp_case_return_o:Nw \c_zero_fp \_\_fp_case_return_same_o:w
\else: \_\_fp_case_return_o:Nw \c_minus_zero_fp
\fi:
\_\_fp \_\_fp_chk:w #2 #3 #4;}
\end{verbatim}

(End of definition for \_\_fp_acsc_o:w.)
The arcsecant of $\pm 0$ raises an invalid operation exception. The arcsecant of $\pm \infty$ is $\pi/2$ (in degrees, 90). The arcsecant of \texttt{nan} is itself. Otherwise, do some more tests, keeping the function name \texttt{asec} (or \texttt{asecd}) as an argument for invalid operation exceptions, and a \texttt{\_\_fp_reverse_args:Nww} following precisely that appearing in \texttt{\_\_fp_acos_o:w}.

\begin{verbatim}
\cs_new:Npn \_\_fp_acsc_normal_o:NfwNnw #1#2#3 \s__fp \_\_fp_chk:w 1#4#5#6; 
{ 
\int_compare:nNnTF {#5} < 1
{ 
\_\_fp_invalid_operation_o:fw {#2} 
\_\_fp_acsc_normal_o:NfwNnw #1 {#1} {#2} \_\_fp_chk:w 1#4#5#6; }
}
\_\_fp_ep_div:wwwwn 1,{1000}{0000}{0000}{0000}{0000}; #5,#6{0000}{0000}; { \_\_fp_asin_auxi_o:NnNww #1 {#3} #4 }
\_\_fp_acsc_normal_o:NfwNnw #1 #2#3 \_\_fp_chk:w #2 #3; }
\end{verbatim}

\textit{(End of definition for \_\_fp_acsc_normal_o:NfwNnw.)}

If the exponent is non-positive, the operand is less than 1 in absolute value, which is always an invalid operation: complain. Otherwise, compute the inverse of the operand, and feed it to \texttt{\_\_fp_asin_auxi_o:NfwNnw} (with all the appropriate arguments). This computes what we want thanks to $\acsc(x) = \asin(1/x)$ and $\asec(x) = \acos(1/x)$.

\begin{verbatim}
\cs_new:Npn \_\_fp_acsc_normal_o:NfwNnw #1#2#3 \s__fp \_\_fp_chk:w 1#4#5#6; 
{ 
\int_compare:nNnTF {#5} < 1
{ 
\_\_fp_invalid_operation_o:fw {#2} 
\_\_fp_acsc_normal_o:NfwNnw #1 {#1} {#2} \_\_fp_chk:w 1#4#5#6; }
}
\_\_fp_ep_div:wwwwn 1,{1000}{0000}{0000}{0000}{0000}; #5,#6{0000}{0000}; { \_\_fp_asin_auxi_o:NfwNnw #1 {#3} #4 }
\_\_fp_acsc_normal_o:NfwNnw #1 #2#3 \_\_fp_chk:w #2 #3; }
\end{verbatim}

\textit{(End of definition for \_\_fp_acsc_normal_o:NfwNnw.)}
Chapter 78

13fp-convert implementation

The first argument is for instance \_\_fp_to_tl_dispatch:w, which converts any floating point object to the appropriate representation. We loop through all items, putting ,~ between all of them and making sure to remove the leading ,~.

\begin{verbatim}
\cs_new:Npn \__fp_tuple_convert:Nw #1 \s__fp_tuple \__fp_tuple_chk:w #2 ;
{
\int_case:nnF { \__fp_array_count:n {#2} }
{ 0 } { ( ) }
{ 1 } { \__fp_tuple_convert_end:w \@ { #1 #2 , } }
}{
\__fp_tuple_convert_loop:nNw { } #1
#2 { ? \__fp_tuple_convert_end:w } ;
\@ { \use_none:nn }
}
\cs_new:Npn \__fp_tuple_convert_loop:nNw #1#2#3#4; #5 @ #6
{
\use_none:n #3
\exp_args:Nf \__fp_tuple_convert_loop:nNw { #2 #3#4 ; } #2 #5
\@ { #6 , ~ #1 }
}
\cs_new:Npn \__fp_tuple_convert_end:w #1 @ #2
{ \exp_after:wN ( \exp:w \exp_end_continue_f:w #2 ) }
\end{verbatim}

(End of definition for \_\_fp_tuple_convert:Nw, \_\_fp_tuple_convert_loop:nNw, and \_\_fp_tuple_convert_end:w.)

78.2 Trimming trailing zeros

If #1 ends with a 0, the loop auxiliary takes that zero as an end-delimiter for its first argument, and the second argument is the same loop auxiliary. Once the last trailing
zero is reached, the second argument is the dot auxiliary, which removes a trailing dot if any. We then clean-up with the end auxiliary, keeping only the number.

\begin{verbatim}
\cs_new:Npn \__fp_trim_zeros:w #1 ;
{ \__fp_trim_zeros_loop:w #1 ; \__fp_trim_zeros_dot:w . ; \s__fp_stop }
\cs_new:Npn \__fp_trim_zeros_loop:w #1 0; #2 { #2 #1 ; #2 }
\cs_new:Npn \__fp_trim_zeros_dot:w #1 .; { \__fp_trim_zeros_end:w #1 ; }
\cs_new:Npn \__fp_trim_zeros_end:w #1 ; #2 \s__fp_stop { #1 }
\end{verbatim}

(End of definition for \__fp_trim_zeros:w and others.)

### 78.3 Scientific notation

The three public functions evaluate their argument, then pass it to \__fp_to_scientific_dispatch:w.

\begin{verbatim}
\cs_new:Npn \fp_to_scientific:N #1 { \exp_after:wN \__fp_to_scientific_dispatch:w #1 }
\cs_generate_variant:Nn \fp_to_scientific:N { c }
\cs_new:Npn \fp_to_scientific:n { \exp_after:wN \__fp_to_scientific_dispatch:w \exp:w \exp_end_continue_f:w \__fp_parse:n }
\end{verbatim}

(End of definition for \fp_to_scientific:N and \fp_to_scientific:n. These functions are documented on page 262.)

Expressing an internal floating point number in scientific notation is quite easy: no rounding, and the format is very well defined. First cater for the sign: negative numbers \((\#2 = 2)\) start with \(-\); we then only need to care about positive numbers and \texttt{nan}. Then filter the special cases: \(\pm 0\) are represented as \(0\); infinities are converted to a number slightly larger than the largest after an “invalid_operation” exception; \texttt{nan} is represented as \(0\) after an “invalid_operation” exception. In the normal case, decrement the exponent.
and unbrace the 4 brace groups, then in a second step grab the first digit (previously hidden in braces) to order the various parts correctly.

```latex
\cs_new:Npn \__fp_to_scientific:w \s__fp \__fp_chk:w \#1#2
  \{ \if_meaning:w 2 #2 \exp_after:wN - \exp:w \exp_end_continue_f:w \fi:
  \if_case:w #1 \exp_stop_f:
    \__fp_case_return:nw { 0.00000000000000e0 }
  \or: \exp_after:wN \__fp_to_scientific_normal:wnnnn
  \or:
    \__fp_case_use:nw
      \__fp_invalid_operation:nnw
        \fp_to_scientific:N \c__fp_overflowing_fp
        \fp_to_scientific
    \fi:
  \s__fp \__fp_chk:w #1 #2
\}
\cs_new:Npn \__fp_to_scientific_normal:wnnnn
  \s__fp \__fp_chk:w 1 #1 #2 #3 #4 #5 #6 ;
  \exp_after:wN \__fp_to_scientific_normal:wNw
  \exp_after:wN e \int_value:w \__fp_int_eval:w #2 - 1 ; #3 #4 #5 #6 ;
\cs_new:Npn \__fp_to_scientific_normal:wNw #1 ; #2#3 ;
  \{ #2.#3 #1 \}
(End of definition for \__fp_to_scientific:w, \__fp_to_scientific_normal:wnnnn, and \__fp_to_scientific_normal:wNw.)
```

## 78.4 Decimal representation

All three public variants are based on the same \__fp_to_decimal_dispatch:w after evaluating their argument to an internal floating point.

```latex
\cs_new:Npn \fp_to_decimal:N \fp_to_decimal:c \fp_to_decimal:n
  \cs_new:Npn \fp_to_decimal:N \fp_to_decimal:c
    \cs_generate_variant:Nn \fp_to_decimal:N { c }
\cs_new:Npn \fp_to_decimal:n
  \exp_after:wN \__fp_to_decimal_dispatch:w
  \exp:w \exp_end_continue_f:w \__fp_parse:n
(End of definition for \fp_to_decimal:N and \fp_to_decimal:n. These functions are documented on page 261.)
```
We allow tuples.

\begin{verbatim}
\cs_new:Npn \__fp_to_decimal_dispatch:w #1
\{ \__fp_change_func_type:NNN #1 \__fp_to_decimal:w \__fp_to_decimal_recover:w #1 \}
\cs_new:Npn \__fp_to_decimal_recover:w #1 #2 ;
\{ \__fp_error:nffn \{ unknown-type \} \{ \tl_to_str:n \{ #2 ; \} \} \} \}
\cs_new:Npn \__fp_tuple_to_decimal:w
\{ \__fp_tuple_convert:Nw \__fp_to_decimal_dispatch:w \}
\end{verbatim}

(End of definition for \__fp_to_decimal_dispatch:w, \__fp_to_decimal_recover:w, and \__fp_tuple_to_decimal:w.)

The structure is similar to \__fp_to_scientific:w. Insert – for negative numbers. Zero gives 0, ±∞ and nan yield an “invalid operation” exception; note that ±∞ produces a very large output, which we don’t expand now since it most likely won’t be needed. Normal numbers with an exponent in the range [1, 15] have that number of digits before the decimal separator: “decimate” them, and remove leading zeros with \int_value:w, then trim trailing zeros and dot. Normal numbers with an exponent 16 or larger have no decimal separator, we only need to add trailing zeros. When the exponent is non-positive, the result should be 0.(zeros)(digits), trimmed.
\int_compare:nNnTF {#2} < \c__fp_prec_int
{ \__fp_decimate:nNnnn { \c__fp_prec_int - #2 } \__fp_to_decimal_large:Nnnw
}
{ \exp_after:wN \exp_after:wN \exp_after:wN \__fp_to_decimal_huge:wnnnn
\prg_replicate:nn { #2 - \c__fp_prec_int } { 0 } ; }
{ (#3) (#4) (#5) (#6) }
{ }
{ \exp_after:wN \__fp_trim_zeros:w \int_value:w \if_int_compare:w #2 > \c_zero_int
#2 \fi: \exp_stop_f: #3.#4 ; }
\cs_new:Npn \__fp_to_decimal_large:Nnnw #1#2#3#4;
{ \cs_new:Npn \__fp_to_decimal_huge:wnnnn #1; #2#3#4#5 { #2#3#4#5 #1 } }
\cs_new:Npn \__fp_to_decimal_large:Nnnw #1#2#3#4#5 #6
(End of definition for \__fp_to_decimal:w and others.)

78.5 Token list representation

\fp_to_tl:N \fp_to_tl:c \fp_to_tl:n
These three public functions evaluate their argument, then pass it to \__fp_to_tl_dispatch:w.
\cs_new:Npn \fp_to_tl:N #1 { \exp_after:wN \__fp_to_tl_dispatch:w #1 }
\cs_generate_variant:Nn \fp_to_tl:N { c }
\cs_new:Npn \fp_to_tl:n \fp_to_tl:n
{ \exp_after:wN \__fp_to_tl_dispatch:w \exp:w \exp_end_continue_f:w \__fp_tuple_to_tl:w }
\cs_new:Npn \__fp_tuple_to_tl:w
(End of definition for \fp_to_tl:N and \fp_to_tl:n. These functions are documented on page 262.)

\__fp_to_tl_dispatch:w \__fp_to_tl_recovery:w \__fp_tuple_to_tl:w
We allow tuples.
\cs_new:Npn \__fp_to_tl_dispatch:w #1
{ \__fp_change_func_type:NNN #1 \__fp_to_tl:w \__fp_to_tl_recovery:w #1 }
\cs_new:Npn \__fp_to_tl_recovery:w #1 #2
{ \__fp_error:nffn { unknown-type } { \tl_to_str:n { #2 ; } } { } { } }

1214
\cs_new:Npn \__fp_tuple_to_tl:w \s__fp \__fp_chk:w #1#2 {
\if_meaning:w 2 #2 \exp_after:wN - \exp:w \exp_end_continue_f:w \fi:
\if_case:w #1 \exp_stop_f:
\__fp_case_return:nw { 0 }
\or: \__fp_to_tl_normal:nnnnn
\or: \__fp_case_return:nw { inf }
\else: \__fp_case_return:nw { nan }
\fi:
}
\cs_new:Npn \__fp_to_tl_normal:nnnnn #1 {
\int_compare:nTF
{-2 <= #1 <= \c__fp_prec_int }
{ \__fp_to_decimal_normal:wnnnnn }
{ \__fp_to_tl_scientific:wnnnnn }
\s__fp \__fp_chk:w 1 0 {#1}
}
\cs_new:Npn \__fp_to_tl_scientific:wnnnnn \s__fp \__fp_chk:w 1 #1 #2 #3#4#5#6 ; {
\exp_after:wN \__fp_to_tl_scientific:wNw \exp_after:wN e \int_value:w \__fp_int_eval:w #2 - 1 ; #3 #4 #5 #6 ;
}
\cs_new:Npn \__fp_to_tl_scientific:wNw #1 ; #2#3; {
\__fp_trim_zeros:w #2.#3 ; #1 }

(End of definition for \__fp_to_tl:w and others.)

78.6 Formatting

This is not implemented yet, as it is not yet clear what a correct interface would be, for this kind of structured conversion from a floating point (or other types of variables) to a string. Ideas welcome.

78.7 Convert to dimension or integer

\fp_to_dim:N
\fp_to_dim:c
\fp_to_dim:n
\__fp_to_dim_dispatch:w
\__fp_to_dim_recover:w
\__fp_to_dim:w

All three public variants are based on the same \__fp_to_dim_dispatch:w after evaluating their argument to an internal floating point. We only allow floating point numbers,
not tuples.

\begin{verbatim}
\cs_new:Npn \fp_to_dim:N #1
{ \exp_after:wN \__fp_to_dim_dispatch:w #1 }
\cs_generate_variant:Nn \fp_to_dim:N { c }
\cs_new:Npn \fp_to_dim:n
{ \exp_after:wN \__fp_to_dim_dispatch:w \exp:w \exp_end_continue_f:w \__fp_parse:n }
\cs_new:Npn \__fp_to_dim_dispatch:w #1#2 ;
{ \__fp_change_func_type:NNN #1 \__fp_to_dim:w \__fp_to_dim_recover:w #1 #2 ; }
\cs_new:Npn \__fp_to_dim_recover:w #1
{ \__fp_invalid_operation:nnw { 0pt } { fp_to_dim } }
\cs_new:Npn \__fp_to_dim:w #1 ;
{ \__fp_to_decimal:w #1 ; pt }
\end{verbatim}

(End of definition for \fp_to_dim:N and others. These functions are documented on page 261.)

\begin{verbatim}
\fp_to_int:N
\fp_to_int:c
\fp_to_int:n
\__fp_to_int_dispatch:w
\__fp_to_int_recover:w
\end{verbatim}

For the most part identical to \fp_to_dim:N but without pt, and where \__fp_to_int:w does more work. To convert to an integer, first round to 0 places (to the nearest integer), then express the result as a decimal number: the definition of \__fp_to_decimal_dispatch:w is such that there are no trailing dot nor zero.

\begin{verbatim}
\cs_new:Npn \fp_to_int:N #1
{ \exp_after:wN \__fp_to_int_dispatch:w #1 }
\cs_generate_variant:Nn \fp_to_int:N { c }
\cs_new:Npn \fp_to_int:n
{ \exp_after:wN \__fp_to_int_dispatch:w \exp:w \exp_end_continue_f:w \__fp_parse:n }
\cs_new:Npn \__fp_to_int_dispatch:w #1#2 ;
{ \__fp_change_func_type:NNN #1 \__fp_to_int:w \__fp_to_int_recover:w #1 #2 ; }
\cs_new:Npn \__fp_to_int_recover:w #1
{ \__fp_invalid_operation:nnw { 0 } { fp_to_int } }
\cs_new:Npn \__fp_to_int:w #1;
{ \exp_after:wN \__fp_to_decimal:w \exp:w \exp_end_continue_f:w \__fp_round:Nwn \__fp_round_to_nearest:NNN #1; { 0 } }
\end{verbatim}

(End of definition for \fp_to_int:N and others. These functions are documented on page 261.)

\section*{78.8 Convert from a dimension}

The dimension expression (which can in fact be a glue expression) is evaluated, converted to a number (i.e., expressed in scaled points), then multiplied by $2^{-16} = 0.0000152587890625$ to give a value expressed in points. The auxiliary \__fp_mul_-npos_o:Nww expects the desired (final sign) and two floating point operands (of the form \s__fp...;) as arguments. This set of functions is also used to convert dimension
registers to floating points while parsing expressions: in this context there is an additional exponent, which is the first argument of \_\_\_fp_from_dim_test:ww, and is combined with the exponent \(-4\) of \(2^{-16}\). There is also a need to expand afterwards: this is performed by \_\_\_fp_mul_npos_o:Nww, and cancelled by \prg_do_nothing: here.

\begin{verbatim}
\cs_new:Npn \dim_to_fp:n #1
{ \exp_after:wN \__fp_from_dim_test:ww
  \exp_after:wN 0
  \exp_after:wN ,
  \int_value:w \tex_glueexpr:D #1 ; }
\cs_new:Npn \__fp_from_dim_test:ww #1, #2
{ \if_meaning:w 0 #2
  \__fp_case_return:nw { \exp_after:wN \c_zero_fp }
\else:
  \exp_after:wN \__fp_from_dim:wNw
  \int_value:w \__fp_int_eval:w #1 - 4
  \if_meaning:w - #2
    \exp_after:wN , \exp_after:wN 2 \int_value:w
  \else:
    \exp_after:wN , \exp_after:wN 0 \int_value:w #2
  \fi:
\fi: }
\cs_new:Npn \__fp_from_dim:wNw #1,#2#3; #4
{ \__fp_pack_twice_four:wNNNNNNNN \__fp_from_dim:wNNnnnnnn ;
  #3 000 0000 00 {10}; #2 {#1} }
\cs_new:Npn \__fp_from_dim:wnnnnwNn #1; #2#3#4#5#6; #7#8
{ \__fp_mul_npos_o:Nww #7
  \__fp \__fp_chk:w 1 #7 {#5} \prg_do_nothing:
  \s__fp \__fp_chk:w 1 0 {#8} {1525} {8789} {0625} {0000} ;
  \prg_do_nothing:
}

(End of definition for \dim_to_fp:n and others. This function is documented on page 231.)

78.9 Use and eval

\fp_use:N Those public functions are simple copies of the decimal conversions.
\fp_use:c \cs_new_eq:NN \fp_use:N \fp_to_decimal:N
\fp_eval:n \cs_generate_variant:NN \fp_use:N \fp_eq:NN

(End of definition for \fp_use:N and \fp_eval:n. These functions are documented on page 262.)

\fp_sign:n Trivial but useful. See the implementation of \fp_add:Nn for an explanation of why to use \_\_\_fp_parse:n, namely, for better error reporting.

1217
\begin{verbatim}
\cs_new:Npn \fp_sign:n #1
{ \fp_to_decimal:n \{ sign \_fp_parse:n {#1} \} }

(End of definition for \fp_sign:n. This function is documented on page 261.)

\fp_abs:n
Trivial but useful. See the implementation of \fp_add:n for an explanation of why to
use \_fp_parse:n, namely, for better error reporting.
\cs_new:Npn \fp_abs:n #1
{ \fp_to_decimal:n \{ abs \_fp_parse:n {#1} \} }

(End of definition for \fp_abs:n. This function is documented on page 280.)

\fp_max:nn \fp_min:nn
Similar to \fp_abs:n, for consistency with \int_max:nn, etc.
\cs_new:Npn \fp_max:nn #1#2
{ \fp_to_decimal:n \{ max ( \_fp_parse:n {#1} , \_fp_parse:n {#2} ) \} }
\cs_new:Npn \fp_min:nn #1#2
{ \fp_to_decimal:n \{ min ( \_fp_parse:n {#1} , \_fp_parse:n {#2} ) \} }

(End of definition for \fp_max:nn and \fp_min:nn. These functions are documented on page 280.)

78.10 Convert an array of floating points to a comma list
\_fp_array_to_clist:n \_fp_array_to_clist_loop:Nw
Converts an array of floating point numbers to a comma-list. If speed here ends up
irrelevant, we can simplify the code for the auxiliary to become
\cs_new:Npn \_fp_array_to_clist_loop:Nw #1#2;
{ \use_none:n #1
{ , ~ } \fp_to_tl:n \{ #1 #2 ; \}
\_fp_array_to_clist_loop:Nw }

The \use_i:nn function is expanded after \_fp_expand:n is done, and it removes ,~
from the start of the representation.
\cs_new:Npn \_fp_array_to_clist:n #1
{ \tl_if_empty:nF \{#1\}
{ \exp_last_unbraced:Ne \use_i:nn
{ \_fp_array_to_clist_loop:Nw #1 \{ ? \prg_break: \} ;
\prg_break_point:
}
}
\cs_new:Npn \_fp_array_to_clist_loop:Nw #1#2;
{ \use_none:n #1
{ , ~ } \exp_not:f \{ \_fp_to_tl_dispatch:w #1 #2 ; \}
\_fp_array_to_clist_loop:Nw
}
\end{verbatim}

1218
(End of definition for __fp_array_to_clist:n and __fp_array_to_clist_loop:Nw.)

29166 ⟨/package⟩
Chapter 79

l3fp-random implementation

Those functions may receive a variable number of arguments. We won’t use the argument ?.

\cs_new:Npn \__fp_parse_word_rand:N
\cs_new:Npn \__fp_parse_word_randint:N

(End of definition for \__fp_parse_word_rand:N and \__fp_parse_word_randint:N.)

79.1 Engine support

Obviously, every word “random” below means “pseudo-random”, as we have no access to entropy (except a very unreliable source of entropy: the time it takes to run some code).

The primitive random number generator (RNG) is provided as \tex_uniformdeviate:D. Under the hood, it maintains an array of 55 28-bit numbers, updated with a linear recursion relation (similar to Fibonacci numbers) modulo $2^{28}$. When \tex_uniformdeviate:D \langle integer \rangle is called (for brevity denote by $N$ the \langle integer \rangle), the next 28-bit number is read from the array, scaled by $N/2^{28}$, and rounded. To prevent 0 and $N$ from appearing half as often as other numbers, they are both mapped to the result 0.

This process means that \tex_uniformdeviate:D only gives a uniform distribution from 0 to $N-1$ if $N$ is a divisor of $2^{28}$, so we will mostly call the RNG with such power of 2 arguments. If $N$ does not divide $2^{28}$, then the relative non-uniformity (difference between probabilities of getting different numbers) is about $N/2^{28}$. This implies that detecting deviation from $1/N$ of the probability of a fixed value $X$ requires about $2^{56}/N$ random trials. But collective patterns can reduce this to about $2^{56}/N^2$. For instance with $N = 3 \times 2^k$, the modulo 3 repartition of such random numbers is biased with a non-uniformity about $2^k/2^{28}$ (which is much worse than the circa $3/2^{28}$ non-uniformity from taking directly $N = 3$). This is detectable after about $2^{56}/2^{2k} = 9 \times 2^{56}/N^2$ random numbers. For $k = 15$, $N = 98304$, this means roughly $2^{20}$ calls to the RNG (experimentally this takes at the very least 16 seconds on a 2 giga-hertz processor). While this bias is not quite problematic, it is uncomfortably close to being so, and it becomes worse as $N$ is increased. In our code, we shall thus combine several results from the RNG.
The RNG has three types of unexpected correlations. First, everything is linear modulo $2^{28}$, hence the lowest $k$ bits of the random numbers only depend on the lowest $k$ bits of the seed (and of course the number of times the RNG was called since setting the seed). The recommended way to get a number from $0$ to $N - 1$ is thus to scale the raw $28$-bit integer, as the engine’s RNG does. We will go further and in fact typically we discard some of the lowest bits.

Second, suppose that we call the RNG with the same argument $N$ to get a set of $K$ integers in $[0, N - 1]$ (throwing away repeats), and suppose that $N > K^3$ and $K > 55$. The recursion used to construct more $28$-bit numbers from previous ones is linear:

$$ x_n = x_{n-55} - x_{n-24}$$

or

$$ x_n = x_{n-55} - x_{n-24} + 2^{28}. $$

After rescaling and rounding we find that the result $N_n \in [0, N-1]$ is among $N_{n-55} - N_{n-24} + \{-1, 0, 1\}$ modulo $N$ (a more detailed analysis shows that $0$ appears with frequency close to $3/4$). The resulting set thus has more triplets $(a, b, c)$ than expected obeying $a = b + c$ modulo $N$. Namely it will have of order $(K - 55) \times 3/4$ such triplets, when one would expect $K^3/(6N)$. This starts to be detectable around $N = 2^{18} > 55^3$ (earlier if one keeps track of positions too, but this is more subtle than it looks because the array of $28$-bit integers is read backwards by the engine). Hopefully the correlation is subtle enough to not affect realistic documents so we do not specifically mitigate against this. Since we typically use two calls to the RNG per \int_rand:nn we would need to investigate linear relations between the $x_{2n}$ on the one hand and between the $x_{2n+1}$ on the other hand. Such relations will have more complicated coefficients than $\pm 1$, which alleviates the issue.

Third, consider successive batches of $165$ calls to the RNG (with argument $2^{28}$ or with argument $2$ for instance), then most batches have more odd than even numbers. Note that this does not mean that there are more odd than even numbers overall. Similar issues are discussed in Knuth’s TAOCP volume 2 near exercise 3.3.2-31. We do not have any mitigation strategy for this.

Ideally, our algorithm should be:

- **Uniform.** The result should be as uniform as possible assuming that the RNG’s underlying $28$-bit integers are uniform.
- **Uncorrelated.** The result should not have detectable correlations between different seeds, similar to the lowest-bit ones mentioned earlier.
- **Quick.** The algorithm should be fast in \TeX{}, so no “bit twiddling”, but “digit twiddling” is ok.
- **Simple.** The behaviour must be documentable precisely.
- **Predictable.** The number of calls to the RNG should be the same for any \int_-rand:nn, because then the algorithm can be modified later without changing the result of other uses of the RNG.
- **Robust.** It should work even for \int_rand:nn {- \c_max_int} \{ \c_max_int \} where the range is not representable as an integer. In fact, we also provide later a floating-point randint whose range can go all the way up to $2 \times 10^{16} - 1$ possible values.

Some of these requirements conflict. For instance, uniformity cannot be achieved with a fixed number of calls to the RNG.

Denote by random($N$) one call to \texttt{randomdeviate:D} with argument $N$, and by ediv($p,q$) the $\varepsilon$-\TeX{} rounding division giving $\lfloor p/q + 1/2 \rfloor$. Denote by $\langle \text{min} \rangle$, $\langle \text{max} \rangle$
and \( R = \langle \text{max} \rangle - \langle \text{min} \rangle + 1 \) the arguments of \texttt{\textbackslash int\_min:nn} and the number of possible outcomes. Note that \( R \in [1, 2^{32} - 1] \) cannot necessarily be represented as an integer (however, \( R = 2^{31} \) can). Our strategy is to get two 28-bit integers \( X \) and \( Y \) from the RNG, split each into 14-bit integers, as \( X = X_1 \times 2^{14} + X_0 \) and \( Y = Y_1 \times 2^{14} + Y_0 \) then return essentially \( \langle \text{min} \rangle + \lfloor R(X_1 \times 2^{-14} + Y_1 \times 2^{-28} + Y_0 \times 2^{-42} + X_0 \times 2^{-56}) \rfloor \). For small \( R \) the \( X_0 \) term has a tiny effect so we ignore it and we can compute \( R \times Y/2^{28} \) much more directly by \texttt{random}(\( R \)).

- If \( R \leq 2^{17} - 1 \) then return \texttt{ediv}(\( R \text{ random}(2^{14}) + \text{random}(\int_{\text{min}}) + 2^{13}, 2^{14}) - 1 + \langle \text{min} \rangle \)). The shifts by \( 2^{17} \) and \( -1 \) convert \( \varepsilon\)-T\texttt{p}X division to truncated division. The bound on \( R \) ensures that the number obtained after the shift is less than \( \texttt{\textbackslash c\_max\_int} \). The non-uniformity is at most of order \( 2^{17}/2^{42} = 2^{-25} \).

- Split \( R = R_2 \times 2^{28} + R_1 \times 2^{14} + R_0 \), where \( R_2 \in [0, 15] \). Compute \( \langle \text{min} \rangle + R_2 X_1 2^{14} + (R_2 Y_1 + R_1 X_1) + \text{ediv}(R_2 Y_0 + R_1 Y_1 + \text{ediv}(R_2 X_0 + R_0 Y_1 + \text{ediv}(2^{14} R_1 + R_0)(2^{14} Y_0 + X_0), 2^{28}), 2^{14}), 2^{14}) \) then map a result of \( \langle \text{max} \rangle + 1 \) to \( \langle \text{min} \rangle \). Writing each \texttt{ediv} in terms of truncated division with a shift, and using \( \lfloor (p + [r/s])q \rfloor = \lfloor (ps + r)/s \rfloor \), what we compute is equal to \( \langle \text{exact} \rangle + 2^{-29} + 2^{-15} + 2^{-1} \) with \( \langle \text{exact} \rangle = \langle \text{min} \rangle + R \times 0.X_1 Y_0 X_0 \). Given we map \( \langle \text{max} \rangle + 1 \) to \( \langle \text{min} \rangle \), the shift has no effect on uniformity. The non-uniformity is bounded by \( R/2^{56} < 2^{-24} \). It may be possible to speed up the code by dropping tiny terms such as \( R_0 X_0 \), but the analysis of non-uniformity proves too difficult.

To avoid the overflow when the computation yields \( \langle \text{max} \rangle + 1 \) with \( \langle \text{max} \rangle = 2^{31} - 1 \) (note that \( R \) is then arbitrary), we compute the result in two pieces. Compute \( \langle \text{first} \rangle = \langle \text{min} \rangle + R_2 X_1 2^{14} \) if \( R_2 < 8 \) or \( \langle \text{min} \rangle + 8 X_1 2^{14} + (R_2 - 8) X_1 2^{14} \) if \( R_2 \geq 8 \), the expressions being chosen to avoid overflow. Compute \( \langle \text{second} \rangle = R_2 Y_1 + R_1 X_1 + \text{ediv}(...) \), at most \( R_2 2^{14} + R_1 2^{14} + R_0 \leq 2^{28} + 15 \times 2^{14} - 1 \), not at risk of overflowing. We have \( \langle \text{first} \rangle + \langle \text{second} \rangle = \langle \text{max} \rangle + 1 = \langle \text{min} \rangle \) if and only if \( \langle \text{second} \rangle = R_1 2^{14} + R_0 + R_2 2^{14} \) and \( 2^{14} R_2 X_1 = 2^{28} R_2 - 2^{14} R_2 \) (namely \( R_2 = 0 \) or \( X_1 = 2^{14} - 1 \)). In that case, return \( \langle \text{min} \rangle \), otherwise return \( \langle \text{first} \rangle + \langle \text{second} \rangle \), which is safe because it is at most \( \langle \text{max} \rangle \). Note that the decision of what to return does not need \( \langle \text{first} \rangle \) explicitly so we don’t actually compute it, just put it in an integer expression in which \( \langle \text{second} \rangle \) is eventually added (or not).

- To get a floating point number in \([0, 1)\) just call the \( R = 10000 \leq 2^{17} - 1 \) procedure above to produce four blocks of four digits.

- To get an integer floating point number in a range (whose size can be up to \( 2 \times 10^{16} - 1 \)), work with fixed-point numbers: get six times four digits to build a fixed point number, multiply by \( R \) and add \( \langle \text{min} \rangle \). This requires some care because \texttt{\textbackslash 3fp\_extended} only supports non-negative numbers.

\__\_kernel\_randint\_max\_int

\begin{verbatim}
Constant equal to \( 2^{17} - 1 \), the maximal size of a range that \texttt{\textbackslash int\_range:nn} can do with its “simple” algorithm.
2073 \texttt{\textbackslash int\_const:Nn} \texttt{\textbackslash c\_kernel\_randint\_max\_int} { 131071 }

(End of definition for \texttt{\textbackslash c\_kernel\_randint\_max\_int}.)
\end{verbatim}

\__\_kernel\_randint:n

\begin{verbatim}
Used in an integer expression, \texttt{\textbackslash __kernel\_randint:n} \{ \( R \) \} gives a random number \( 1 + \lfloor (R \text{ random}(2^{14}) + \text{random}(\int_{\text{min}}))/2^{14} \rfloor \) that is in \([1, \int_{\text{min}}]\). Previous code was computing \( \lfloor p/2^{14} \rfloor \) as \texttt{ediv}(\( p - 2^{13}, 2^{14} \)) but that wrongly gives \(-1\) for \( p = 0 \).
\end{verbatim}

1222
\cs_new:Npn \__kernel_randint:n #1
\{
\(#1 * \tex_uniformdeviate:D 16384 + \tex_uniformdeviate:D \#1 + 8192 \) / 16384 \}
\endinput (End of definition for \__kernel_randint:n.)

\__fp_rand_myriads:n
\__fp_rand_myriads_loop:w
\__fp_rand_myriads_get:w

Used as \__fp_rand_myriads:n \{XXX\} with one letter X (specifically) per block of four digit we want; it expands to ; followed by the requested number of brace groups, each containing four (pseudo-random) digits. Digits are produced as a random number in [10000,19999] for the usual reason of preserving leading zeros.

\cs_new:Npn \__fp_rand_myriads:n #1
\{
\__fp_rand_myriads_loop:w #1 \prg_break: X \prg_break_point: ; \}
\cs_new:Npn \__fp_rand_myriads_loop:w #1 X
\{
\#1 \exp_after:wN \__fp_rand_myriads_get:w \int_value:w \__fp_int_eval:w 9999 + \_\kernel_randint:n \{ 10000 \} \__fp_rand_myriads_loop:w \}
\cs_new:Npn \__fp_rand_myriads_get:w 1 #1 ; { ; {#1} }
\endinput (End of definition for \__fp_rand_myriads:n, \__fp_rand_myriads_loop:w, and \__fp_rand_myriads_get:w.)

\__fp_rand_o:Nw \__fp_rand_o:w

First we check that \texttt{rand} was called without argument. Then get four blocks of four digits and convert that fixed point number to a floating point number (this correctly sets the exponent). This has a minor bug: if all of the random numbers are zero then the result is correctly \texttt{0} but it raises the \texttt{underflow} flag; it should not do that.

\cs_new:Npn \__fp_rand_o:Nw ? #1 @
\{
\tl_if_empty:nTF {#1}
\{
\exp_after:wN \__fp_rand_o:w \exp:w \exp_end_continue_f:w \__fp_rand_myriads:n \{ XXXX \} \{ 0000 \} \{ 0000 \} ; 0 \}
\}
\msg_expandable_error:nnnnn \{ fp \} \{ num-args \} \{ rand() \} \{ 0 \} \{ 0 \} \exp_after:wN \c_nan_fp
\}
\exp_after:wN \__fp_sanitize:Nw
\exp_after:wN 0 \int_value:w \__fp_int_eval:w \c_zero_int \__fp_fixed_to_float_o:wN
\}
\cs_new:Npn \__fp_rand_o:w \\
\{
\exp_after:wN \__fp_rand_o:w \exp:w \exp_end_continue_f:w \__fp_rand_myriads:n \\{ XXXX \} \{ 0000 \} \{ 0000 \} ; 0 \}
\}
\cs_new:Npn \__fp_rand_o:w \\
\{
\exp_after:wN \__fp_sanitze:Nw \exp_after:wN 0 \int_value:w \__fp_int_eval:w \c_zero_int \__fp_fixed_to_float_o:wN \}

79.2 Random floating point

1223
Random integer

Enforce that there is one argument (then add first argument) or two arguments. Call \__fp_randint_badarg: w on each; this function inserts \exp_stop_f: to end the \if_case:w statement if either the argument is not an integer or if its absolute value is \geq 10^{16}. Also bail out if \__fp_compare_back:ww yields 1, meaning that the bounds are not in the right order. Otherwise an auxiliary converts each argument times 10^{-16} (hence the shift in exponent) to a 24-digit fixed point number (see \l3fp-extended). Then compute the number of choices, \langle \max \rangle + 1 - \langle \min \rangle. Create a random 24-digit fixed-point number with \__fp_randint_myriads:n, then use a fused multiply-add instruction to multiply the number of choices to that random number and add it to \langle \min \rangle. Then truncate to 16 digits (namely select the integer part of 10^{16} times the result) before converting back to a floating point number (\__fp-sanitize:Nw takes care of zero). To avoid issues with negative numbers, add 1 to all fixed point numbers (namely 10^{16} to the integers they represent), except of course when it is time to convert back to a float.
\cs_new:Npn \__fp_randint_auxi_o:ww \#1 ; \#2 ; \#3 \exp_end:
  \{ 
    \fi:
    \__fp_randint_auxii:wn \#2 ;
    \{ \__fp_randint_auxii:wn \#1 ; \__fp_randint_auxiii_o:ww \} 
  \}
\cs_new:Npn \__fp_randint_auxii:wn \s__fp \__fp_chk:w \#1\#2\#3\#4 
  \{ 
    \if_meaning:w 0 \#1
    \exp_after:wN \use_i:nn 
    \else:
    \exp_after:wN \use_ii:nn 
    \fi:
    \{ \exp_after:wN \__fp_fixed_continue:wn \c__fp_one_fixed_tl \} 
  \}
\cs_new:Npn \__fp_randint_auxiii_o:ww \#1 ; \#2 ;
  \{ 
    \__fp_fixed_add:wwn \#2 ;
    \{0000} {0000} {0000} {0000} ;
    \__fp_fixed_sub:wwn \#1 ;
    \{ \exp_after:wN \use_i:nn \exp_after:wN \__fp_fixed_mul_add:wwwn
      \exp:w \exp_end_continue_f:w \__fp_rand_myriads:n \{ XXXXXX \}
    \} 
    \__fp_randint_auxiv_o:ww \#1 ; \#2 ;
    \__fp_randint_auxv_o:w \#1 ; \#2 ;
  \}
\cs_new:Npn \__fp_randint_auxiv_o:ww \#1\#2\#3\#4\#5 ; \#6\#7\#8\#9 
  \{ 
    \if_int_compare:w \#1\#2 > \#6\#7 \exp_stop_f: \1 \else:
      \if_int_compare:w \#1\#2 < \#6\#7 \exp_stop_f: - \fi: \fi:
    \__fp_randint_auxv_o:w \#1\#2\#3\#4\#5 ; \#6 \0 
  \}
\cs_new:Npn \__fp_randint_auxv_o:w \#1\#2\#3\#4\#5 ; \#6 \0 

1225
\begin{verbatim}
{ \exp_after:wN \__fp_sanitize:Nw
 \int_value:w
 \if_int_compare:w \#1 < 10000 \exp_stop_f:
  2
 \else:
  0
 \exp_after:wN \exp_after:wN
 \exp_after:wN \__fp_reverse_args:Nww
 \fi:
 \exp_after:wN \__fp_fixed_sub:wwn \c__fp_one_fixed_tl
 \{\#1\} \{\#2\} \{\#3\} \{\#4\} \{0000\} \{0000\} ;
 \}
 \exp_after:wN \exp_stop_f:
 \int_value:w \__fp_int_eval:w \c__fp_prec_int
 \__fp_fixed_to_float_o:wN}
0
\exp:w \exp_after:wN \exp_end:
}\end{verbatim}

(End of definition for \__fp_randint:o:Nw and others.)

\texttt{\int_rand:nn \__fp_randint:ww}

Evaluate the argument and filter out the case where the lower bound \#1 is more than the upper bound \#2. Then determine whether the range is narrower than \c__kernel_randint_max_int; \#2-\#1 may overflow for very large positive \#2 and negative \#1. If the range is narrow, call \__kernel_randint:n \{\langle \textit{choices} \rangle\} where \langle \textit{choices} \rangle is the number of possible outcomes. If the range is wide, use somewhat slower code.

\begin{verbatim}
\cs_new:Npn \int_rand:nn \#1\#2
{ \int_eval:n
  { \exp_after:wN \__fp_randint:ww
    \int_value:w \int_eval:n {\#1} \exp_after:wN ;
    \int_value:w \int_eval:n {\#2} ;
  }
}\cs_new:Npn \__fp_randint:ww \#1; \#2;
{ \if_int_compare:w \#1 > \#2 \exp_stop_f:
    \msg_expandable_error:nnnn { kernel } { randint-backward-range } {\#1} \{\#2\}
    \__fp_randint:ww \#2; \#1;
  \else:
    \if_int_compare:w \#1 > \c_zero_int
      \__fp_int_eval:w - \#1 < \__fp_int_eval:w
    \else:
      < \__fp_int_eval:w \#1 +
    \fi:
    \c__kernel_randint_max_int
    \__fp_int_eval_end:
    \__kernel_randint:n
    \__fp_int_eval:w \#2 - \#1 + 1 \__fp_int_eval_end:
}\end{verbatim}

\texttt{1226}
Any $n \in [-2^{31} + 1, 2^{31} - 1]$ is uniquely written as $2^{14}n_1 + n_2$ with $n_1 \in [-2^{17}, 2^{17} - 1]$ and $n_2 \in [0, 2^{14} - 1]$. Calling \_fp_randint_split_o:W n gives $n_1, n_2$; and expands the next token once. We do this for two random numbers and apply \_fp_randint_split_o:W twice to fully decompose the range $R$. One subtlety is that we compute $R - 2^{31} = \langle \text{max} \rangle - \langle \text{min} \rangle - (2^{31} - 1) \in [-2^{31} + 1, 2^{31} - 1]$ rather than $R$ to avoid overflow.

Then we have \_fp_randint_wide_aux:w $(X_1);(X_0);(Y_1);(Y_0);(R_2);(R_1);(R_0)$, and we apply the algorithm described earlier.

(End of definition for \texttt{\int_rand:nn} and \texttt{\_fp_randint:ww}. This function is documented on page 177.)
\input{package}

\begin{verbatim}
( #5 * #4 + #6 * #3 + #7 * #1 +
  #5 * #2 + #7 * #3 +
  (16384 * #6 + #7) * (16384 * #4 + #2) / 268435456) / 16384
  \int_value:w \__fp_int_eval:w (5 + #6) * 16384 + #7;
#1; #5;
}
\cs_new:Npn \__fp_randint_wide_auxii:w #1; #2; #3; #4;
{\if_int_odd:w 0
  \if_int_compare:w #1 = #2 \else: \exp_stop_f: \fi:
    \if_int_compare:w #4 = \c_zero_int 1 \fi:
    \exp_stop_f:
  \exp_after:wN \prg_break:
\fi:
\if_int_compare:w #4 < 8 \exp_stop_f:
  + #4 * #3 * 16384
\else:
  + 8 * #3 * 16384 + (#4 - 8) * #3 * 16384
\fi:
  + #1
\prg_break_point:
}
\end{verbatim}

(End of definition for \__kernel_randint:nn and others.)

\begin{verbatim}
\int_rand:n \__fp_randint:n
Similar to \int_rand:nn, but needs fewer checks.
\cs_new:Npn \int_rand:n #1
{\int_eval:n { \exp_args:Nf \__fp_randint:n { \int_eval:n {#1} } }
}
\cs_new:Npn \__fp_randint:n #1
{\if_int_compare:w #1 < \c_one_int
  \msg_expandable_error:nnnn { kernel } { randint-backward-range } { 1 } {#1}
  \__fp_randint:ww #1; 1;
\else:
  \if_int_compare:w #1 > \c__kernel_randint_max_int
    \__kernel_randint:nn { 1 } {#1}
\else:
  \__kernel_randint:n {#1}
\fi:
\fi:
\end{verbatim}

(End of definition for \int_rand:n and \__fp_randint:n. This function is documented on page 177.)

\end{verbatim}
Chapter 80

\textbf{l3fp-types implementation}

80.1 Support for types

Despite lack of documentation, the l3fp internals support types. Each additional type must define

- \_\_fp\_\_\texttt{type}\_chk:w;
- \_\_fp\_exp_after\_\texttt{type}\_f:nw;
- \_\_fp\_\texttt{type}\_to\_\texttt{out}:w for \texttt{out} among decimal, scientific, tl;

and may define

- \_\_fp\_\texttt{type}\_to_int:w and \_\_fp\_\texttt{type}\_to_dim:w;
- \_\_fp\_\texttt{op}\_\texttt{type}:w for any of the \texttt{op} that the type implements, among \texttt{acos}, \texttt{acsc}, \texttt{asec}, \texttt{asin}, \texttt{cos}, \texttt{cot}, \texttt{csc}, \texttt{exp}, \texttt{ln}, \texttt{not}, \texttt{sec}, \texttt{set_sign}, \texttt{sin}, \texttt{tan};
- \_\_fp\_\texttt{type1}\_\texttt{op}\_\texttt{type2}:ww for \texttt{op} among \^{\texttt{*/+-&|}} and for every pair of types;
- \_\_fp\_\texttt{type1}\_\texttt{bcmp}\_\texttt{type2}:ww for every pair of types.

The latter is set up in l3fp-logic.

80.2 Dispatch according to the type

From \_\_fp\_\texttt{op}:w produce \texttt{op}, otherwise \texttt{?}.
\cs_new:Npn \_fp_types_binary_auxii:NNww #1#2
    {
        \token_if_eq_meaning:NNTF \scan_stop: #1
            { \_fp_invalid_operation_o:Nww #2 } #1
    }

(End of definition for \_fp_types_binary:Nww, \_fp_types_binary_auxi:Nww, and \_fp_types_binary_auxii:Nww.)

//package
Chapter 81

l3fp-symbolic implementation

81.1 Misc

\l__fp_symbolic_fp

Scratch floating point.

\fp_new:N \l__fp_symbolic_fp

(End of definition for \l__fp_symbolic_fp.)

81.2 Building blocks for expressions

Every symbolic expression has the form \s__fp_symbolic \__fp_symbolic_chk:w
⟨operation⟩, {⟨operands⟩} ⟨junk⟩; where the ⟨operation⟩ is a list of N-type tokens,
the ⟨operands⟩ is an array of floating point objects, and the ⟨junk⟩ is to be discarded.
If the outermost operator (last to be evaluated) is unary, the expression has the form

\s__fp_symbolic \__fp_symbolic_chk:w
\__fp_types_unary:Nww \__fp_⟨op⟩_o:w ⟨token⟩ ,
{ ⟨operand⟩ } ⟨junk⟩;

where the ⟨op⟩ is a unary operation (set_sign, cos, ...), and the ⟨token⟩ and ⟨operand⟩
are used as arguments for \__fp_⟨op⟩_o:w (or the type-specific analog of this function).
If the outermost operator is binary, the expression has the form

\s__fp_symbolic \__fp_symbolic_chk:w
\__fp_types_binary:Nww \__fp_⟨op⟩_o:ww ,
{ ⟨operand1⟩ ⟨operand2⟩ } ⟨junk⟩ ;

where the ⟨op⟩ is an operation (+, &, ...), and \__fp_⟨op⟩_o:ww receives the ⟨operands⟩
as arguments. If the expression consists of a single variable, it is stored as

\s__fp_symbolic \__fp_symbolic_chk:w
\__fp_variable_o:w ⟨identifier⟩ ,
{ } ⟨junk⟩ ;
Symbolic expressions are stored in a prefix form. When encountering a symbolic expression in a floating point computation, we attempt to evaluate the operands as much as possible, and if that yields floating point numbers rather than expressions, we apply the operator which follows (if the function is known).

For instance, the expression \( a + b \times \sin(c) \) is stored as

\[
\text{__fp_symbolic} \quad \text{__fp_symbolic_chk:w} \\
\text{__fp_types_binary:Nww} \quad \text{__fp+_o:ww} , \\
\{ \\
\text{__fp_symbolic} \quad \text{__fp_symbolic_chk:w} \\
\text{__fp_variable_o:w} \quad a , \{ \} ; \\
\text{__fp_symbolic} \quad \text{__fp_symbolic_chk:w} \\
\text{__fp_types_binary:Nww} \quad \text{__fp_*_o:ww} , \\
\{ \\
\text{__fp_symbolic} \quad \text{__fp_symbolic_chk:w} \\
\text{__fp_variable_o:w} \quad b , \{ \} ; \\
\text{__fp_symbolic} \quad \text{__fp_symbolic_chk:w} \\
\text{__fp_types_unary:NNw} \quad \text{__fp_sin_o:w} \quad \text{use_i:nn} , \\
\{ \\
\text{__fp_symbolic} \quad \text{__fp_symbolic_chk:w} \\
\text{__fp_variable_o:w} \quad c , \{ \} ; \\
\} ; \\
\}
\]

\text{Scan mark indicating the start of a symbolic expression.}

\text{End of definition for} \text{__fp_symbolic.} 

\text{__fp_symbolic_chk:w} 

\text{Analog of} \text{__fp_chk:w} \text{ for symbolic expressions.}

\text{__fp_if_has_symbolic:nTF} \text{\_fp_if_has_symbolic_aux:w} 

\text{Tests if} \#1 \text{ contains} \text{__fp_symbolic} \text{ at top-level. This test should be precise enough to determine if a given array contains a symbolic expression or only consists of floating points. Used in} \text{__fp_exp_after_symbolic_f:nu.}

\text{End of definition for} \text{__fp_symbolic_chk:w.} 

\textbf{81.3 Expanding after a symbolic expression}
This function does two things: trigger an f-expansion of the argument #1 after the following symbolic expression, and evaluate all pieces of the expression which can be evaluated.

81.4 Applying infix operators to expressions

\__fp_symbolic_binary_o:Nww

Used when applying infix operators to expressions.
81.5 Applying prefix functions to expressions

Used when applying infix operators to expressions.

(End of definition for \_fp_symbolic_unary_o:NNw.)
81.6 Conversions

Symbolic expressions cannot be converted to decimal, integer, or scientific notation unless they can be reduced to

\[
\text{\texttt{cs_set_protected:Npn \_\_fp_tmp:w #1#2#3}}
\]

\[
\begin{align*}
\text{\texttt{\_\_fp_symbolic_to_decimal:w}} & \quad \text{\texttt{\_\_fp_symbolic_to_int:w}} \\
\text{\texttt{\_\_fp_symbolic_to_scientific:w}} & \quad \text{\texttt{\_\_fp_symbolic_convert:wnnN}}
\end{align*}
\]

Converting a symbolic expression to a token list is possible.

\[
\text{\texttt{\_\_fp_symbolic_to_tl:w}}
\]

\[
\begin{align*}
\text{\texttt{\_\_fp_symbolic_to_int:NNw}} & \quad \text{\texttt{\_\_fp_symbolic_to_scientific:NNw}} \\
\text{\texttt{\_\_fp_symbolic_convert:vnw}} & \quad \text{\texttt{\_\_fp_symbolic_convert:vnw}}
\end{align*}
\]
\_{\__fp_types_binary:Nww } \{ \_\_fp_symbolic_binary_to_tl:Nww \} \\
\{ \_\_fp_function_o:w \} \{ \_\_fp_symbolic_function_to_tl:Nw \} 

\{ #2, #3 @ \}

\{ \tl_to_str:n {#2} \}

\cs_new:Npn \_\_fp_symbolic_unary_to_tl:NNw #1#2 , #3 @
\{ \use:e \{ \_\_fp_symbolic_cs_arg_to_fn:NN #1#2 \( \_\_fp_to_tl_dispatch:w #3 \) \} \}

\cs_new:Npn \_\_fp_symbolic_binary_to_tl:Nww #1, #2; #3; @
\{ \use:e \{ \_\_fp_types_cs_to_op:N #1 \( \_\_fp_to_tl_dispatch:w #2; \) \_\_fp_types_cs_to_op:N #1 \( \_\_fp_to_tl_dispatch:w #3; \) \} \}

\cs_new:Npn \_\_fp_symbolic_function_to_tl:Nw #1, #2@
\{ \use:e \{ \_\_fp_types_cs_to_op:N #1 \( \_\_fp_array_to_clist:n {#2} \) \} \}

(End of definition for \_\_fp_symbolic_to_tl:w and others.)

81.7 Identifiers

Functions defined here are not necessarily tied to symbolic expressions.

If \#1 contains a space, it is not a valid identifier. Otherwise, loop through letters in \#1: if it is not a letter, break the loop and return true. If the end of the loop is reached without finding any non-letter, return false. Note \#1 must be a str (i.e., resulted from \_\_fp_to_tl_dispatch:w).
\cs_new:Npn \__fp_id_if_invalid_aux:N #1 
{ 
\use_none:n #1 
\int_compare:nF { 'a <= '#1 <= 'z } 
{ 
\int_compare:nF { 'A <= '#1 <= 'Z } 
{ \prg_break:n \prg_return_true: } 
} 
\__fp_id_if_invalid_aux:N 
} 

(End of definition for \__fp_id_if_invalid:nTF and \__fp_id_if_invalid_aux:N.)

\__fp_variable_o:w
We do not use \exp_last_unbraced:Nv to extract the value of \l__fp_variable_#1_fp because in \fp_set_variable:nn we define this fp variable to be something which f-expands to an actual floating point, rather than a genuine floating point.

\cs_new:Npn \__fp_variable_o:w #1 @ #2
{ 
\fp_if_exist:cTF { l__fp_variable_#1_fp } 
{ \exp_last_unbraced:Nf \__fp_exp_after_array_f:w 
\{ \use:c { l__fp_variable_#1_fp } \} \s__fp_expr_stop 
\exp_after:wN \exp_stop_f: #2 
} 
\token_if_eq_meaning:NNTF #2 \prg_do_nothing: 
{ 
\s__fp_symbolic \__fp_symbolic_chk:w 
\__fp_variable_o:w #1 , { } ; 
} 
\{ 
\exp_after:wN \s__fp_symbolic 
\exp_after:wN \__fp_symbolic_chk:w 
\__fp_variable_o:w \#1 , { } ; 
} 
\{ 
\exp_after:wN \s__fp_symbolic 
\exp_after:wN \__fp_symbolic_chk:w 
\exp_after:wN \__fp_variable_o:w 
\exp:w 
\__fp_exp_after_symbolic_loop:N \#1 
{ , \exp:w \use_none:nn } 
\exp_after:wN \exp_end: 
\exp_after:wN \{ \exp_after:wN \} \exp_after:wN ; 
#2 
} 

(End of definition for \__fp_variable_o:w.)
\cs_new_protected:Npn \__fp_variable_set_parsing:Nn #1#2
  {
    \cs_set:Npn \__fp_tmp:w
    {
      \__fp_exp_after_symbolic_f:nw { \__fp_parse_infix:NN }
      \s__fp_symbolic \__fp_symbolic_chk:w
      \__fp_variable_o:w #2 , { } ;
    }
    \exp_args:NNc \__fp_variable_set_parsing_aux:NNn #1
    { __fp_parse_word_#2:N } {#2}
  }
\cs_new_protected:Npn \__fp_variable_set_parsing_aux:NNn #1#2#3
  {
    \cs_if_eq:NNF #2 \__fp_tmp:w
      {
        \cs_if_exist:NTF #2
          {
            \msg_warning:nnnn { fp } { id-used-elsewhere } {#3} { variable }
            #1 #2 \__fp_tmp:w
          }
          {
            \cs_new_eq:NN #2 \scan_stop: % to declare the function
            #1 #2 \__fp_tmp:w
          }
      }
  }
\cs_new_protected:Npn \fp_clear_variable:n #1
  {
    \exp_args:No \__fp_clear_variable:n { \tl_to_str:n {#1} }
  }
\cs_new_protected:Npn \__fp_clear_variable:n #1
  {
    \__fp_id_if_invalid:nTF {#1}
      {
        \msg_error:nnn { fp } { id-invalid } {#1}
        \__fp_clear_variable_aux:n {#1}
      }
  }
\cs_new_protected:Npn \__fp_clear_variable_aux:n #1
  {
    \cs_set_eq:cN { l__fp_variable_#1_fp } \tex_undefined:D
    \__fp_variable_set_parsing:Nn \cs_set_eq:NN {#1}
  }
\fp_clear_variable:n \__fp_clear_variable:n
\__fp_clear_variable_aux:n
\fp_new_variable:n \__fp_new_variable:n
\__fp_new_variable_aux:n
\fp_new_variable:n
\__fp_new_variable:n
\__fp_new_variable_aux:n
\fp_new_variable:n
\__fp_new_variable:n
\__fp_new_variable_aux:n
\fp_new_variable:n
\__fp_new_variable:n
\__fp_new_variable_aux:n
\fp_new_variable:n
\__fp_new_variable:n
\__fp_new_variable_aux:n
(End of definition for \__fp_variable_set_parsing:Nn and \__fp_variable_set_parsing_aux:Nnn.)
(End of definition for \__fp_variable_set_parsing:Nn and \__fp_variable_set_parsing_aux:Nnn.)
(End of definition for \__fp_variable_set_parsing:Nn and \__fp_variable_set_parsing_aux:Nnn. This function is documented on page 268.)
\exp_args:No \__fp_new_variable:n \tl_to_str:n \#1
\exp_args:No \__fp_new_variable:n \#1
\__fp_id_if_invalid:NT \#1
\msg_error:nnn \fp \id-invalid \#1
\cs_if_exist:cT \__fp_parse_word_#1:N
\msg_error:nnn \fp \id-already-defined \#1
\cs_undefine:c \__fp_parse_word_#1:N \tex_undefined:D
\__fp_variable_set_parsing:Nn \cs_gset_eq:NN \#1
\flag_new:N \l__fp_symbolic_flag
\cs_new_protected:Npn \fp_set_variable:nn \#1 \#2
\exp_args:No \__fp_set_variable:nn \tl_to_str:n \#1
\__fp_id_if_invalid:nTF \#1
\msg_error:nnn \fp \id-invalid \#1
\cs_if_exist:cT \__fp_parse_word_\#1:N
\msg_error:nnn \fp \id-already-defined \#1
\cs_undefine:c \__fp_parse_word_\#1:N \tex_undefined:D
\__fp_variable_set_parsing:Nn \cs_gset_eq:NN \#1
\flag_new:N \l__fp_symbolic_flag
\cs_new_protected:Npm \fp_set_variable:nn \#1
\exp_args:No \__fp_set_variable:nn \tl_to_str:n \#1
\cs_new_protected:Npm \fp_set_variable:nn \#1\#2
\__fp_id_if_invalid:NT \#1
\msg_error:nnn \fp \id-invalid \#1
\__fp_variable_set_parsing:Nn \cs_gset_eq:NN \#1
\cs_set_eq:NN \l__fp_variable_\#1_fp \tex_undefined:D
\flag_ensure_raised:N \l__fp_symbolic_flag \c_nan_fp
\flag_clear:N \l__fp_symbolic_flag
\fp_set:cn \l__fp_variable_\#1_fp \l__fp_symbolic_fp
\flag_if_raised:NT \l__fp_symbolic_flag
\msg_error:nnee \fp \id-loop \#1
\msg_error:nnee \fp \id-loop \#2

\l__fp_symbolic_flag
\fp_set_variable:nn
\__fp_set_variable:nn

Refuse invalid identifiers. If the variable does not exist yet, define it just as in \fp_new_variable:n (but without unnecessary checks). Then evaluate \#2. If the result contains the identifier \#1, we would later get a loop in cases such as

\fp_set_variable:nn \{A\} \{A\}
\fp_show:n \{A\}

To detect this, define \l__fp_variable_\#1_fp to raise an internal flag and evaluate to nan. Then re-evaluate \l__fp_symbolic_fp, and store the result in \#1. If the flag is raised, \#1 was present in \l__fp_symbolic_fp. In all cases, the \#1-free result ends up in \l__fp_variable_\#1_fp.

\flag_new:N \l__fp_symbolic_flag
\cs_new_protected:Npm \fp_set_variable:nn \#1
\exp_args:No \__fp_set_variable:nn \tl_to_str:n \#1
\cs_new_protected:Npm \fp_set_variable:nn \#1\#2
\__fp_id_if_invalid:NT \#1
\msg_error:nnn \fp \id-invalid \#1
\__fp_variable_set_parsing:Nn \cs_gset_eq:NN \#1
\cs_set_eq:NN \l__fp_variable_\#1_fp \tex_undefined:D
\flag_ensure_raised:N \l__fp_symbolic_flag \c_nan_fp
\flag_clear:N \l__fp_symbolic_flag
\fp_set:cn \l__fp_variable_\#1_fp \l__fp_symbolic_fp
\flag_if_raised:NT \l__fp_symbolic_flag
\msg_error:nnee \fp \id-loop \#1
\msg_error:nnee \fp \id-loop \#2
\tl_to_str:n \#2
(End of definition for \_\_fp_symbolic_flag, \_\_fp_set_variable:nn, and \_\_fp_set_variable:na. This variable is documented on page \texttt{267}.)

### 81.9 Messages

\begin{verbatim}
\msg_new:nnnn { fp } { id-invalid } { Floating-point-identifier-'#1'-invalid. }
\{ \LaTeX\-has\-been\-asked\-to\-create\-a\-new\-floating\-point\-identifier-'#1'-but\-this\-may\-only\-contain-ASCII\-letters. \}
\msg_new:nnnn { fp } { id-already-defined } { Floating-point-identifier-'#1'-already-defined. }
\{ \LaTeX\-has\-been\-asked\-to\-create\-a\-new\-floating\-point\-identifier-'#1'-but\-this\-name\-has\-already\-been\-used\-elsewhere. \}
\msg_new:nnnn { fp } { id-used-elsewhere } { Floating-point-identifier-'#1'-already-used-for-something-else. }
\{ \LaTeX\-has\-been\-asked\-to\-create\-a\-new\-floating\-point\-identifier-'#1'-but\-this\-name\-is\-used,-and\-is\-not\-a\-user-defined-#2. \}
\msg_new:nnnn { fp } { id-loop } { Variable-'#1'-used-in-the-definition-of-'#1'. }
\{ \LaTeX\-has\-been\-asked\-to\-set\-the\-floating\-point\-identifier-'#1'-to\-the\-expression-'#2'.\-Evaluating\-this\-expression\-yields-'#3',-which\-contains-'#1'-itself. \}
\end{verbatim}

### 81.10 Road-map

The following functions are not implemented: \texttt{min}, \texttt{max}, \texttt{?:}, comparisons, \texttt{round}, \texttt{atan}, \texttt{acot}.
Chapter 82

l3fp-functions implementation

82.1 Declaring functions

\fp_new_function:n
\__fp_new_function:n
\cs_new_protected:Npn \fp_new_function:n #1
{ \exp_args:No \__fp_new_function:n { \tl_to_str:n {#1} } }
\cs_new_protected:Npn \__fp_new_function:n #1
{ \__fp_id_if_invalid:nTF {#1}
  \msg_error:nnn { fp } { id-invalid } {#1}
  \cs_if_exist:cT { __fp_parse_word_#1:N }
  \msg_error:nnn { fp } { id-already-defined } {#1}
  \cs_undefine:c { __fp_parse_word_#1:N }
  \cs_undefine:c { __fp_#1_o:w }
  \__fp_function_set_parsing:Nn \cs_gset_eq:NN {#1} }
\cs_new_protected:Npn \__fp_function_set_parsing:Nn #1#2
{ \exp_args:NNc \__fp_function_set_parsing_aux:NNn #1
  { __fp_parse_word_#2:N } {#2} }
\cs_new_protected:Npn \__fp_function_set_parsing_aux:NNn #1#2#3
{ \cs_set:Npe \__fp_tmp:w
  { \exp_not:N \__fp_parse_function:NNN }
  \cs_new_protected:Npm \__fp_function:NNN #1#2#3
}\exp_not:N \__fp_parse_function:NNN

(End of definition for \fp_new_function:n and \__fp_new_function:n. This function is documented on page 268.)
\def\__fp_function_o:wn #1#2 @ {
  \ifcs_exists:nTF #1
    { #1 #2 @ }
  { \exp_after:wN \s__fp_symbolic \exp_after:wN \__fp_symbolic_chk:w \exp_after:wN \__fp_function_o:w \exp_after:wN , \exp:w \exp_end_continue_f:w \__fp_exp_after_array_f:w #2 \s__fp_expr_stop \exp_after:wN ; }
}\__fp_function_o:w

\def\__fp_function_arg_int
\int_new:N \__fp_function_arg_int

\__fp_set_function:Nnnn {⟨identifier⟩} {⟨comma-list of variables⟩} {⟨expression⟩}

\def\fp_set_function:nnn
\__fp_set_function:Nnnn {⟨identifier⟩} {⟨comma-list of variables⟩} {⟨expression⟩}

82.2 Defining functions by their expression

\def\__fp_function_arg_int
\int_new:N \__fp_function_arg_int

(End of definition for \__fp_function_o:w.)

(End of definition for \__fp_function_set_parsing:Nn and \__fp_function_set_parsing_aux:Nnn.)

(End of definition for \__fp_function_o:w.)
\cs_new_protected:Npn \fp_set_function:nnn \l_{\text{\texttt{#1}}} \\
\{ \exp_args:NNo \__fp_set_function:Nnnn \cs_set_eq:cN \\\n{ \tl_to_str:n {#1} } \}
\cs_new_protected:Npn \__fp_set_function:Nnnn \l_{\text{\texttt{#1}#2#3#4}} \\
{ \__fp_id_if_invalid:nTF {#2} \\
{ \msg_error:n { \text{\texttt{fp}}} { id-invalid } {#2} } \\
{ \cs_if_exist:cF { \_\_\_\fp_parse_word_#2:N } \\
{ \_\_\_\fp_function_set_parsing:Nn \cs_set_eq:NN {#2} } \\
\group_begin: \\
\int_zero:N \l_{\text{\texttt{#3}}} \\
\exp_args:No \clist_map_inline:nn { \tl_to_str:n {#3} } \\
\group_end: \\
\int_incr:N \l_{\text{\texttt{#3}}} \\
\exp_args:Ne \__fp_clear_variable_aux:n \\
{ \c_underscore_str \tex_romannumeral:D \l_{\text{\texttt{#3}}} } \\
\fp_clear_variable:n {##1} \\
\cs_set_nopar:cpe { l_{\text{\texttt{variable##1}}}} \\
{ \exp_not:N \s__fp_symbolic \\
\exp_not:N \_\_\_\fp_function_arg_o:w \\
\int_use:N \l_{\text{\texttt{#3}}} } \\
##1 , { } ; \\
\exp_after:wN \c_zero_int \\
\exp_after:wN { \exp_after:wN } \exp_after:wN ; \\
\fp_set:Nn \l_{\text{\texttt{symbolic_fp}}} {#4} \\
\exp_after:wN \cs_gset:Npn \__\_\_\_\_\text{\texttt{tmp:w}} {#4} \\
{ \exp_not:o \exp_not:n { \__\_\_\_\_\text{\texttt{tmp:w}}{#4} } } \\
\use:e \\
{ \exp_not:n \cs_gset:Npn \__\_\_\_\_\text{\texttt{tmp:w}} {#4} } \\
{ \exp_not:N \__\_\_\_\fp_exp_after_symbolic_f:nw \\
\text{\texttt{stop_f:}} } \\
1244
\_\_fp\_function\_arg\_o:w
\_\_fp\_function\_arg\_few:w
\_\_fp\_function\_arg\_get:w
\cs\_new:Npn \_\_fp\_function\_arg\_o:w \#1. \#2 {
    \if\meaning:w @ \#2
        \exp_after:wN \_\_fp\_function\_arg\_few:w
    \fi:
    \if_int_compare:w \#1 = \c\_one_int
        \exp_after:wN \_\_fp\_function\_arg\_get:w
    \fi:
    \_\_fp\_use\_i\_until\_s:nw
    { \exp_after:wN \_\_fp\_function\_arg\_o:w \int_value:w \text{int_eval:n} \{ \#1 - 1 \} . }
\}
\cs\_new:Npn \_\_fp\_function\_arg\_few:w \#1 @ { \exp_after:wN \c\_\text{nan}\_fp }
\cs\_new:Npn \_\_fp\_function\_arg\_get:w \#1\#2\#3; \#4 @ {
    \_\_fp\_exp\_after\_array\_f:w \#3; \s\_\_fp\_expr\_stop
    \exp_after:wN \text{exp\_stop:f}:
}

(End of definition for \texttt{fp\_set\_function:nnn} and others. This function is documented on page 268.)
\texttt{fp\_clear\_function:n}
\_\_fp\_clear\_function:n
\cs\_new\_protected:Npn \texttt{fp\_clear\_function:n} \#1 {
    \exp\_args:No \_\_fp\_clear\_function:n \{ \tl\_to\_str:n \{\#1\} \} }
\cs\_new\_protected:Npn \_\_fp\_clear\_function:n \#1 {
    \_\_fp\_id\_if\_invalid:nTF \#1
    { \msg\_error:nnn \{fp\} \{id\text{\textunderscore}invalid\} \{\#1\} }
    {\cs\_set\_eq:cN \_\_fp\_#1:o:w \text{\textunderscore}\text{\textunderscore}define:D
    \_\_fp\_function\_set\_parsing:Nn \cs\_set\_eq:NN \{\#1\} }
}

(End of definition for \texttt{fp\_clear\_function:n} and \_\_fp\_clear\_function:n. This function is documented on page 268.)
Chapter 83

\textbf{l3fparray implementation}

In analogy to \texttt{l3intarray} it would make sense to have \texttt{<@@=fparray>}, but we need direct access to \texttt{\_\_fp_parse:n} from \texttt{l3fp-parse}, and a few other (less crucial) internals of the \texttt{l3fp} family.

\section{Allocating arrays}

There are somewhat more than \((2^{31} - 1)^2\) floating point numbers so we store each floating point number as three entries in integer arrays. To avoid having to multiply indices by three or to add 1 etc, a floating point array is just a token list consisting of three tokens: integer arrays of the same size.

\begin{verbatim}
\\_\_fp\_array\_int
Used to generate unique names for the three integer arrays.
\end{verbatim}

\begin{verbatim}
\int_new:N \_\_fp\_array\_int
(End of definition for \_\_fp\_array\_int.)
\end{verbatim}

\begin{verbatim}
\_\_fp\_array\_loop\_int
Used to loop in \_\_fp\_array\_gzero:n.
\end{verbatim}

\begin{verbatim}
\int_new:N \_\_fp\_array\_loop\_int
(End of definition for \_\_fp\_array\_loop\_int.)
\end{verbatim}

\begin{verbatim}
\_fparray\_new:Nn
\_fparray\_new:cn
\_fparray\_new:nNNN
Build a three-token token list, then define all three tokens to be integer arrays of the same size. No need to initialize the data: the integer arrays start with zeros, and three zeros denote precisely \texttt{\_c\_zero\_fp}, as we want.
\end{verbatim}

\begin{verbatim}
\cs_new_protected:Npn \_fparray\_new:Nn #1#2
\tl_new:N #1
\prg_replicate:nn { 3 }
\int_gincr:N \_fp\_array\_int
\exp_args:NNc \tl_gput_right:Nn #1 { g\_fp\_array\_\_\_fp\_int\_to\_roman:w \_fp\_array\_int\_intarray }
\exp_last_unbraced:Nfo \_fp\_array\_new:nNNN
\end{verbatim}

\begin{verbatim}
\int_eval:n {#2} #1 #1
\end{verbatim}


83.2 Array items

See the \texttt{l3intarray} analogue: only names change. The functions \texttt{fparray_gset:Nnn} and \texttt{fparray_item:Nn} share bounds checking. The T branch is used if \texttt{#3} is within bounds of the array \texttt{#2}.

\begin{verbatim}
\cs_new:Npn \__fp_array_bounds:NNnTF #1#2#3#4#5
    {\if_int_compare:w 1 > #3 \exp_stop_f:
        \__fp_array_bounds_error:NNn #1 #2 {#3}
        #5
      \else:
        \if_int_compare:w #3 > \fparray_count:N #2 \exp_stop_f:
            \__fp_array_bounds_error:NNn #1 #2 {#3}
            #5
        \else:
            #4
        \fi:
        \fi:
    }
\cs_new:Npn \__fp_array_bounds_error:NNn #1#2#3
    {#1 { kernel } { out-of-bounds } \token_to_str:N #2 }{#3} { \fparray_count:N #2 }
\end{verbatim}

(End of definition for \texttt{fparray_count:N} and \texttt{__fp_array_bounds:NNnTF}. This function is documented on page 284.)

\section*{83.2 Array items}

See the \texttt{l3intarray} analogue: only names change. The functions \texttt{fparray_gset:Nnn} and \texttt{fparray_item:Nn} share bounds checking. The T branch is used if \texttt{#3} is within bounds of the array \texttt{#2}.

\begin{verbatim}
\cs_new:Npn \__fp_array_bounds:NNnTF #1#2#3#4#5
    {\if_int_compare:w 1 > #3 \exp_stop_f:
        \__fp_array_bounds_error:NNn #1 #2 {#3}
        #5
      \else:
        \if_int_compare:w #3 > \fparray_count:N #2 \exp_stop_f:
            \__fp_array_bounds_error:NNn #1 #2 {#3}
            #5
        \else:
            #4
        \fi:
        \fi:
    }
\cs_new:Npn \__fp_array_bounds_error:NNn #1#2#3
    {#1 { kernel } { out-of-bounds } \token_to_str:N #2 }{#3} { \fparray_count:N #2 }
\end{verbatim}

(End of definition for \texttt{fparray_count:N} and \texttt{__fp_array_bounds:NNnTF}. This function is documented on page 284.)
\fparray_gset:NNn
\fparray_gset:cn
\__fp_array_gset:NNNNww
\__fp_array_gset_recover:Nw
\__fp_array_gset_special:nnNN
\__fp_array_gset_normal:w

Evaluate, then store exponent in one intarray, sign and 8 digits of mantissa in the next, and 8 trailing digits in the last.

\cs_new_protected:Npn \fparray_gset:Nnn #1#2#3
\exp_after:wN \exp_after:wN \exp_after:wN \__fp_array_gset:NNNNww \exp_after:wN \exp_after:wN \exp_after:wN \exp_after:wN \int_value:w \int_eval:n {#2} \exp_after:wN ;\exp:w \exp_end_continue_f:w \__fp_parse:n {#3}
\cs_generate_variant:Nn \fparray_gset:Nnn { c }
\cs_new_protected:Npn \__fp_array_gset:NNNNww #1#2#3#4#5 ; #6 ;
\__fp_array_bounds:NNnTF \msg_error:nneee #4 {#5}
\exp_after:wN \__fp_change_func_type:NNN \__fp_use_i_until_s:nw #6 ;\__fp_array_gset:w \__fp_array_gset_recover:Nw #6 ; {#5} #1 #2 #3
\cs_new_protected:Npn \__fp_array_gset_recover:Nw #1#2 ;
\__fp_error:nffn { unknown-type } { \tl_to_str:n { #2 ; } } { } { }
\exp_after:wN #1 \c_nan_fp
\cs_new_protected:Npn \__fp_array_gset:w \s__fp \__fp_chk:w #1#2
\if_case:w #1 \exp_stop_f:
\__fp_case_return:nw { \__fp_array_gset_special:nnNNN {#2} }
or: \exp_after:wN \__fp_array_gset_normal:w
\__fp_case_return:nw { \__fp_array_gset_special:nnNNN { #2 3 } }
or: \exp_after:wN \__fp_array_gset_normal:w
\__fp_case_return:nw { \__fp_array_gset_special:nnNNN { 1 } }
fi:
\s__fp \__fp_chk:w #1 #2
\cs_new_protected:Npn \__fp_array_gset_normal:w
\s__fp \__fp_chk:w 1 #1 #2 #3#4#5 ; #6#7#8#9
\cs_new_protected:Npn \__kernel_intarray_gset:Nnn #7 {#6} {#2}
\cs_new_protected:Npn \__kernel_intarray_gset:Nnn #8 {#6}
\if_meaning:w 2 #1 3 \else: 1 \fi: #3#4 }
\__kernel_intarray_gset:Nnn #9 {#6} { 1 \use:nn #5 }
\cs_new_protected:Npn \__kernel_intarray_gset:Nnn #3 {#2} {#1}
\__kernel_intarray_gset:Nnn #4 {#2} {0}
\__kernel_intarray_gset:Nnn #5 {#2} {0}
1248
(End of definition for \texttt{\fparray_gzero:Nn} and others. This function is documented on page 283.)

\begin{verbatim}
\cs_new_protected:Npn \fparray_gzero:N #1 {
\int_zero:N \l__fp_array_loop_int
\prg_replicate:nn { \fparray_count:N #1 } {
\int_incr:N \l__fp_array_loop_int
\exp_after:wN \__fp_array_gset_special:nnNNN
\exp_after:wN 0
\exp_after:wN \l__fp_array_loop_int
#1 }
}
\cs_generate_variant:Nn \fparray_gzero:N { c }
\end{verbatim}

(End of definition for \texttt{\fparray_gzero:N}. This function is documented on page 283.)

\begin{verbatim}
\cs_new:Npn \fparray_item:Nn #1#2 {
\exp_after:wN \__fp_array_item:NwN
\exp_after:wN #1
\int_value:w \int_eval:n {#2} ;
\__fp_to_decimal:w
}
\cs_generate_variant:Nn \fparray_item:Nn { c }
\cs_new:Npn \fparray_item_to_tl:Nn #1#2 {
\exp_after:wN \__fp_array_item:NwN
\exp_after:wN #1
\int_value:w \int_eval:n {#2} ;
\__fp_to_tl:w
}
\cs_generate_variant:Nn \fparray_item_to_tl:Nn { c }
\cs_new:Npn \__fp_array_item:NwN #1#2 ; #3 {
\if_meaning:w 0 #1 \exp_after:wN \__fp_array_item_special:w \fi:
\__fp_array_item:w #1
\end{verbatim}
\cs_new:Npn \_fp_array_item:w #1 #2#3#4#5 #6 ; 1 #7 ;
{ \exp_after:wN \_fp_array_item_normal:w \int_value:w \if_meaning:w #1 1 0 \else: 2 \fi: \exp_stop_f: #7 ; {#2#3#4#5} {#6} ; }
\cs_new:Npn \_fp_array_item_special:w #1 ; #2 ; #3 ; #4
{ \exp_after:wN #4 \exp:w \exp_end_continue_f:w \if_case:w #3 \exp_stop_f: \exp_after:wN \c_zero_fp \or: \exp_after:wN \c_nan_fp \or: \exp_after:wN \c_minus_zero_fp \or: \exp_after:wN \c_inf_fp \else: \exp_after:wN \c_minus_inf_fp \fi: \fi: }
\cs_new:Npn \_fp_array_item_normal:w #1 #2#3#4#5 #6 ; #7 ; #8 ; #9
{ #9 \s__fp \__fp_chk:w 1 #1 {#8} #7 {#2#3#4#5} {#6} ; }

\fparray_if_exist_p:N Copies of the cs functions defined in l3basics.
\fparray_if_exist:p\c\fparray_if_exist:NTF\fparray_if_exist:c\TF

(End of definition for \fparray_item:Nn and others. These functions are documented on page 284.)

\fparray_if_exist:p\N
\fparray_if_exist:p\c
\fparray_if_exist:NTF
\fparray_if_exist:c\TF

(End of definition for \fparray_if_exist:NTF. This function is documented on page 284.)
Chapter 84

\l3bitset implementation

\bitset_new:N
\bitset_new:c
\bitset_new:Nn
\cs_new_protected:Npn \bitset_new:N #1
\cs_gset_eq:NN #1 \c_zero_str
\prop_new:c { g__bitset_ \cs_to_str:N #1 _name_prop }
\prop_gset_from_keyval:cn
\prop_gput_from_keyval:cn
\bitset_addto_named_index:Nn
\cs_new_protected:Npn \bitset_addto_named_index:Nn #1\#2
\prop_gput_from_keyval:cn
\bitset_new:Nn \bitset_new:Nn \bitset_new:Nn { c } { #2 }

A bitset is a string variable.

End of definition for \bitset_new:N and \bitset_new:Nn. These functions are documented on page 286.
(End of definition for \bitset_addto_named_index:N. This function is documented on page 286.)

\bitset_if_exist_p:N
\bitset_if_exist_p:c
\bitset_if_exist:NTF
\bitset_if_exist:cTF

Existence tests.

\bitset_if_exist:N \str_if_exist:N { p , T , F , TF }
\bitset_if_exist:c \str_if_exist:c { p , T , F , TF }

(End of definition for \bitset_if_exist:NTF. This function is documented on page 287.)

\bitset_if_exist:NTF
\bitset_if_exist:TF
\bitset_if_exist:N
\bitset_if_exist:c

The internal command uses only numbers (integer expressions) for the position. A bit is set by either extending the string or by splitting it and then inserting an 1. It is not checked if the value was already 1.

\cs_new_protected:Npn \__bitset_set_true:Nn #1#2
{ \__bitset_set:NNnN \str_set:Ne #1 {#2} 1 }
\cs_new_protected:Npn \__bitset_gset_true:Nn #1#2
{ \__bitset_set:NNnN \str_gset:Ne #1 {#2} 1 }
\cs_new_protected:Npn \__bitset_set_false:Nn #1#2
{ \__bitset_set:NNnN \str_set:Ne #1 {#2} 0 }
\cs_new_protected:Npn \__bitset_gset_false:Nn #1#2
{ \__bitset_set:NNnN \str_gset:Ne #1 {#2} 0 }
\cs_new_protected:Npn \__bitset_set:NNnN #1#2#3#4
{ \int_compare:nNnTF {\str_count:N #2} < {#3} 
  {\str_range:Nnn #2 { 1 } { -1 - (#3) } {#4} 
   \str_range:Nnn #2 { 1 - (#3) } { -1 } 
  }
}
\__bitset_set true
\__bitset_set false
\__bitset_set true
\__bitset_set false

(End of definition for \__bitset_set_true:Nn and others.)

\l__bitset_internal_int
\l__bitset_set true
\l__bitset_set false

(End of definition for \l__bitset_internal_int.)
The user commands must first translate the argument to an index number.

```latex
\bitset_set_true:Nn \bitset_set_true:cn \bitset_gset_true:Nn \bitset_gset_true:cn \bitset_set_false:Nn \bitset_set_false:cn \bitset_gset_false:Nn \bitset_gset_false:cn
\bitset_set_aux:NNn
```

The user commands must first translate the argument to an index number.
End of definition for \bitset_set_true:Nn and others. These functions are documented on page 287.

\bitset_clear:N
\bitset_clear:c
\bitset_gclear:N
\bitset_gclear:c

\cs_new_protected:Npn \bitset_clear:N #1
\cs_new_protected:Npn \bitset_gclear:N #1
\cs_generate_variant:Nn \bitset_clear:N { c }
\cs_generate_variant:Nn \bitset_gclear:N { c }

End of definition for \bitset_clear:N and \bitset_gclear:N. These functions are documented on page 287.

\bitset_to_arabic:N
\bitset_to_arabic:c
\bitset_to_bin:N
\bitset_to_bin:c
\__bitset_to_int:nN

The naming of the commands follow the names in the int module. \bitset_to_arabic:N uses \int_from_bin:n if the string is shorter than 32 and the slower \fp_eval:n for larger bitsets.

\cs_new:Npn \bitset_to_arabic:N #1
\cs_new:Npn \__bitset_to_int:nN #1#2
\cs_new:Npn \bitset_to_bin:N #1
\cs_generate_variant:Nn \bitset_to_arabic:N { c }
\cs_generate_variant:Nn \bitset_to_bin:N { c }

End of definition for \bitset_to_arabic:N, \bitset_to_bin:N, and \__bitset_to_int:nN. These functions are documented on page 288.

\bitset_item:Nn
\bitset_item:cn

All bits that have been set at anytime have an entry in the prop, so we can take everything else as 0.

\cs_new:Npn \bitset_item:Nn #1#2
\cs_new:Npn \bitset_item:cn #1#2

\bitset_show:N  \bitset_show:c  \
\bitset_log:N  \bitset_log:c  
\cs_new_protected:Npn \bitset_show:N { \__bitset_show:NN \msg_show:nneeee }  
\cs_generate_variant:Nn \bitset_show:N { c }  
\cs_new_protected:Npn \bitset_log:N { \__bitset_show:NN \msg_log:nneeee }  
\cs_generate_variant:Nn \bitset_log:N { c }  
\cs_new_protected:Npn \__bitset_show:NN #1#2  
{  
\__kernel_chk_defined:NT #2  
{  
#1 { bitset } { show }  
{ \token_to_str:N #2 }  
{ \bitset_to_bin:N #2 }  
{ \bitset_to_arabic:N #2 }  
{ }  
}  
}  

(End of definition for \bitset_show:N and \bitset_log:N. These functions are documented on page 288.)

\bitset_show_named_index:N  \bitset_show_named_index:c  
\bitset_log_named_index:N  \bitset_log_named_index:c  
\cs_new_protected:Npn \bitset_show_named_index:N { \__bitset_show_named_index:NN \msg_show:nneeee }  
\cs_generate_variant:Nn \bitset_show_named_index:N { c }  
\cs_new_protected:Npn \bitset_log_named_index:N { \__bitset_show_named_index:NN \msg_log:nneeee }  
\cs_generate_variant:Nn \bitset_log_named_index:N { c }  
\cs_new_protected:Npn \__bitset_show_named_index:NN #1#2  
{  
\__kernel_chk_defined:NT #2  
{  
#1 { bitset } { show-names }  
{ \token_to_str:N #2 }  
{ \bitset_to_arabic:N #2 }  
{ }  
}  
}  

(End of definition for \bitset_show_named_index:N and \bitset_log_named_index:N. These functions are documented on page 288.)
84.1 Messages

\msg_new:nnn { bitset } { show }
{
  \textbf{The-bitset-#1-has-the-representation:} \ \ \n  \textbf{-binary:-}#2 \ \ \n  \textbf{-arabic:-}#3 .
}
\msg_new:nnn { bitset } { show-names }
{
  \textbf{The-bitset-#1-}
  \tl_if_empty:nTF {#2}
  { \textbf{knows-no-names-yet} \ \textbf{-} . }
  { \textbf{knows-the-name/index-pairs-(without-outer-braces):} #2 . }
}
\msg_new:nnn { bitset } { unknown-name }
{ \textbf{The-name-}'#2'-is-unknown-for-bitset-}\tl_to_str:n {#1} }
\prop_gput:Nnn \g_msg_module_name_prop { bitset } { LaTeX }
\prop_gput:Nnn \g_msg_module_type_prop { bitset } { }
{/package}
Chapter 85

\texttt{l3cctab} implementation

As Lua\TeX{} offers engine support for category code tables, and this is entirely lacking from the other engines, we need two complementary approaches. (Some future Xe\TeX{} may add support, at which point the conditionals below would be different.)

85.1 Variables

\begin{itemize}
  \item \texttt{\g__cctab_stack_seq}\hspace{1em}\texttt{\g__cctab_unused_seq}

  List of catcode tables saved by nested \texttt{\cctab_begin:N}, to restore catcodes at the matching \texttt{\cctab_end:}. When popped from the \texttt{\g__cctab_stack_seq} the table numbers are stored in \texttt{\g__cctab_unused_seq} for later reuse.

  \begin{itemize}
    \item \texttt{\seq_new:N \g__cctab_stack_seq}
    \item \texttt{\seq_new:N \g__cctab_unused_seq}
  \end{itemize}

  (End of definition for \texttt{\g__cctab_stack_seq} and \texttt{\g__cctab_unused_seq}.)

  \item \texttt{\g__cctab_group_seq}

    A stack to store the group level when a catcode table started.

    \begin{itemize}
      \item \texttt{\seq_new:N \g__cctab_group_seq}
    \end{itemize}

    (End of definition for \texttt{\g__cctab_group_seq}.)

  \item \texttt{\g__cctab_allocate_int}

    Integer to keep track of what category code table to allocate. In Lua\TeX{} it is only used in format mode to implement \texttt{\cctab_new:N}. In other engines it is used to make csnames for dynamic tables.

    \begin{itemize}
      \item \texttt{\int_new:N \g__cctab_allocate_int}
    \end{itemize}

    (End of definition for \texttt{\g__cctab_allocate_int}.)

  \item \texttt{\l__cctab_internal_a_tl} \hspace{1em} \texttt{\l__cctab_internal_b_tl}

    Scratch space. For instance, when popping \texttt{\g__cctab_stack_seq/\g__cctab_unused_seq}, consists of the catcodetable number (integer denotation) in Lua\TeX{}, or of an intarray variable (as a single token) in other engines.

    \begin{itemize}
      \item \texttt{\tl_new:N \l__cctab_internal_a_tl}
      \item \texttt{\tl_new:N \l__cctab_internal_b_tl}
    \end{itemize}

    (End of definition for \texttt{\l__cctab_internal_a_tl} and \texttt{\l__cctab_internal_b_tl}.)
\end{itemize}
In LuaTeX we store the \texttt{endlinechar} associated to each \texttt{catcodetable} in a property list, unless it is the default value 13.

\begin{verbatim}
\prop_new:N \g__cctab_endlinechar_prop
(End of definition for \g__cctab_endlinechar_prop.)
\end{verbatim}

\section{Allocating category code tables}

The \texttt{\_\_cctab\_new:N} auxiliary allocates a new catcode table but does not attempt to set its value consistently across engines. It is used both in \texttt{\cctab\_new:N}, which sets catcodes to init\TeX\ values, and in \texttt{\cctab\_begin:N/\cctab\_end:} for dynamically allocated tables.

First, the Lua\TeX\ case. Creating a new category code table is done like other registers. In Con\TeX t, \texttt{newcatcodetable} does not include the initialisation, so that is added explicitly.

\begin{verbatim}
\sys_if_engine_luatex:TF
{ \cs_new_protected:Npn \cctab\_new:N #1
{ \__kernel_chk_if_free_cs:N #1
  \__cctab\_new:N #1
} \cs_new_protected:Npn \__cctab\_new:N #1
{ \newcatcodetable \cctab\_new:N #1
  \tex_initcatcodetable:D #1 }
} \cs_new_protected:Npn \__cctab\_gstore:Nnn #1#2#3
{ \intarray_gset:Nnn #1 { #2 + 1 } {#3} }
\cs_new_protected:Npn \cctab\_new:N #1
{ \__kernel_chk_if_free_cs:N #1
  \__cctab\_new:N #1
  \int_step_inline:nn { 256 }
  { \__kernel_intarray_gset:Nnn #1 {##1} { 12 } }
  \__kernel_intarray_gset:Nnn #1 { 257 } { 13 }
  \__cctab\_gstore:Nnn #1 { 0 } { 9 }
  \__cctab\_gstore:Nnn #1 { 13 } { 5 }
}
\end{verbatim}

Now the case for other engines. Here, each table is an integer array. Following the Lua\TeX\ pattern, a new table starts with init\TeX\ codes. The \texttt{\debug\_suspend:} and \texttt{\debug\_resume:} functions prevent errors and logging from \texttt{\debug} commands which are either duplicate or false when \texttt{\_\_cctab\_new:N} is used by \texttt{\cctab\_new:N} or \texttt{\cctab\_const:Nn}. The index base is out-by-one, so we have an internal function to handle that. The init\TeX\ \texttt{\endlinechar} is 13.

\begin{verbatim}
{ \cs_new_protected:Npn \__cctab\_new:N \cctab\_new:N #1
{ \debug\_suspend:
  \intarray\_new:Nn #1 { 257 }
  \debug\_resume:
} \cs_new_protected:Npn \__cctab\_gstore:Nnn #1#2#3
{ \intarray\_gset:Nnn #1 { #2 + 1 } {#3} }
\cs_new_protected:Npn \cctab\_new:N \cctab\_gstore:Nnn #1
{ \__kernel\_chk\_if\_free\_cs:N \cctab\_new:N #1
  \int\_step\_inline:n { 256 }
  { \__kernel\_intarray\_gset:Nnn #1 {##1} { 12 } }
  \__kernel\_intarray\_gset:Nnn #1 { 257 } { 13 }
  \__cctab\_gstore:Nnn #1 { 0 } { 9 }
  \__cctab\_gstore:Nnn #1 { 13 } { 5 }
}
\end{verbatim}
\__cctab_gstore:Nnn \#1 \{ 32 \} \{ 10 \}
\__cctab_gstore:Nnn \#1 \{ 37 \} \{ 14 \}
\int_step_inline:nnn \{ 65 \} \{ 90 \}
\{ \__cctab_gstore:Nnn \#1 \{##1\} \{ 11 \} \}
\__cctab_gstore:Nnn \#1 \{ 92 \} \{ 0 \}
\int_step_inline:nnn \{ 97 \} \{ 122 \}
\{ \__cctab_gstore:Nnn \#1 \{##1\} \{ 11 \} \}
\__cctab_gstore:Nnn \#1 \{ 127 \} \{ 15 \}
\}
\cs_generate_variant:Nn \cctab_new:N \{ c \}
\sys_if_engine_luatex:TF
\cs_new_protected:Npn \__cctab_gset:n #1
\{ \exp_args:Nf \__cctab_gset_aux:n \{ \int_eval:n \{#1\} \} \}
\cs_new_protected:Npn \__cctab_gset_aux:n #1
\{ \tex_savecatcodetable:D #1 \scan_stop:
\int_compare:nNnTF { \tex_endlinechar:D } = { 13 }
\{ \prop_gremove:Nn \g__cctab_endlinechar_prop \{#1\} \}
\}{ \prop_gput:NnV \g__cctab_endlinechar_prop \{#1\}
\tex_endlinechar:D \}
\}
\}
\cs_new_protected:Npn \__cctab_gset:n #1
\{ \int_step_inline:nn \{ 256 \}
\{ \__kernel_intarray_gset:Nnn \#1 \{##1\}
\{ \char_value_catcode:n \{ ##1 - 1 \} \}
\}
\__kernel_intarray_gset:Nnn \#1 \{ 257 \}
\{ \tex_endlinechar:D \}
\}
\cs_generate_variant:Nn \cctab_new:N \{ c \}

(End of definition for \cctab_new:N, \__cctab_new:N, and \__cctab_gstore:Nnn. This function is documented on page 289.)

85.3 Saving category code tables

In various functions we need to save the current catcodes (globally) in a table. In LuaTeX, saving the catcodes is a primitive, but the \endlinechar needs more work: to avoid filling \g__cctab_endlinecharprop with many entries we special-case the default value 13. In other engines we store 256 current catcodes and the \endlinechar in an intarray variable.

\sys_if_engine_luatex:TF
\{ \cs_new_protected:Npn \__cctab_gset:n \#1
\{ \exp_args:Nf \__cctab_gset_aux:n \{ \int_eval:n \{#1\} \} \}
\cs_new_protected:Npn \__cctab_gset_aux:n \#1
\{ \tex_savecatcodetable:D \#1 \scan_stop:
\int_compare:nNnTF { \tex_endlinechar:D } = { 13 }
\{ \prop_gremove:Nn \g__cctab_endlinechar_prop \{#1\} \}
\}{ \prop_gput:NnV \g__cctab_endlinechar_prop \{#1\}
\tex_endlinechar:D \}
\}
\}
\cs_new_protected:Npn \__cctab_gset:n \#1
\{ \int_step_inline:nn \{ 256 \}
\{ \__kernel_intarray_gset:Nnn \#1 \{##1\}
\{ \char_value_catcode:n \{ ##1 - 1 \} \}
\}
\__kernel_intarray_gset:Nnn \#1 \{ 257 \}
\{ \tex_endlinechar:D \}
\}
\cs_generate_variant:Nn \cctab_new:N \{ c \}

(End of definition for \__cctab_gset:n and \__cctab_gset_aux:n.)
Category code tables are always global, so only one version of assignments is needed. Simply run the setup in a group and save the result in a category code table \#1, provided it is valid. The internal function is defined above depending on the engine.

```latex
\begin{verbatim}
\cs_new_protected:Npn \cctab_gset:Nn #1 #2
\__cctab_chk_if_valid:NT #1
\{\group_begin:
\cctab_select:N \c_initex_cctab
#2 \scan_stop:
\__cctab_gset:n {#1}
\group_end:
\}
\cs_generate_variant:Nn \cctab_gset:Nn { c }
\end{verbatim}
\end{quote}

(End of definition for \cctab_gset:Nn. This function is documented on page 289.)

\begin{verbatim}
\cs_new_protected:Npn \cctab_gsave_current:N #1
\__cctab_chk_if_valid:NT #1
\{ \__cctab_gset:n {#1} \}
\cs_generate_variant:Nn \cctab_gsave_current:N { c }
\end{verbatim}
\end{quote}

(End of definition for \cctab_gsave_current:N. This function is documented on page 289.)

85.4 Using category code tables

In \LaTeX{}X, we must ensure that the saved tables are read-only. This is done by applying the saved table, then switching immediately to a scratch table. Any later catcode assignment will affect that scratch table rather than the saved one. If we simply switched to the saved tables, then \char_set_catcode_other:N \_ in the example below would change \c_document_cctab and a later use of that table would give the wrong category code to _:

\begin{verbatim}
\use:n
\{\cctab_begin:N \c_document_cctab
\char_set_catcode_other:N \_\cctab_end:
\cctab_begin:N \c_document_cctab
\int_compare:nTF { \char_value_catcode:n { \_ } = 8 }
\{ \TRUE \} \{ \ERROR \}
\cctab_end:
\}
\end{verbatim}

We must also make sure that a scratch table is never reused in a nested group: in the following example, the scratch table used by the first \cctab_begin:N would be changed globally by the second one issuing \savecatcodetable, and after \group_end: the wrong
category codes (those of \c_str_cctab) would be imposed. Note that the inner \cctab_end: restores the correct catcodes only locally, so the problem really comes up because of the different grouping level. The simplest is to use a scratch table labeled by the \currentgrouplevel. We initialize one of them as an example.

\use:n
{\cctab_begin:N \c_document_cctab
 \group_begin:
  \cctab_begin:N \c_str_cctab
  \cctab_end:
  \group_end:
  \cctab_end:
}

\sys_if_engine_luatex:T
{\__cctab_new:N \g__cctab_internal_cctab
 \cs_new:Npn \__cctab_internal_cctab_name:
 {g__cctab_internal\tex_romannumeral:D \tex_currentgrouplevel:D_cctab}
}

\end of definition for \g__cctab_internal_cctab and \__cctab_internal_cctab_name:.)

\cctab_select:N \cctab_select:c \__cctab_select:N
\cctab_select:N
The public function simply checks the \langle cctab var \rangle exists before using the engine-dependent \__cctab_select:N. Skipping these checks would result in low-level engine-dependent errors. First, the LuaTEX case. In other engines, selecting a catcode table is a matter of doing 256 catcode assignments and setting the \endlinechar.

\cs_new_protected:Npn \cctab_select:N #1 \cs_generate_variant:Nn \cctab_select:N { c }
\sys_if_engine_luatex:TF
{ \cs_new_protected:Npn \__cctab_select:N #1
 { \tex_catcodetable:D #1
  \prop_get:NVNTF \g__cctab_endlinechar_prop #1 \l__cctab_internal_a_tl
  { \int_set:Nn \tex_endlinechar:D { \l__cctab_internal_a_tl } }
  { \int_set:Nn \tex_endlinechar:D { 13 } }
  \cs_if_exist:cF \__cctab_internal_cctab_name:
  { \exp_args:Nc \__cctab_new:N \__cctab_internal_cctab_name: }
  \exp_args:Nc \tex_savecatcodetable:D \__cctab_internal_cctab_name: }
 \exp_args:Nc \tex_catcodetable:D \__cctab_internal_cctab_name: }
}

\end of definition for \g__cctab_internal_cctab and \__cctab_internal_cctab_name:.)

\cs_new_protected:Npn \cctab_select:N #1
{ \int_step_inline:nn { 256 }
 }
\begin{verbatim}
\char_set_catcode:nn { #1 - 1 }
    \__kernel_intarray_item:Nn #1 {#1}
\int_set:Nn \tex_endlinechar:D
    \__kernel_intarray_item:Nn #1 { 257 } \}
\end{verbatim}

(End of definition for \cctab_select:N and \__cctab_select:N. This function is documented on page 290.)

For \cctab_begin:N/\cctab_end: we will need to allocate dynamic tables. This is done here by \_\_cctab_begin_aux:, which puts a table number (in LuaTeX) or name (in other engines) into \_\_cctab_internal_a_tl. In LuaTeX this simply calls \_\_cctab_new:N and uses the resulting catcodetable number; in other engines we need to give a name to the intarray variable and use that. In LuaTeX, to restore catcodes at \cctab_end: we cannot just set \catcodetable to its value before \cctab_begin:N, because that table may have been altered by other code in the mean time. So we must make sure to save the catcodes in a table we control and restore them at \cctab_end:.

\begin{verbatim}
\sys_if_engine_luatex:TF
    \{
        \cs_new_protected:Npn \_\_cctab_begin_aux:
            \__cctab_new:N \g__cctab_next_cctab
            \tl_set:NV \l__cctab_internal_a_tl \g__cctab_next_cctab
            \cs_undefine:N \g__cctab_next_cctab }
    \}
\end{verbatim}

(End of definition for \g__cctab_next_cctab and \_\_cctab_begin_aux:.)

Check the ⟨cctab var⟩ exists, to avoid low-level errors. Get in \_\_cctab_internal_a_tl the number/name of a dynamic table, either from \g__cctab_unused_seq where we save tables that are not currently in use, or from \_\_cctab_begin_aux: if none are available. Then save the current catcodes into the table (pointed to by) \_\_cctab_internal_a_tl and save that table number in a stack before selecting the desired catcodes.

\begin{verbatim}
\cs_new_protected:Npn \cctab_begin:N #1
    \__cctab_chk_if_valid:NT #1
    \__cctab_chk_group_begin:e
\end{verbatim}
\cctab_end:  Make sure a \cctab_begin:N was used some time earlier, get in \l__cctab_internal_a_tl the catcode table number/name in which the prevailing catcodes were stored, then restore these catcodes. The dynamic table is now unused hence stored in \g__cctab_unused_seq for recycling by later \cctab_begin:N.

\cs_new_protected:Npn \cctab_end: { \seq_gpop:NNTF \g__cctab_stack_seq \l__cctab_internal_a_tl { \seq_gpush:NV \g__cctab_unused_seq \l__cctab_internal_a_tl \exp_args:Ne \__cctab_chk_group_end:n { \__cctab_nesting_number:N \l__cctab_internal_a_tl } \__cctab_select:N \l__cctab_internal_a_tl } \msg_error:nn { cctab } { extra-end } }

(End of definition for \cctab_end:. This function is documented on page 290.)

\__cctab_chk_group_begin:n \__cctab_chk_group_begin:e \__cctab_chk_group_end:n Catcode tables are not allowed to be intermixed with groups, so here we check that they are properly nested regarding TEX groups. \__cctab_chk_group_begin:n stores the current group level in a stack, and locally defines a dummy control sequence \__cctab_group⟨cctab-level⟩_chk:. \__cctab_chk_group_end:n pops the stack, and compares the returned value with \tex_currentgrouplevel:D. If they differ, \cctab_end: is in a different grouping level than the matching \cctab_begin:N. If they are the same, both happened at the same level, however a group might have ended and another started between \cctab_begin:N and \cctab_end::

\group_begin: \cctab_begin:N \c_document_cctab \group_end: \cctab_end: \group_end:

In this case checking \tex_currentgrouplevel:D is not enough, so we locally define \__cctab_group⟨cctab-level⟩_chk:, and then check if it exist in \cctab_end:. If it doesn’t, we know there was a group end where it shouldn’t.

The ⟨cctab-level⟩ in the sentinel macro above cannot be replaced by the more convenient \tex_currentgrouplevel:D because with the latter we might be tricked. Suppose:
The line marked with A would start a \texttt{cctab} with a sentinel token named \texttt{\_\_cctab\_group\_1\_chk:}, which would disappear at the \texttt{\group_end:} that follows. But B would create the same sentinel token, since both are at the same group level. Line C would end the \texttt{cctab} from line B correctly, but so would line D because line B created the same sentinel token. Using \texttt{\langle cctab-level \rangle} works correctly because it signals that certain \texttt{cctab} level was activated somewhere, but if it doesn’t exist when the \texttt{\cctab_end:} is reached, we had a problem.

Unfortunately these tests only flag the wrong usage at the \texttt{\cctab_end:}, which might be far from the \texttt{\cctab_begin:N}. However it isn’t possible to signal the wrong usage at the \texttt{\group_end:} without using \texttt{\tex_aftergroup:D}, which is unsafe in certain types of groups.

The three cases checked here just raise an error, and no recovery is attempted: usually interleaving groups and catcode tables will work predictably.

\begin{verbatim}
\cs_new_protected:Npn \__cctab_chk_group_begin:n #1
\group_begin:
   \cctab_begin:N \c_code_cctab % A
\group_end:
\group_begin:
   \cctab_begin:N \c_code_cctab % B
   \cctab_end: % C
   \cctab_end: % D
\group_end:
\end{verbatim}

(End of definition for \texttt{\__cctab_chk_group_begin:n} and \texttt{\__cctab_chk_group_end:n}.)

\texttt{\__cctab_nesting_number:N} This macro returns the numeric index of the current catcode table. In LuaTeX this is just the argument, which is a count reference to a \texttt{\catcodetable} register. In other engines, the number is extracted from the \texttt{cctab} variable.
\sys_if_engine_luatex:TF
{ \cs_new:Npn \_\_cctab_nesting_number:N #1 \ (#1 \ }  }
{ \cs_new:Npn \_\_cctab_nesting_number:N #1
{ \exp_after:wN \exp_after:wN \exp_after:wN \_\_cctab_nesting_number:w
\exp_after:wN \token_to_str:N #1 } }
\use:e
{ \cs_new:Npn \exp_not:N \_\_cctab_nesting_number:w
#1 \tl_to_str:n { g__cctab_ } #2 \tl_to_str:n { _cctab } \{ #2 \} }
(End of definition for \_\_cctab_nesting_number:N and \_\_cctab_nesting_number:w.)
Finally, install some code at the end of the TEX run to check that all \cctab_begin:N
were ended by some \cctab_end:.
\cs_if_exist:NT \hook_gput_code:nnn
{ \hook_gput_code:nnn \{ enddocument/end \} \ cctab }
{ \seq_if_empty:NF \g__cctab_stack_seq
{ \msg_error:nn \{ cctab \} \{ missing-end \} }
\cs_generate_variant:Nn \cctab_item:Nn { c }
(End of definition for \_\_cctab_nesting_number:N. This function is documented on page 290.)

85.5 Category code table conditionals

\cctab_if_exist_p:N \cctab_if_exist_p:c
\cctab_if_exist:NTF \cctab_if_exist:cTF

Evaluate the integer argument only once. In most engines the \cctab variable only has
256 entries so we only look up the catcode for these entries, otherwise we use the current
catcode. In particular, for out-of-range values we use whatever fall-back \char_value_-
catcode:n. In LuaTEX, we use the \tex.getcatcode function.
\cs_if_exist:NT \cctab_item:Nn #1\#2
{ \exp_args:Nf \_\_cctab_item:nN \{ \int_eval:n { #2 } \} \#1 }
\sys_if_engine_luatex:TF
{ \cs_new:Npn \_\_cctab_item:nN #1#2
{ \lua_now:e \{ tex.print(-2, tex.getcatcode(\int_use:N #2, #1)) \} }
\cs_new:Npn \_\_cctab_item:nN #1#2
{ \int_compare:nNnTF {#1} < \{ 256 \}
{ \intarray_item:Nn #2 \#1 + 1 \}
{ \char_value_catcode:n {#1} } }
\cs_generate_variant:Nn \cctab_item:Nn \{ c \}
(End of definition for \cctab_item:Nn. This function is documented on page 290.)
Checks whether the argument is defined and whether it is a valid ⟨cctab var⟩. In \LaTeX{} the validity of the ⟨cctab var⟩ is checked by the engine, which complains if the argument is not a \texttt{\chardef}'ed constant. In other engines, check if the given command is an intarray variable (the underlying definition is a copy of the \texttt{cmr10} font).

\begin{verbatim}
\prg_new_protected_conditional:Npnn \__cctab_chk_if_valid:N #1 { TF , T , F } { \cctab_if_exist:NTF #1 { \__cctab_chk_if_valid_aux:NTF #1 { \prg_return_true: } { \msg_error:nne { cctab } { invalid-cctab } { \token_to_str:N #1 } \prg_return_false: } } { \msg_error:nne { kernel } { command-not-defined } { \token_to_str:N #1 } \prg_return_false: } \sys_if_engine_luatex:TF { \cs_new_protected:Npn \__cctab_chk_if_valid_aux:NTF #1 \__cctab_chk_if_valid:N \prg_return_true: } \cs_if_exist:NT \cs_syst_catcodes_n { \cs_gset_protected:Npn \__cctab_chk_if_valid_aux:NTF #1 \int_compare:nTF {#1-1} < { \e@alloc@ccodetable@count } \int_compare:nTF { #1 <= \cs_syst_catcodes_n } \prg_return_true: \prg_return_false: } \end{verbatim}

(End of definition for \texttt{\__cctab_chk_if_valid:NTF} and \texttt{\__cctab_chk_if_valid_aux:NTF}.)
85.6 Constant category code tables

\cctab Const category code tables

\cctab\texttt{\texttt{const:}Nn} Creates a new \texttt{cctab var} then sets it with the iniTeX and user-supplied codes. To avoid false debug errors, we write out implementation of \texttt{cctab new:N} and \texttt{cctab get:N} instead of directly using them here. The initialization part in \texttt{cctab new:N} in non-LuaTeX is omitted as it’s covered by the iniTeX settings.

\begin{verbatim}
\cs_new_protected:Npn \cctab_const:Nn #1#2
{ \__kernel_chk_if_free_cs:N #1 \__cctab_new:N #1 \group_begin:
\cctab_select:N \c_initex_cctab #2 \scan_stop:
\__cctab_gset:n {#1} \group_end:
}
\cs_generate_variant:Nn \cctab_const:Nn { c }
\end{verbatim}

(End of definition for \texttt{cctab const:}Nn. This function is documented on page 289.)

\cctab\texttt{c initex cctab}\cctab\texttt{c other cctab}\cctab\texttt{c str cctab}

Creating category code tables means thinking starting from iniTeX. For all-other and the standard “string” tables that’s easy.

\begin{verbatim}
\cctab new:N \c_initex_cctab \cctab_const:Nn \c_other_cctab
{ \cctab select:N \c_initex_cctab \int_set:Nn \tex_endlinechar:D { -1 }
\int_step_inline:nnn { 0 } { 127 }
{ \char_set_catcode_other:n {#1} }
\cctab const:Nn \c_str_cctab
{ \cctab select:N \c_other_cctab \char_set_catcode_space:n { 32 } }
\end{verbatim}

(End of definition for \texttt{c initex cctab}, \texttt{c other cctab}, and \texttt{c str cctab}. These variables are documented on page 291.)

\cctab\texttt{c code cctab}\cctab\texttt{c document cctab}

To pick up document-level category codes, we need to delay set up to the end of the format, where that’s possible. Also, as there are a lot of category codes to set, we avoid using the official interface and store the document codes using internal code. Depending on whether we are in the hook or not, the catcodes may be code or document, so we explicitly set up both correctly.

\begin{verbatim}
\cs_if_exist:NTF \@expl@finalise@setup@@
{ \tl_gput_right:Nn \@expl@finalise@setup@@ }
{ \use:n }
{ \__cctab_new:N \c_code_cctab \group_begin:
\int_set:Nn \tex_endlinechar:D { 32 }
\bool_lazy_or:nnTF
{ \sys_if_engine_xetex_p: } { \sys_if_engine_luatex_p: }
1267 \{ \sys_if_engine_xetex_p: } { \sys_if_engine_luatex_p: }
\end{verbatim}

1267
85.7 Messages

\msg_new:nnnn { cctab } { stack-full } { The-category-code-table-stack-is-exhausted. }

(End of definition for \g_tmpa_cctab and \g_tmpb_cctab. These variables are documented on page 291.)
LaTeX has been asked to switch to a new category code table, but there is no more space to do this!

\msg_new:nnnn { cctab } { extra-end }
{ Extra-\iow_char:N\cctab_end: ignored-\msg_line_context:. }
\msg_new:nnnn { cctab } { missing-end }
{ Missing-\iow_char:N\cctab_end: before end of TeX run. }
\msg_new:nnnn { cctab } { invalid-cctab }
{ You can only switch to a \iow_char:N\catcode\table that is initialized using \iow_char:N\cctab_new:N or -}
\msg_new:nnnn { cctab } { group-mismatch }
{ \iow_char:N\cctab_end: occurred in a-}
\int_case:nn {#1}
{ \iow_char:N\cctab_end: occurred in a-
\iow_char:N\cctab_begin:N.
\iow_char:N\cctab_const:Nn.
\iow_char:N\cctab_end:.}
\prop_gput:Nnn \g_msg_module_name_prop { cctab } { LaTeX }
\prop_gput:Nnn \g_msg_module_type_prop { cctab } { }
⟨/package⟩
Chapter 86

l3unicode implementation

86.1 User functions

Conversion of a codepoint to a character (Unicode engines) or to one or more bytes (8-bit engines) is required. For loading the data, all that is needed is the form which creates strings: these are outside the group as they will also be used when looking up data in the hash table storage at point-of-use. Later, we will also need functions that can generate character tokens for document use: those are defined below, in the data recovery setup.

\cs_new:Npn \codepoint_str_generate:n #1
  { \int_compare:nNnTF {#1} = { ' } { \char_generate:nn {#1} { 12 } } }
\cs_new:Npn \codepoint_generate:nn #1#2
  { \int_compare:nNnTF {#1} = { ' } { \__kernel_exp_not:w \exp_after:wN \exp_after:wN \exp_after:wN \char_generate:nn {#1} {#2} } }
\cs_new:Npn \codepoint_str_generate:n #1
  { \int_compare:nNnTF {#1} = { ' } { \use:e } }
This code converts a codepoint into the correct UTF-8 representation. In terms of the algorithm itself, see https://en.wikipedia.org/wiki/UTF-8 for the octet pattern.
\__codepoint_to_bytes_outputiv:nw
    { \__codepoint_to_bytes_auxiii:n {#1} }
\fi:
\fi:
\else:
\__codepoint_to_bytes_outputi:nw {#1}
\fi:
\__codepoint_to_bytes_end: { } { } { } { }
}
\cs_new:Npn \__codepoint_to_bytes_auxii:Nnn #1#2#3
{ "#10 + \int_div_truncate:nn {#2} {#3} }
\cs_new:Npn \__codepoint_to_bytes_auxiii:n #1
{ \int_mod:nn {#1} { 64 } + 128 }
\cs_new:Npn \__codepoint_to_bytes_outputi:nw
#1 #2 \__codepoint_to_bytes_end: #3
{ \__codepoint_to_bytes_output:fnn { \int_eval:n {#1} } { } {#2} }
\cs_new:Npn \__codepoint_to_bytes_outputii:nw
#1 #2 \__codepoint_to_bytes_end: #3#4
{ \__codepoint_to_bytes_output:fnn { \int_eval:n {#1} } { {#3} } {#2} }
\cs_new:Npn \__codepoint_to_bytes_outputiii:nw
#1 #2 \__codepoint_to_bytes_end: #3#4#5
{ \__codepoint_to_bytes_output:fnn
    { \int_eval:n {#1} } { {#3} {#4} } {#2} }
\cs_new:Npn \__codepoint_to_bytes_outputiv:nw
#1 #2 \__codepoint_to_bytes_end: #3#4#5#6
{ \__codepoint_to_bytes_output:fnn
    { \int_eval:n {#1} } { {#3} {#4} {#5} } {#2} }
\cs_new:Npn \__codepoint_to_bytes_output:nnn #1#2#3
{ #3
\__codepoint_to_bytes_end: #2 {#1} }
\cs_generate_variant:Nn \__codepoint_to_bytes_output:nnn { f }
\cs_new:Npn \__codepoint_to_bytes_output:n { #1}
\__codepoint_to_category:n
Get the value and convert back to the string.
\cs_new:Npn \codepoint_to_category:n #1
{ \cs:w
c__codepoint_category_
\rex_romannumeral:D
\__kernel_codepoint_data:nn { category } {#1}
\str
\cs_end:
}
(End of definition for \__kernel_codepoint_to_bytes:n and others.)

\codepoint_to_category:n
Get the value and convert back to the string.
\cs_new:Npn \codepoint_to_category:n #1
{ \cs:w
c__codepoint_category_
\rex_romannumeral:D
\__kernel_codepoint_data:nn { category } {#1}
\str
\cs_end:
}
(End of definition for \codepoint_to_category:n. This function is documented on page 295.)
Converted to NFD is a potentially-recursive process: the key is to check if we get the input codepoint back again. As far as possible, we use the same path for all engines.

\begin{verbatim}
\cs_new:Npn \codepoint_to_nfd:n #1
{ \exp_args:Ne \__codepoint_to_nfd:n { \int_eval:n {#1} } }
\cs_new:Npn \__codepoint_to_nfd:n #1
{ \__codepoint_to_nfd:nn {#1} { \char_value_catcode:n {#1} } }
\bool_lazy_or:nnF
{ \sys_if_engine_luatex_p: } { \sys_if_engine_xetex_p: }
{ \cs_gset:Npn \__codepoint_to_nfd:n #1
{ \int_compare:nNnTF {#1} > { "80 }
{ \__codepoint_to_nfd:nn {#1} { 12 } }
{ \__codepoint_to_nfd:nn {#1} { \char_value_catcode:n {#1} } } }
}\cs_new:Npn \__codepoint_to_nfd:nn #1#2
{ \exp_args:Ne \__codepoint_to_nfd:nnn { \__codepoint_nfd:n {#1} } {#1} {#2} }
\cs_new:Npn \__codepoint_to_nfd:nnn #1#2#3
{ \int_compare:nNnTF {#1} = {#3}
{ \codepoint_generate:nn {#1} {#4} }
{ \__codepoint_to_nfd:nn {#1} {#4}
{ \tl_if_blank:nF {#2}
\tl_if_blank:nF {#2}
\__codepoint_to_nfd:nn {#2} {#4} }
}
}
\end{verbatim}

(End of definition for \codepoint_to_nfd:n and others. This function is documented on page 295.)

\section{86.2 Data loader}

Text operations requires data from the Unicode Consortium. Data read into Unicode engine formats is at best a small part of what we need, so there is a loader here to set up the appropriate data structures.

Where we need data for most or all of the Unicode range, we use the two-stage table approach recommended by the Unicode Consortium and demonstrated in a model implementation in Python in \url{https://www.strchr.com/multi-stage_tables}. This approach uses the \intarray (fontdimen-based) data type as it is fast for random access and avoids significant hash table usage. In contrast, where only a small subset of codepoints are required, storage as macros is preferable. There is also some consideration of the effort needed to load data: see for example the grapheme breaking information, which would be problematic to convert into a two-stage table but which can be used with reasonable performance in a small number of comma lists (at the cost that breaking at higher codepoint Hangul characters will be slightly slow).
Choosing the block size for the blocks in the two-stage approach is non-trivial: depending on the data stored, the optimal size for memory usage will vary. At the same time, for us there is also the question of load-time: larger blocks require longer comma lists as intermediates, so are slower. As this is going to be needed to use the data, we set it up outside of the group for clarity.

\begin{verbatim}
\int_const:Nn \c__codepoint_block_size_int { 64 }
\end{verbatim}

Parsing the data files can be the same way for all engines, but where they are stored as character tokens, the construction method depends on whether they are Unicode or 8-bit internally. Parsing is therefore done by common functions, with some data storage using engine-specific auxiliaries.

As only the data needs to remain at the end of this process, everything is set up inside a group. The only thing that is outside is creating a stream: they are global anyway and it is best to force a stream for all engines.

\begin{verbatim}
\ior_new:N \g__codepoint_data_ior
\end{verbatim}

We need some setup for the two-part table approach. The number of blocks we need will be variable, but the resulting size of the stage one table is predictable. For performance reasons, we therefore create the stage one tables now so they can be used immediately, and will later rename them as a constant tables. For each two-stage table construction, we need a comma list to hold the partial block and a couple of integers to track where we are up to. To avoid burning registers, the latter are stored in macros and are “fake” integers. We also avoid any \texttt{new} functions, keeping as much as possible local.

As we need both positive and negative values, case data requires one two-stage table for each transformation. In contrasts, general Unicode properties could be stored in one table with appropriate combination rules: that is not done at present but is likely to be added over time. Here, all that is needed is additional entries into the comma-list to create the structures.

Notice that in the standard \texttt{expl3} way we are indexes position not offset: that does mean a little work later.

\begin{verbatim}
\group_begin:
\clist_map_inline:nn
{ category , uppercase , lowercase }
{
  \cs_set_nopar:cpn { l__codepoint_ #1 _block_clist } { }
  \cs_set_nopar:cpn { l__codepoint_ #1 _block_tl } { 1 }
  \cs_set_nopar:cpn { l__codepoint_ #1 _pos_tl } { 0 }
  \intarray_new:cn { g__codepoint_ #1 _index_intarray }
  \int_array__init:nn { { #1 - 110000 } } \c__codepoint_block_size_int
}
\group_end:
\end{verbatim}

We need an integer value when matching the current block to those we have already seen, and a way to track codepoints for handling ranges. Again, we avoid using up registers or creating global names.

\begin{verbatim}
\cs_set_nopar:Npn \l__codepoint_next_codepoint_fint_tl { 0 }
\cs_set_nopar:Npn \l__codepoint_matched_block_tl { 0 }
\end{verbatim}
For Unicode general category, there needs to be numerical representation of each possible value. As we need to go from string to number here, but the other way elsewhere, we set up fast mappings both ways, but one set local and the other as constants.

\cs_set_protected:Npn \_codepoint_data_auxi:w \#1\#2
{ \quark_if_recursion_tail_stop:n \{#2\}
  \cs_set_nopar:cpn { l__codepoint_category_ \#2 _tl } \{#1\}
  \str_const:cn \{ c__codepoint_category_ \tex_romannumeral:D \#1 _str \} \{#2\}
  \exp_args:N e \_codepoint_data_auxi:w \{ \int_eval:n \{ \#1 + 1 \} \}
}
\_codepoint_data_auxi:w \{ 1 \}
{ \Lu } { \Ll } { \Lt } { \Lm } { \Lo }
{ \Mn } { \Me } { \Mc }
{ \Nd } { \Nl } { \No }
{ \Zs } { \Zl } { \Zp }
{ \Cc } { \Cf } { \Co } { \Cs } { \Cn }
{ \Pd } { \Ps } { \Pe } { \Pc } { \Po } { \Pf }
{ \Sm } { \Sc } { \Sk } { \So }
\q_recursion_tail
\q_recursion_stop
}

Parse the main Unicode data file and pull out the NFD and case changing data. The NFD data is stored on using the hash table approach and can yield a predictable number of codepoints: one or two. We also need the case data, which will be modified further below. To allow for finding ranges, the description of the codepoint needs to be carried forward.

\cs_set_protected:Npn \_codepoint_data_auxi:w
\{ #1 ; #2 ; #3 ; #4 ; #5 ; #6 ; #7 ; #8 ; #9 ;
{ \tl_if_blank:nF \{#6\}
  \{ \tl_if_head_eq_charcode:nNF \{#6\} < \% >
    \{ \_codepoint_data_auxii:w \#1 ; #6 - \q_stop \}
  \}
  \_codepoint_data_auxiii:w \#1 ; #2 ; #3 ;
}
\cs_set_protected:Npn \_codepoint_data_auxii:w \#1 ; #2 - #3 \q_stop
{ \tl_const:ce
  \{ \_codepoint_nfd_ \codepoint_str_generate:n \{"#1\} _tl \}
  \{ "#2\}
  \{ \tl_if_blank:nF \{#3\} \{"#3\} \}
}
\_codepoint_data_auxiii:w \#1 ; #2 - #3 \q_stop

The category data needs to be converted from a string to the numerical equivalent: a simple operation. The case data is going to be stored as an offset from the parent character, rather than an absolute value. We therefore deal with that plus the situation where a codepoint has no mapping data in one shot.

\cs_set_protected:Npn \_codepoint_data_auxiii:w
\{ #1 ; #2 ; #3 ; #4 ; #5 ; #6 ; #7 ; #8 ; #9 - \q_stop
\{ \use:e

1276
To deal with ranges, we track the position of the next codepoint expected. If there is a gap, we deal with that separately: it could be a range or an unused part of the Unicode space. As such, we deal with the current codepoint here whether or not there is range to fill in. Upper- and lowercase data go into the two-stage table, any titlecase exception is just stored in a macro. The data for the codepoint is added to the current block, and if that is now complete we move on to save the block. The case exceptions are all stored as codepoints, with a fixed number of balanced text as we know that there are never more than three.

Distinguish between a range and a gap, and pass on the appropriate value(s). The general
category for unassigned characters is Cn, so we find the correct value once and then use that.

\begin{verbatim}
\cs_set_protected:Npe \_\_codepoint_data_auxv:nmnw #1#2#3#4#5 Last> #6 \q_stop
  \{ \exp_not:N \tl_if_blank:nTF {#6} 
  \{ \exp_not:N \_\_codepoint_range:nnn {#1} \{ category \}
  \{ \exp_not:N \l_\_codepoint_category_Cn_tl \}
  \exp_not:N \_\_codepoint_range:nnn {#1} \{ uppercase \} { 0 }
  \exp_not:N \_\_codepoint_range:nnn {#1} \{ lowercase \} { 0 }
  \}
  \}
  \exp_not:N \_\_codepoint_range:nnn {#1} \{ category \} \{#2\}
  \exp_not:N \_\_codepoint_range:nnn {#1} \{ uppercase \} \{#3\}
  \exp_not:N \_\_codepoint_range:nnn {#1} \{ lowercase \} \{#4\}
}\end{verbatim}

Calculated the length of the range and the space remaining in the current block.

\begin{verbatim}
\cs_set_protected:Npn \_\_codepoint_range:nnn #1#2#3#4
  \{ \prg_replicate:nn {#1}
  \{ \clist_put_right:cn { l_\_codepoint_ #3_block_clist } {#4} \}
  \int_compare:nNnT { \clist_count:c { l_\_codepoint_ #3_block_clist } } = \c_\_codepoint_block_size_int
  \{ \_\_codepoint_save_blocks:nn {#3} { 1 } \}
  \int_compare:nNnF { \int_div_truncate:nn { #2 - #1 } \c_\_codepoint_block_size_int } = 0
  \{ \_\_codepoint_block_size_int \}
\end{verbatim}

Here we want to do three things: add to and possibly complete the current block, add complete blocks quickly, then finish up the range in a final open block. We need to avoid as far as possible avoid dealing with every single codepoint, so the middle step is optimised.

\begin{verbatim}
\cs_set_protected:Npn \_\_codepoint_range:nnn #1#2#3#4
  \{ \prg_replicate:nn {#1}
  \{ \clist_put_right:cn { l_\_codepoint_ #3_block_clist } \} \#4 \}
  \\_\_codepoint_block_size_int \#6 \}\end{verbatim}
To allow rapid comparison, each completed block is stored locally as a comma list: once all of the blocks have been created, they are converted into an \texttt{intarray} in one step. The aim here is to check the current block against those we've already used, and either match to an existing block or save a new block.

Close out the final block, rename the first stage table, then combine all of the block comma-lists into one large second-stage table with offsets. As we use an index not an offset, there is a little back-and-forward to do.
With the setup done, read the main data file: it’s easiest to do that as a token list with spaces retained.

```
\ior_open:Nn \g__codepoint_data_ior { UnicodeData.txt }
\group_begin:
\char_set_catcode_space:n { ' \ }
\ior_map_variable:NNn \g__codepoint_data_ior \l__codepoint_tmpa_tl
\exp_after:wN \ior_map_break:
\if_meaning:w \l__codepoint_tmpa_tl \c_space_tl
    \exp_after:wN \ior_map_break:
\fi:
\exp_after:wN \ior_map_break:
```

Recover data from a two-stage table: entirely generic as this applies to all tables (as we
use the same block size for all of them). Notice that as we use indices not offsets we have to shuffle out-by-one issues. This function is needed before loading the special casing data, as there we need to be able to check the standard case mappings.

\cs_new:Npn \__kernel_codepoint_data:nn #1#2
\exp_args:Nf \__codepoint_data:nnn
\int_eval:n
\c__codepoint_block_size_int *
\intarray_item:cn { c__codepoint_ #1 _index_intarray }
\int_div_truncate:nn {#2}
\c__codepoint_block_size_int + 1
- 1
\}
\}
\}
\cs_new:Npn \__codepoint_data:nnn #1#2#3
\intarray_item:cn { c__codepoint_ #3 _blocks_intarray }
\{ #1 + \int_mod:nn {#2} \c__codepoint_block_size_int + 1 }

(End of definition for \__kernel_codepoint_data:nn and \__codepoint_data:nnn.)

The other data files all use C-style comments so we have to worry about # tokens (and reading as strings). The set up for case folding is in two parts. For the basic (core) mappings, folding is the same as lower casing in most positions so only store the differences. For the more complex foldings, always store the result, splitting up the two or three code points in the input as required.

\group_begin:
\ior_open:Nn \g__codepoint_data_ior { CaseFolding.txt }
\cs_set_protected:Npn \__codepoint_data_auxi:w #1 ;~ #2 ;~ #3 ; #4 \q_stop
\if:w \tl_head:n { #2 ? } C \reverse_if:N \if_int_compare:w
\int_eval:n { \__kernel_codepoint_data:nn { lowercase } {"#1} + "#1 }
= "#3 - \tl_const:ce
\{ c__codepoint_casefold_ \codepoint_str_generate:n {"#1} _tl 
\{ {"#3} \} \} \}
\fi:
\else:
\if:w \tl_head:n { #2 ? } F
\__codepoint_data_auxii:w #1 - #3 - \q_stop
\fi:
\fi:
}

1281
Here, \#4 can have a trailing space, so we tidy up a bit at the cost of speed for these small number of cases it applies to.

\begin{verbatim}
\cs_set_protected:Npn \__codepoint_data_auxii:w #1 - #2 - #3 - #4 \q_stop
  {
    \tl_const:ce { c__codepoint_casefold_ \codepoint_str_generate:n {"#1"} _tl }
    {"#2"}
    {"#3"}
    \tl_if_blank:nF {#4} { " \int_to_hex:n {"#4"} }
  }
\ior_str_map_inline:Nn \g__codepoint_data_ior
}{
  \reverse_if:N \if:w \c_hash_str \tl_head:w #1 \c_hash_str \q_stop
  \__codepoint_data_auxi:w #1 \q_stop
  \fi:
}\ior_close:N \g__codepoint_data_ior
\end{verbatim}

For upper- and lowercasing special situations, there is a bit more to do as we also have titlecasing to consider, plus we need to stop part-way through the file.

\begin{verbatim}
\ior_open:Nn \g__codepoint_data_ior { SpecialCasing.txt }
\cs_set_protected:Npn \__codepoint_data_auxi:w #1 - #2 - #3 - #4 \q_stop
  \use:n { \__codepoint_data_auxii:w #1 lower #2 - } \q_stop
  \use:n { \__codepoint_data_auxii:w #1 upper #4 - } \q_stop
  \str_if_eq:nnF {#3} {#4}
    { \use:n { \__codepoint_data_auxii:w #1 title #3 - } \q_stop }
}\cs_set_protected:Npn \__codepoint_data_auxii:w #1 - #2 - #3 - #4 - #5 \q_stop
  \use:n { \__codepoint_data_auxii:w #1 \c_hash_str \q_stop }
  \use:n { \__codepoint_data_auxii:w #1 \c_space_tl Conditional-Mappings }
  \ior_map_break: }
}\ior_str_map_inline:Nn \g__codepoint_data_ior
\end{verbatim}
With the core data files loaded, there is now a need to provide access to this information for other modules. That is done here such that case folding can also be covered. At this level, all that needs to be returned is the

\begin{verbatim}
\cs_new:Npn \__kernel_codepoint_case:nn #1#2
{ \exp_args:Ne \__codepoint_case:nnn { \codepoint_str_generate:n {#2} } {#1} {#2} }
\end{verbatim}

\begin{verbatim}
\cs_new:Npn \__codepoint_uppercase:n { \__codepoint_case:nn { uppercase } }
\cs_new:Npn \__codepoint_lowercase:n { \__codepoint_case:nn { lowercase } }
\cs_new:Npn \__codepoint_titlecase:n { \__codepoint_case:nn { uppercase } }
\cs_new:Npn \__codepoint_casefold:n { \__codepoint_case:nn { lowercase } }
\cs_new:Npn \__codepoint_case:nn #1#2
{ \int_eval:n { \__kernel_codepoint_data:nn {#1} {#2} + #2 } { } { } }
\end{verbatim}

(End of definition for \____kernel_codepoint_case:nn and others.)

\begin{verbatim}
\cs_new:Npn \__codepoint_nfd:n #1
{ \exp_args:Ne \__codepoint_nfd:nn { \codepoint_str_generate:n {#1} } {#1} }
\cs_new:Npn \__codepoint_nfd:nn #1#2
{ \tl_if_exist:cTF { c__codepoint_nfd_ #1 _tl } { \tl_use:c { c__codepoint_nfd_ #1 _tl } } { \use:c { __codepoint_nfd_ #2 :n } {#3} } }
\end{verbatim}

(End of definition for \____codepoint_nfd:n and \____codepoint_nfd:nn.)

A simple interface.

\begin{verbatim}
\cs_new:Npn \__codepoint_nfd:n #1
{ \exp_args:Ne \__codepoint_nfd:nn { \codepoint_str_generate:n {#1} } {#1} }
\cs_new:Npn \__codepoint_nfd:nn #1#2
{ \tl_if_exist:cTF { c__codepoint_nfd_ #1 _tl } { \tl_use:c { c__codepoint_nfd_ #1 _tl } } { \use:c { __codepoint_nfd_ #2 :n } {#3} } }
\end{verbatim}

(End of definition for \____codepoint_nfd:n and \____codepoint_nfd:nn.)

\begin{verbatim}
\ior_new:N \g__text_data_ior
\group_begin:
\ior_open:Nn \g__text_data_ior \{ GraphemeBreakProperty.txt \}
\cs_set_nopar:Npn \l__text_tmpa_str { }
\end{verbatim}

Read the Unicode grapheme data. This is quite easy to handle and we only need codepoints, not characters, so there is no need to worry about the engine in use. As reading as a string is most convenient, we have to do some work to remove spaces: the hardest part of the entire process!
Chapter 87

\textbf{l3text implementation}

\begin{verbatim}
\s__text_stop  Internal scan marks.
\scan_new:N \s__text_stop
(End of definition for \s__text_stop.)
\q__text_nil  Internal quarks.
\quark_new:N \q__text_nil
(End of definition for \q__text_nil.)
\__text_quark_if_nil_p:n  Branching quark conditional.
\__text_quark_if_nil:nTF
(End of definition for \__text_quark_if_nil:nTF.)
\q__text_recursion_tail
\q__text_recursion_stop  Internal recursion quarks.
\quark_new:N \q__text_recursion_tail
\quark_new:N \q__text_recursion_stop
(End of definition for \q__text_recursion_tail and \q__text_recursion_stop.)
\__text_use_i_delimit_by_q_recursion_stop:nw
Functions to gobble up to a quark.
\cs_new:Npn \__text_use_i_delimit_by_q_recursion_stop:nw
\#1 \#2 \q__text_recursion_stop {#1}
(End of definition for \__text_use_i_delimit_by_q_recursion_stop:nw.)
\__text_if_q_recursion_tail_stop_do:Nn
\__text_if_q_recursion_tail_stop_do:nn
Functions to query recursion quarks.
\__kernel_quark_new_test:N \__text_if_q_recursion_tail_stop_do:Nn
\__kernel_quark_new_test:N \__text_if_q_recursion_tail_stop_do:nn
\end{verbatim}

1285
Internal scan marks quarks.

Functions to gobble up to a scan mark.

Functions to query recursion scan marks. Slower than a quark test but needed to avoid issues in the outer expansion loop with unterminated \romannumeral primitives.

The idea here is to take a token and ensure that if it's an implicit char, we output the explicit version. Otherwise, the token needs to be unchanged. First, we have to split between control sequences and everything else.
For control sequences, we can check for macros versus other cases using `\if_meaning:w`, then explicitly check for `\chardef` and `\mathchardef`.

```latex
\cs_new:Npn \__text_token_to_explicit_cs:N #1
\exp_after:wN \if_meaning:w \exp_not:N #1 #1
\exp_after:wN \use:nn \exp_after:wN \__text_token_to_explicit_cs_aux:N
\else:
\exp_after:wN \exp_not:n \fi:
\exp_after:wN \exp_not:n \fi:
\exp_after:wN \exp_not:n \fi:
\exp_after:wN \exp_not:n \fi:
\exp_after:wN \exp_not:n \fi:
\exp_after:wN \exp_not:n \fi:
\exp_after:wN \exp_not:n \fi:
\exp_after:wN \exp_not:n \fi:
\exp_after:wN \exp_not:n \fi:
\exp_after:wN \exp_not:n \fi:
\exp_after:wN \exp_not:n \fi:
\exp_after:wN \exp_not:n \fi:
\exp_after:wN \exp_not:n \fi:
\exp_after:wN \exp_not:n \fi:
\exp_after:wN \exp_not:n \fi:
\exp_after:wN \exp_not:n \fi:
\exp_after:wN \exp_not:n \fi:
\exp_after:wN \exp_not:n \fi:
\exp_after:wN \exp_not:n \fi:
\exp_after:wN \exp_not:n \fi:
\exp_after:wN \exp_not:n \fi:
\exp_after:wN \exp_not:n \fi:
\exp_after:wN \exp_not:n \fi:
\exp_after:wN \exp_not:n \fi:
\exp_after:wN \exp_not:n \fi:
```

For character tokens, we need to filter out the implicit characters from those that are explicit. That’s done here, then if necessary we work out the category code and generate the char. To avoid issues with alignment tabs, that one is done by elimination rather than looking up the code explicitly. The trick with finding the charcode is that the \TeX{} messages are either the \texttt{(something)} character \texttt{char} or the \texttt{\langle type \rangle char}.

```latex
\cs_new:Npn \__text_token_to_explicit_char:N #1
\if:w \if_catcode:w ^ \exp_args:No \str_tail:n { \token_to_str:N #1 } ^ \else:
AB \else:
\exp_after:wN \exp_not:n \fi:
\exp_after:wN \exp_not:n \fi:
\exp_after:wN \exp_not:n \fi:
\exp_after:wN \exp_not:n \fi:
\exp_after:wN \exp_not:n \fi:
\exp_after:wN \exp_not:n \fi:
\exp_after:wN \exp_not:n \fi:
\exp_after:wN \exp_not:n \fi:
\exp_after:wN \exp_not:n \fi:
\exp_after:wN \exp_not:n \fi:
\exp_after:wN \exp_not:n \fi:
\exp_after:wN \exp_not:n \fi:
\exp_after:wN \exp_not:n \fi:
\exp_after:wN \exp_not:n \fi:
\exp_after:wN \exp_not:n \fi:
\exp_after:wN \exp_not:n \fi:
```

1287
An idea from l3char: we need to get the category code of a specific token, not the general case.
\__text_if_expandable:N\textbf{TF}
Test for tokens that make sense to expand here: that is more restrictive than the engine view.
\begin{verbatim}
\prg_new_conditional:Npn \__text_if_expandable:N #1 { T , F , TF }
{ \token_if_expandable:NTF #1 {
  \bool_lazy_any:nTF {
    \token_if_protected_macro_p:N #1 
    \token_if_protected_long_macro_p:N #1 
    \token_if_eq_meaning_p:NN \q__text_recursion_tail #1 
  }
  \prg_return_false: }
  \prg_return_true: }
}{ \prg_return_false: }
\end{verbatim}
(End of definition for \texttt{\_text_if_expandable:NTF}.)

87.3 Codepoint utilities
For working with codepoints in an engine-neutral way.
\begin{verbatim}
\textbf{bool_lazy_or:nnTF}
{ \sys_if_engine_luatex_p: }
{ \sys_if_engine_xetex_p: }
{ \sys_if_engine_pdftex:TF }
\exp_not:N \int_compare:nNnTF {'#2} > { "80 }
\end{verbatim}

Grab a codepoint and apply some code to it: here \texttt{#1} should expect one following \emph{balanced text}.
\text{End of definition for \_\_text_codepoint_process:nN and others.}

Allows comparison for all engines using a first “character” followed by a codepoint.
(End of definition for \__text_codepoint_compare:nNnTF and others.)
87.4 Configuration variables

\l_text_accents_tl  Used to be used for excluding these ideas from expansion: now deprecated.

\l_text_letterlike_tl

(End of definition for \l_text_accents_tl and \l_text_letterlike_tl.)

\l_text_case_exclude_arg_tl  Non-text arguments, including covering the case of \protected@edef applied to \cite.

\l_text_math_arg_tl  Math mode as arguments.

\l_text_math_delims_tl  Paired math mode delimiters.

\l_text_expand_exclude_tl  Commands which need not to expand. We start with a somewhat historical list, and tidy up if possible.

87.5 Expansion to formatted text

Markers for implicit char handling.

\begin{verbatim}
\text_expand:n
  \text_expand_result:n
  \text_expand_store:n
  \text_expand_store:o
  \text_expand_store:nw
  \text_expand_loop:w
  \text_expand_group:n
  \text_expand_space:w
  \text_expand_math_loop:Nw
  \text_expand_math_group:Nn
  \text_expand_math_space:Nw
  \text_expand_math_accent:N
  \text_expand_math_letterlike:N
  \text_expand_accent:N
  \text_expand_accent:NN
  \text_expand_encoding:N
  \text_expand_encoding_escape:N
  \text_expand_replace:N
  \text_expand_replace:n
  \text_expand_cs:N
  \text_expand_protect:w
  \text_expand_protect:N
  \text_expand_protect:nN
  \text_expand_protect:Nw
  \text_expand_testopt:N
  \text_expand_testopt:NNn
  \text_expand_encoding:N
  \text_expand_encoding_escape:N
  \text_expand_replace:N
  \text_expand_replace:n
  \text_expand_cs_expand:N
  \text_expand_unexpanded:w
  \text_expand_unexpanded_test:w
  \text_expand_unexpanded:N
  \text_expand_unexpanded:n
\end{verbatim}

After precautions against & tokens, start a simple loop: that of course means that “text” cannot contain the two recursion quarks. The loop here must be f-type expandable; we have arbitrary user commands which might be protected and take arguments, and if the expansion code is used in a typesetting context, that will otherwise explode. (The same issue applies more clearly to case changing: see the example there.) The outer loop has to use scan marks as delimiters to protect against unterminated \romannumeral usage in the input.
The approach to making the code \( f \)-type expandable is to use a marker result token and to shuffle the collected tokens.

The main loop is a standard “\texttt{tl} action”: groups are handled recursively, while spaces are just passed through. Thus all of the action is in handling \( N \)-type tokens.

The first step in dealing with \( N \)-type tokens is to look for math mode material: that needs to be left alone. The starting function has to be split into two as we need \texttt{quark-if_recursion_tail_stop:N} first before we can trigger the search. We then look for matching pairs of delimiters, allowing for the case where math mode starts but does not end. Within math mode, we simply pass all the tokens through unchanged, just checking the \( N \)-type ones against the end marker.
At this stage, either we have a control sequence or a simple character: split and handle. The need to check for non-protected actives arises from handling of legacy input encodings: they need to end up in a representation we can deal with in further processing. The tests for explicit parts of the \LaTEX2ε UTF-8 mechanism cover the case of bookmarks, where definitions change and are no longer protected. The same is true for babel shorthands.

\new:Npn \_text_expand_explicit:N #1
\{ 
\token_if_cs:NTF #1
\{ \_text_expand_exclude:N #1 \}
\{
 bool_lazy_and:nnTF
\{ \token_if_active_p:N #1 \}
\{
 ! bool_lazy_any_p:n
\{
 \token_if_protected_macro_p:N #1
\{
 \token_if_protected_long_macro_p:N #1
\{
 \tl_if_head_eq_meaning_p:oN {#1} \UTFviii@two@octets 
\{
 \tl_if_head_eq_meaning_p:oN {#1} \UTFviii@three@octets 
\{
 \tl_if_head_eq_meaning_p:oN {#1} \UTFviii@four@octets 
\{
 \tl_if_head_eq_meaning_p:oN {#1} \active@prefix 
\}
\}
\}
\}
\}
\}
\}
\exp_after:wN \_text_expand_loop:w #1
\}
\}
\_text_expand_store:n {#1}
\_text_expand_loop:w
\}
\}
Next we exclude math commands: this is mainly as there might be an \ensuremath. The switching command for case needs special handling as it has to work by meaning.

\new:Npn \_text_expand_exclude:N #1
\{ \cs_if_eq:NNTF #1 \text_case_switch:nnnn
\{ \_text_expand_exclude_switch:Nnnnn #1 \}
\{
 \exp_args:Ne \_text_expand_exclude:nN
\{
 \_text_expand_store:n {#1}
\_text_expand_loop:w
\}
\}
\}
\_text_expand_store:n { #1 {#2} {#3} {#4} {#5} }
\_text_expand_loop:w
Accents.

Another list of exceptions: these ones take no arguments so are easier to handle.
\LaTeX's `\protect` makes life interesting. Where possible, we simply remove it and replace with the "parent" command; of course, the `\protect` might be explicit, in which case we need to leave it alone. That includes the case where it's not even followed by an N-type token. There is also the case of a straight `\@protected@testopt` to cover.
Deal with encoding-specific commands

Finally, expand any macros which can be: this then loops back around to deal with what they produce. The only issue is if the token is \texttt{\exp_not:n}, as that must apply to the following balanced text.
Since \texttt{\exp_not:n} is actually a primitive, it allows a strange syntax and it particular
the primitive expands what follows and discards spaces and \texttt{\scan_stop:} until finding
a braced argument (the opening brace can be implicit but we will not support this
here). Here, we repeatedly \texttt{f}-expand after such an \texttt{\exp_not:n}, and test what follows.
If it is a brace group, then we found the intended argument of \texttt{\exp_not:n}. If it is a
space, then the next \texttt{f}-expansion will eliminate it. If it is an \texttt{N}-type token then \texttt{\text_
expand_unexpanded:N} leaves the token to be expanded if it is expandable, and otherwise
removes it, assuming that it is \texttt{\scan_stop:}. This silently hides errors when \texttt{\exp_not:n}
is incorrectly followed by some non-expandable token other than \texttt{\scan_stop:}, but this
should be pretty rare, and there is no good error recovery anyways.

(End of definition for \texttt{\text_expand:n} and others. This function is documented on page 296.)

\texttt{\text_declare_expand_equivalent:Nn}
\texttt{\text_declare_expand_equivalent:cn}
Create equivalents to allow replacement.
\texttt{\tl_clear_new:c \{ l\_text\_expand\ _\token\_to\_str:N \#1 \_tl \}}
\texttt{\tl_set:cn \{ l\_text\_expand\ _\token\_to\_str:N \#1 \_tl \}} \{\#2\}
\texttt{\cs\_generate\_variant:Nn \text\_declare\_expand\_equivalent:Nn \{ \texttt{c} \}}

\textit{(End of definition for \text\_declare\_expand\_equivalent:Nn. This function is documented on page 296.)}

Prevent expansion of various standard values.

\texttt{\tl\_map\_inline:nn}
\texttt{\{ \texttt{"\ #\ } \texttt{\catcode\^=12\ u\ .\ v\ \d\ c\ k\ b\ t\} \}}
\texttt{\{ \text\_declare\_expand\_equivalent:Nn \#1 \{ \exp\_not:n \{\#1\} \} \}}
\texttt{\tl\_map\_inline:nn}
\texttt{\{ \texttt{	ext\_AA\ } \texttt{\aa} \}}
\texttt{\text\_AE\ } \texttt{\ae}
\texttt{\text\_DH\ } \texttt{\dh}
\texttt{\text\_DJ\ } \texttt{\dj}
\texttt{\text\_IJ\ } \texttt{\ij}
\texttt{\text\_L\ } \texttt{\l}
\texttt{\text\_NG\ } \texttt{\ng}
\texttt{\text\_O\ } \texttt{\o}
\texttt{\text\_OE\ } \texttt{\oe}
\texttt{\text\_SS\ } \texttt{\ss}
\texttt{\text\_TH\ } \texttt{\th}
\texttt{\} \}}
\texttt{\text\_declare\_expand\_equivalent:Nn \#1 \{ \exp\_not:n \{\#1\} \} \}}
\texttt{\} \}}
Chapter 88

\text-case implementation

88.1 Case changing

Needed to determine the route used in titlecasing.

\bool_new:N \l_text_titlecase_check_letter_bool
\bool_set_true:N \l_text_titlecase_check_letter_bool

(End of definition for \l_text_titlecase_check_letter_bool. This variable is documented on page 299.)

The user level functions here are all wrappers around the internal functions for case changing.

\cs_new:Npn \text_lowercase:n #1 { \__text_change_case:nnn { lower } { } {#1} }
\cs_new:Npn \text_uppercase:n #1 { \__text_change_case:nnn { upper } { } {#1} }
\cs_new:Npn \text_titlecase_all:n #1 { \__text_change_case:nnn { title } { } {#1} }
\cs_new:Npn \text_titlecase_first:n #1 { \__text_change_case:nnnn { title } { break } { } {#1} }
\cs_new:Npn \text_lowercase:nn #1#2 { \__text_change_case:nnn { lower } {#1} {#2} }
\cs_new:Npn \text_uppercase:nn #1#2 { \__text_change_case:nnn { upper } {#1} {#2} }
\cs_new:Npn \text_titlecase_all:nn #1#2 { \__text_change_case:nnn { title } {#1} {#2} }
\cs_new:Npn \text_titlecase_first:nn #1#2 { \__text_change_case:nnnn { title } { break } {#1} {#2} }
\cs_new:Npn \__text_change_case:nnn #1#2#3 { \__text_change_case:nnnn {#1} {#1} {#2} {#3} }

(End of definition for \text_lowercase:n and others. These functions are documented on page 297.)

As for the expansion code, the business end of case changing is the handling of N-type tokens. First, we expand the input fully (so the loops here don’t need to worry about awkward look-aheads and the like). Then we split into the different paths.
The code here needs to be \_f-type expandable to deal with the situation where case changing is applied in running text. There, we might have case changing as a document command and the text containing other non-expandable document commands.

\cs_set_eq:NN \MakeLowercase \text_lowercase
\...
\MakeLowercase{\enquote*{A} text}

If we use an \e-type expansion and wrap each token in \exp_not:n, that would explode: the document command grabs \exp_not:n as an argument, and things go badly wrong. So we have to wrap the entire result in exactly one \exp_not:n, or rather in the kernel version.

\begin{verbatim}
\cs_new:Npn \__text_change_case:nnnn #1#2#3#4
  {\__kernel_exp_not:w \exp_after:wN
   \exp:w \exp_args:Ne \__text_change_case_auxi:nnnn
   { \text_expand:n {#4} }
   {#1} {#2} {#3}
  }
\cs_new:Npn \__text_change_case_auxi:nnnn #1#2#3#4
  {\exp_args:No \__text_change_case_BCP:nnnn
   { \tl_to_str:n {#4} } {#1} {#2} {#3}
  }
\cs_new:Npe \__text_change_case_BCP:nnnn #1#2#3#4
  {\exp_not:N \__text_change_case_BCP:nnnw
   #1#2#3#4 \tl_to_str:n { -x- -x- } #5 \tl_to_str:n { -x- } #6
   \exp_not:N \q__text_stop
  }
\cs_new:Npn \__text_change_case_BCP:nnnnw #1#2#3#4#5#6 - #7 \q__text_stop
  { \bool_lazy_or:nnTF
    { \cs_if_exist_p:c { __text_change_case_ #2 _ #6 -x- #4 :nnnnn } }
    { \tl_if_exist_p:c { l__text_ #2 case_special_ #6 -x- #4 _tl } }
    { \__text_change_case_auxii:nnnn {#1} {#2} {#3} {#6} }
    { \cs_if_exist:cTF { __text_change_case_ #2 _ #6 :nnnnn }
      { \__text_change_case_auxii:nnnn {#1} {#2} {#3} {#6} }
      { \__text_change_case_auxii:nnnn {#1} {#2} {#3} {#5} }
    }
  }
\cs_new:Npn \__text_change_case_auxii:nnnn #1#2#3#4
  { \group_align_safe_begin:
\end{verbatim}

1303
As for expansion, collect up the tokens for future use.

The main loop is the standard \texttt{tl} action type.

For a group, we \textit{could} worry about whether this contains a character or not. However, that would make life very complex for little gain: exactly what a first character is is rather weakly-defined anyway. So if there is a group, we simply assume that a character has been seen, and for title case we switch to the “rest of the tokens” situation. To avoid having too much testing, we use a two-step process here to allow the titlecase functions to be separate.
The first step of handling \(N\)-type tokens is to filter out the end-of-loop. That has to be
done separately from the first real step as otherwise we pick up the wrong delimiter. The
loop here is the same as the \texttt{expand} one, just passing the additional data long. If no
close-math token is found then the final clean-up is forced (i.e. there is no assumption
of “well-behaved” input in terms of math mode).
Once potential math-mode cases are filtered out the next stage is to test if the token grabbed is a control sequence: the two routes the code may take are then very different.
To deal with a control sequence there is first a need to test if it is on the list which indicate that case changing should be skipped. That’s done using a loop as for the other special cases. If a hit is found then the argument is grabbed and passed through as-is.

Deal with any specialist replacement for case changing.
Allow for manually-controlled case switching.

Skip over material quickly after titlecase-first-only initials
Letter-like commands may still be present: they are set up using a simple lookup approach, so can easily be handled with no loop. If there is no hit, we are at the end of the process: we loop around. Letter-like chars are all available only in upper- and lowercase, so titlecasing maps to the uppercase version.

Check for a customised codepoint result.
For upper- and lowercase changes, once we get to this stage there are only a couple of questions remaining: is there a language-specific mapping and is there the special case of a terminal sigma. If not, then we pass to a simple codepoint mapping.

If the current character is an uppercase sigma, the a check is made on the next item in the input. If it is N-type and not a control sequence then there is a look-ahead phase: the logic here is simply based on letters or actives (to cover 8-bit engines).
For titlecasing, we need to obtain the general category of the current codepoint.

\cs_new:Npn \_text_change_case_codepoint_title:nnn #1#2#3
\{ #4 \}
\cs_new:Npn \_text_change_case_codepoint_title:nnnn #1#2#3#4
\{ #5 \}
We need to ensure that only valid catcode-extraction is attempted. That’s fine with
Unicode engines but needs a bit of work with 8-bit ones. The logic is that if the original codepoint was in the ASCII range, we keep the catcode. Otherwise, if the target is in the ASCII range, we use the standard catcode. If neither are true, we set as 13 on the grounds that this will be what is used anyway!

```latex
\begin{verbatim}
\bool_lazy_or:nnTF
  \sys_if_engine_luatex_p:
  \sys_if_engine_xetex_p:
  { \cs_new:Npn \__text_change_case_catcode:nn #1#2
      \__text_char_catcode:N #1 }
  { \cs_new:Npn \__text_change_case_catcode:nn #1#2
      { \__text_codepoint_compare:nNnTF {#1} < { "80 }
      \__text_char_catcode:N #1 }
      { \int_compare:nNnTF {#2} < { "80 }
      \char_value_catcode:n {#2} }
      { 13 } }
  }

\cs_new:Npn \__text_change_case_next_lower:nnn #1#2#3
  \__text_change_case_loop:nnnw {#1} {#2} {#3}
\cs_new_eq:NN \__text_change_case_next_upper:nnn
  \__text_change_case_next_lower:nnn
\cs_new_eq:NN \__text_change_case_next_title:nnn
  \__text_change_case_next_lower:nnn
\cs_new:Npn \__text_change_case_next_end:nnn #1#2#3
  \__text_change_case_skip:nnw {#2} {#3}
\end{verbatim}

(End of definition for \textit{\_text_change_case:nnnn} and others.)

\textit{\text declare case equivalent:Nn}

Create equivalents to allow replacement.

```latex
\begin{verbatim}
\cs_new_protected:Npn \text Declare_case_equivalent:Nn #1#2
  { \__text_declare_case_equivalent:nnn { lower } {#1} {#2} }
\cs_new_protected:Npn \text Declare_titlecase_mapping:nn #1#2
  { \__text_declare_case_mapping:nnn { title } {#1} {#2} }
\cs_new_protected:Npn \text Declare_uppercase_mapping:nn #1#2
  { \__text_declare_case_mapping:nnn { upper } {#1} {#2} }
\end{verbatim}

(End of definition for \textit{\text declare case equivalent:Nn}. This function is documented on page 298.)

\textit{\text declare lowercase mapping:nn}
\textit{\text declare titlecase mapping:nn}
\textit{\text declare uppercase mapping:nn}
\textit{\text declare lowercase mapping:nnn}
\textit{\text declare titlecase mapping:nnn}
\textit{\text declare uppercase mapping:nnn}
\textit{\text declare lowercase mapping:nnn}
\textit{\text declare case mapping_aux:nnn}
\textit{\text declare case mapping_aux:nnn}
\textit{\text declare case mapping_aux:nnn}

Codepoint customisation.

```latex
\begin{verbatim}
\exp_args:Ne \__text_declare_case_mapping_aux:nnn
  \codepoint_str_generate:n {#2} ) {#1} {#3}
\end{verbatim}

1313
\text_case_switch:nnnn
\__text_case_switch_marker:

Set up the mechanism for manual case switching.

\cs_new:Npn \__text_case_switch_marker: #1
\cs_new:Npn \text_case_switch:nnnn #1#2#3#4
\__text_case_switch_marker: #1

(End of definition for \text_case_switch:nnnn and \__text_case_switch_marker:. This function is documented on page 298.)

\__text_change_case_generate:n
\__text_change_case_upper_de-x-eszett:nnnnnn
\__text_change_case_upper_de-alt:nnnnnn

A utility.

\cs_new:Npn \__text_change_case_generate:n #1
\text_change_case_upper_de-x-eszett:nnnnnn
\__text_change_case_upper_de-alt:nnnnnn

(End of definition for \__text_change_case_generate:n.)

A simple alternative version for German.

\cs_new:c { __text_change_case_upper_de-x-eszett:nnnnnn } #1#2#3#4#5
\__text_codepoint_compare:nNnTF {#5} = { "00DF }
\__text_change_case_store:e
\use:c { __text_change_case_next_ #2 :nnn }

\UseOf{1314}
For Greek uppercasing, we need to know if characters in the Greek range have accents. That means doing a NFD conversion first, then starting a search. As described by the Unicode CLDR, Greek accents need to be found after any U+0308 (diaeresis) and are done in two groups to allow for the canonical ordering. The implementation here follows the data and examples from ICU (https://icu.unicode.org/design/case/greek-upper), although necessarily the implementation is somewhat different. The ὑπογράμμην is filtered out here as it is not actually in the Greek range, so gets lost if we leave until later. The one Greek codepoint we skip is the numeral sign and question mark: the first has an awkward NFD for pdfTeX so is best left unchanged, and the latter has issues concerning how LGR outputs the input and output (differently!).
At this stage we have the first NFD codepoint as \#3. What we need to know is whether after that we have another character, either from the NFD or directly in the input. If not, we store the changed character at this stage.

Now, we check the detail of the next codepoint: again we filter out the not-a-char cases, before checking if it’s an dialytika, accent or diacritic. (The latter do not have the same hiatus behavior as accents.) There is additional work if the codepoint can take a ypogegrammeni: there, we need to move any ypogegrammeni to after accents (in case the input is not normalised). The ypogegrammeni itself is handled separately.

Now, we check the detail of the next codepoint: again we filter out the not-a-char cases, before checking if it’s an dialytika, accent or diacritic. (The latter do not have the same hiatus behavior as accents.) There is additional work if the codepoint can take a ypogegrammeni: there, we need to move any ypogegrammeni to after accents (in case the input is not normalised). The ypogegrammeni itself is handled separately.
We handle \textit{dialytika} in parts as it’s also needed for the hiatus. We know only two letters take it, so we can shortcut here on the second part of the tests.

Adding a hiatus needs some of the same ideas, but if there is not one we skip this code point, hence needing a separate function.
Handling the ὑπογεγραμμενὶ output depends on the selected approach

\cs_new:Npn \__text_change_case_upper_el_ypogegrammeni:n #1
\exp_args:Nm { \__text_change_case_generate:n \int_case:n \__text_codepoint_from_chars:Nw \int_case:n } { "0391" \{"1FBC \} \"03B1" \{"1FBC \} \"0397" \{"1FCC \} \"03B7" \{"1FCC \} \"03A9" \{"1FFC \} \"03C9" \{"1FFC \} }
\cs_new:cpn { \__text_change_case_upper_el-x-iota_ypogegrammeni:n } #1
We choose to retain stress diacritics, but we also need to recombine them for pdfTEX. That is handled here.

\cs_new:Npn \__text_change_case_upper_el_stress:nn #1#2
{ \exp_args:Ne \__text_change_case_generate:n
{ \int_case:nn
{ \__text_codepoint_from_chars:Nw #2 }
{ "0304 
{ \int_case:nn
{ \__text_codepoint_from_chars:Nw #1 }
{ "0391 } { "1FB9 }
{ "03B1 } { "1FB9 }
{ "0399 } { "1FD9 }
{ "03B9 } { "1FD9 }
{ "03A5 } { "1FE9 }
{ "03C5 } { "1FE9 }
}
{ "0306 }
{ \int_case:nn
{ \__text_codepoint_from_chars:Nw #1 }
{ "0391 } { "1FB8 }
{ "03B1 } { "1FB8 }
{ "0399 } { "1FD8 }
{ "03B9 } { "1FD8 }
{ "03A5 } { "1FE8 }
{ "03C5 } { "1FE8 }
}
}
\tl_if_head_is_N_type:nTF {#4}
{ \__text_change_case_upper_el_gobble:nnnW
{ \__text_change_case_upper_el_gobble:nnnW }
\__text_change_case_loop:nnnW
{#1} {#2} {#3} #4 \q__text_recursion_stop
}
\cs_new:Npn \__text_change_case_upper_el_gobble:nnnW
{ \__text_change_case_upper_el_gobble:nnnW
\__text_change_case_upper_el_gobble:nnnW }
\q__text_recursion_stop
}
\cs_new:Npn \__text_change_case_upper_el_gobble:nnnW
{ \__text_change_case_upper_el_gobble:nnnW
\__text_change_case_upper_el_gobble:nnnW }
\q__text_recursion_stop
\token_if_cs:NTF #4
\endinput
Luckily the Greek range is limited and clear.

We follow ICU in adding a few extras to the accent list here.
\if_int_compare:w \l_\text_change_case_if_greek_spacing_diacritic:n \space #1  = "0300 \exp_stop_f:
\prg_return_true:
\else:
\if_int_compare:w \l_\text_change_case_if_greek_spacing_diacritic:n \space #1  = "0301 \exp_stop_f:
\prg_return_true:
\else:
\if_int_compare:w \l_\text_change_case_if_greek_spacing_diacritic:n \space #1  = "0342 \exp_stop_f:
\prg_return_true:
\else:
\if_int_compare:w \l_\text_change_case_if_greek_spacing_diacritic:n \space #1  = "0302 \exp_stop_f:
\prg_return_true:
\else:
\if_int_compare:w \l_\text_change_case_if_greek_spacing_diacritic:n \space #1  = "0303 \exp_stop_f:
\prg_return_true:
\else:
\if_int_compare:w \l_\text_change_case_if_greek_spacing_diacritic:n \space #1  = "0311 \exp_stop_f:
\prg_return_true:
\else:
\prg_return_false:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\else:
\if_int_compare:w \l_\text_change_case_if_greek_spacing_diacritic:n \space #1  < "1FBD \exp_stop_f:
\if_int_compare:w \l_\text_change_case_if_greek_spacing_diacritic:n \space #1  = "030A \exp_stop_f:
\prg_return_true:
\else:
\prg_return_false:
\fi:
\else:
\if_int_compare:w \l_\text_change_case_if_greek_spacing_diacritic:n \space #1  = "1FBD \exp_stop_f:
\prg_return_true:
\else:
\if_int_compare:w \l_\text_change_case_if_greek_spacing_diacritic:n \space #1  = "1FBF \exp_stop_f:
\prg_return_true:
\else:
\if_int_compare:w \l_\text_change_case_if_greek_spacing_diacritic:n \space #1  = "1FC0 \exp_stop_f:
\prg_return_true:
\else:
\if_int_compare:w \l_\text_change_case_if_greek_spacing_diacritic:n \space #1  = "1FC1 \exp_stop_f:
\prg_return_true:
\else:
\if_int_compare:w \l_\text_change_case_if_greek_spacing_diacritic:n \space #1  = "1FCD \exp_stop_f:
\begin{verbatim}
{ \int_eval:n { \__text_codepoint_from_chars:Nw #1 } }
\cs_new:Npn \__text_change_case_if_greek_breathing:n #1
{ \if_int_compare:w #1 = "0313 \exp_stop_f:
  \prg_return_true:
  \else:
  \if_int_compare:w #1 = "0314 \exp_stop_f:
    \prg_return_true:
    \else:
    \prg_return_false:
    \fi:
  \fi:
\cs_new:Npn \__text_change_case_if_greek_stress:n #1
{ \exp_args:Nf \__text_change_case_if_greek_stress:n
  \{ \int_eval:n { \__text_codepoint_from_chars:Nw #1 } \} }
\cs_new:Npn \__text_change_case_if_takes_dialytika:n #1
{ \exp_args:Nf \__text_change_case_if_takes_dialytika:n
  \{ \int_eval:n { \__text_codepoint_from_chars:Nw #1 } \} }
\cs_new:Npn \__text_change_case_if_takes_dialytika:n #1
{ \if_int_compare:w #1 = "0304 \exp_stop_f:
  \prg_return_true:
  \else:
  \if_int_compare:w #1 = "0306 \exp_stop_f:
    \prg_return_true:
    \else:
    \prg_return_false:
    \fi:
  \fi:
\prg_new_conditional:Npnn \__text_change_case_if_takes_dialytika:n #1 { TF }
{ \exp_args:Nf \__text_change_case_if_takes_dialytika:n
  \{ \int_eval:n { \__text_codepoint_from_chars:Nw #1 } \} }
\cs_new:Npn \__text_change_case_if_takes_dialytika:n #1
{ \if_int_compare:w #1 = "0399 \exp_stop_f:
  \prg_return_true:
  \else:
  \if_int_compare:w #1 = "03B9 \exp_stop_f:
    \prg_return_true:
    \else:
    \if_int_compare:w #1 = "03A5 \exp_stop_f:
      \prg_return_true:
      \else:
      \if_int_compare:w #1 = "03C5 \exp_stop_f:
        \prg_return_true:
        \else:
        \prg_return_false:
        \fi:
      \fi:
    \fi:
  \fi:
\prg_new_conditional:Npnn \__text_change_case_if_takes_dialytika:n #1 { TF }
{ \exp_args:Nf \__text_change_case_if_takes_dialytika:n
  \{ \int_eval:n { \__text_codepoint_from_chars:Nw #1 } \} }
\cs_new:Npn \__text_change_case_if_takes_dialytika:n #1
{ \if_int_compare:w #1 = "0399 \exp_stop_f:
  \prg_return_true:
  \else:
  \if_int_compare:w #1 = "03B9 \exp_stop_f:
    \prg_return_true:
    \else:
    \if_int_compare:w #1 = "03A5 \exp_stop_f:
      \prg_return_true:
      \else:
      \if_int_compare:w #1 = "03C5 \exp_stop_f:
        \prg_return_true:
        \else:
        \prg_return_false:
        \fi:
      \fi:
    \fi:
  \fi:
\end{verbatim}
There is one thing that needs special treatment at the start of words in Greek. For an
isolated accent eta, which is handled by seeing if we have exactly one of the affected
codepoints followed by a space or brace group.
In Greek, breathing diacritics are normally dropped when uppercasing; see the code for the general case. However, for the first character of a word, if there is a breather and the next character takes a dialytika, it needs to be added. We start by checking if the current codepoint is in the Greek range, then decomposing.

Normal form decomposition will always give between one and three codepoints. Luckily, the two breathing marks (psilli and dasia) will be in a predictable position: last. So we can quickly establish first that there was a change on decomposition, and second if the final resulting codepoint is one of the two we care about.
Now the lookahead can be fired: check the next codepoint and assess whether it takes a dialytika. Drop the breathing mark or generate the dialytika: the latter is code shared with the general mechanism.
Titlecasing retains accents, but to prevent the uppercasing code from kicking in, there
has to be an explicit function here.

\cs_new:Npn \__text_change_case_title_el:nnnnn #1#2#3#4#5
\{ \__text_change_case_codepoint:nnnnn {#1} {#2} {#3} {#4} {#5} \}
(End of definition for \__text_change_case_title_el:nnnnn.)


\cs_new:Npn \__text_change_case_upper_by:nnnnn #1#2#3#4#5
\{ \__text_codepoint_compare:nNnTF {#5} = { "0587 } \}
\{ \__text_change_case_codepoint:mnna {#1} {#2} {#3} {#4} {#5} \}
\cs_new:Npn \__text_change_case_title_by:nnnnn #1#2#3#4#5
\{ \__text_codepoint_compare:nNnTF {#5} = { "0587 } \}
\{ \__text_change_case_codepoint:mnna {#1} {#2} {#3} {#4} {#5} \}
\cs_new:Npn \__text_change_case_upper_by-x-yiwn:nnnnn #1#2#3#4#5
\{ \__text_codepoint_compare:nNnTF {#5} = { "0587 } \}
\{ \__text_change_case_codepoint:mnna {#1} {#2} {#3} {#4} {#5} \}
\cs_new_eq:cc \__text_change_case_title_by-x-yiwn:nnnnn \__text_change_case_upper_by-x-yiwn:nnnnn
(End of definition for \__text_change_case_upper_by:nnnnn and others.)
Simply swaps of characters.

For Lithuanian, the issue to be dealt with is dots over lower case letters: these should be present if there is another accent. The first step is a simple match attempt: look for the three uppercase accented letters which should gain a dot-above char in their lowercase form.

If there was a hit, output the result with the dot-above and move on. Otherwise, look for one of the three letters that can take a combining accent: I, J, and I-ogonek.
Again, branch depending on a hit. If there is one, we output the character then need to look for a combining accent: as usual, we need to be aware of the loop situation.

\cs_new:Npn \__text_change_case_lower_lt_auxii:nnnnn #1#2#3#4#5
\tl_if_blank:nTF {#1}
\__text_change_case_codepoint:nnnnn {#2} {#2} {#3} {#4} {#5}
\__text_change_case_store:e
\codepoint_generate:nn { "0069 }
\__text_change_case_catacode:nn {#5} { "0069 }
\codepoint_generate:nn { "0307 }
\__text_change_case_catacode:nn {#5} { "0307 }
\codepoint_generate:nn {#1}
\__text_change_case_catacode:nn {#5} {#1}
\__text_change_case_loop:nnnw {#2} {#3} {#4}
}
\cs_new:Npn \__text_change_case_lower_lt:nnnw #1#2#3#4 \q__text_recursion_stop
\tl_if_head_is_N_type:nTF {#4}
\__text_change_case_lower_lt:nnnN #1#2#3#4 \q__text_recursion_stop
\cs_new:Npn \__text_change_case_lower_lt:nnnN #1#2#3#4
\__text_codepoint_process:nN
\__text_change_case_lower_lt:nnnn {#1} {#2} {#3} #4
\cs_new:Npn \__text_change_case_lower_lt:nnnn #1#2#3#4
\bool_lazy_and:nnT
\bool_lazy_or_p:nn #4
\bool_lazy_and:nnT
\bool_lazy_or_p:nn
\begin{verbatim}
\__text_change_case_upper_lt:nnnw #1 #2 #3 #4 
  \exp_args:Ne \__text_change_case_upper_lt_aux:nnnnn #1 #2 #3 #4 #5
  \int_case:nn \__text_change_case_upper_lt:nnnn \#1 \#2 \#3 \#4 \#5
  \exp_args:Ne \__text_change_case_upper_lt:nnnnn #1 #2 #3 #4 #5
  \tl_if_blank:nTF {#1}
    \__text_change_case_upper_lt:nnnn { upper } {#2} {#3} {#4} {#5}
    \text_change_case_store:e
    \codepoint_generate:nn {#4}
    \__text_change_case_upper_lt:nnnN {#1} {#2} {#3} #4
  
  \tl_if_single_p:n {#4}
  \token_if_cs_p:N #4
  \bool_lazy_any_p:n
  \__text_codepoint_compare_p:nNn {#4} = { "0300 }
  \__text_codepoint_compare_p:nNn {#4} = { "0301 }
  \__text_codepoint_compare_p:nNn {#4} = { "0303 }

\__text_change_case_store:e
\codepoint_generate:nn { "0307 }
\__text_change_case_catcode:nn {#4} { "0307 }
\__text_change_case_loop:nnnw {#1} {#2} {#3} #4
\end{verbatim}

The uppercasing version: first find i/j/i-ogonek, then look for the combining char: drop it if present.

\begin{verbatim}
\cs_new:Npn \__text_change_case_upper_lt:nnnnn #1 #2 #3 #4 #5
  \exp_args:Ne \__text_change_case_upper_lt_aux:nnnnn #1 #2 #3 #4 #5
  \int_case:nn \__text_codepoint_from_chars:Nw #5 \#5
  \{ "0069" \"0049"
  \{ "006A" \"004A"
  \{ "012F" \"012E"
  \}
  \{#2} \{#3} \{#4} \{#5\}
\end{verbatim}
For Dutch, there is a single look-ahead test for ij when title casing. If the appropriate letters are found, produce IJ and gobble the j/J.
The Turkic languages need special treatment for dotted-i and dotless-i. The lower casing rule can be expressed in terms of searching first for either a dotless-I or a dotted-I. In the latter case the mapping is easy, but in the former there is a second stage search.

After a dotless-I there may be a dot-above character. If there is then a dotted-i should be produced, otherwise output a dotless-i. When the combination is found both the dotless-I and the dot-above char have to be removed from the input.

1333
\cs_new:Npn \__text_change_case_lower_tr:Nnnn \#1\#2\#3\#4\#5
\text_codepoint_process:nN
{ \__text_change_case_lower_tr:Nnnn \#1 \{\#2\} \{\#3\} \{\#4\} } \#5
\cs_new:Npn \__text_change_case_lower_tr:Nnnn \#1\#2\#3\#4\#5
{ \bool_lazy_or:nnTF
    { \bool_lazy_and_p:nn
        { \tl_if_single_p:n \{\#5\} }
        { \token_if_cs_p:N \#5 }
    }
    { \use:c { __text_change_case_next_ \#2 } \{\#2\} \{\#3\} \{\#4\} \#5 }
}
{ \__text_change_case_store:e
    { \codepoint_generate:nn \{\"0131\} }
    { \__text_change_case_catcode:nn \{\#1\} \{\"0131\} }
}
\__text_change_case_loop:nnnw \{\#2\} \{\#3\} \{\#4\} 
\__text_change_case_upper_tr:nnnnn
Uppercasing is easier: just one exception with no context.
\cs_new:Npn \__text_change_case_upper_tr:nnnnn \#1\#2\#3\#4\#5
{ \__text_codepoint_compare:nNnTF \{\#5\} \{\"0069\} }
{ \__text_change_case_store:e
    { \codepoint_generate:nn \{\"0130\} }
    { \__text_change_case_catcode:nn \{\#1\} \{\"0130\} }
}
\__text_change_case_loop:nnnw \{\#2\} \{\#3\} \{\#4\} \{\#5\}
\__text_change_case_lower_az:nnnnn
\__text_change_case_upper_az:nnnnn
Straight copies.
The (fixed) look-up mappings for letter-like control sequences.

\begin{verbatim}
\cs_new_eq:NN \__text_change_case_upper_az:nnnnn
\__text_change_case_upper_tr:nnnnn

(End of definition for \__text_change_case_lower_az:nnnnn and \__text_change_case_upper_az:nnnnn.)

The (fixed) look-up mappings for letter-like control sequences.

\begin{verbatim}
\group_begin:
\cs_set_protected:Npn \__text_change_case_setup:NN #1#2
{
  \quark_if_recursion_tail_stop:N #1
  \tl_const:cn { c__text_lowercase_ \token_to_str:N #1 _tl } { #2 }
  \tl_const:cn { c__text_uppercase_ \token_to_str:N #2 _tl } { #1 }
}\__text_change_case_setup:NN
\__text_change_case_setup:NN
\AA \aa
\AE \ae
\DH \dh
\DJ \dj
\IJ \ij
\L \l
\NG \ng
\O \o
\OE \oe
\SS \ss
\TH \th
\q_recursion_tail ?
\q_recursion_stop
\tl_const:cn { c__text_uppercase_ \token_to_str:N \i _tl } { I }
\tl_const:cn { c__text_uppercase_ \token_to_str:N \j _tl } { J }
\group_end:

To deal with possible encoding-specific extensions to \@uclclist, we check at the end of the preamble. This will therefore only apply to \texttt{\LaTeX}\textsuperscript{2e} package mode.

\begin{verbatim}
\tl_if_exist:NT \@expl@finalise@setup@@
{
  \tl_gput_right:Nn \@expl@finalise@setup@@
  {\tl_gput_right:Nn \@kernel@after@begindocument
    {\group_begin:
      \cs_set_protected:Npn \__text_change_case_setup:Nn #1#2
      {
        \quark_if_recursion_tail_stop:N #1
        \tl_if_single_token:nT {#2}
        {\tl_if_exist:cF
          { c__text_uppercase_ \token_to_str:N #1 _tl }
          \tl_const:cn { c__text_uppercase_ \token_to_str:N #1 _tl }
          { #2 }
        }
      }
    \group_end:
  }
}
\end{verbatim}
\end{verbatim}

1335
A few adjustments to case mapping for combining chars: these are not needed for the Unicode engines

Chapter 89

\text-map implementation

\begin{verbatim}
\text_map_function:nN
  \text_map_function:nN
  \text_map_loop:Nnw
  \text_map_group:Nnn
  \text_map_space:Nnw
  \text_map\_N\_type:NnN
  \text_map\_codepoint:Nnn
  \text_map\_CR:Nnw
  \text_map\_CR:NnN
  \text_map\_class:Nnnn
  \text_map\_class:nNnnn
  \text_map\_class_loop:Nnnnw
  \text_map\_Prepend:Nnn
  \text_map\_Prepend\_aux:Nnn
  \text_map\_Prepend:nNnn
  \text_map\_Prepend\_loop:Nnnw
  \text_map\_not\_Control:Nnn
  \text_map\_not\_Prepend:Nnn
  \text_map\_not\_SpacingMark:Nnn
  \text_map\_not\_Prepend\_loop:Nnn
  \text_map\_not\_C\_type:NnN
  \text_map\_not\_LV:Nnn
  \text_map\_not\_V\_type:Nnn
  \text_map\_not\_T\_type:Nnn
  \text_map\_L\_type:Nnn
  \text_map\_LV\_type:Nnn
  \text_map\_V\_type:Nnn
  \text_map\_T\_type:Nnn
  \text_map\_hangul\_N\_type:Nnn
  \text_map\_hangul\_aux:Nnnnn
  \text_map\_hangul\_loop:Nnnnn
\end{verbatim}

89.1 Mapping to text

The standard lead-off for an action loop.

\begin{verbatim}
\cs_new:Npn \text_map\_function:nN \#1\#2
  \exp_args:Ne \__text_map\_function:nN \text\_expand:n {\#1} \#2
\cs_new:Npn \__text_map\_function:nN \#1\#2
  \__text_map\_loop:Nnw \#2 {} \#1
  \q__text\_recursion\_tail \q__text\_recursion\_stop
  \prg\_break\_point:Nn \text\_map\_break: {} \#1
\end{verbatim}

The standard set up for an “action” loop. Groups are handled by recursion, spaces are treated similarly: both count as grapheme boundaries. For \text{-type} tokens, we filter out control sequences (again a boundary), then move on to further analysis.

\begin{verbatim}
\cs_new:Npn \__text_map\_loop:Nnw \#1\#2\#3 \q__text\_recursion\_stop
  \tl\_if\_head\_is\_N\_type:nTF \#3
    \__text_map\_N\_type:NnN
  \else
    \tl\_if\_head\_is\_group:nTF \#3
      \__text_map\_group:Nnn
      \__text_map\_space:Nnw
    \fi
  \fi
  \#1 {#2} \#3 \q__text\_recursion\_stop
\end{verbatim}

\begin{verbatim}
\cs_new:Npn \__text_map\_group:Nnn
  \__text_map\_output:Nn \#1 \#2
  \__text_map\_loop:Nnw \#2 {} \#1
  \q__text\_recursion\_tail \q__text\_recursion\_stop
\end{verbatim}

1337
We pull out a few special cases here. Carriage returns case needs a bit of context handling so has an auxiliary. Codepoint U+200D is the zero-width joiner, which has no context to concern us: just don’t break.

A carriage return is a boundary unless it is immediately followed by a line feed, in which case that pair is a boundary.
There are various classes of character, and we deal with them all in the same general way. We need to example the relevant list of codepoints: if we get a hit, then we do whatever the relevant action is. Otherwise we loop, but only if the current codepoint could still match: the loop stops early otherwise and we move forward.
Keep collecting.

\cs_new:Npn \__text_map_Extend:Nnn \#1 \#2 \#3
{ \__text_map_loop:Nnw \#1 \#2 \#3 }
\cs_new_eq:NN \__text_map_SpacingMark:Nnn \__text_map_Extend:Nnn

Outputting anything earlier, the combine with what follows. The only exclusions are control characters.

\cs_new:Npn \__text_map_Prepend:Nnn \#1 \#2 \#3
{ \__text_map_output:Nn \#1 \#2
\__text_map_lookahead:Nnn \#1 \#3 \__text_map_Prepend_aux:Nnn
}
\cs_new:Npn \__text_map_Prepend_aux:Nnn \#1 \#2 \#3
{ \bool_lazy_or:nnTF
{ \__text_codepoint_compare_p:nNn { \#3 } = { "0A } }
{ \__text_codepoint_compare_p:nNn { \#3 } = { "0D } }
{ \#1 \#2
\__text_map_loop:Nnw \#1 \#3 }
{ \__text_map_Prepend:nNnn \c__text_grapheme_Control_clist
\#1 \#2 \#3 }
}
\cs_new:Npn \__text_map_Prepend:nNnn \#1 \#2 \#3 \#4
{ \__text_map_Prepend_loop:Nnn \#2 \#3 \#4
\#1 , \q__text_recursion_tail .. , \q__text_recursion_stop
}
\cs_new:Npn \__text_map_Prepend_loop:Nnn \#1 \#2 \#3 \#4
{ \__text_if_q_recursion_tail_stop_do:nn { \#4 }
\__text_map_loop:Nnw \#1 \#2 \#3
\__text_codepoint_compare:nNnTF { \#3 } < { "#4 } 
{ \__text_map_class_end:nw
\__text_map_loop:Nnw \#1 \#2 \#3 }
{ \__text_map_class_end:nw
\__text_codepoint_compare:nNnTF { \#3 } > { "#5 } 
{ \__text_map_Prepend_loop:Nnn \#1 \#2 \#3 }
{ \__text_map_class_end:nw
\__text_map_loop:Nnw \#1 \#2 \#3 }
}
}
Dealing with end-of-class is done such that we can be flexible.

Hangul needs additional treatment. First we have to deal with the start-of-Hangul position: output what we had up to now, then move the specialist handler. The idea here is to pick off the different codepoint types one at a time, tracking what else can be considered at each stage until we hit the end of the viable types. Other than that, we just keep building up the Hangul codepoints using a dedicated version of the loop from above.
\cs_new:Npn \__text_map_hangul:NnnN #1 {#2} {#3} {#4}
\{ 
  #4 \q__text_recursion_stop
\}
\cs_new:Npn \__text_map_hangul:Nnnn #1 {#2} {#3} {#4}
\{ 
  \__text_map_hangul_aux:Nnnw #1 {#2} {#4} #3 ; \q_recursion_tail ; \q_recursion_stop
\}
\cs_new:Npn \__text_map_hangul_loop:Nnnnnw #1 #2 #3 #4 #5 #6 .. #7 ,
\{ 
  \__text_if_q_recursion_tail_stop_do:nn {#6}
  \{ \__text_map_hangul_next:Nnnn #1 {#2} {#3} {#5} \}
  \__text_codepoint_compare:nNnTF {#3} < { "#6} 
  \{ \__text_map_hangul_end:ns \}
  \} 
  \__text_codepoint_compare:nNnTF {#3} > { "#7} 
  \{ \__text_map_hangul_next:Nnnn #1 {#2} {#3} {#4} {#5} \} 
  \}
\}
The Regional Indicator rule means looking ahead and dealing with the case where there are two in a row. So we use a look ahead to pick them off. As there is only one range the values are hard-coded.

A generic loop-ahead setup.
\text_map:nn The standard non-expandable inline version.

\text_map_inline:nn (End of definition for \text_map_function:nn and others. These functions are documented on page 300.)

\text_map_inline:nn (End of definition for \text_map_inline:nn. This function is documented on page 300.)

\package
Chapter 90

13text-purify implementation

90.1 Purifying text

Functions to query recursion quarks.

As in the other parts of the module, we start off with a standard “action” loop, with
expansion applied up-front.

As for expansion, collect up the tokens for future use.
The main loop is a standard “tl action”. Unlike the expansion or case changing, here any groups have to be run inline. Most of the business end is as before in the \texttt{N}-type token processing.

\begin{verbatim}
\cs_new:Npn \__text_purify_loop:w #1 \q__text_recursion_stop
{\tl_if_head_is_N_type:nTF {#1}
 { \__text_purify_N_type:N }
 { \tl_if_head_is_group:nTF {#1}
   { \__text_purify_group:n }
   { \__text_purify_space:w }
 }
 #1 \q__text_recursion_stop
}
\cs_new:Npn \__text_purify_group:n #1 { \__text_purify_loop:w #1 }
\exp_last_unbraced:NNo \cs_new:Npn \__text_purify_space:w \c_space_tl
{ \__text_purify_store:n { ~ }
 \__text_purify_loop:w
}
\cs_new:Npn \__text_purify_N_type:N #1
{\__text_if_q_recursion_tail_stop_do:Nn #1 { \__text_purify_end:w }
 \__text_purify_N_type_aux:N #1 }
\cs_new:Npn \__text_purify_N_type_aux:N #1
{\exp_after:wN \__text_purify_math_search:NNN \exp_after:wN #1 \l_text_math_delims_tl
 \q__text_recursion_tail ?
 \q__text_recursion_stop
}
\cs_new:Npn \__text_purify_math_search:NNN #1#2#3
{\__text_if_q_recursion_tail_stop_do:Nn #2
 { \__text_purify_math_start:NNw #2 #3 }
 \token_if_eq_meaning:NNTF \texttt{#1} \#1
 { \__text_use_i_delimit_by_q_recursion_stop:nw
   { \__text_purify_math_start:NNw \#2 \#3 }
 }
 { \__text_purify_math_search:NNN #1 #2 \#3 }
}
\cs_new:Npn \__text_purify_math_start:NNw #1 #2 #3
{ \__text_purify_math_loop:NNw \#1#2#3 \q__text_recursion_stop
 \__text_purify_math_result:n { }
}
\cs_new:Npn \__text_purify_math_loop:NNw #1#2#3 \q__text_recursion_stop
\__text_purify_math_result:n #1
\end{verbatim}

The first part of handling math mode is exactly the same as in the other functions: look for a start-of-math mode token and if found start a new loop tracking the closing token.
Then handle math mode as an argument: same outcomes, different input syntax.
For \textsc{N}-type tokens, we first look for a string-context replacement before anything else: this can therefore cover anything. Assuming we don’t find one, check to see if we can expand control sequences: if not, they have to be dropped. We also allow for \LaTeX \texttt{\protect} \{ \texttt{\oops} \} or similar, but that’s also in the expansion code and seems like a reasonable balance.
Handle encoding commands, as detailed for expansion.

\cs_new:Npn \__text_purify_encoding:N #1
\{ \bool_lazy_or:nnTF
\{ \cs_if_eq_p:NN #1 \@current@cmd \}
\{ \cs_if_eq_p:NN #1 \@changed@cmd \}
\{ \__text_purify_encoding_escape:NN \}
\} \__text_if_expandable:NTF #1
\{ \exp_after:wN \__text_purify_loop:w #1 \}
\{ \__text_purify_loop:w \}
\}
\cs_new:Npn \__text_purify_encoding_escape:NN #1#2
\{ \__text_purify_store:n {#1} \__text_purify_loop:w \}

(End of definition for \text_purify:n and others. This function is documented on page 299.)

\text_declare_purify_equivalent:Nn \text_declare_purify_equivalent:Ne
\cs_new_protected:Npn \text_declare_purify_equivalent:Nn #1#2
\{ \tl_clear_new:c { l__text_purify_ \token_to_str:N #1 _tl } \tl_set:cn { l__text_purify_ \token_to_str:N #1 _tl } {#2} \}
\cs_generate_variant:Nn \text_declare_purify_equivalent:Nn { Ne }

(End of definition for \text_declare_purify_equivalent:Nn. This function is documented on page 299.)

Now pre-define a range of standard commands that need dedicated definitions in purified text. First handle font-related stuff: all of this needs to be disabled.

\tl_map_inline:nn
\{ \fontencoding \fontfamily \fontseries \fontshape \}
\{ \text_declare_purify_equivalent:Nn #1 \{ \use_none:n \} \}
\text_declare_purify_equivalent:Nn \fontsize \{ \use_none:nn \}
\text_declare_purify_equivalent:Nn \selectfont \{ \}
\text_declare_purify_equivalent:Nn \usefont \{ \use_none:nnnn \}
\tl_map_inline:nn
\{ \textit \text \textnormal \texttt \}

1349
Environments have to be handled by pure expansion.

\_\_text\_end\_env:n

\begin {\text \_\_text\_end\_env:n} \text \_\_text\_end\_env:n \end {\_\_text\_end\_env:n} \cs_new:Npn \_\_text\_end\_env:n \#1 { \cs:w end #1 \cs_end: }

(End of definition for \_\_text\_end\_env:n.)

Some common symbols and similar ideas.

\text \_\_text\_end\_env:n \label { \use: none:n }

Cross-referencing.
Spaces.

\begin{verbatim}
\group_begin:
\char_set_catcode_active:N \-
\use:n
{ \group_end:
  \char_set_catcode_active:N { \c_space_tl }
}
\text_declare_purify_equivalent:Ne ~ { \c_space_tl }
\text_declare_purify_equivalent:Nn \nobreakspace { ~ }
\text_declare_purify_equivalent:Nn \ { ~ }
\text_declare_purify_equivalent:Nn \, { ~ }
\end{verbatim}

90.2 Accent and letter-like data for purifying text

In contrast to case changing, both 8-bit and Unicode engines need information for text purification to handle accents and letter-like functions: these all need to be removed. However, the results are of course engine-dependent.

For the letter-like commands, life is relatively easy: they are all simply added as standard exceptions. The only oddity is \SS, which gets converted to two letters. (At some stage an alternative version can presumably be added to babel or similar.)

\begin{verbatim}
\cs_set_protected:Npn \__text_loop:Nn #1#2
{ \quark_if_recursion_tail_stop:N #1
  \text_declare_purify_equivalent:Ne #1
  { \codepoint_generate:nn {"#2} { \char_value_catcode:n {"#2} } }
\__text_loop:Nn }
\__text_loop:Nn \AA { 00C5 }
\AE { 00C6 }
\DH { 00D0 }
\DJ { 0110 }
\IJ { 0132 }
\L { 0141 }
\NG { 014A }
\O { 00D8 }
\OE { 0152 }
\TH { 00DE }
\aa { 00E5 }
\ae { 00E6 }
\dh { 00F0 }
\dj { 0111 }
\l { 0131 }
\j { 0237 }
\lj { 0132 }
\l { 0142 }
\ng { 014B }
\o { 00F8 }
\oe { 0153 }
\end{verbatim}

1351
Accent ligatures handling is a little more complex. Accents may exist as pre-composed codepoints or as independent glyphs. The former are all saved as single token lists, whilst for the latter the combining accent needs to be re-ordered compared to the character it applies to.

\text_purify_equivalent:Nn \SS { SS } \__text_purify_accent:NN

First set up the combining accents.
\group_begin:
\cs_new:Npn \__text_purify_accent:NN #1#2
{\cs_if_exist:cTF { c__text_purify_ \token_to_str:N #1 _ \token_to_str:N #2 _tl } { \exp_not:v { c__text_purify_ \token_to_str:N #1 _ \token_to_str:N #2 _tl } } { \exp_not:n {#2} \exp_not:v { c__text_purify_ \token_to_str:N #1 _tl } } \tl_map_inline:nn { \' \^ \~ \= \u \. " \r \H \v \d \c \k \b \t } { \text_purify_equivalent:Nn #1 { \__text_purify_accent:NN { #2 } } }
\group_end:
Now we handle the pre-composed accents: the list here is taken from puenc.def. All of the precomposed cases take a single letter as their second argument. We do not try to cover the case where an accent is added to a “real” dotless-i or -j, or a æ/Æ. Rather, we assume that if the UTF-8 character is used, it will have the real accent character too.

\cs_set_protected:Npn \__text_loop:NNn #1#2#3
\quark_if_recursion_tail_stop:N #1
\tl_const:ce
\{ \c__text_purify_ \token_to_str:N #1 _ \token_to_str:N #2 _tl \}
\{ \codepoint_generate:nn {"#3} \{ \char_value_catcode:n { "#3 } \} \}
\__text_loop:NNn
\__text_loop:NNn
\textcircled{A} { 00C0 }
\textcircled{A} { 00C1 }
\textcircled{A} { 00C2 }
\textcircled{A} { 00C3 }
\textcircled{A} { 00C4 }
\textcircled{A} { 00C5 }
\textcircled{C} { 00C7 }
\textcircled{E} { 00C8 }
\textcircled{E} { 00C9 }
\textcircled{E} { 00CA }
\textcircled{E} { 00CB }
\textcircled{I} { 00CC }
\textcircled{I} { 00CD }
\textcircled{I} { 00CE }
\textcircled{I} { 00CF }
\textcircled{N} { 00D0 }
\textcircled{O} { 00D2 }
\textcircled{O} { 00D3 }
\textcircled{O} { 00D4 }
\textcircled{O} { 00D5 }
\textcircled{O} { 00D6 }
\textcircled{U} { 00D9 }
\textcircled{U} { 00DA }
\textcircled{U} { 00DB }
\textcircled{U} { 00DC }
\textcircled{Y} { 00DD }
\textcircled{a} { 00E0 }
\textcircled{a} { 00E1 }
\textcircled{a} { 00E2 }
\textcircled{a} { 00E3 }
\textcircled{a} { 00E4 }
\textcircled{a} { 00E5 }
\textcircled{c} { 00E7 }
\textcircled{e} { 00E8 }
\textcircled{e} { 00E9 }
\textcircled{e} { 00EA }
\textcircled{e} { 00EB }
\textcircled{i} { 00EC }
\textcircled{i} { 00ED }
\textcircled{i} { 00ED }
\group_end:

(End of definition for \__text_purify_accents.)

\package
Chapter 91

l3box implementation

91.1 Support code

\__box_dim_eval:w \mbox{Evaluating a dimension expression expandably. The only difference with \texttt{\dim_eval:n} is the lack of \texttt{\dim_use:N}, to produce an internal dimension rather than expand it into characters.}
\begin{verbatim}
\cs_new_eq:NN \__box_dim_eval:w \tex_dimexpr:D
\cs_new:Npn \__box_dim_eval:n #1 { \__box_dim_eval:w #1 \scan_stop: }
\end{verbatim}

(End of definition for \__box_dim_eval:w and \__box_dim_eval:n.)

\__kernel_kern:n \mbox{We need kerns in a few places. At present, we don’t have a module for this concept, so it goes in at first use: here. The idea is to avoid repeated use of the bare primitive.}
\begin{verbatim}
\cs_new_protected:Npn \__kernel_kern:n #1 { \tex_kern:D \__box_dim_eval:n {#1} }
\end{verbatim}

(End of definition for \__kernel_kern:n.)

91.2 Creating and initialising boxes

The following test files are used for this code: \texttt{m3box001.l3t}.
\begin{verbatim}
\box_new:N \mbox{Defining a new \texttt{(box)} register: remember that box 255 is not generally available.}
\box_new:c
\end{verbatim}
\begin{verbatim}
\cs_new_protected:Npm \__kernel_chk_if_free_cs:N \mbox{\texttt{\__kernel_chk_if_free_cs:N} \#1}
{ \tex_kern:D \__box_dim_eval:n \mbox{\texttt{(#1)} } }
\end{verbatim}

\begin{verbatim}
\cs_generate_variant:Nn \box_new:N \mbox{\texttt{\cs_generate_variant:Nn} \box_new:N { c }}
\end{verbatim}
Clear a ⟨box⟩ register.
\begin{verbatim}
cs_new_protected:N \box_clear:N #1
\box_clear:c
\cs_new_protected:N \box_gclear:N #1
\box_gclear:c
\cs_generate_variant:Nn \box_clear:N { c }
\cs_generate_variant:Nn \box_gclear:N { c }
\end{verbatim}

Clear or new.
\begin{verbatim}
cs_new_protected:N \box_clear_new:N #1
\cs_new_protected:N \box_gclear_new:N #1
\cs_generate_variant:Nn \box_clear_new:N { c }
\cs_generate_variant:Nn \box_gclear_new:N { c }
\end{verbatim}

Assigning the contents of a box to be another box.
\begin{verbatim}
cs_new_protected:N \box_set_eq:NN \box_set_eq:c \box_set_eq:Nc
\box_set_eq:cc
\cs_new_protected:N \box_gset_eq:NN \box_gset_eq:c \box_gset_eq:Nc
\box_gset_eq:cc
\end{verbatim}

Assigning the contents of a box to be another box, then drops the original box.
\begin{verbatim}
cs_new_protected:N \box_set_eq_drop:NN \box_set_eq_drop:c \box_set_eq_drop:Nc
\box_set_eq_drop:cc
\cs_new_protected:N \box_gset_eq_drop:NN \box_gset_eq_drop:c \box_gset_eq_drop:Nc
\box_gset_eq_drop:cc
\end{verbatim}

Copies of the cs functions defined in \texttt{l3basics}.
\begin{verbatim}
prg_new_eq_conditional:NNN \box_if_exist:N \cs_if_exist:N \box_if_exist:p
\box_if_exist:c
\box_if_exist:N TP
\end{verbatim}

1359

\section{Measuring and setting box dimensions}

Accessing the height, depth, and width of a ⟨box⟩ register.
\begin{verbatim}
cs_new_protected:N \box_ht:N \text{ht}:D
\box_hc
\cs_new_protected:N \box_dp:N \text{dp}:D
\box_dp:c
\cs_new_protected:N \box_wd:N \text{wd}:D
\box_wd:c
\cs_generate_variant:Nn \box_ht:N { c }
\cs_generate_variant:Nn \box_dp:N { c }
\cs_generate_variant:Nn \box_wd:N { c }
\end{verbatim}

The \texttt{\box_ht:N} and \texttt{\box_dp:N} primitives do not expand but rather are suitable for use after \texttt{\the} or inside dimension expressions. Here we obtain the same behaviour by using \texttt{\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_dim_eval:n} (basically \texttt{\dimexpr}) rather than \texttt{\dim_eval:n} (basically \texttt{\the\dimexpr}).
Setting the size whilst respecting local scope requires copying; the same issue does not come up when working globally. When debugging, the dimension expression #2 is surrounded by parentheses to catch early termination.

Using boxes

Using a \texttt{⟨box⟩}. These are just \TeX primitives with meaningful names.

Move box material in different directions. When debugging, the dimension expression #1 is surrounded by parentheses to catch early termination.
91.5 Box conditionals

The primitives for testing if a ⟨box⟩ is empty/void or which type of box it is.

\if_hbox:N \if_vbox:N \if_box_empty:N

\if_box_empty_p:N \if_box_empty:c

\box_if_horizontal_p:N \box_if_horizontal_p:c \box_if_horizontal:N \box_if_horizontal:c

\box_if_vertical_p:N \box_if_vertical_p:c \box_if_vertical:N \box_if_vertical:c

\box_if_empty_p:N \box_if_empty_p:c \box_if_empty:N \box_if_empty:c

(End of definition for \box_new:N and others. These functions are documented on page 302.)

91.6 The last box inserted

\box_set_to_last:N \box_set_to_last:c \box_set_to_last:N \box_set_to_last:c

(End of definition for \box_set_to_last:N and \box_set_to_last:N. These functions are documented on page 305.)

91.7 Constant boxes

\c_empty_box

(End of definition for \c_empty_box. This variable is documented on page 305.)
91.8 Scratch boxes

Scratch boxes.

(End of definition for \l_tmpa_box and others. These variables are documented on page 305.)

91.9 Viewing box contents

\TeX’s \showbox is not really that helpful in many cases, and it is also inconsistent with other \TeX3 show functions as it does not actually shows material in the terminal. So we provide a richer set of functionality.

\box_show:N
\box_show:c
\box_show:Nnn
\box_show:cnn

Essentially a wrapper around the internal function, but evaluating the breadth and depth arguments now outside the group.

(End of definition for \box_show:N and \box_show:Nnn. These functions are documented on page 306.)

\box_log:N
\box_log:c
\box_log:Nnn
\box_log:cnn
\__box_log:nNnn

Getting \TeX to write to the log without interruption the run is done by altering the interaction mode. For that, the \#\-\TeX extensions are needed.

\__box_show:NNnn
\__box_show:NNff

The internal auxiliary to actually do the output uses a group to deal with breadth and depth values. The use:n here gives better output appearance. Setting \tracingonline and \errorcontextlines is used to control what appears in the terminal.
91.10 Horizontal mode boxes

\hbox:n

(\textit{The test suite for this command, and others in this file, is \texttt{m3box002.ltx}.})

Put a horizontal box directly into the input stream.

\hbox_set:Nn
\hbox_set:cn
\hbox_gset:Nn
\hbox_gset:cn

Storing material in a horizontal box with a specified width. Again, put the dimension expression in parentheses when debugging.

\hbox_set_to_wd:Nnn
\hbox_set_to_wd:cnn
\hbox_gset_to_wd:Nnn
\hbox_gset_to_wd:cnn

(End of definition for \_\_\_box_show:NNNn.)
Storing material in a horizontal box. This type is useful in environment definitions.

\hspace{\hbox_set:Nw \hbox_set:cw \hbox_gset:Nw \hbox_gset:cw \hbox_set_end: \hbox_gset_end:}

Combining the above ideas.

\hspace{\hbox_set_to_wd:Nnw \hbox_set_to_wd:cnw \hbox_gset_to_wd:Nnw \hbox_gset_to_wd:cnw}

Put a horizontal box directly into the input stream.

\hspace{\hbox_to_wd:nn \hbox_to_zero:n}

(End of definition for \hbox_set_to_wd:NNn and \hbox_gset_to_wd:NNn. These functions are documented on page 307.)
34848 \{ \texttt{color\_group\_begin: \#1 \texttt{color\_group\_end: }} \}
34849 }

(End of definition for \hbox\_to\_wd:n and \hbox\_to\_zero:n. These functions are documented on page 306.)

\hbox\_overlap\_center:n Put a zero-sized box with the contents pushed against one side (which makes it stick out on the other) directly into the input stream.

34850 \cs\_new\_protected:Npn \hbox\_overlap\_center:n \#1
34851 \{ \hbox\_to\_zero:n \{ \texttt{tex\_hss:D \#1 \texttt{tex\_hss:D}} \} \}
34852 \cs\_new\_protected:Npn \hbox\_overlap\_left:n \#1
34853 \{ \hbox\_to\_zero:n \{ \texttt{tex\_hss:D \#1} \} \}
34854 \cs\_new\_protected:Npn \hbox\_overlap\_right:n \#1
34855 \{ \hbox\_to\_zero:n \{ \#1 \texttt{tex\_hss:D} \} \}

(End of definition for \hbox\_overlap\_center:n, \hbox\_overlap\_left:n, and \hbox\_overlap\_right:n. These functions are documented on page 307.)

\hbox\_unpack:N Unpacking a box and if requested also clear it.

34856 \cs\_new\_eq:NN \hbox\_unpack:N \tex\_unhcopy:D
34857 \cs\_new\_eq:NN \hbox\_unpack\_drop:N \tex\_unhbox:D
34858 \cs\_generate\_variant:Nn \hbox\_unpack\_drop:N \{ c \}
34859 \cs\_generate\_variant:Nn \hbox\_unpack\_drop:N \{ c \}

(End of definition for \hbox\_unpack:N and \hbox\_unpack\_drop:N. These functions are documented on page 307.)

91.11 Vertical mode boxes

\TeX \ ends \ these \ boxes \ directly \ with \ the \ internal \ \begin{tt}end\_graf\end{tt} routine. \ This \ means \ that \ there \ is \ no \ \begin{tt}par\end{tt} \ at \ the \ end \ of \ vertical \ boxes \ unless \ we \ insert \ one. \ Thus \ all \ vertical \ boxes \ include \ a \ \begin{tt}par\end{tt} \ just \ before \ closing \ the \ color \ group.

\vbox:n The following test files are used for this code: \texttt{m3box003.lvt}.

\vbox_top:n Put a vertical box directly into the input stream.

34860 \cs\_new\_protected:Npn \vbox:n \#1
34861 \{ \texttt{tex\_vbox:D} \{ \texttt{color\_group\_begin: \#1 \par \texttt{color\_group\_end: }} \} \}
34862 \cs\_new\_protected:Npn \vbox\_top:n \#1
34863 \{ \texttt{tex\_vtop:D} \{ \texttt{color\_group\_begin: \#1 \par \texttt{color\_group\_end: }} \} \}

(End of definition for \vbox:n and \vbox\_top:n. These functions are documented on page 308.)

\vbox\_to\_ht:nn Put a vertical box directly into the input stream.

34864 \cs\_new\_protected:Npn \vbox\_to\_ht:nn \#1\#2
34865 \{ \texttt{tex\_vbox:D} \texttt{to \_\_box\_dim\_eval:n} \#1 \}
34866 \{ \texttt{color\_group\_begin: \#2 \par \texttt{color\_group\_end: }} \}
34867 \}
34868 \}
34869 \cs\_new\_protected:Npn \vbox\_to\_zero:n \#1
34870 \{ \texttt{tex\_vbox:D} \texttt{to \_c\_zero\_dim} \}
34871 \{ \texttt{color\_group\_begin: \#1 \par \texttt{color\_group\_end: }} \}
34872 \}
34873 \}

1365
Storing material in a vertical box with a natural height.

\begin{verbatim}
\cs_new_protected:Npn \vbox_set:Nn \#1#2
\tex_setbox:D #1 \tex_vbox:D
\color_group_begin: #2 \par \color_group_end: }
\end{verbatim}

\begin{verbatim}
\cs_new_protected:Npn \vbox_gset:Nn \#1#2
\tex_global:D \tex_setbox:D #1 \tex_vbox:D
\color_group_begin: #2 \par \color_group_end: }
\end{verbatim}

\begin{verbatim}
\cs_generate_variant:Nn \vbox_set:Nn { c }
\cs_generate_variant:Nn \vbox_gset:Nn { c }
\end{verbatim}

(End of definition for \vbox_set:Nn and \vbox_gset:Nn. These functions are documented on page 308.)

Storing material in a vertical box with a natural height and reference point at the baseline of the first object in the box.

\begin{verbatim}
\cs_new_protected:Npn \vbox_set_top:Nn \#1#2
\tex_setbox:D #1 \tex_vtop:D
\color_group_begin: #2 \par \color_group_end: }
\end{verbatim}

\begin{verbatim}
\cs_new_protected:Npn \vbox_gset_top:Nn \#1#2
\tex_global:D \tex_setbox:D #1 \tex_vtop:D
\color_group_begin: #2 \par \color_group_end: }
\end{verbatim}

\begin{verbatim}
\cs_generate_variant:Nn \vbox_set_top:Nn { c }
\cs_generate_variant:Nn \vbox_gset_top:Nn { c }
\end{verbatim}

(End of definition for \vbox_set_top:Nn and \vbox_gset_top:Nn. These functions are documented on page 308.)

Storing material in a vertical box with a specified height.

\begin{verbatim}
\cs_new_protected:Npn \vbox_set_to_ht:Nnn \#1#2#3
\tex_setbox:D #1 \tex_vbox:D to \__box_dim_eval:n {#2}
\color_group_begin: #3 \par \color_group_end: }
\end{verbatim}

\begin{verbatim}
\cs_new_protected:Npn \vbox_gset_to_ht:Nnn \#1#2#3
\tex_global:D \tex_setbox:D #1 \tex_vbox:D to \__box_dim_eval:n {#2}
\color_group_begin: #3 \par \color_group_end: }
\end{verbatim}

\begin{verbatim}
\cs_generate_variant:Nn \vbox_set_to_ht:Nnn { c }
\cs_generate_variant:Nn \vbox_gset_to_ht:Nnn { c }
\end{verbatim}

(End of definition for \vbox_set_to_ht:Nnn and \vbox_gset_to_ht:Nnn. These functions are documented on page 308.)
Storing material in a vertical box. This type is useful in environment definitions.

```
\cs_new_protected:Npn \vbox_set:Nw #1
  \tex_setbox:D #1 \tex_vbox:D

\cs_new_protected:Npn \vbox_gset:Nw #1
  \tex_global:D \tex_setbox:D #1 \tex_vbox:D

\cs_generate_variant:Nn \vbox_set:Nw { c }
\cs_generate_variant:Nn \vbox_gset:Nw { c }
```

(End of definition for \vbox_set:Nw and others. These functions are documented on page 308.)

```
\cs_new_protected:Npm \vbox_set_to_ht:Nnw
\cs_new_protected:Npm \vbox_set_to_ht:cnw
\cs_new_protected:Npm \vbox_gset_to_ht:Nnw
\cs_new_protected:Npm \vbox_gset_to_ht:cnw
```

A combination of the above ideas.

```
\cs_new_protected:Npm \vbox_set_to_ht:Nnw \#1 \#2
  \tex_setbox:D \#1 \tex_vbox:D to \_\_box_dim_eval:n \#2

\cs_generate_variant:Nn \vbox_set_to_ht:Nnw { c }
```

(End of definition for \vbox_set_to_ht:Nnw and \vbox_gset_to_ht:Nnw. These functions are documented on page 309.)

```
\cs_new_eq:NN \vbox_unpack:N \tex_unvcopy:D
\cs_new_eq:NN \vbox_unpack_drop:N \tex_unvbox:D
```

Unpacking a box and if requested also clear it.

```
\cs_generate_variant:NN \vbox_unpack:N \vbox_unpack_drop:N
\cs_generate_variant:NN \vbox_unpack_drop:N \vbox_unpack_drop:cnw
```

(End of definition for \vbox_unpack:N and \vbox_unpack_drop:N. These functions are documented on page 309.)
91.12 Affine transformations

\l__box_angle_fp When rotating boxes, the angle itself may be needed by the engine-dependent code. This is done using the \textit{fp} module so that the value is tidied up properly.

\fp_new:N \l__box_angle_fp

(End of definition for \l__box_angle_fp.)

\l__box_cos_fp \l__box_sin_fp These are used to hold the calculated sine and cosine values while carrying out a rotation.

\fp_new:N \l__box_cos_fp
\fp_new:N \l__box_sin_fp

(End of definition for \l__box_cos_fp and \l__box_sin_fp.)

\l__box_top_dim \l__box_bottom_dim \l__box_left_dim \l__box_right_dim These are the positions of the four edges of a box before manipulation.

\dim_new:N \l__box_top_dim
\dim_new:N \l__box_bottom_dim
\dim_new:N \l__box_left_dim
\dim_new:N \l__box_right_dim

(End of definition for \l__box_top_dim and others.)

\l__box_top_new_dim \l__box_bottom_new_dim \l__box_left_new_dim \l__box_right_new_dim These are the positions of the four edges of a box after manipulation.

\dim_new:N \l__box_top_new_dim
\dim_new:N \l__box_bottom_new_dim
\dim_new:N \l__box_left_new_dim
\dim_new:N \l__box_right_new_dim

(End of definition for \l__box_top_new_dim and others.)

\l__box_internal_box Scratch space, but also needed by some parts of the driver.

\box_new:N \l__box_internal_box

(End of definition for \l__box_internal_box.)
Figure 1: Coordinates of a box prior to rotation.

Rotation of a box starts with working out the relevant sine and cosine. The actual rotation is in an auxiliary to keep the flow slightly clearer.

\[ \boxed{\text{box_rotate:Nn}} \quad \boxed{\text{box_rotate:cn}} \quad \boxed{\text{box_grotate:Nn}} \quad \boxed{\text{box_grotate:cn}} \]
\[ \boxed{\text{\_box_rotate:NnN}} \quad \boxed{\text{\_box_rotate:N}} \]
\[ \boxed{\text{\_box_rotate_xdir:nnN}} \quad \boxed{\text{\_box_rotate_ydir:nnN}} \]
\[ \boxed{\text{\_box_rotate_quadrant_one:}} \quad \boxed{\text{\_box_rotate_quadrant_two:}} \quad \boxed{\text{\_box_rotate_quadrant_three:}} \quad \boxed{\text{\_box_rotate_quadrant_four:}} \]

Rotation of the four edges then takes place: this is most efficiently done on a quadrant by quadrant basis.

\[ \boxed{\text{\cs_new_protected:Npn \text{\_box_rotate:N}} \#1 \#2} \]
\[ \boxed{\text{\cs_generate_variant:Nn \text{\_box_rotate:N} { c }}} \]

The next step is to work out the \( x \) and \( y \) coordinates of vertices of the rotated box in relation to its original coordinates. The box can be visualized with vertices \( B, C, D \) and \( E \) is illustrated (Figure 1). The vertex \( O \) is the reference point on the baseline, and in this implementation is also the centre of rotation. The formulae are, for a point \( P \) and angle \( \alpha \):

\[
\begin{align*}
P'_x &= P_x - O_x \\
P'_y &= P_y - O_y \\
P''_x &= (P'_x \cos(\alpha)) - (P'_y \sin(\alpha)) \\
P''_y &= (P'_x \sin(\alpha)) + (P'_y \cos(\alpha)) \\
P'''_x &= P''_x + O_x + L_x \\
P'''_y &= P''_y + O_y
\end{align*}
\]
The “extra” horizontal translation $L_2$ at the end is calculated so that the leftmost point of the resulting box has x-coordinate 0. This is desirable as \TeX boxes must have the reference point at the left edge of the box. (As $O$ is always $(0,0)$, this part of the calculation is omitted here.)

\begin{verbatim}
\fp_compare:nNnTF \l__box_sin_fp > \c_zero_fp
{ \fp_compare:nNnTF \l__box_cos_fp > \c_zero_fp
  { \__box_rotate_quadrant_one: }
  { \__box_rotate_quadrant_two: }
}
{ \fp_compare:nNnTF \l__box_cos_fp < \c_zero_fp
  { \__box_rotate_quadrant_three: }
  { \__box_rotate_quadrant_four: }
}

The position of the box edges are now known, but the box at this stage be misplaced relative to the current \TeX reference point. So the content of the box is moved such that the reference point of the rotated box is in the same place as the original.

\cs_new_protected:Npn \__box_rotate_xdir:nnN #1#2#3
{ \dim_set:Nn #3
  { \fp_to_dim:n
    { \l__box_cos_fp \dim_to_fp:n {#1}
      - \l__box_sin_fp \dim_to_fp:n {#2} } } }

Tidy up the size of the box so that the material is actually inside the bounding box. The result can then be used to reset the original box.

\cs_new_protected:Npn \__box_rotate_xdir:nnN #1#2#3
{ \dim_set:Nn #3
  { \fp_to_dim:n
    { \l__box_cos_fp \dim_to_fp:n {#1}
      - \l__box_sin_fp \dim_to_fp:n {#2} } } }

These functions take a general point $(#1,#2)$ and rotate its location about the origin, using the previously-set sine and cosine values. Each function gives only one component of the location of the updated point. This is because for rotation of a box each step needs only one value, and so performance is gained by avoiding working out both $x'$ and $y'$ at the same time. Contrast this with the equivalent function in the \texttt{l3coffins} module, where both parts are needed.
Rotation of the edges is done using a different formula for each quadrant. In every case, the top and bottom edges only need the resulting $y$-values, whereas the left and right edges need the $x$-values. Each case is a question of picking out which corner ends up at with the maximum top, bottom, left and right value. Doing this by hand means a lot less calculating and avoids lots of comparisons.

```latex
\cs_new_protected:Npm \_\_\_box_rotate_ydir:nnN #1#2#3
\{\dim_set:Nn #3 \dim_to_dim:\{
  \l__box_sin_fp * \dim_to_fp:n {#1}
  + \l__box_cos_fp * \dim_to_fp:n {#2}\}
\}
```

```latex
\cs_new_protected:Npn \_\_\_box_rotate_quadrant_one:
\{\_\_\_box_rotate_ydir:nnN \l__box_right_dim \l__box_top_dim \l__box_right_new_dim
\_\_\_box_rotate_ydir:nnN \l__box_left_dim \l__box_bottom_dim \l__box_left_new_dim
\_\_\_box_rotate_xdir:nnN \l__box_left_dim \l__box_top_dim \l__box_left_new_dim
\_\_\_box_rotate_xdir:nnN \l__box_right_dim \l__box_bottom_dim \l__box_right_new_dim
\}
```

```latex
\cs_new_protected:Npn \_\_\_box_rotate_quadrant_two:
\{\_\_\_box_rotate_ydir:nnN \l__box_right_dim \l__box_bottom_dim \l__box_right_new_dim
\_\_\_box_rotate_ydir:nnN \l__box_left_dim \l__box_top_dim \l__box_left_new_dim
\_\_\_box_rotate_xdir:nnN \l__box_right_dim \l__box_top_dim \l__box_left_new_dim
\_\_\_box_rotate_xdir:nnN \l__box_left_dim \l__box_bottom_dim \l__box_right_new_dim
\}
```

```latex
\cs_new_protected:Npn \_\_\_box_rotate_quadrant_three:
\{\_\_\_box_rotate_ydir:nnN \l__box_left_dim \l__box_bottom_dim \l__box_right_new_dim
\_\_\_box_rotate_ydir:nnN \l__box_right_dim \l__box_top_dim \l__box_left_new_dim
\_\_\_box_rotate_xdir:nnN \l__box_right_dim \l__box_bottom_dim \l__box_right_new_dim
\_\_\_box_rotate_xdir:nnN \l__box_left_dim \l__box_top_dim \l__box_left_new_dim
\}
```

```latex
\cs_new_protected:Npn \_\_\_box_rotate_quadrant_four:
```

1371
{\_box_rotate_ydir:nnN \l__box_left_dim \l__box_top_dim
\l__box_top_new_dim
\l__box_rotate_ydir:nnN \l__box_right_dim \l__box_bottom_dim
\l__box_bottom_new_dim
\l__box_rotate_xdir:nnN \l__box_left_dim \l__box_bottom_dim
\l__box_left_new_dim
\l__box_rotate_xdir:nnN \l__box_right_dim \l__box_top_dim
\l__box_right_new_dim
}

(End of definition for \box_rotate:Nn and others. These functions are documented on page 313.)

\l__box_scale_x_fp
\l__box_scale_y_fp

Scaling is potentially different in the two axes.
\fp_new:N \l__box_scale_x_fp
\fp_new:N \l__box_scale_y_fp

(End of definition for \l__box_scale_x_fp and \l__box_scale_y_fp.)

Resizing a box starts by working out the various dimensions of the existing box.
\cs_new_protected:Npn \box_resize_to_wd_and_ht_plus_dp:Nnn #1#2#3
\{\__box_resize_to_wd_and_ht_plus_dp:NnnN #1 {#2} {#3}
\hbox_set:Nn\}
\cs_generate_variant:Nn \box_resize_to_wd_and_ht_plus_dp:Nnn { c }
\cs_new_protected:Npn \__box_resize_to_wd_and_ht_plus_dp:NnnN #1#2#3#4
\{\__box_resize_set_corners:N #1
\}
\cs_new_protected:Npn \__box_resize_set_corners:N \l__box_resize_to_wd_and_ht_plus_dp:NnnN \l__box_resize:NNN
\l__box_resize:N
\l__box_resize:NNN

The x-scaling and resulting box size is easy enough to work out: the dimension is that given as \#2, and the scale is simply the new width divided by the old one.
\fp_set:Nn \l__box_scale_x_fp
\{\dim_to_fp:n \l__box_right_dim \} / \dim_to_fp:n { \l__box_right_dim }

The y-scaling needs both the height and the depth of the current box.
\fp_set:Nn \l__box_scale_y_fp
\{\dim_to_fp:n \l__box_top_dim - \l__box_bottom_dim \}

Hand off to the auxiliary which does the rest of the work.
\cs_new_protected:Npn \__box_resize:N \l__box_resize:NNN
\{\l__box_resize_to_wd_and_ht_plus_dp:NnnN \l__box_resize_set_corners:N \l__box_resize:N #1
\}
With at least one real scaling to do, the next phase is to find the new edge coordinates. In the \( x \) direction this is relatively easy: just scale the right edge. In the \( y \) direction, both dimensions have to be scaled, and this again needs the absolute scale value. Once that is all done, the common resize/rescale code can be employed.

\[
\begin{align*}
\text{\box_resize_to_wd_and_ht_plus_dp:Nn} & \quad \text{Scaling to a (total) height or to a width is a simplified version of the main resizing operation, with the scale simply copied between the two parts. The internal auxiliary is called using the scaling value twice, as the sign for both parts is needed (as this allows the same internal code to be used as for the general case).}
\end{align*}
\]
\begin{verbatim}
\cs_generate_variant:Nn \box_resize_to_ht_plus_dp:Nn { c }
\cs_new_protected:Npn \box_gresize_to_ht_plus_dp:Nn #1#2
{ \__box_resize_to_ht_plus_dp:NnN #1 {#2} \hbox_gset:Nn }
\cs_generate_variant:Nn \box_gresize_to_ht_plus_dp:Nn { c }
\cs_new_protected:Npn \__box_resize_to_ht_plus_dp:NnN #1#2#3
{ #3 #1
\__box_resize_set_corners:N #1
\fp_set:Nn \l__box_scale_y_fp
\dim_to_fp:n {#2} \hbox_set:Nn \l__box_top_dim - \l__box_bottom_dim }
\fp_set_eq:NN \l__box_scale_x_fp \l__box_scale_y_fp
\__box_resize:N #1 }
\cs_new_protected:Npn \box_resize_to_wd:Nn #1#2
{ \__box_resize_to_wd:NnN #1 {#2} \hbox_set:Nn }
\cs_generate_variant:Nn \box_resize_to_wd:Nn { c }
\cs_new_protected:Npn \box_gresize_to_wd:Nn #1#2
{ \__box_resize_to_wd:NnN #1 {#2} \hbox_gset:Nn }
\cs_generate_variant:Nn \box_gresize_to_wd:Nn { c }
\cs_new_protected:Npn \__box_resize_to_wd:NnN #1#2#3
{ #3 #1
\__box_resize_set_corners:N #1
\fp_set:Nn \l__box_scale_x_fp
\dim_to_fp:n {#2} \hbox_set:Nn \l__box_right_dim}
\fp_set_eq:NN \l__box_scale_x_fp \l__box_scale_y_fp
\__box_resize:N #1 }
\cs_new_protected:Npn \box_resize_to_wd_and_ht:Nnn #1#2#3
{ \__box_resize_to_wd_and_ht:NnnN #1 {#2} {#3} \hbox_set:Nn }
\cs_generate_variant:Nn \box_resize_to_wd_and_ht:Nnn { c }
\cs_new_protected:Npn \box_gresize_to_wd_and_ht:Nnn #1#2#3
{ \__box_resize_to_wd_and_ht:NnnN #1 {#2} {#3} \hbox_gset:Nn }
\cs_generate_variant:Nn \box_gresize_to_wd_and_ht:Nnn { c }
\cs_new_protected:Npn \__box_resize_to_wd_and_ht:NnnN #1#2#3#4
{ #4 #1
\__box_resize_set_corners:N #1
\fp_set:Nn \l__box_scale_x_fp
\dim_to_fp:n {#2} \hbox_set:Nn \l__box_right_dim}
\fp_set_eq:NN \l__box_scale_x_fp \l__box_scale_y_fp
\dim_to_fp:n \l__box_top_dim
\end{verbatim}

1374
When scaling a box, setting the scaling itself is easy enough. The new dimensions are also relatively easy to find, allowing only for the need to keep them positive in all cases. Once that is done then after a check for the trivial scaling a hand-off can be made to the common code. The code here is split into two as this allows sharing with the auto-resizing functions.

\[
\text{\textbackslash box\_scale:Nnn} \quad \text{\textbackslash box\_scale:cnm} \\
\text{\textbackslash box\_gscale:cnm} \quad \text{\_box\_scale:NNnN} \\
\text{\_box\_scale:N} \\
\]

Although autosizing a box uses dimensions, it has more in common in implementation with scaling. As such, most of the real work here is done elsewhere.
\hbox_set:Nn
\cs_generate_variant:Nn \box_autosize_to_wd_and_ht_plus_dp:Nnn { c }
\cs_new_protected:Nm \box_gautosize_to_wd_and_ht_plus_dp:Nnn #1#2#3
\{
  \_\__box_autosize:NnnN #1 {#2} {#3} { \\box_ht:N #1 + \\box_dp:N #1 }
\\box_gset:Nn
\}
\cs_generate_variant:Nn \box_gautosize_to_wd_and_ht_plus_dp:Nnn { c }
\cs_new_protected:Npn \__box_autosize:NnnnN #1#2#3#4#5
\{
  #5 #1
  \fp_set:Nn \l__box_scale_x_fp { ( \dim_to_fp:n {#2} ) / \box_wd:N #1 }
  \fp_set:Nn \l__box_scale_y_fp
  \fp_compare:nNnTF \l__box_scale_x_fp > \c_zero_fp
  { \fp_set_eq:NN \l__box_scale_x_fp \l__box_scale_y_fp }
  { \fp_set_eq:NN \l__box_scale_y_fp \l__box_scale_x_fp }
  \__box_scale:N #1
\}
\(\text{End of definition for } \box_autosize_to_wd_and_{ht}:Nnn \text{ and others. These functions are documented on page 311.}\)
\_\__box_resize_common:N
The main resize function places its input into a box which start off with zero width, and includes the handles for engine rescaling.
\cs_new_protected:Nm \_\__box_resize_common:N #1
\{
  \hbox_set:Nn \l__box_internal_box
  \{
    \_\__box_backend_scale:Nnn #1
    \l__box_scale_x_fp
    \l__box_scale_y_fp
  \}
\}
The new height and depth can be applied directly.
\fp_compare:nNnTF \l__box_scale_y_fp > \c_zero_fp
\{
  \box_set_ht:Nn \l__box_internal_box \l__box_top_new_dim
  \box_set_dp:Nn \l__box_internal_box \l__box_bottom_new_dim
\}
\fp_compare:nNnTF \l__box_scale_y_fp < \c_zero_fp
\{
  \box_set_dp:Nn \l__box_internal_box \l__box_top_new_dim
  \box_set_ht:Nn \l__box_internal_box \l__box_bottom_new_dim
\}
Things are not quite as obvious for the width, as the reference point needs to remain unchanged. For positive scaling factors resizing the box is all that is needed. However, for case of a negative scaling the material must be shifted such that the reference point ends up in the right place.
\fp_compare:nNnTF \l__box_scale_x_fp < \c_zero_fp
\{
91.13 Viewing part of a box

A wrapper around the driver-dependent code.

Trimming from the left- and right-hand edges of the box is easy: kern the appropriate parts off each side.

For the height and depth, there is a need to watch the baseline is respected. Material always has to stay on the correct side, so trimming has to check that there is enough material to trim. First, the bottom edge. If there is enough depth, simply set the depth, or if not move down so the result is zero depth. \texttt{\box_move_down:nn} is used in both...
cases so the resulting box always contains a \lower primitive. The internal box is used here as it allows safe use of \box_set_dp:Nn.

\dim_compare:nNnTF { \box_dp:N \#1 } > {#3}
{ \hbox_set:N \l__box_internal_box
\box_move_down:nn \c_zero_dim
 \{ \box_use_drop:N \l__box_internal_box \}
}\box_set_dp:Nn \l__box_internal_box \{ \box_dp:N \#1 - (#3) \}
\box_set_dp:Nn \l__box_internal_box \{ \box_move_down:nn \c_zero_dim
\{ \box_use_drop:N \l__box_internal_box \}
}\box_set_dp:Nn \l__box_internal_box \c_zero_dim

Same thing, this time from the top of the box.

\dim_compare:nNnTF { \box_ht:N \l__box_internal_box } > {#5}
{ \hbox_set:N \l__box_internal_box
\box_move_up:nn \c_zero_dim
 \{ \box_use_drop:N \l__box_internal_box \}
}\box_set_ht:Nn \l__box_internal_box \{ \box_ht:N \l__box_internal_box - (#5) \}
\box_set_ht:Nn \l__box_internal_box \c_zero_dim
\box_set_ht:Nn \l__box_internal_box \c_zero_dim
\box_set_ht:Nn \l__box_internal_box \c_zero_dim
\#6 \#1 \l__box_internal_box

(End of definition for \box_set_trim:Nnnn, \box_gset_trim:Nnnn, and \l__box_set_trim:NnnnN. These functions are documented on page 313.)

The same general logic as for the trim operation, but with absolute dimensions. As a result, there are some things to watch out for in the vertical direction.

\cs_new_protected:Npn \box_set_viewport:Nnnn \l__box_viewport:Nnnn
\box_set_viewport:cnnn \__box_viewport:NnnnN
\box_gset_viewport:Nnnn \l__box_viewport:Nnnn
\box_gset_viewport:cnnn \__box_viewport:NnnnN
\cs_generate_variant:Nn \box_set_viewport:Nnnn \{ c \}
\cs_generate_variant:Nn \box_gset_viewport:Nnnn \{ c \}
\cs_new_protected:Npn \__box_viewport:NnnnN
\box_set_viewport:Nnnn \l__box_viewport:NnnnN \#1\#2\#3\#4\#5
\box_gset_viewport:Nnnn \l__box_viewport:NnnnN \#1\#2\#3\#4\#5
\box_set_viewport:Nnnn \l__box_viewport:NnnnN \#1\#2\#3\#4\#5
\box_gset_viewport:Nnnn \l__box_viewport:NnnnN \#1\#2\#3\#4\#5

1378
(End of definition for \_\_box_viewport:NnnnnN, \_\_box_gset_viewport:NnnnnN, and \_\_box_set_viewport:Nnnnn. These functions are documented on page 313.)
Chapter 92

\texttt{l3coffins} implementation

\section{Coffins: data structures and general variables}

Scratch variables.

\begin{verbatim}
\l__coffin_internal_box
\l__coffin_internal_dim
\l__coffin_internal_tl
\end{verbatim}

The “corners”; of a coffin define the real content, as opposed to the \TeX{} bounding box. They all start off in the same place, of course.

\begin{verbatim}
\c__coffin_corners_prop
\end{verbatim}

Pole positions are given for horizontal, vertical and reference-point based values.

\begin{verbatim}
\c__coffin_poles_prop
\end{verbatim}
92.2 Basic coffin functions

There are a number of basic functions needed for creating coffins and placing material in them. This all relies on the following data structures.

\texttt{\_coffin_to_value:N}  Coffins are a two-part structure and we rely on the internal nature of box allocation to make everything work. As such, we need an interface to turn coffin identifiers into numbers. For the purposes here, the signature allowed is \texttt{N} despite the nature of the underlying primitive.
Several of the higher-level coffin functions would give multiple errors if the coffin does not exist. A cleaner way to handle this is provided here: both the box and the coffin structure are checked.

\prg_new_conditional:Nn \coffin_if_exist:N #1 { p , T , F , TF } \begin{verbatim}
{ \cs_if_exist:NTF #1
  { \cs_if_exist:cTF { coffin ~ \__coffin_to_value:N #1 ~ poles }
    { \prg_return_true: }
    { \prg_return_false: }
  }
  { \prg_return_false: }
}
\prg_generate_conditional_variant:Nnn \coffin_if_exist:N { c } { p , T , F , TF }
\end{verbatim}

(End of definition for \coffin_if_exist:NTF. This function is documented on page 315.)

\__coffin_if_exist:NT Several of the higher-level coffin functions would give multiple errors if the coffin does not exist. So a wrapper is provided to deal with this correctly, issuing an error on erroneous use.

\cs_new_protected:Npn \__coffin_if_exist:NT #1#2 \begin{verbatim}
{ \cs_if_exist:NT #1
  { \__coffin_to_value:N #1 ~ poles }
  { \msg_error:nne { coffin } { unknown }
    { \token_to_str:N #1 }
  }
}
\end{verbatim}

(End of definition for \__coffin_if_exist:NT.)

\coffin_clear:N Clearing coffins means emptying the box and resetting all of the structures.

\cs_new_protected:Npn \__coffin_if_exist:NT #1#2 \begin{verbatim}
{ \cs_if_exist:NT #1
  { \__coffin_reset_structure:N #1 }
  { \msg_error:nne { coffin } { unknown }
    { \token_to_str:N #1 }
  }
}
\cs_generate_variant:Nn \coffin_clear:N { c }
\end{verbatim}

(End of definition for \__coffin_if_exist:NT.)

\coffin_gclear:N Clearing coffins means emptying the box and resetting all of the structures.

\cs_new_protected:Npn \__coffin_if_exist:NT #1#2 \begin{verbatim}
{ \cs_if_exist:NT #1
  { \__coffin_greset_structure:N #1 }
  { \msg_error:nne { coffin } { unknown }
    { \token_to_str:N #1 }
  }
}
\cs_generate_variant:Nn \coffin_gclear:N { c }
\end{verbatim}

(End of definition for \__coffin_if_exist:NT.)

\coffin_clear:N and \coffin_gclear:N. These functions are documented on page 315.)
Creating a new coffin means making the underlying box and adding the data structures. The `\debug_suspend:` and `\debug_resume:` functions prevent `\prop_gclear_new:c` from writing useless information to the log file.

```
\cs_new_protected:Npn \coffin_new:N #1
  { \box_new:N #1 \debug_suspend:
    \prop_gclear_new:c { coffin \_\coffin_to_value:N #1 \_corners }
    \prop_gclear_new:c { coffin \_\coffin_to_value:N #1 \_poles }
    \c__coffin_corners_prop
    \prop_gset_eq:cN { coffin \_\coffin_to_value:N #1 \_corners }
    \c__coffin_poles_prop
    \prop_gset_eq:cN { coffin \_\coffin_to_value:N #1 \_poles }
    \debug_resume: }
\cs_generate_variant:Nn \coffin_new:N { c }
```

(End of definition for \coffin_new:N. This function is documented on page 315.)

Horizontal coffins are relatively easy: set the appropriate box, reset the structures then update the handle positions.

```
\hcoffin_set:Nn \hcoffin_set:cn
\hcoffin_gset:Nn \hcoffin_gset:cn
```

```
\__coffin_set_vertical:NnnNNN
\__coffin_set_vertical_aux:
```

Setting vertical coffins is more complex. First, the material is typeset with a given width. The default handles and poles are set as for a horizontal coffin, before finding the top baseline using a temporary box. No `\color_ensure_current:` here as that would add a
whatsit to the start of the vertical box and mess up the location of the T pole (see \textit{\LaTeX} by Topic for discussion of the \texttt{\vtop} primitive, used to do the measuring).

\begin{verbatim}
\cs_new_protected:Npn \vcoffin_set:Nnn #1#2#3
\__coffin_set_vertical:NnnNNN #1 {#2} {#3}
\vbox_set:Nn \coffin_reset_poles:N \__coffin_set_pole:Nnn
\cs_generate_variant:Nn \vcoffin_set:Nnn { c }
\cs_new_protected:Npn \vcoffin_gset:Nnn #1#2#3
\__coffin_set_vertical:NnnNNN #1 {#2} {#3}
\vbox_gset:Nn \coffin_greset_poles:N \__coffin_gset_pole:Nnn
\cs_generate_variant:Nn \vcoffin_gset:Nnn { c }
\cs_new_protected:Npe \__coffin_set_vertical_aux:
\bool_lazy_and:nnT \cs_if_exist_p:N \fmtname
\str_if_eq_p:Vn \fmtname { LaTeX2e }
\dim_set_eq:NN \exp_not:N \linewidth \tex_hsize:D
\dim_set_eq:NN \exp_not:N \columnwidth \tex_hsize:D
\end{verbatim}

(End of definition for \vcoffin_set:Nnn and others. These functions are documented on page 316.)

These are the “begin”/“end” versions of the above: watch the grouping!

\begin{verbatim}
\hcoffin_set:Nw \hcoffin_set:cv
\hcoffin_gset:Nw \hcoffin_gset:cv
\hcoffin_set_end:
\hcoffin_gset_end:
\end{verbatim}

1384
\hbox_set:Nw #1 \color_ensure_current: \cs_set_protected:Npn \hcoffin_set_end: {
\hbox_set_end: \coffin_reset_poles:N #1
}
\cs_generate_variant:Nn \hcoffin_set:Nw { c }
\cs_new_protected:Npn \hcoffin_gset:Nw #1 {
\__coffin_if_exist:NT #1 {
\hbox_gset:Nw #1 \color_ensure_current: \cs_set_protected:Npn \hcoffin_gset_end: {
\hbox_gset_end: \coffin_greset_poles:N #1
}
}
\cs_generate_variant:Nn \hcoffin_gset:Nw { c }
\cs_new_protected:Npn \hcoffin_set_end: { }
\cs_new_protected:Npn \hcoffin_gset_end: { }
(End of definition for \hcoffin_set:Nw and others. These functions are documented on page 316.)

\vcoffin_set:Nnw \vcoffin_set:cnw \vcoffin_gset:Nnw \vcoffin_gset:cnw \__coffin_set_vertical:NnNNNNNw \vcoffin_set_end: \vcoffin_gset_end:
The same for vertical coffins.
\cs_new_protected:Npn \vcoffin_set:Nnw #1#2 {
\__coffin_set_vertical:NnNNNNNw #1 {#2} \vbox_set:Nw \\
\vbox_set_end: \coffin_reset_poles:N \__coffin_set_pole:Nnn
}
\cs_generate_variant:Nn \vcoffin_set:Nnw { c }
\cs_new_protected:Npn \vcoffin_gset:Nnw #1#2 {
\__coffin_set_vertical:NnNNNNNw #1 {#2} \vbox_gset:Nw \\
\vbox_gset_end: \coffin_greset_poles:N \__coffin_gset_pole:Nnn
}
\cs_generate_variant:Nn \vcoffin_gset:Nnw { c }
\cs_new_protected:Npn \__coffin_set_vertical:NNnNNNNNw #12#3#4#5#6#7 {
\__coffin_if_exist:NT #1
#3 #1
\dim_set:Nn \tex_hsize:D {#2} \__coffin_set_vertical_aux: \\
\cs_set_protected:Npn #4
}
Setting two coffins equal is just a wrapper around other functions.

Special coffins: these cannot be set up earlier as they need \texttt{\coffin_new:N}. The empty coffin is set as a box as the full coffin-setting system needs some material which is not yet available. The empty coffin is created entirely by hand: not everything is in place yet.
92.3 Measuring coffins

Coffins are just boxes when it comes to measurement. However, semantically a separate set of functions are required.

\cs_new_protected:Npn \__coffin_get_pole:NnN #1#2#3
\begin{verbatim}
prop_get:cnNF { coffin ~ \__coffin_to_value:N #1 ~ poles } {#2} #3
\end{verbatim}

(End of definition for \__coffin_get_pole:NnN. These variables are documented on page 318.)

92.4 Coffins: handle and pole management

A simple wrapper around the recovery of a coffin pole, with some error checking and recovery built-in.

\cs_new_protected:Npm \__coffin_reset_structure:N #1#2
\begin{verbatim}
\prop_set_eq:cN { coffin ~ \__coffin_to_value:N #1 ~ corners } \c__coffin_corners_prop
\prop_set_eq:cN { coffin ~ \__coffin_to_value:N #1 ~ poles } \c__coffin_poles_prop
\end{verbatim}

(End of definition for \__coffin_reset_structure:N.)
Setting the pole of a coffin at the user/designer level requires a bit more care. The idea here is to provide a reasonable interface to the system, then to do the setting with full expansion. The three-argument version is used internally to do a direct setting.
Simple shortcuts.

Updating the corners of a coffin is straight-forward as at this stage there can be no
rotation. So the corners of the content are just those of the underlying TeX box.
This function is called when a coffin is set, and updates the poles to reflect the nature of size of the box. Thus this function only alters poles where the default position is dependent on the size of the box. It also does not set poles which are relevant only to vertical coffins.

\cs_new_protected:Npm \__coffin_update_poles:N #1
\exp_args:Nc \__coffin_to_value:N #1 { hc }
\dim_eval:n { 0.5 \box_wd:N \#2 } { 0pt } { 1000pt }
\dim_eval:n { ( \box_ht:N \#2 - \box_dp:N \#2 ) / 2 } { 1000pt } { 0pt }
\dim_eval:n { -\box_dp:N \#2 } { 1000pt } { 0pt }
\cs_new_protected:Npm \__coffin_update_poles:NN #1 { r }
\dim_eval:n { \box_wd:N \#2 } { 0pt } { 1000pt }
\cs_new_protected:Npm \__coffin_update_poles:NNN #1 #2 #3
\dim_eval:n { \box_dp:N \#2 } { 1000pt } { 0pt }
\dim_eval:n { \box_ht:N \#2 } { 1000pt } { 0pt }
\dim_eval:n { -\box_dp:N \#2 }
92.5 Coffins: calculation of pole intersections

The lead off in finding intersections is to recover the two poles and then hand off to the auxiliary for the actual calculation. There may of course not be an intersection, for which an error trap is needed.

\begin{verbatim}
\cs_new_protected:Npn \_coffin_calculate_intersection:Nnn #1#2#3
\__coffin_get_pole:NnN #1 {#2} \l__coffin_pole_a_tl
\__coffin_get_pole:NnN #1 {#3} \l__coffin_pole_b_tl
\bool_set_false:N \l__coffin_error_bool
\exp_last_two_unbraced:Noo
\__coffin_calculate_intersection:nnnnnnnn
\l__coffin_pole_a_tl \l__coffin_pole_b_tl
\bool_if:NT \l__coffin_error_bool
{ \msg_error:nn { coffin } { no-pole-intersection }
\dim_zero:N \l__coffin_x_dim
\dim_zero:N \l__coffin_y_dim }
\end{verbatim}

The two poles passed here each have four values (as dimensions), \((a, b, c, d)\) and \((a', b', c', d')\). These are arguments 1–4 and 5–8, respectively. In both cases \(a\) and \(b\) are the coordinates of a point on the pole and \(c\) and \(d\) define the direction of the pole. Finding the intersection depends on the directions of the poles, which are given by \(d/c\) and \(d'/c'\). However, if one of the poles is either horizontal or vertical then one or more of \(c\), \(d\), \(c'\) and \(d'\) are zero and a special case is needed.

\begin{verbatim}
\cs_new_protected:Npm \_coffin_calculate_intersection:Nnnn
\_coffin_calculate_intersection:nnnnnnn
\dim_compare:nNnTF {#3} = \c_zero_dim
\{ \__coffin_calculate_intersection:nnn \}
\end{verbatim}

The case where the first pole is vertical. So the \(x\)-component of the interaction is at \(a\). There is then a test on the second pole: if it is also vertical then there is an error.

\begin{verbatim}
\dim_compare:nNnTF {#3} = \c_zero_dim
{ \dim_set:Nn \l__coffin_x_dim {#1}
\dim_compare:nNnTF {#7} = \c_zero_dim
{ \bool_set_true:N \l__coffin_error_bool }

The second pole may still be horizontal, in which case the \(y\)-component of the intersection is \(b'\). If not,

\begin{equation*}
y = \frac{d'}{c'} (a - a') + b'
\end{equation*}

with the \(x\)-component already known to be \(a\).

\begin{verbatim}
\dim_set:Nn \l__coffin_y_dim
{ \dim_compare:nNnTF {#8} = \c_zero_dim
\end{verbatim}

\end{verbatim}
If the first pole is not vertical then it may be horizontal. If so, then the procedure is essentially the same as that already done but with the \textit{x}- and \textit{y}-components interchanged.

\begin{verbatim}
{ \dim_compare:nNnTF {#4} = \c_zero_dim
  { \dim_set:Nn \l__coffin_y_dim {#2}
    \dim_compare:nNnTF {#8} = \c_zero_dim 
    { \bool_set_true:N \l__coffin_error_bool }
  }
}
\end{verbatim}

Now we deal with the case where the second pole may be vertical, or if not we have

\begin{equation}
x = \frac{c'}{d'} (b - b') + a'
\end{equation}

which is again handled by the same auxiliary.

\begin{verbatim}
\dim_set:Nn \l__coffin_x_dim
{ \dim_compare:nNnTF {#7} = \c_zero_dim
  { \fp_to_dim:n
    { ( \dim_to_fp:n {#7} / \dim_to_fp:n {#8} ) 
      * ( \dim_to_fp:n {#4} - \dim_to_fp:n {#6} ) 
      + \dim_to_fp:n {#5} 
    }
  }
}
\end{verbatim}

The first pole is neither horizontal nor vertical. To avoid even more complexity, we now work out both slopes and pass to an auxiliary.

\begin{verbatim}
{ \use:e
  { \_coffin_calculate_intersection:nnnnnn
    { \dim_to_fp:n {#4} / \dim_to_fp:n {#3} }
    { \dim_to_fp:n {#8} / \dim_to_fp:n {#7} }
    {#1} {#2} {#5} {#6}
  }
}
\end{verbatim}
Assuming the two poles are not parallel, then the intersection point is found in two steps. First we find the $x$-value with

$$x = \frac{sa - s'a' - b + b'}{s - s'}$$

and then finding the $y$-value with

$$y = s(x - a) + b$$

(End of definition for $\textbackslash _{\text{coffin}}\text{\_calculate\_intersection}:nnnn$, $\textbackslash _{\text{coffin}}\text{\_calculate\_intersection}:nnnnn$, $\textbackslash _{\text{coffin}}\text{\_calculate\_intersection}:nnnnnn$, and $\textbackslash _{\text{coffin}}\text{\_calculate\_intersection}:nnnnnnn$.)

## 92.6 Affine transformations

$\textbackslash _{\text{coffin}}\text{\_sin\_fp}$ $\textbackslash _{\text{coffin}}\text{\_cos\_fp}$

Used for rotations to get the sine and cosine values.

(End of definition for $\textbackslash _{\text{coffin}}\text{\_sin\_fp}$ and $\textbackslash _{\text{coffin}}\text{\_cos\_fp}$.)
\l__coffin_bounding_prop
A property list for the bounding box of a coffin. This is only needed during the rotation, so there is just the one.
\prop_new:N \l__coffin_bounding_prop
(End of definition for \l__coffin_bounding_prop.)

\l__coffin_corners_prop
\l__coffin_poles_prop
Used to avoid needing to track scope for intermediate steps.
\prop_new:N \l__coffin_corners_prop
\prop_new:N \l__coffin_poles_prop
(End of definition for \l__coffin_corners_prop and \l__coffin_poles_prop.)

\l__coffin_bounding_shift_dim
The shift of the bounding box of a coffin from the real content.
\dim_new:N \l__coffin_bounding_shift_dim
(End of definition for \l__coffin_bounding_shift_dim.)

\l__coffin_left_corner_dim
\l__coffin_right_corner_dim
\l__coffin_bottom_corner_dim
\l__coffin_top_corner_dim
These are used to hold maxima for the various corner values: these thus define the minimum size of the bounding box after rotation.
\dim_new:N \l__coffin_left_corner_dim
\dim_new:N \l__coffin_right_corner_dim
\dim_new:N \l__coffin_bottom_corner_dim
\dim_new:N \l__coffin_top_corner_dim
(End of definition for \l__coffin_left_corner_dim and others.)

\coffin_rotate:Nn
\coffin_rotate:cn
\coffin_grotate:Nn
\coffin_grotate:cn
\__coffin_rotate:NnNNN
Rotating a coffin requires several steps which can be conveniently run together. The sine and cosine of the angle in degrees are computed. This is then used to set \l__coffin_sin_fp and \l__coffin_cos_fp, which are carried through unchanged for the rest of the procedure.
\cs_new_protected:Npn \coffin_rotate:Nn #1#2
{ \__coffin_rotate:NnNNN #1 {#2} \box_rotate:Nn \prop_set_eq:cN \hbox_set:Nn }
\cs_generate_variant:Nn \coffin_rotate:Nn { c }
\cs_new_protected:Npn \coffin_grotate:Nn #1#2
{ \__coffin_rotate:NnNNN #1 {#2} \box_grotate:Nn \prop_gset_eq:cN \hbox_gset:Nn }
\cs_generate_variant:Nn \coffin_grotate:Nn { c }
\cs_new_protected:Npn \__coffin_rotate:NnNNN #1#2#3#4#5
{ \fp_set:Nn \l__coffin_sin_fp { sind ( #2 ) } \fp_set:Nn \l__coffin_cos_fp { cosd ( #2 ) } }
Use a local copy of the property lists to avoid needing to pass the name and scope around.
\prop_set_eq:Nc \l__coffin_corners_prop
{ coffin ~ \__coffin_to_value:N #1 ~ corners }
\prop_set_eq:Nc \l__coffin_poles_prop
{ coffin ~ \__coffin_to_value:N #1 ~ poles }
The corners and poles of the coffin can now be rotated around the origin. This is best achieved using mapping functions.
\prop_map_inline:Nn \l__coffin_corners_prop
{ \__coffin_rotate_corner:Nnnn #1 {##1} #2 }
\prop_map_inline:Nn \l__coffin_poles_prop
{ \__coffin_rotate_pole:Nnnnnn #1 {##1} #2 }

1394
The bounding box of the coffin needs to be rotated, and to do this the corners have to be found first. They are then rotated in the same way as the corners of the coffin material itself.

\_\_coffin_set_bounding:N \ #1
\prop_map_inline:Nn \_\_coffin_bounding_prop
{ \_\_coffin_rotate_bounding:nnn \#1 \#2 }

At this stage, there needs to be a calculation to find where the corners of the content and the box itself will end up.

\_\_coffin_find_corner_maxima:N \ #1
\_\_coffin_find_bounding_shift:
\_\_coffin_find_bounding_shift:
#3 \#1 \{\#2\}

The correction of the box position itself takes place here. The idea is that the bounding box for a coffin is tight up to the content, and has the reference point at the bottom-left. The x-direction is handled by moving the content by the difference in the positions of the bounding box and the content left edge. The y-direction is dealt with by moving the box down by any depth it has acquired. The internal box is used here to allow for the next step.

\hbox_set:Nn \l__coffin_internal_box
{ \_\kernel_kern:n
{ \_\_coffin_bounding_shift_dim - \_\_coffin_left_corner_dim }
\box_move_down:nn \{ \_\_coffin_bottom_corner_dim }
\}

If there have been any previous rotations then the size of the bounding box will be bigger than the contents. This can be corrected easily by setting the size of the box to the height and width of the content. As this operation requires setting box dimensions and these transcend grouping, the safe way to do this is to use the internal box and to reset the result into the target box.

\box_set_ht:Nn \l__coffin_internal_box
{ \_\_coffin_top_corner_dim - \_\_coffin_bottom_corner_dim }
\box_set_dp:Nn \l__coffin_internal_box \{ 0pt \}
\box_set_wd:Nn \l__coffin_internal_box
{ \_\_coffin_right_corner_dim - \_\_coffin_left_corner_dim }
\box_set_drop:N \l__coffin_internal_box

The final task is to move the poles and corners such that they are back in alignment with the box reference point.

\prop_map_inline:Nn \_\_coffin_corners_prop
{ \_\_coffin_shift_corner:NNnnn \#1 \#2 \#3 \#4 \#5
\prop_map_inline:Nn \_\_coffin_poles_prop
{ \_\_coffin_shift_pole:NNnnn \#1 \#2 \#3 \#4 \#5

Update the coffin data.

\_\_coffin_to_value:N \_\_coffin_corners\_prop
\_\_coffin_to_value:N \_\_coffin_poles\_prop

(End of definition for \coffin_rotate:Nn, \coffin_grotate:Nn, and \_\_coffin_rotate:NnNNN. These functions are documented on page 317.)
The bounding box corners for a coffin are easy enough to find: this is the same code as for the corners of the material itself, but using a dedicated property list.

\begin{verbatim}
\cs_new_protected:Npn \__coffin_set_bounding:N #1
\prop_put:Nne \l__coffin_bounding_prop { tl }
\prop_put:Nne \l__coffin_bounding_prop { tr }
\prop_put:Nne \l__coffin_bounding_prop { bl }
\prop_put:Nne \l__coffin_bounding_prop { br }
\end{verbatim}

Rotating the position of the corner of the coffin is just a case of treating this as a vector from the reference point. The same treatment is used for the corners of the material itself and the bounding box.

\begin{verbatim}
\cs_new_protected:Npn \__coffin_rotate_bounding:nnn #1#2#3
\__coffin_rotate_vector:nnNN {#2} {#3} \l__coffin_x_dim \l__coffin_y_dim
\prop_put:Nne \l__coffin_bounding_prop {#1}
\end{verbatim}

Rotating a single pole simply means shifting the coordinate of the pole and its direction. The rotation here is about the bottom-left corner of the coffin.

\begin{verbatim}
\cs_new_protected:Npn \__coffin_rotate_pole:Nnnnnn #1#2#3#4#5#6
\__coffin_rotate_vector:mmNN {#2}{#3} \l__coffin_x_dim \l__coffin_y_dim
\__coffin_rotate_vector:mmNN {#4}{#5} \l__coffin_x_prime_dim \l__coffin_y_prime_dim
\prop_put:Nne \l__coffin_poles_prop {#2}
\end{verbatim}
A rotation function, which needs only an input vector (as dimensions) and an output space. The values $\l__coffin\cos\_fp$ and $\l__coffin\sin\_fp$ should previously have been set up correctly. Working this way means that the floating point work is kept to a minimum: for any given rotation the sin and cosine values do no change, after all.

\begin{verbatim}
\cs_new_protected:Npn \__coffin_rotate_vector:nnNN #1#2#3#4
\{\dim_set:Nn #3 \{\dim_to_fp:n {#1} * \l__coffin\cos\_fp
- \dim_to_fp:n {#2} * \l__coffin\sin\_fp \} \dim_set:Nn #4 \{\dim_to_fp:n {#1} * \l__coffin\sin\_fp
+ \dim_to_fp:n {#2} * \l__coffin\cos\_fp \}
\}
\end{verbatim}

The idea here is to find the extremities of the content of the coffin. This is done by looking for the smallest values for the bottom and left corners, and the largest values for the top and right corners. The values start at the maximum dimensions so that the case where all are positive or all are negative works out correctly.

\begin{verbatim}
\cs_new_protected:Npn \__coffin_find_corner_maxima:N #1
\{\dim_set:Nn \l__coffin\top\_corner\_dim \{-\c_max\_dim\}
\dim_set:Nn \l__coffin\right\_corner\_dim \{-\c_max\_dim\}
\dim_set:Nn \l__coffin\bottom\_corner\_dim \{\c_max\_dim\}
\dim_set:Nn \l__coffin\left\_corner\_dim \{\c_max\_dim\}
\prop_map_inline:Nn \l__coffin\corners\_prop
\{ \__coffin_find_corner_maxima\_aux:nn \#2 \}
\}
\cs_new_protected:Npn \__coffin_find_corner_maxima\_aux:nn #1#2
\{\dim_set:Nn \l__coffin\left\_corner\_dim \{\dim_min:nn \{\l__coffin\left\_corner\_dim\} \#1\}
\dim_set:Nn \l__coffin\right\_corner\_dim \{\dim_min:nn \{\l__coffin\right\_corner\_dim\} \#1\}
\dim_set:Nn \l__coffin\bottom\_corner\_dim \{\dim_max:nn \{\l__coffin\bottom\_corner\_dim\} \#2\}
\dim_set:Nn \l__coffin\top\_corner\_dim \{\dim_max:nn \{\l__coffin\top\_corner\_dim\} \#2\}
\}
\end{verbatim}

(End of definition for \__coffin\rotate\_pole:Nnnnnn.)
The approach to finding the shift for the bounding box is similar to that for the corners. However, there is only one value needed here and a fixed input property list, so things are a bit clearer.

```latex
\cs_new_protected:Npn \__coffin_find_bounding_shift: \#
\dim_set:Nn \l__coffin_bounding_shift_dim { \c_max_dim }
\prop_map_inline:Nn \l__coffin_bounding_prop
{ \__coffin_find_bounding_shift_aux:nn #2 }
\cs_new_protected:Npn \__coffin_find_bounding_shift_aux:nn #1#2
{ \dim_set:Nn \l__coffin_bounding_shift_dim
{ \dim_min:nn { \l__coffin_bounding_shift_dim } {#1} } }
```

(End of definition for \__coffin_find_bounding_shift: and \__coffin_find_bounding_shift_aux:nn.)

Shifting the corners and poles of a coffin means subtracting the appropriate values from the $x$- and $y$-components. For the poles, this means that the direction vector is unchanged.

```latex
\cs_new_protected:Npn \__coffin_shift_corner:Nnnn #1#2#3#4
{ \prop_put:Nne \l__coffin_corners_prop {#2}
{ \dim_eval:n { #3 - \l__coffin_left_corner_dim } }
{ \dim_eval:n { #4 - \l__coffin_bottom_corner_dim } }
}
\cs_new_protected:Npn \__coffin_shift_pole:Nnnnnn #1#2#3#4#5#6
{ \prop_put:Nne \l__coffin_poles_prop {#2}
{ \dim_eval:n { #3 - \l__coffin_left_corner_dim } }
{ \dim_eval:n { #4 - \l__coffin_bottom_corner_dim } }
{#5} {#6}
}
```

(End of definition for \__coffin_shift_corner:Nnnn and \__coffin_shift_pole:Nnnnnn.)

Storage for the scaling factors in $x$ and $y$, respectively.

```latex
\fp_new:N \l__coffin_scale_x_fp
\fp_new:N \l__coffin_scale_y_fp
```

(End of definition for \l__coffin_scale_x_fp and \l__coffin_scale_y_fp.)

When scaling, the values given have to be turned into absolute values.

```latex
\dim_new:N \l__coffin_scaled_total_height_dim
\dim_new:N \l__coffin_scaled_width_dim
```

(End of definition for \l__coffin_scaled_total_height_dim and \l__coffin_scaled_width_dim.)
Resizing a coffin begins by setting up the user-friendly names for the dimensions of the coffin box. The new sizes are then turned into scale factor. This is the same operation as takes place for the underlying box, but that operation is grouped and so the same calculation is done here.

\cs_new_protected:Npn \coffin_resize:Nnn #1#2#3
\__coffin_resize:NnnNN #1 {#2} {#3}
\box_resize_to_wd_and_ht_plus_dp:Nnn
\prop_set_eq:cN
\cs_generate_variant:Nn \coffin_resize:Nnn { c }

\cs_new_protected:Npn \coffin_gresize:Nnn #1#2#3
\__coffin_resize:NnnNN #1 {#2} {#3}
\box_gresize_to_wd_and_ht_plus_dp:Nnn
\prop_gset_eq:cN
\cs_generate_variant:Nn \coffin_gresize:Nnn { c }

\cs_new_protected:Npn \__coffin_resize:NnnNN #1#2#3#4#5
\fp_set:Nn \l__coffin_scale_x_fp { \dim_to_fp:n {#2} / \dim_to_fp:n { \coffin_wd:N #1 } }
\fp_set:Nn \l__coffin_scale_y_fp { \dim_to_fp:n {#3} / \dim_to_fp:n { \coffin_ht:N #1 + \coffin_dp:N #1 } }
#4 #1 {#2} {#3}
\__coffin_resize_common:NnnN #1 {#2} {#3} #5

(End of definition for \coffin_resize:Nnn, \coffin_gresize:Nnn, and \__coffin_resize:NnnNN. These functions are documented on page 317.)

The poles and corners of the coffin are scaled to the appropriate places before actually resizing the underlying box.

\cs_new_protected:Npn \__coffin_resize_common:NnnN #1#2#3#4
\prop_set_eq:Nc \l__coffin_corners_prop { coffin ~ \__coffin_to_value:N #1 ~ corners }
\prop_set_eq:Nc \l__coffin_poles_prop { coffin ~ \__coffin_to_value:N #1 ~ poles }
\prop_map_inline:Nn \l__coffin_corners_prop { \__coffin_scale_corner:Nnnn #1 {##1} ##2 }
\prop_map_inline:Nn \l__coffin_poles_prop { \__coffin_scale_pole:Nnnnnn #1 {##1} ##2 }

Negative x-scaling values place the poles in the wrong location: this is corrected here.

\fp_compare:nNnT \l__coffin_scale_x_fp < \c_zero_fp
\prop_map_inline:Nn \l__coffin_corners_prop { \__coffin_x_shift_corner:Nnnn #1 {##1} ##2 }
\prop_map_inline:Nn \l__coffin_poles_prop { \__coffin_x_shift_pole:Nnnnnn #1 {##1} ##2 }
For scaling, the opposite calculation is done to find the new dimensions for the coffin. Only the total height is needed, as this is the shift required for corners and poles. The scaling is done the \TeX way as this works properly with floating point values without needing to use the `fp` module.

\begin{Verbatim}
\cs_new_protected:Npn \coffin_scale:Nnn \#1 \#2 \#3
{ \__coffin_scale:NnnNN \#1 { \#2 } { \#3 } \box_scale:Nnn \prop_set_eq:cN }
\cs_generate_variant:Nn \coffin_scale:Nnn { c }\end{Verbatim}

\begin{Verbatim}
\cs_new_protected:Npn \coffin_gscale:Nnn \#1 \#2 \#3
{ \__coffin_scale:NnnNN \#1 { \#2 } { \#3 } \box_gscale:Nnn \prop_gset_eq:cN }
\cs_generate_variant:Nn \coffin_gscale:Nnn { c }\end{Verbatim}

\begin{Verbatim}
\cs_new_protected:Npn \__coffin_scale:NnnNN \#1 \#2 \#3 \#4 \#5
{ \fp_set:Nn \l__coffin_scale_x_fp { \#2 }\end{Verbatim}

\begin{Verbatim}
\fp_set:Nn \l__coffin_scale_y_fp { \#3 }\end{Verbatim}

\begin{Verbatim}
\#4 \#1 \l__coffin_to_value:N \#1 \corners \end{Verbatim}

\begin{Verbatim}
\l__coffin_corners_prop \end{Verbatim}

\begin{Verbatim}
\#4 \#1 \l__coffin_to_value:N \#1 \poles \end{Verbatim}

\begin{Verbatim}
\l__coffin_poles_prop \end{Verbatim}

\begin{Verbatim}
\}
\end{Verbatim}

\begin{Verbatim}
(End of definition for \__coffin_resize_common:NnnN.)\end{Verbatim}

\begin{Verbatim}
\coffin_scale:Nnn \coffin_scale:cnn \coffin_gscale:Nnn \coffin_gscale:cnn \__coffin_scale:NnnNN \end{Verbatim}

This function scales a vector from the origin using the pre-set scale factors in $x$ and $y$. This is a much less complex operation than rotation, and as a result the code is a lot clearer.

\begin{Verbatim}
\cs_new_protected:Npm \__coffin_scale_vector:nnNN \#1 \#2 \#3 \#4
{ \fp_set:Nn \l__coffin_scaled_x_fp { \fp_to_dim:n { \dim_to_fp:n { \#1 } * \l__coffin_scale_x_fp } }\end{Verbatim}

\begin{Verbatim}
\fp_set:Nn \l__coffin_scaled_y_fp { \fp_to_dim:n { \dim_to_fp:n { \#2 } * \l__coffin_scale_y_fp } }\end{Verbatim}

\begin{Verbatim}
\dim_set:Nn \l__coffin_scaled_total_height_dim { \fp_abs:n { \l__coffin_scaled_x_fp } \l__coffin_internal_dim }\end{Verbatim}

\begin{Verbatim}
\dim_set:Nn \l__coffin_scaled_width_dim { \fp_abs:n { \l__coffin_scaled_y_fp } \l__coffin_internal_dim }\end{Verbatim}

\begin{Verbatim}
\__coffin_resize_common:NnnNN \#1 \#2 \#3 \#4 \#5\end{Verbatim}

\begin{Verbatim}
(End of definition for \__coffin_scale:Nnn, \__coffin_gscale:Nnn, and \__coffin_scale:NnnNN. These functions are documented on page 317.)\end{Verbatim}

\begin{Verbatim}
\__coffin_scale_vector:nnNN \end{Verbatim}

\begin{Verbatim}
(End of definition for \__coffin_scale_vector:NnnN.)\end{Verbatim}

\begin{Verbatim}
\__coffin_scale_corner:Nnnn \__coffin_scale_pole:Nnnnnn \end{Verbatim}

Scaling both corners and poles is a simple calculation using the preceding vector scaling.
These functions correct for the \( x \) displacement that takes place with a negative horizontal scaling.

92.7 Aligning and typesetting of coffins

This command joins two coffins, using a horizontal and vertical pole from each coffin and making an offset between the two. The result is stored as the as a third coffin, which has all of its handles reset to standard values. First, the more basic alignment function is used to get things started.
Correct the placement of the reference point. If the $x$-offset is negative then the reference point of the second box is to the left of that of the first, which is corrected using a kern.

On the right side the first box might stick out, which would show up if it is wider than the sum of the $x$-offset and the width of the second box. So a second kern may be needed.

The coffin structure is reset, and the corners are cleared: only those from the two parent coffins are needed.

The structures of the parent coffins are now transferred to the new coffin, which requires that the appropriate offsets are applied. That then depends on whether any shift was needed.

(End of definition for $\texttt{\coffin\_join:NnnNnnnn}$, $\texttt{\coffin\_gjoin:NnnNnnnn}$, and $\texttt{\__\coffin\_join:NnnNnnnnN}$. These functions are documented on page 318.)
A more simple version of the above, as it simply uses the size of the first coffin for the new one. This means that the work here is rather simplified compared to the above code. The function used when marking a position is hear also as it is similar but without the structure updates.

\cs_new_protected:Npn \coffin_attach:NnnNnnnn #1#2#3#4#5#6#7#8
\{ \__coffin_attach:NnnNnnnn \#1 \{#2\} \{#3\} \{#4\} \{#5\} \{#6\} \{#7\} \{#8\} \coffin_set_eq:NN \}
\cs_generate_variant:Nn \coffin_attach:NnnNnnnn { c , Nnnc , cnnc }
\cs_new_protected:Npn \coffin_gattach:NnnNnnnn #1#2#3#4#5#6#7#8
\{ \__coffin_attach:NnnNnnnn \#1 \{#2\} \{#3\} \{#4\} \{#5\} \{#6\} \{#7\} \{#8\} \coffin_gset_eq:NN \}
\cs_generate_variant:Nn \coffin_gattach:NnnNnnnn { c , Nnnc , cnnc }
\cs_new_protected:Npn \__coffin_attach:NnnNnnnnN \#1\#2\#3\#4\#5\#6\#7\#8
\{ \__coffin_align:NnnNnnnnN \#1 \{#2\} \{#3\} \{#4\} \{#5\} \{#6\} \{#7\} \{#8\} \l__coffin_aligned_coffin \box_set_ht:Nn \l__coffin_aligned_coffin { \box_ht:N #1 } \box_set_dp:Nn \l__coffin_aligned_coffin { \box_dp:N #1 } \box_set_wd:Nn \l__coffin_aligned_coffin { \box_wd:N #1 } \__coffin_reset_structure:N \l__coffin_aligned_coffin \prop_set_eq:cc \coffin \__coffin_to_value:N \l__coffin_aligned_coffin \c_space_tl \l__coffin_to_value:N \l__coffin_aligned_coffin \}
\cs_new_protected:Npn \__coffin_attach_mark:NnnNnnnn #1#2#3#4#5#6#7#8
\{ \__coffin_align:NnnNnnnnN \#1 \{#2\} \{#3\} \{#4\} \{#5\} \{#6\} \{#7\} \{#8\} \l__coffin_aligned_coffin \box_set_eq:NN #1 \l__coffin_aligned_coffin \}

(End of definition for \coffin_attach:NnnNnnnn and others. These functions are documented on page 318.)

\__coffin_align:NnnNnnnnN

The internal function aligns the two coffins into a third one, but performs no corrections on the resulting coffin poles. The process begins by finding the points of intersection for the poles for each of the input coffins. Those for the first coffin are worked out after those for the second coffin, as this allows the ‘primed’ storage area to be used for the
second coffin. The ‘real’ box offsets are then calculated, before using these to re-box the input coffins. The default poles are then set up, but the final result depends on how the bounding box is being handled.

\cs_new_protected:Npn \c__coffin_align:NnnNnnnnN \#1\#2\#3\#4\#5\#6\#7\#8\#9
\begin{verbatim}
\__coffin_calculate_intersection:Nnn \#4 \{\#5\} \{\#6\}
\dim_set:Nn \l__coffin_x_prime_dim \{ \l__coffin_x_dim \}
\dim_set:Nn \l__coffin_y_prime_dim \{ \l__coffin_y_dim \}
\__coffin_calculate_intersection:Nnn \#1 \{\#2\} \{\#3\}
\dim_set:Nn \l__coffin_offset_x_dim
{ \l__coffin_x_dim \- \l__coffin_x_prime_dim \+ \#7 }
\dim_set:Nn \l__coffin_offset_y_dim
{ \l__coffin_y_dim \- \l__coffin_y_prime_dim \+ \#8 }
\hbox_set:Nn \l__coffin_aligned_internal_coffin
{ \box_use:N \#1 \__kernel_kern:n { -\box_wd:N \#1 }
\__kernel_kern:n { \l__coffin_offset_x_dim }
\box_move_up:nn { \l__coffin_offset_y_dim } \{ \box_use:N \#4 \}
}
\coffin_set_eq:NN \#9 \l__coffin_aligned_internal_coffin
\end{verbatim}

\end{definition}

Transferring structures from one coffin to another requires that the positions are updated by the offset between the two coffins. This is done by mapping over the property list of the source coffins, moving as appropriate and saving to the new coffin data structures. The test for a \- means that the structures from the parent coffins are uniquely labelled and do not depend on the order of alignment. The pay off for this is that \- should not be used in coffin pole or handle names, and that multiple alignments do not result in a whole set of values.

\cs_new_protected:Npn \c__coffin_offset_poles:Nnn \#1 \#2 \#3
\begin{verbatim}
\prop_map_inline:cn { coffin ~ \c__coffin_to_value:N \#1 ~ poles }
\{ \c__coffin_offset_pole:Nnnnnnn \#1 \{\#1\} \#2 \{\#2\} \#3 \}
\end{verbatim}

\cs_new_protected:Npn \c__coffin_offset_pole:Nnnnnnn \#1 \#2 \#3 \#4 \#5 \#6 \#7 \#8
\begin{verbatim}
\dim_set:Nn \l__coffin_x_dim \#3 \+ \#7
\dim_set:Nn \l__coffin_y_dim \#4 \+ \#8
\tl_if_in:nTF \#2 \{ - \}
{ \tl_set:Nn \l__coffin_internal_tl \{\#2\} }
{ \tl_set:Nn \l__coffin_internal_tl \{ \#1 \- \#2 \} }
\exp_last_unbraced:NNo \c__coffin_set_pole:Nnn \l__coffin_aligned_coffin
\{ \l__coffin_internal_tl \}
{ \dim_use:N \l__coffin_x_dim \{ \dim_use:N \l__coffin_y_dim }
{ \dim_use:N \l__coffin_y_dim }
{ \#5 \#6}
\end{verbatim}
\end{definition}

\(End \ of \ definition \ for \ \c__coffin_align:NnnnnnnN\).

\(End \ of \ definition \ for \ \c__coffin_offset_poles:Nnn\ \text{and} \ \c__coffin_offset_pole:Nnnnnnn\).

1404
Saving the offset corners of a coffin is very similar, except that there is no need to worry about naming: every corner can be saved here as order is unimportant.

```latex
\cs_new_protected:Npn \__coffin_offset_corners:Nnn #1#2#3
\prop_map_inline:cn { coffin - \__coffin_to_value:N #1 - corners }
\{ \__coffin_offset_corner:Nnnnn #1 {##1} ##2 {#2} {#3} \}
\cs_new_protected:Npn \__coffin_offset_corner:Nnnnn #1#2#3#4#5#6
\prop_put:cne
{ coffin - \__coffin_to_value:N \l__coffin_aligned_coffin \c_space_tl corners }
{ #1 - #2 }
{ \dim_eval:n { #3 + #5 } }
{ \dim_eval:n { #4 + #6 } }
\exp_last_two_unbraced:Noo \__coffin_update_T:nnnnnnnnN
\l__coffin_pole_a_tl \l__coffin_pole_b_tl
\exp_last_two_unbraced:Noo \__coffin_update_B:nnnnnnnnN
\l__coffin_pole_a_tl \l__coffin_pole_b_tl
(End of definition for \__coffin_offset_corners:Nnn and \__coffin_offset_corner:Nnnnn.)
```

The T and B poles need to be recalculated after alignment. These functions find the larger absolute value for the poles, but this is of course only logical when the poles are horizontal.

```latex
\cs_new_protected:Npn \__coffin_update_vertical_poles:NNN
\__coffin_update_T:nnnnnnnnN
\__coffin_update_B:nnnnnnnnN
\cs_new_protected:Npn \__coffin_update_T:nnnnnnnnN #1#2#3#4#5#6#7#8#9
\dim_compare:nNnTF {#2} < {#6}
{ \__coffin_set_pole:Nnn #9 { T } \{ 0pt \} {#6} \{ 1000pt \} \{ 0pt \} }
{ \__coffin_set_pole:Nnn #9 { T } \{ 0pt \} \{#2} \{ 1000pt \} \{ 0pt \} }
\exp_last_two_unbraced:Noo \__coffin_update_B:nnnnnnnnN
\l__coffin_pole_a_tl \l__coffin_pole_b_tl
\cs_new_protected:Npn \__coffin_update_B:nnnnnnnnN #1#2#3#4#5#6#7#8#9
\dim_compare:nNnTF {#2} < {#6}
{ \__coffin_set_pole:Nnn #9 { B } \{ 0pt \} \{#2} \{ 1000pt \} \{ 0pt \} }
```

1405
\c__coffin_empty_coffin
An empty-but-horizontal coffin.
\coffin_new:N \c__coffin_empty_coffin
\tex_setbox:D \c__coffin_empty_coffin = \tex_hbox:D { }
(End of definition for \c__coffin_empty_coffin.)
\coffin_typeset:Nnnnn
\coffin_typeset:cnnnn
Typesetting a coffin means aligning it with the current position, which is done using a
coffin with no content at all. As well as aligning to the empty coffin, there is also a need
to leave vertical mode, if necessary.
\cs_new_protected:Npn \coffin_typeset:Nnnnn #1#2#3#4#5
{ \mode_leave_vertical:
  \__coffin_align:Nnnn \c__coffin_empty_coffin { H } { l } #1 {#2} {#3} {#4} {#5} \l__coffin_aligned_coffin
  \box_use_drop:N \l__coffin_aligned_coffin }
\cs_generate_variant:Nn \coffin_typeset:Nnnnn { c }
(End of definition for \coffin_typeset:Nnnnn. This function is documented on page 318.)

92.8 Coffin diagnostics
\l__coffin_display_coffin
\l__coffin_display_coord_coffin
\l__coffin_display_pole_coffin
Used for printing coffins with data structures attached.
\l__coffin_display_coffin
\l__coffin_display_coord_coffin
\l__coffin_display_pole_coffin
(End of definition for \l__coffin_display_coffin, \l__coffin_display_coord_coffin, and \l__coffin_display_pole_coffin.)
\l__coffin_display_handles_prop
This property list is used to print coffin handles at suitable positions. The offsets are
expressed as multiples of the basic offset value, which therefore acts as a scale-factor.
\prop_new:N \l__coffin_display_handles_prop
\prop_put:Nnn \l__coffin_display_handles_prop { tl } { b } { r } { -1 } { 1 }
\prop_put:Nnn \l__coffin_display_handles_prop { thc } { b } { hc } { 0 } { 1 }
\prop_put:Nnn \l__coffin_display_handles_prop { tr } { b } { l } { 1 } { 1 }
\prop_put:Nnn \l__coffin_display_handles_prop { vcl } { vc } { r } { -1 } { 0 }
\prop_put:Nnn \l__coffin_display_handles_prop { vchc } { vc } { hc } { 0 } { 0 }

1406
\prop_put:Nnn \l__coffin_display_handles_prop { vcr } \{{ \{ vc \} \{ 1 \} \{ 1 \} \{ 0 \} \} \}
\prop_put:Nnn \l__coffin_display_handles_prop { bl } \{{ \{ t \} \{ x \} \{ -1 \} \{ -1 \} \} \}
\prop_put:Nnn \l__coffin_display_handles_prop { bhc } \{{ \{ t \} \{ hc \} \{ 0 \} \{ -1 \} \} \}
\prop_put:Nnn \l__coffin_display_handles_prop { br } \{{ \{ t \} \{ 1 \} \{ 1 \} \{ -1 \} \} \}
\prop_put:Nnn \l__coffin_display_handles_prop { Tl } \{{ \{ t \} \{ r \} \{ -1 \} \{ -1 \} \} \}
\prop_put:Nnn \l__coffin_display_handles_prop { Thc } \{{ \{ t \} \{ hc \} \{ 0 \} \{ -1 \} \} \}
\prop_put:Nnn \l__coffin_display_handles_prop { Tr } \{{ \{ t \} \{ 1 \} \{ 1 \} \{ -1 \} \} \}
\prop_put:Nnn \l__coffin_display_handles_prop { HL } \{{ \{ vc \} \{ x \} \{ -1 \} \{ 1 \} \} \}
\prop_put:Nnn \l__coffin_display_handles_prop { Hhc } \{{ \{ vc \} \{ hc \} \{ 0 \} \{ 1 \} \} \}
\prop_put:Nnn \l__coffin_display_handles_prop { Hr } \{{ \{ vc \} \{ l \} \{ 1 \} \{ 1 \} \} \}
\prop_put:Nnn \l__coffin_display_handles_prop { Bl } \{{ \{ b \} \{ x \} \{ -1 \} \{ -1 \} \} \}
\prop_put:Nnn \l__coffin_display_handles_prop { Bhc } \{{ \{ b \} \{ hc \} \{ 0 \} \{ -1 \} \} \}
\prop_put:Nnn \l__coffin_display_handles_prop { Br } \{{ \{ b \} \{ 1 \} \{ 1 \} \{ -1 \} \} \}
\prop_put:Nnn \l__coffin_display_handles_prop { Tl } \{{ \{ b \} \{ r \} \{ -1 \} \{ -1 \} \} \}
\prop_put:Nnn \l__coffin_display_handles_prop { Thc } \{{ \{ b \} \{ hc \} \{ 0 \} \{ -1 \} \} \}
\prop_put:Nnn \l__coffin_display_handles_prop { Tr } \{{ \{ b \} \{ l \} \{ 1 \} \{ -1 \} \} \}
\prop_put:Nnn \l__coffin_display_handles_prop { HL } \{{ \{ vc \} \{ x \} \{ -1 \} \{ 1 \} \} \}
\prop_put:Nnn \l__coffin_display_handles_prop { Hhc } \{{ \{ vc \} \{ hc \} \{ 0 \} \{ 1 \} \} \}
\prop_put:Nnn \l__coffin_display_handles_prop { Hr } \{{ \{ vc \} \{ l \} \{ 1 \} \{ 1 \} \} \}
\prop_put:Nnn \l__coffin_display_handles_prop { Bl } \{{ \{ b \} \{ x \} \{ -1 \} \{ -1 \} \} \}
\prop_put:Nnn \l__coffin_display_handles_prop { Bhc } \{{ \{ b \} \{ hc \} \{ 0 \} \{ -1 \} \} \}
\prop_put:Nnn \l__coffin_display_handles_prop { Br } \{{ \{ b \} \{ 1 \} \{ 1 \} \{ -1 \} \} \}

(End of definition for \l__coffin_display_handles_prop.)

\l__coffin_display_offset_dim The standard offset for the label from the handle position when displaying handles.
\dim_new:N \l__coffin_display_offset_dim \dim_set:Nn \l__coffin_display_offset_dim { 2pt }
(End of definition for \l__coffin_display_offset_dim.)

\l__coffin_display_x_dim As the intersections of poles have to be calculated to find which ones to print, there is a need to avoid repetition. This is done by saving the intersection into two dedicated values.
\dim_new:N \l__coffin_display_x_dim
\dim_new:N \l__coffin_display_y_dim
(End of definition for \l__coffin_display_x_dim and \l__coffin_display_y_dim.)

\l__coffin_display_poles_prop A property list for printing poles: various things need to be deleted from this to get a "nice" output.
\prop_new:N \l__coffin_display_poles_prop
(End of definition for \l__coffin_display_poles_prop.)

\l__coffin_display_font_tl Stores the settings used to print coffin data: this keeps things flexible.
\tl_new:N \l__coffin_display_font_tl
\bool_lazy_and:nnT
\cs_if_exist_p:N \fmtname
\str_if_eq_p:Vn \fmtname { LaTeX2e }
\arr_if_eq_p:Vn \fmtname { LaTeX2e }
\{$

Abstract out creation of rules here until there is a higher-level interface.

Marking a single handle is relatively easy. The standard attachment function is used, meaning that there are two calculations for the location. However, this is likely to be okay given the load expected. Contrast with the more optimised version for showing all handles which comes next.
\coffin_display_handles:Nn
\coffin_display_handles:cn
\coffin_display_handles:aux:nnnn
\coffin_display_handles:aux:nnnn
\coffin_display_attach:NNnn
\coffin_display_handles

Printing the poles starts by removing any duplicates, for which the H poles is used as the definitive version for the baseline and bottom. Two loops are then used to find the combinations of handles for all of these poles. This is done such that poles are removed during the loops to avoid duplication.

\coffin_display_handles

For each pole there is a check for an intersection, which here does not give an error if none is found. The successful values are stored and used to align the pole coffin with the main coffin for output. The positions are recovered from the preset list if available.
This is a dedicated version of \coffin_attach:Nnnnn with a hard-wired first coffin. As the intersection is already known and stored for the display coffin the code simply uses it directly, with no calculation.
\coffin{
\l__coffin_display_x_dim = \l__coffin_x_prime_dim + #4
\dim_set:Nn \l__coffin_offset_y_dim
\l__coffin_display_y_dim = \l__coffin_y_prime_dim + #5
\box_set:Nn \l__coffin_aligned_coffin
{
  \box_use:N \l__coffin_display_coffin
  \__kernel_kern:n { -\box_wd:N \l__coffin_display_coffin }
  \__kernel_kern:n { \l__coffin_offset_x_dim }
  \box_move_up:nn { \l__coffin_offset_y_dim } { \box_use:N #1 }
}
\box_set_ht:Nn \l__coffin_aligned_coffin
{ \box_ht:N \l__coffin_display_coffin }
\box_set_dp:Nn \l__coffin_aligned_coffin
{ \box_dp:N \l__coffin_display_coffin }
\box_set_wd:Nn \l__coffin_aligned_coffin
{ \box_wd:N \l__coffin_display_coffin }
\box_set_eq:NN \l__coffin_display_coffin \l__coffin_aligned_coffin
}

(End of definition for \coffin_display_handles:Nn and others. This function is documented on page 319.)

\coffin_show_structure:N
\coffin_show_structure:c
\coffin_log_structure:N
\coffin_log_structure:c
\__coffin_show_structure:NN

\cs_new_protected:Nnpn \coffin_show_structure:N
{ \__coffin_show_structure:NN \msg_show:nneeee }
\cs_generate_variant:Nn \coffin_show_structure:N { c }
\cs_new_protected:Npn \coffin_log_structure:N
{ \__coffin_show_structure:NN \msg_log:nneeee }
\cs_generate_variant:Nn \coffin_log_structure:N { c }
\cs_new_protected:Npn \__coffin_show_structure:NN #1#2
{\__coffin_if_exist:NT #2
{ #1 { coffin } { show }
{ \token_to_str:N #2 }
{ \iow_newline: >~ ht ~=~ \dim_eval:n { \coffin_ht:N #2 }
\iow_newline: >~ dp ~=~ \dim_eval:n { \coffin_dp:N #2 }
\iow_newline: >~ wd ~=~ \dim_eval:n { \coffin_wd:N #2 }
}
{ \prop_map_function:cN
  { coffin ~ \__coffin_to_value:N #2 ~ poles }
  \msg_show_item_unbraced:nn
}
{ }
}

(End of definition for \coffin_show_structure:N, \coffin_log_structure:N, and \__coffin_show_structure:NN. These functions are documented on page 319.)

\coffin_show:N
\coffin_show:c
\coffin_log:N
\coffin_log:c
\coffin_show:Nnn
\coffin_log:Nnn
\coffin_log:cnn
\__coffin_show:NNnn

Essentially a combination of \coffin_show_structure:N and \box_show:Nnn, but we need to avoid having two prompts, so we use \msg_term:nneeee instead of \msg_show:nneeee in the show case.
92.9 Messages

\msg_new:nnnn { coffin } { no-pole-intersection } { No-intersection-between-coffin-poles. }
\msg_new:nnnn { coffin } { unknown } { Unknown-coffin-‘#1’. }
\msg_new:nnnn { coffin } { unknown-pole } { Pole-‘#1’-unknown-for-coffin-‘#2’. }
\msg_new:nnnn { coffin } { show } { Size-of-coffin-#1 : #2 \ " Poles-of-coffin-#1 : #3 .
\msg_new:nnnn { coffin } { show } { Size-of-coffin-#1 : #2 \ " Poles-of-coffin-#1 : #3 .
\msg_new:nnnn { coffin } { show } { Size-of-coffin-#1 : #2 \ " Poles-of-coffin-#1 : #3 .
\msg_new:nnnn { coffin } { show } { Size-of-coffin-#1 : #2 \ " Poles-of-coffin-#1 : #3 .
\msg_new:nnnn { coffin } { show } { Size-of-coffin-#1 : #2 \ " Poles-of-coffin-#1 : #3 .
\msg_new:nnnn { coffin } { show } { Size-of-coffin-#1 : #2 \ " Poles-of-coffin-#1 : #3 .
\msg_new:nnnn { coffin } { show } { Size-of-coffin-#1 : #2 \ " Poles-of-coffin-#1 : #3 .
\msg_new:nnnn { coffin } { show } { Size-of-coffin-#1 : #2 \ " Poles-of-coffin-#1 : #3 .
\msg_new:nnnn { coffin } { show } { Size-of-coffin-#1 : #2 \ " Poles-of-coffin-#1 : #3 .
\msg_new:nnnn { coffin } { show } { Size-of-coffin-#1 : #2 \ " Poles-of-coffin-#1 : #3 .
\msg_new:nnnn { coffin } { show } { Size-of-coffin-#1 : #2 \ " Poles-of-coffin-#1 : #3 .
\msg_new:nnnn { coffin } { show } { Size-of-coffin-#1 : #2 \ " Poles-of-coffin-#1 : #3 .
\msg_new:nnnn { coffin } { show } { Size-of-coffin-#1 : #2 \ " Poles-of-coffin-#1 : #3 .
\msg_new:nnnn { coffin } { show } { Size-of-coffin-#1 : #2 \ " Poles-of-coffin-#1 : #3 .
Chapter 93

\texttt{l3color implementation}

\texttt{\_color_current_tl} The color currently active for foreground (text, \textit{etc.}) material. This is stored in the form of a color model followed by one or more values. There are four pre-defined models, three of which take numerical values in the range $[0, 1]$:

- \texttt{gray} \langle\texttt{gray}\rangle Grayscale color with the \langle\texttt{gray}\rangle value running from 0 (fully black) to 1 (fully white)
- \texttt{cmyk} \langle\texttt{cyan}\rangle \langle\texttt{magenta}\rangle \langle\texttt{yellow}\rangle \langle\texttt{black}\rangle
- \texttt{rgb} \langle\texttt{red}\rangle \langle\texttt{green}\rangle \langle\texttt{blue}\rangle

Notice that the value are separated by spaces. There is a fourth pre-defined model using a string value and a numerical one:

- \texttt{spot} \langle\texttt{name}\rangle \langle\texttt{tint}\rangle A pre-defined spot color, where the \langle\texttt{name}\rangle should be a pre-defined string color name and the \langle\texttt{tint}\rangle should be in the range $[0, 1]$.

Additional models may be created to allow mixing of spot colors. The number of data entries these require will depend on the number of colors to be mixed.

\textbf{\LaTeX}{}hackers note: The content of \texttt{\_color_current_tl} comprises two brace groups, the first containing the color model and the second containing the value(s) applicable in that model.

(\textit{End of definition for \_color_current_tl.})

\texttt{\color_group_begin:} \texttt{\color_group_end:} Grouping for color is the same as using the basic \texttt{\group_begin:} and \texttt{\group_end:} functions. However, for semantic reasons, they are renamed here.

(\textit{End of definition for \color_group_begin: and \color_group_end:. These functions are documented on page 321.})
\color\_ensure\_current: A driver-independent wrapper for setting the foreground color to the current color "now".

\begin{verbatim}
\cs_new_protected:Npn \color\_ensure\_current:
\{ \__color\_select:N \l__color\_current\_tl \}
\end{verbatim}

(End of definition for \color\_ensure\_current:. This function is documented on page 321.)

\s\_color\_stop Internal scan marks.

\begin{verbatim}
\scan\_new:N \s\_color\_stop
\end{verbatim}

(End of definition for \s\_color\_stop.)

\__color\_select:N Take an internal color specification and pass it to the driver. This code is needed to ensure the current color but will also be used by the higher-level material.

\begin{verbatim}
\cs_new_protected:Npn \__color\_select:N #1
\{ \exp\_after:wN \__color\_select:nn #1
\group\_insert\_after:N \__color\_backend\_reset:\ }
\end{verbatim}

\begin{verbatim}
\cs_new_protected:Npn \__color\_select\_math:N #1
{ \exp\_after:wN \__color\_select:nn \#1 }
\end{verbatim}

\begin{verbatim}
\cs_new_protected:Npn \__color\_select:nn #1#2
{ \use:c { \_\_color\_backend\_select\_ #1 :n } {#2} }
\end{verbatim}

(End of definition for \__color\_select:N, \__color\_select\_math:N, and \__color\_select:nn.)

\l\_color\_current\_tl The current color, with the model and

\begin{verbatim}
\tl\_new:N \l\_color\_current\_tl
\tl\_set:Nn \l\_color\_current\_tl \{ \{ \gray \} \{ 0 \} \}
\end{verbatim}

(End of definition for \l\_color\_current\_tl.)

93.2 Predefined color names

The ability to predefine colors with a name is a key part of this module and means there has to be a method for storing the results. At first sight, it seems natural to follow the usual expl3 model and create a color variable type for the process. That would then allow both local and global colors, constant colors and the like. However, these names need to be accessible in some form at the user level, for selection of colors either simply by name or as part of a more complex expression. This does not require that the full name is exposed but does require that they can be looked up in a predictable way. As such, it is more useful to expose just the color names as part of the interface, with the result that only local color names can be created. (This is also seen for example in key creation in l3keys.) As a result, color names are declarative (no new functions).

Since there is no need to manipulate colors en masse, each is stored in a two-part structure: a prop for the colors themselves, and a tl for the default model for each color.
\l__color_internal_int
\l__color_internal_tl
\s__color_mark
\l__color_ignore_error_bool
\color_if_exist_p:n
\color_if_exist:nTF
\__color_model:N
\__color_values:N
\__color_extract:nNN
\__color_extract:VNN

93.3 Setup

(End of definition for \l__color_internal_int and \l__color_internal_tl.)

\s__color_mark
Internal scan marks. \s__color_stop is already defined in l3color-base.

(End of definition for \s__color_mark.)

\l__color_ignore_error_bool
Used to avoid issuing multiple errors if there is a change-of-model with input container an error.

(End of definition for \l__color_ignore_error_bool.)

93.4 Utility functions

\color_if_exist_p:n
A simple wrapper to avoid needing to have the lookup repeated in too many places. To guard against a color created in a group, we need to test for entries in the prop.

(End of definition for \color_if_exist:nTF. This function is documented on page 324.)

\__color_model:N
\__color_values:N
Simple abstractions.

(End of definition for \__color_model:N and \__color_values:N.)

\__color_extract:nNN
\__color_extract:VNN
Recover the values for the standard model for a color.

(End of definition for \__color_extract:nNN.)
Model conversion is carried out using standard formulae for base models, as described in the manual for xcolor (see also the PostScript Language Reference Manual). For other models direct conversion might not be defined, so we go through the fallback models if necessary.

These rather odd values are based on NTSC television: the set are used for the cmyk conversion.
The conversion from rgb to cmyk is the most complex: a two-step procedure which requires black generation and undercolor removal functions. The PostScript reference describes them as device-dependent, but following xcolor we assume they are linear. Moreover, as the likelihood of anyone using a non-unitary matrix here is tiny, we simplify and treat those two concepts as no-ops. To allow code sharing with parsing of cmy values, we have an intermediate function here (\__color_convert_rgb_cmyk:nnn) which actually takes cmy values as input.

\cs_new:Npn \__color_convert_rgb_cmyk:w #1 ~ #2 ~ #3 \s__color_stop
{ \exp_args:Neee \__color_convert_rgb_cmyk:nnn
\exp_args:Ne \__color_convert_rgb_cmyk:nnnn
\exp_args:Ne \__color_convert_rgb_cmyk_gray:w #1 - #2 - #3 - #4 \s__color_stop
\exp_args:Ne \__color_convert_rgb_cmyk_cmyk:w #1 \s__color_stop
}(End of definition for \__color_convert:nnnn and others.)

93.6 Color expressions

Working space to store the color data whilst doing calculations: keeping it on the stack is attractive but gets tricky (return is non-trivial).
The main function for parsing color expressions removes actives but otherwise expands, then starts working through the expression itself. At the end, we apply the payload.

36807 \cs_new_protected:Npe \__color_parse:nN #1#2
36808 { \tl_set:Ne \exp_not:c { l__color_named_ . _tl } 
36809 { \exp_not:N \__color_model:N \exp_not:N \l__color_current_tl }
36810 \prop_put:NVe \exp_not:c { l__color_named_ . _prop } 
36811 \exp_not:c { l__color_named_ . _tl } 
36812 { \exp_not:N \__color_values:N \exp_not:N \l__color_current_tl } 
36813 \exp_not:N \exp_args:Ne \exp_not:N \__color_parse_aux:nN 
36814 { \exp_not:N \tl_to_str:n {#1} } #2 
36815 
36816 

Before going to all of the effort of parsing an expression, these two precursor functions look for a pre-defined name, either on its own or with a trailing ! (which is the same thing).

36817 \cs_new_protected:Npn \__color_parse_aux:nN #1#2
36818 { \color_if_exist:nTF {#1} 
36819 { \__color_parse_set_eq:Nn #2 {#1} } 
36820 { \__color_parse:Nw #2#1 ! \s__color_stop } 
36821 \__color_check_model:N #2
36822 }

36823 \cs_new_protected:Npn \__color_parse_set_eq:Nn #1#2#3
36824 { \prop_get:cnNTF 
36825 { l__color_named_ #3 _prop } {#1} 
36826 \l__color_value_tl 
36827 { \tl_set:Ne #2 { {#1} { \l__color_value_tl } } } 
36828 \l__color_model_tl 
36829 \prop_get:cnV { l__color_named_ #3 _prop } \l__color_model_tl 
36830 \l__color_value_tl 
36831 \__color_convert:nN 
36832 \l__color_model_tl {#1} \l__color_value_tl 
36833 \tl_set:Ne #2 
36834 { (#1) 
36835 { \l__color_value_tl } 
36836 }

Here, we have to allow for the case where there is a fixed model: that can’t be swept up by generic conversion as we are dealing with a named color.

36837 \cs_new_protected:Npm \__color_parse_set_eq:nNn #1#2#3
36838 { \tl_if_empty:NTF \l__color_fixed_model_tl 
36839 { \exp_args:Nv \__color_parse_set_eq:nNn { l__color_named_ #2 _tl } } 
36840 { \exp_args:Nv \__color_parse_set_eq:nNn \l__color_fixed_model_tl } 
36841 #1 (#2)
36842 }

1418
Once we establish that a full parse is needed, the next job is to get the detail of the first color. That will determine the model we use for the calculation: splitting here makes checking that a bit easier.

\cs_new_protected:Npn \__color_parse_loop_init:Nnn #1#2#3
\group_begin:
\__color_extract:nNN {#2} \l__color_model_tl \l__color_value_tl
\__color_parse_loop:w #3 ! ! ! ! \s__color_stop
\tl_set:Ne \l__color_internal_tl { \l__color_model_tl } { \l__color_value_tl }
\exp_args:NNNV \group_end:
\tl_set:Nn #1 \l__color_internal_tl
\__color_parse_end:

This is the loop proper: there can be an open-ended set of colors to parse, separated by \texttt{!} tokens. There are a few cases to look out for. At the end of the expression and with we find a mix of \texttt{100} then we simply skip the next color entirely (we can’t stop the loop as there might be a further valid color to mix in). On the other hand, if we get a mix of \texttt{0} then drop everything so far and start again. There is also a trailing \texttt{white} to “read in” if the final explicit data is a mix. Those conditions are separate from actually looping, which is therefore sorted out by checking if we have further data to process: in contrast to \texttt{xcolor}, we don’t allow \texttt{!!} so the test can be simplified.

\cs_new_protected:Npn \__color_parse_loop:w #1 ! #2 ! #3 ! #4 ! #5 \s__color_stop
\group_begin:
\tl_if_blank:nF {#1}
\bool_lazy_and:nnTF { \fp_compare_p:nNn {#1} > { 0 } } { \fp_compare_p:nNn {#1} < { 100 } }
\use:e
{ \__color_parse_loop:nn {#1} { \tl_if_blank:nTF {#2} \texttt{white} {#2} } }
\tl_if_blank:nF {#3}
{ \__color_parse_loop:nn {#1} { #2 } }
\tl_if_blank:nF {#3}
{ \__color_parse_loop:nn {#1} { #2 } }
\tl_if_blank:nF {#3}
{ \__color_parse_loop:w {#3} ! #4 ! #5 \s__color_stop }
\__color_parse_end:
As these are unusual cases, we accept slower performance here for clearer code: check for the error conditions, handle the boundary cases after that.

\begin{verbatim}
\cs_new_protected:Npn \__color_parse_loop_check:nn #1#2
\begin{verbatim}
\bool_if:NF \l__color_ignore_error_bool
\bool_lazy_or:nnT
{ \fp_compare_p:nNn {#1} < { 0 } }
{ \fp_compare_p:nNn {#1} > { 100 } }
{ \msg_error:nnnn { color } \{ out-of-range \} \{#1\} \{ 0 \} \{ 100 \} }
\end{verbatim}
\fp_compare:nNnF {#1} > \c_zero_fp
\begin{verbatim}
\tl_if_blank:nTF {#2}
{ \__color_extract:nNN { white } }
{ \__color_extract:nNN {#2} }
\end{verbatim}
\l__color_model_tl \l__color_value_tl
\end{verbatim}
\begin{verbatim}
\__color_parse_loop:nn #1#2
\end{verbatim}
\end{verbatim}

The “payload” of calculation in the loop first. If the model for the upcoming color is different from that of the existing (partial) color, convert the model. For gray the two are flipped round so that the outcome is something with “real” color. We are then in a position to do the actual calculation itself. The two auxiliaries here give us a way to break the loop should an invalid name be found.

\begin{verbatim}
\cs_new_protected:Npn \__color_parse_loop:nn #1#2
\begin{verbatim}
\color_if_exist:nTF {#2}
\begin{verbatim}
\__color_extract:nNN {#2} \l__color_next_model_tl \l__color_next_value_tl
\tl_if_eq:NNF \l__color_model_tl \l__color_next_model_tl
\begin{verbatim}
\str_if_eq:VnTF \l__color_model_tl { gray }
\begin{verbatim}
\__color_parse_gray:n {#2}
\end{verbatim}
\end{verbatim}
\begin{verbatim}
\__color_parse_std:n {#2}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\begin{verbatim}
\tl_set:Ne \l__color_value_tl
\begin{verbatim}
\__color_parse_mix:NVVn
\l__color_model_tl \l__color_value_tl \l__color_next_value_tl {#1}
\end{verbatim}
\end{verbatim}
\begin{verbatim}
\msg_error:nmn { color } \{ unknown-color \} \{#2}
\__color_extract:nNN { black } \l__color_model_tl \l__color_value_tl
\__color_parse_break:w
\end{verbatim}
\end{verbatim}

The gray model needs special handling: the models need to be swapped: we do that using a dedicated function.

\begin{verbatim}
\cs_new_protected:Npn \__color_parse_gray:n #1
\begin{verbatim}
\tl_set_eq:NN \l__color_model_tl \l__color_next_model_tl
\l__color_next_value_tl \l__color_value_tl {#1}
\end{verbatim}
\end{verbatim}

1420
\tl_set:Nn \l__color_next_model_tl { gray }
\exp_args:NnV \__color_convert:nnN { gray } \l__color_model_tl
\prop_get:cVN { l__color_named_ #1 _prop } \l__color_model_tl
\l__color_next_value_tl
}
\cs_new_protected:Npn \__color_parse_std:n #1

\prop_get:cVNF { l__color_named_ #1 _prop }
\l__color_model_tl
\l__color_next_value_tl
\__color_convert:VVN
\l__color_next_model_tl
\l__color_model_tl
\l__color_next_value_tl
}
\cs_new_protected:Npn \__color_parse_mix:Nnnn #1#2#3#4
\exp_args:Nf \__color_parse_mix:nNnn
{ \fp_eval:n { #4 / 100 } }
#1 {#2} {#3}
\cs_generate_variant:Nn \__color_parse_mix:Nnnn {NVV}
\cs_new:Npn \__color_parse_mix:nNnn #1#2#3#4
\use:c { __color_parse_mix_ #2 :nw } {#1}
#3 \s__color_mark #4 \s__color_stop
\cs_new:Npn \__color_parse_mix_gray:nw #1#2 \s__color_mark #3 \s__color_stop
{ \fp_eval:n { #2 * #1 + #3 * ( 1 - #1 ) } }
\cs_new:Npn \__color_parse_mix_rgb:nw
#1#2 ~ #3 ~ #4 \s__color_mark #5 ~ #6 ~ #7 \s__color_stop
{ \fp_eval:n { #2 * #1 + #5 * ( 1 - #1 ) } \c_space_tl
\fp_eval:n { #3 * #1 + #6 * ( 1 - #1 ) } \c_space_tl
\fp_eval:n { #4 * #1 + #7 * ( 1 - #1 ) } }
\cs_new:Npn \__color_parse_mix_cmyk:nw
#1#2 ~ #3 ~ #4 ~ #5 \s__color_mark #6 ~ #7 ~ #8 ~ #9 \s__color_stop
{ \fp_eval:n { #2 * #1 + #6 * ( 1 - #1 ) } \c_space_tl
\fp_eval:n { #3 * #1 + #7 * ( 1 - #1 ) } \c_space_tl
\fp_eval:n { #4 * #1 + #8 * ( 1 - #1 ) } \c_space_tl
\fp_eval:n { #5 * #1 + #9 * ( 1 - #1 ) } }
(End of definition for \__color_parse:nN and others.)
\_color_parse_model_gray:w
\_color_parse_model_rgb:w
\_color_parse_model_cmyk:w
\_color_parse_number:n
\_color_parse_number:w

Turn the input into internal form, also tidying up the number quickly.

\cs_new:Npn \__color_parse_model_gray:w #1 , #2 \s__color_stop
\cs_new:Npn \__color_parse_model_rgb:w #1 , #2 , #3 , #4 \s__color_stop
\cs_new:Npn \__color_parse_model_cmyk:w #1 , #2 , #3 , #4 , #5 \s__color_stop
\cs_new:Npn \__color_parse_number:n #1
\cs_new:Npn \__color_parse_number:w #1 . #2 . #3 \s__color_stop

\cs_new:Npn \_color_parse_model_Gray:w #1 , #2 \s__color_stop
\cs_new:Npn \_color_parse_model_hsb:w #1 , #2 , #3 , #4 \s__color_stop
\cs_new:Npn \_color_parse_model_Hsb:w #1 , #2 , #3 , #4 \s__color_stop
\cs_new:Npn \_color_parse_model_hsb:nnn #1 #2 #3
\cs_new:Npn \_color_parse_model_hsb_0:nnnn #1 #2 #3 #4
\cs_new:Npn \_color_parse_model_hsb_1:nnnn #1 #2 #3 #4
\cs_new:Npn \_color_parse_model_hsb_2:nnnn #1 #2 #3 #4
\cs_new:Npn \_color_parse_model_hsb_3:nnnn #1 #2 #3 #4
\cs_new:Npn \_color_parse_model_hsb_4:nnnn #1 #2 #3 #4
\cs_new:Npn \_color_parse_model_hsb_5:nnnn #1 #2 #3 #4
\cs_new:Npn \_color_parse_model_wave:w \_color_parse_model_wave_suri:nn \_color_parse_model_wave_auxi:nn \_color_parse_model_wave_rhn:n

The conversion here is non-trivial but is described at length in the xcolor manual. For ease, we calculate the integer and fractional parts of the hue first, then use them to work out the possible values for \( r \), \( g \) and \( b \) before putting them in the correct places.

End of definition for \_color_parse_model_gray:w and others.
Following the description in the xcolor manual. As we always use rgb, there is no need to find the sixth, we just pass the information straight to the hsb auxiliary defined earlier.
\cs_new:Npn \__color_parse_model_wave:w #1 , #2 \s__color_stop
{
  \rgb
{
    \fp_compare:nNnTF {#1} < { 420 }
    { \__color_parse_model_wave_auxi:nn {#1} { 0.3 + 0.7 * (#1 - 380) / 40 } }
    { \fp_compare:nNnTF {#1} > { 700 }
      { \__color_parse_model_wave_auxi:nn {#1} { 0.3 + 0.7 * (#1 - 780) / -80 } }
      { \__color_parse_model_wave_auxi:nn {#1} { 1 } }
    }
  }
\cs_new:Npn \__color_parse_model_wave_auxi:nn #1#2
{
  \fp_compare:nNnTF {#1} < { 440 }
  { \__color_parse_model_wave_auxii:nn
    { 4 + \__color_parse_model_wave_rho:n { (#1 - 440) / -60 } }
    {#2}
  }
  { \fp_compare:nNnTF {#1} < { 490 }
    { \__color_parse_model_wave_auxii:nn
      { 4 - \__color_parse_model_wave_rho:n { (#1 - 440) / 50 } }
      {#2}
    }
    { \fp_compare:nNnTF {#1} < { 510 }
      { \__color_parse_model_wave_auxii:nn
        { 2 + \__color_parse_model_wave_rho:n { (#1 - 510) / -20 } }
        {#2}
      }
      { \fp_compare:nNnTF {#1} < { 580 }
        { \__color_parse_model_wave_auxii:nn
          { 2 - \__color_parse_model_wave_rho:n { (#1 - 510) / 70 } }
          {#2}
        }
        { \fp_compare:nNnTF {#1} < { 645 }
          { \__color_parse_model_wave_auxii:nn
            { \__color_parse_model_wave_rho:n { (#1 - 645) / -65 } }
            {#2}
          }
          { \__color_parse_model_wave_auxii:nn { 0 } {#2} }
        }
      }
    }
  }
}

1424
\_\color\_parse\_model\_wave\_auxii:nn \#1\#2
\{ }
\exp\_args:Neee \_\color\_parse\_model\_hsb\_aux:nnn
\{ \fp\_eval:n \{#1\} \}
\{ 1 \}
\{ \_\color\_parse\_model\_wave\_rho:n \{#2\} \}
\cs\_new:Npn \_\color\_parse\_model\_wave\_rho:n \#1
\{ \fp\_eval:n \{ \min(1, \max(0,\#1)) \} \}

\_\color\_parse\_model\_Gray:w
\_\color\_parse\_model\_cmy:w
\_\color\_parse\_model\_cmy:w #1 , #2 , #3 , #4 \_s\_\color\_stop
\{ \_color\_convert\_rgb\_cmyk:nnn \#1 \#2 \#3 \}

\_\color\_parse\_model\_Hsb:w
\_\color\_parse\_model\_hsb:n
\_\color\_parse\_model\_tHsb:nw
\_\color\_parse\_model\_tHsb:w #1 , #2 , #3 , #4 \_s\_\color\_stop
\{ \exp\_args:Ne \_\color\_parse\_model\_hsb:nnn
\{ \_\color\_parse\_model\_hsb:n \#1 \#2 \#3 \}
\}\cs\_new:Npn \_\color\_parse\_model\_hsb:n \#1
\{ \_\color\_parse\_model\_hsb:nnw \#1 \#2 , \#3 ; \#4 , \#5 ;
\quark\_if\_recursion\_tail\_stop\_do:nn \#4 \{ 0 \}
\fp\_compare:nNnTF \#1 > \#4 \{ \_\color\_parse\_model\_hsb:nnw \#1 \#4 , \#5 ; \}
\use\_i\_delimit\_by\_q\_recursion\_stop:nn

(End of definition for \_\color\_parse\_model\_Gray:w and others.)

Simply pass data to the conversion functions.

There are three stages to the process here: bring the tH argument into the normal range, divide through to get to hsb and finally convert that to rgb. The final stage can be delegated to the parsing function for hsb, and the conversion from Hsb to rgb is trivial, so the main focus here is the first stage. We use a simple expandable loop to do the work, and we implement the equation given in the xcolor manual (number 85 there) as a simple expression.

(End of definition for \_\color\_parse\_model\_cmy:w.)

1425
We cannot extract data here from that passed by xcolor, so we fall back on a black tint.

\cs_new:cpn { __color_parse_model_&spot:w } #1 , #2 \s__color_stop

\cs_new_protected:Npn \__color_check_model:N #1
\tl_set:Nn \l__color_model_tl {#1}
\tl_set:Nn \l__color_value_tl
\cs_new_protected:Npe \__color_finalise_current:
\tl_set:Ne \exp_not:c { l__color_named_ . _tl }
\prop_clear:N \exp_not:c { l__color_named_ . _prop }
\prop_put:NVe \exp_not:c { l__color_named_ . _prop }
\exp_not:c { l__color_named_ . _tl }

93.7 Selecting colors (and color models)

\l_color_fixed_model_tl
For selecting a single fixed model.

\tl_new:N \l_color_fixed_model_tl

\__color_check_model:N
\__color_check_model:nn
Check that the model in use is the one required.

\cs_new_protected:Npm \__color_check_model:N #1
\tl_if_empty:NF \l_color_fixed_model_tl
\tl_if_eq:NNF \l__color_model_tl \l_color_fixed_model_tl
\__color_convert:VVN \l__color_model_tl \l_color_fixed_model_tl
\tl_set:Ne #1
{ \l_color_fixed_model_tl } \{ \l__color_value_tl } }\)
\cs_new_protected:Npm \__color_check_model:nn #1#2
\tl_set:Nn \l__color_model_tl {#1}
\tl_set:Nn \l__color_value_tl {#2}

\__color_finalise_current:
A backend-neutral location for “last minute” manipulations before handing off to the
backend code. We set the special . syntax here: this will therefore always be available.
The finalisation is separate from the main function so it can also be applied to e.g. page
color.
\cs_new_protected:Npe \__color_finalise_current:
\tl_set:Ne \exp_not:c { l__color_named_ . _tl }
\exp_not:c { l__color_named_ . _tl }

Parse the input expressions then get the backend to actually activate them. The main complexity here is the need to check through multiple models. That is done “locally” here as the approach is subtly different to when different models are being stored.

If the first color model is the fixed one, or if there is no fixed model, we don’t need most of the data: just set up and apply the backend function.

If a fixed model applies, we need to check each possible value in order. If there is no hit at all, fall back on the generic formula-based interchange.
\l_color_math_active_tl Tokens representing active sub/superscripts.
\g__color_math_seq Not all engines have multiple color stacks, and at the same time we are not expecting breaking within a colored math fragment. So we track the color stack ourselves.
\color_math:nn The basic set up here is relatively simple: store the current color, parse the new color as-normal, then switch color before inserting the tokens we are asked to change. The tricky part is right at the end, handling the reset.
\color_math:nnn
\__color_math:nn

(End of definition for \color_math:nn and others. These functions are documented on page 324.)

93.8 Math color

The approach here is the same as for the \LaTeX \texttt{\mathcolor} command, but as we are working at the expl3 level we can make some minor changes.
The complication when changing the color back is due to the fact that the \texttt{\color{math:nn}(n)} may be followed by \texttt{^} or \texttt{_} or the hidden superscript (for example \texttt{'}), and its argument may end in a \texttt{\mathop} in which case the sub- and superscripts may be attached as \texttt{limits} instead of after the material. All cases need separate treatment. To avoid repeatedly collecting the same token, we first check for an alignment tab: assuming we don’t have one of those, we can “recycle” \texttt{l_peek_token} safely. As we have an explicit \texttt{\c_{alignment_token}}, there needs to be an align-safe group present.

Dealing with literal \texttt{_} and \texttt{^} is easy, and as we have exactly two cases, we can hard-code this. We use a hard-coded list for limits: these are all primitives. The \texttt{\use{none}} here also removes the test token so it is left just in the right place.

The one final case to handle is math-active tokens, most obviously \texttt{\}'}, as these won’t be covered earlier.
The tricky part of handling sub and superscripts is that we have to reset color to the one that is on the stack but reset it back to what it was before to allow for cases like

\[
\color_math:n { \text{red} } { a + \sum } _ { i = 1 } ^ { n }
\]

Here, \TeX{} constructs a \texttt{vbox} stacking subscript, summation sign, and superscript. So technically the superscript comes first and the \texttt{\sum} that should get colored red is the middle.

The approach here is to set up a brace group immediately after the script token, then to set the color appropriately in that argument. We need an extra group to keep the color contained, and as we need to allow for an explicit closing brace in the source, the inner group also is a brace one rather than \texttt{\group_begin:}-based. At the end of the outer group we need to insert \texttt{\__color_math_scan:w} to continue the search for a second script token.

Notice that here we don’t need to use the math-specific color selector as we can allow the \texttt{\group_insert_after:N} \texttt{\__backend_reset:} to operate normally.

\[
\begin{array}{l}
\color_math:n \{ \text{red} \} \{ a + \sum \} _ { i = 1 } ^ { n } \end{array}
\]
Deal with the case where we do not have an explicit brace pair in the source.

\cs_new_protected:Npn \__color_math_script_aux:N \#1 \{ \#1 \c_group_end_token \}

(End of definition for \__color_math_scripts:Nw and \__color_math_script_aux:N.)

### 93.9 Fill and stroke color

\color_fill:n
\color_stroke:n
\color_fill:nn
\color_stroke:nn
\__color_draw:nnn

\cs_new_protected:Npn \color_fill:n \#1 \{
\__color_parse:nN {#1} \l__color_current_tl
\exp_after:wN \__color_draw:nnn \l__color_current_tl \{ fill \}
\}

\cs_new_protected:Npn \color_stroke:n \#1 \{
\__color_parse:nN {#1} \l__color_current_tl
\exp_after:wN \__color_draw:nnn \l__color_current_tl \{ stroke \}
\}

\cs_new_protected:Npn \color_fill:nn \#1\#2 \{
\__color_select_main:Nw \l__color_current_tl
\#1 / / \s__color_mark \#2 / / \s__color_stop
\exp_after:wN \__color_draw:nnn \l__color_current_tl \{ fill \}
\}

\cs_new_protected:Npn \color_stroke:nn \#1\#2 \{
\__color_select_main:Nw \l__color_current_tl
\#1 / / \s__color_mark \#2 / / \s__color_stop
\exp_after:wN \__color_draw:nnn \l__color_current_tl \{ stroke \}
\}

\cs_new_protected:Npn \__color_draw:nnn \#1\#2\#3 \{
\use:c { __color_backend_ \#3 _ #1 :n } \{#2\}
\exp_args:Nc \group_insert_after:N \{ __color_backend_ \#3 _ reset: \}
\}

(End of definition for \color_fill:n and others. These functions are documented on page 325.)

### 93.10 Defining named colors

\l__color_named_tl Space to store the detail of the named color.

\tl_new:N \l__color_named_tl

(End of definition for \l__color_named_tl.)

\color_set:nn
\color_set:nnn
\color_set:nnn
\color_set:nn
\color_set:nnn
\__color_set_aux:nnn
\__color_set_colon:nnn
\__color_set_loop:nn
\color_set_eq:nn

Defining named colors means working through the model list and saving both the “main” color and any equivalents in other models. Even if there is only one model, we store a prop as well as a tl, as there could be grouping weirdness, etc. When setting using an expression, we need to avoid any fixed model issues, which is done without a group as in l3keys.

\cs_new_protected:Npn \color_set:nn \#1\#2 \{

When setting an expression-based color, there could be multiple model data available for one or more of the input colors. Where that is true for the first named color in an expression, we re-parse the expression when they are also parameter-based: only cmyk, gray and rgb make any sense here. There is a bit of a performance hit but this should be rare and taking place during set-up.

```latex
\cs_new_protected:Npn \__color_set:nnnv #1#2#3 ! #4 \s__color_stop
\group_begin:
\bool_set_true:N \l__color_ignore_error_bool
\tl_set:cn { l__color_named_ #1 _tl } {##1}
\__color_parse:nN {#2} \l__color_named_tl
\exp_args:NNNV \group_end:
\tl_set:Nn \l__color_named_tl \l__color_internal_tl
\__color_model:N \l__color_named_tl
\__color_values:N \l__color_named_tl
\prop_get:cnNT { l__color_named_ #3 _prop } {##1} \l__color_internal_tl
\prop_if_in:cnF { l__color_named_ #1 _prop } {##1}
\group_begin:
\bool_set_true:N \l__color_ignore_error_bool
\exp_args:NV \group_end:
\prop_put:cee { l__color_named_ #1 _prop }
{ \__color_model:N \l__color_internal_tl }
{ \__color_values:N \l__color_internal_tl }
\tl_set:nn { \l__color_fixed_model_tl } {##1} {##2}
\tl_set:Nn \l__color_fixed_model_tl {##1}
\tl_clear:N \l__color_fixed_model_tl
\__color_set:nn {##2} {##3}
\group_end:
```

```latex
\cs_new_protected:Npn \__color_set:nn #1#2
\group_begin:
\str_if_eq:nnF {#1} { . }
\group_end:
\__color_parse:nN {#2} \l__color_named_tl
\prop_clear_new:c { l__color_named_ #1 _prop }
\prop_put:cve { l__color_named_ #1 _prop } { l__color_named_ #1 _tl }
{ \__color_model:N \l__color_named_tl }
\prop_clear_new:c { l__color_named_ #1 _prop }
\prop_put:cee { l__color_named_ #1 _prop }
{ \__color_model:N \l__color_named_tl }
\__color_set:nnw {##1} {##2} ! \s__color_stop
```

```latex
\cs_new_protected:Npn \color_set:nnn #1#2#3
\exp_args:NNNV \__color_set:nnn #1#2#3
\tl_clear:N \l__color_fixed_model_tl
\__color_set:nn {##2} {##3}
\tl_set:Nn \l__color_fixed_model_tl {##1}
```

```latex
\__color_set:nnn \l__color_fixed_model_tl {##1} {##2} {##3}
```

When setting an expression-based color, there could be multiple model data available for one or more of the input colors. Where that is true for the first named color in an expression, we re-parse the expression when they are also parameter-based: only cmyk, gray and rgb make any sense here. There is a bit of a performance hit but this should be rare and taking place during set-up.
\tl_clear_new:c { l__color_named_ #1 _tl }
\prop_clear_new:c { l__color_named_ #1 _prop }
\exp_args:Ne \__color_set_aux:nnn { \tl_to_str:n {#2} }
  {#1} {#3}
\endinput
A small set of colors are always defined.

\begin{verbatim}
\color_set:nnn { black } { gray } { 0 }
\color_set:nnn { white } { gray } { 1 }
\color_set:nnn { cyan } { cmyk } { 1 , 0 , 0 , 0 }
\color_set:nnn { magenta } { cmyk } { 0 , 1 , 0 , 0 }
\color_set:nnn { yellow } { cmyk } { 0 , 0 , 1 , 0 }
\color_set:nnn { red } { rgb } { 1 , 0 , 0 }
\color_set:nnn { green } { rgb } { 0 , 1 , 0 }
\color_set:nnn { blue } { rgb } { 0 , 0 , 1 }
\end{verbatim}

A special named color: this is always defined though not fixed in definition.

\begin{verbatim}
\prop_new:c { l__color_named__.prop }
\tl_new:c { l__color_named__.tl }
\tl_set:ce { l__color_named__.tl } { \__color_model:N \l__color_current_tl }
\end{verbatim}

\subsection{Exporting colors}

\begin{verbatim}
\cs_new_protected:Npn \color_export:nnN #1#2#3
{ \group_begin:
\tl_if_exist:cT { c__color_export_ #2 _tl }
{ \tl_set_eq:Nc \l_color_fixed_model_tl { c__color_export_ #2 _tl } }
\__color_parse:nN {#1} #3
\__color_export:nN {#2} #3
\exp_args:NNNV \group_end:
\tl_set:Nn #3 #3 }
\cs_new_protected:Npn \color_export:nnnN #1#2#3#4
{ \group_begin:
\tl_if_exist:cT { c__color_export_ #2 _tl }
{ \tl_set_eq:Nc \l_color_fixed_model_tl { c__color_export_ #2 _tl } }
\__color_parse:nN \__color_parse:nN {#1} #3
\__color_parse:nN \__color_parse:nN {#4} #3
\exp_args:NNNV \group_end:
\tl_set:Nn #3 #3 #4 }
\cs_new_protected:Npn \color_export:nnnN #1#2#3#4
{ \group_begin:
\tl_if_exist:cT { c__color_export_ #2 _tl }
{ \tl_set_eq:Nc \l_color_fixed_model_tl { c__color_export_ #2 _tl } }
\__color_parse:nN \__color_parse:nN {#1} #3
\__color_parse:nN \__color_parse:nN {#4} #3
\exp_args:NNNV \group_end:
\tl_set:Nn #3 #3 #4 #4 }
\cs_if_exist_use:cF { __color_export_format_ #3 :nnN }
{ \msg_error:nnn { color } { unknown-export-format } {#3}
\use_none:nnn }
{#1} {#2} {#4 }
\end{verbatim}

(End of definition for $\color_set:mm$ and others. These functions are documented on page 324.)

93.11 Exporting colors

\begin{verbatim}
\color_export:nnN \color_export:nnnN \__color_export:nnN \__color_export:nnnN
\cs_new_protected:Npn \color_export:nnN #1#2#3
{ \group_begin:
\tl_if_exist:cT { c__color_export_ #2 _tl }
{ \tl_set_eq:Nc \l_color_fixed_model_tl { c__color_export_ #2 _tl } }
\__color_parse:nN {#1} #3
\__color_export:nN {#2} #3
\exp_args:NNNV \group_end:
\tl_set:Nn #3 #3 }
\cs_new_protected:Npn \color_export:nnnN #1#2#3#4
{ \group_begin:
\tl_if_exist:cT { c__color_export_ #2 _tl }
{ \tl_set_eq:Nc \l_color_fixed_model_tl { c__color_export_ #2 _tl } }
\__color_parse:nN \__color_parse:nN {#1} #3
\__color_parse:nN \__color_parse:nN {#4} #3
\exp_args:NNNV \group_end:
\tl_set:Nn #3 #3 #4 }
\cs_new_protected:Npn \color_export:nnnN #1#2#3#4
{ \group_begin:
\tl_if_exist:cT { c__color_export_ #2 _tl }
{ \tl_set_eq:Nc \l_color_fixed_model_tl { c__color_export_ #2 _tl } }
\__color_parse:nN \__color_parse:nN {#1} #3
\__color_parse:nN \__color_parse:nN {#4} #3
\exp_args:NNNV \group_end:
\tl_set:Nn #3 #3 #4 #4 }
\cs_if_exist_use:cF { __color_export_format_ #3 :nnN }
{ \msg_error:nnn { color } { unknown-export-format } {#3}
\use_none:nnn }
{#1} {#2} {#4 }
\end{verbatim}
A generic auxiliary for cases where only one model is appropriate.

\cs_new_protected:Npn \__color_export:nnnNN #1#2#3#4#5
\{\str_if_eq:nnTF{#2}{#1}{#5 \#4 \#3 \s__color_stop}
\{\s__color_convert:nnnN{#2}{#1}{#3}{#4}\exp_after:wN\exp_after:wN\exp_after:wN\s__color_stop\}
\}

\tl_const:cn{\c__color_export_comma-sep-cmyk_tl}{cmyk}
\tl_const:cn{\c__color_export_comma-sep-rgb_tl}{rgb}
\tl_const:Nn{\c__color_export_HTML_tl}{rgb}
\tl_const:cn{\c__color_export_space-sep-cmyk_tl}{cmyk}
\tl_const:cn{\c__color_export_space-sep-rgb_tl}{rgb}

\__color_export_format_comma-sep-cmyk:nnN
\__color_export_format_comma-sep-rgb:nnN
\__color_export_format_space-sep-cmyk:nnN
\__color_export_format_space-sep-rgb:nnN
\group_begin:
\cs_set_protected:Npn \__color_tmp:w #1#2
\{\cs_new_protected:cpe{\__color_export_format_#1:nnN}{##1}{##2}\}
\__color_tmp:w{comma-sep-cmyk}{cmyk}
\__color_tmp:w{comma-sep-rgb}{rgb}
\__color_tmp:w{HTML}{rgb}
\__color_tmp:w{space-sep-cmyk}{cmyk}
\__color_tmp:w{space-sep-rgb}{rgb}
\group_end:

(End of definition for \__color_export_comma-sep-cmyk_tl and others.)
93.12 Additional color models
Mapping from names to colorants.
\prop_new:N \g__color_colorants_prop
\prop_gput:Nnn \g__color_colorants_prop { black } { Black }
\prop_gput:Nnn \g__color_colorants_prop { blue } { Blue }
\prop_gput:Nnn \g__color_colorants_prop { cyan } { Cyan }
\prop_gput:Nnn \g__color_colorants_prop { green } { Green }
\prop_gput:Nnn \g__color_colorants_prop { magenta } { Magenta }
\prop_gput:Nnn \g__color_colorants_prop { none } { None }
\prop_gput:Nnn \g__color_colorants_prop { red } { Red }
\prop_gput:Nnn \g__color_colorants_prop { yellow } { Yellow }

Whitepoint data for the CIELAB profiles.
\tl_const:Nn \c__color_model_whitepoint_CIELAB_a_tl { 1.0985 - 1 - 0.3558 }
\tl_const:Nn \c__color_model_whitepoint_CIELAB_b_tl { 0.9807 - 1 - 1.8222 }
\tl_const:Nn \c__color_model_whitepoint_CIELAB_e_tl { 1 - 1 - 1 }
\tl_const:cn { \c__color_model_whitepoint_CIELAB_d50_tl } { 0.9642 - 1 - 0.8251 }
\tl_const:cn { \c__color_model_whitepoint_CIELAB_d55_tl } { 0.9568 - 1 - 0.9214 }
\tl_const:cn { \c__color_model_whitepoint_CIELAB_d65_tl } { 0.9504 - 1 - 1.0888 }
\tl_const:cn { \c__color_model_whitepoint_CIELAB_d75_tl } { 0.9497 - 1 - 1.2261 }

The range for CIELAB color spaces.
\tl_const:Nn \c__color_model_range_CIELAB_tl { 0 - 100 - -128 - 127 - -128 - 127 }

For tracking the alternative model set up for separations, etc.
\clist_map_inline:nn { cyan , magenta , yellow , black }
{ \prop_gput:Nnn \g__color_alternative_model_prop {#1} { cmyk } }
\clist_map_inline:nn { red , green , blue }
{ \prop_gput:Nnn \g__color_alternative_model_prop {#1} { rgb } }

Same for the values: a bit more involved.
\clist_map_inline:nn { cyan , magenta , yellow , black }
{ \prop_gput:Nnn \g__color_alternative_values_prop {#1} { cmyk } }
\clist_map_inline:nn { red , green , blue }
{ \prop_gput:Nnn \g__color_alternative_values_prop {#1} { rgb } }
Set up a new model: in general this has to be handled by a family-dependent function. To avoid some “interesting” questions with casing, we fold the case of the family name. The key–value list should always be present, so we convert it up-front to a prop, then deal with the detail on a per-family basis.

\begin{verbatim}
\cs_new_protected:Npn \color_model_new:nnn #1#2#3
{ \exp_args:Nee \__color_model_new:nnn
{ \tl_to_str:n {#1} }
{ \str_casefold:n {#2} } {#3}
}
\cs_new_protected:Npn \__color_model_new:nnn #1#2#3
{ \cs_if_exist:cTF { __color_parse_model_ #1 :w }
{ \msg_error:nnn { color } { model-already-defined } {#1} }
{ \cs_if_exist:cTF { __color_model_ #2 :n }
{ \prop_set_from_keyval:Nn \l__color_internal_prop {#3}
\use:c { __color_model_ #2 :n } {#1} }
{ \msg_error:nnn { color } { unknown-model-type } {#2} }
}
}
\end{verbatim}

(End of definition for \color_model_new:nnn and \__color_model_new:nnn. This function is documented on page 327.)

A shared auxiliary to do the basics of setting up a new model: reserve a number, create a white-equivalent, set up links to the backend.

\begin{verbatim}
\cs_new_protected:Npn \__color_model_init:nnn #1#2#3
{ \int_gincr:N \g__color_model_int
\clist_map_inline:nn { fill , stroke , select }
{ \cs_new_protected:cpe { __color_backend_ ##1 _ #1 :n } ###1
  { \exp_not:c { __color_backend_ ##1 _ #1 :n } ###1
    { color \int_use:N \g__color_model_int } ###1
  }
}
\cs_new_protected:cpe { __color_model_ #1 _white: }
{ \prop_put:Nnn \exp_not:N \l__color_named_white_prop {#1}
  { \exp_not:n {#3} }
  \exp_not:N \int_compare:nNnF { \tex_currentgrouplevel:D } = 0
  { \group_insert_after:N \exp_not:c { __color_model_ #1 _ white: } }
}
\use:c { __color_model_ #1 _white: }
\cs_generate_variant:Nn \__color_model_init:nnn { nne }
\end{verbatim}

\end{document}
Separations must have a “real” name, which is pretty easy to find.

We have two keys to find at this stage: the alternative space model and linked values.

As each alternative space leads to a different requirement for conversion, and as there are only a small number of choices, we manually split the data and then set up. Notice that mixing tints is really just the same as mixing gray. The white color is special, as it
allows tints to be adjusted without an additional color space. To make sure the data is set for that at all group levels, we need to work on a per-level basis. Within the output, only the set-up needs the “real” name of the colorspace: we use a simple tracking number for general usage as this is a clear namespace without issues of escaping chars.

\cs_new_protected:Npn \__color_model_separation:w
\begin{verbatim}
  \__color_model_init:nnn {#6} { separation } { 0 }
  \cs_new_eq:cN { __color_parse_mix_#6 :nw } \__color_parse_mix_gray:nw
  \cs_new:cN { __color_parse_model_#6 :w } ##1 \__s__color_stop
  \prop_gput:Nnn \g__color_alternative_model_prop {#6} {#8}
  \prop_gput:Nne \g__color_colorants_prop {#6}
  { \str_convert_pdfname:n {#7} }
\end{verbatim}

\cs_new_protected:Npn \__color_model_separation_cmyk:nnnnnn #1#2#3#4#5#6
\begin{verbatim}
  \tl_const:cn { c__color_fallback_ #1 _tl } { cmyk }
  \cs_new:cN { __color_convert_#1_cmyk:w } ##1 \__s__color_stop
  \fp_eval:n {##1 * #3} ~ \fp_eval:n {##1 * #4} ~ \fp_eval:n {##1 * #5} ~ \fp_eval:n {##1 * #6}
  \cs_new:cN { __color_convert_cmyk_ #1 :w } ##1 \__s__color_stop { 1 }
  \prop_gput:Nnn \g__color_alternative_values_prop {#1} { #3 , #4 , #5 , #6 }
  \__color_backend_separation_init:nnnnn {#2} { /DeviceCMYK } { } { 0 ~ 0 ~ 0 ~ 0 } { #3 ~ #4 ~ #5 ~ #6 }
\end{verbatim}

\cs_new_protected:Npn \__color_model_separation_rgb:nnnnnn #1#2#3#4#5#6
\begin{verbatim}
  \tl_const:cn { c__color_fallback_ #1 _tl } { rgb }
  \cs_new:cN { __color_convert_#1_rgb:w } ##1 \__s__color_stop
  \fp_eval:n {##1 * #3} ~ \fp_eval:n {##1 * #4} ~ \fp_eval:n {##1 * #5}
  \cs_new:cN { __color_convert_rgb_ #1 :w } ##1 \__s__color_stop { 1 }
  \prop_gput:Nnn \g__color_alternative_values_prop {#1} { #3 , #4 , #5 }
  \__color_backend_separation_init:nnnnn {#2} { /DeviceRGB } { } { 0 ~ 0 ~ 0 ~ 0 } { #3 ~ #4 ~ #5 ~ #6 }
\end{verbatim}

\cs_new_protected:Npn \__color_model_separation_gray:nnnnnn #1#2#3#4#5#6
\begin{verbatim}
  \tl_const:cn { c__color_fallback_ #1 _tl } { gray }
  \cs_new:cN { __color_convert_#1_gray:w } ##1 \__s__color_stop
  \fp_eval:n {##1 * #3}
  \cs_new:cN { __color_convert_gray_ #1 :w } ##1 \__s__color_stop { 1 }
  \prop_gput:Nnn \g__color_alternative_values_prop {#1} {#3}
  \__color_backend_separation_init:nnnnn {#2} { /DeviceGray } { } { 0 } {#3}
\end{verbatim}
Generic model conversion \textit{via} an alternative intermediate.

Setting up for CIELAB needs a bit more work: there is the illuminant and the need for an appropriate object.

If a CIELAB space is being set up, we need the illuminant, then create the appropriate set up. At present, this doesn't include BlackPoint or Range data, but that may be added later. As CIELAB colors cannot be converted to anything else, we fallback to producing black in the gray colorspace: the user should set up a second model for colors set up this way.

We require a list of component names here: one might call them colorants, but it's convenient to use \TeX{} names instead so we slightly adjust the terminology.
All valid models will have an alternative listed, either hard-coded for the core device ones, or dynamically added for Separations, etc.

We now complete the data we require by first finding out how many colorants there are, then moving on to begin constructing the function required to map to the alternative...
color space.

```
\cs_new_protected:Npn \__color_model_devicen:nnn #1#2#3
\exp_args:Ne \__color_model_devicen:nnnn
\clist_count:n {#2} {#1} {#2} {#3}
}
```

At this stage, we have checked everything is in place, so we can set up the \TeX\ and backend data structures. As for separations, it’s not really possible in general to have a fallback, so we simply provide “black” for each element.

```
\cs_new_protected:Npn \__color_model_devicen:nnnn #1#2#3#4
\__color_model_init:nne {#4} { devicen }
\prg_replicate:nn { #1 - 1 } { ~ 0 }
\cs_if_exist_use:cF { __color_model_devicen_parse_ #1 :nn }
\__color_model_devicen_parse_generic:nn
\__color_model_devicen_init:nnn {#1} {#2} {#3}
\__color_model_devicen_convert:nnne {#4} {#2} {#3}
\prg_replicate:nn { #1 - 1 } { - 1 }
```

For short lists of DeviceN colors, we can use hand-tuned parsing. This lines up with other models, where we allow for up to four components. For larger spaces, rather than limit artificially, we use a somewhat slow approach based on open-ended commas-lists.

```
\cs_new_protected:cpn { __color_model_devicen_parse_1:nn } #1#2
\cs_new:cpn { __color_parse_model_ #1 :w } ##1 , ##2 \s__color_stop
\__color_parse_number:n {##1} ~ \__color_parse_number:n {##2}
\cs_new_eq:cN { __color_parse_mix_ #1 :nw } \__color_parse_mix_gray:nw
```

```
\cs_new_protected:cpn { __color_model_devicen_parse_2:nn } #1#2
\cs_new:cpn { __color_parse_model_ #1 :w } ##1 , ##2 , ##3 \s__color_stop
\__color_parse_number:n {##1} ~ \__color_parse_number:n {##2} ~ \__color_parse_number:n {##3}
\fp_eval:n { ##2 * ##1 + ##4 * ( 1 - ##1 ) } \c_space_tl
\fp_eval:n { ##3 * ##1 + ##5 * ( 1 - ##1 ) }
```

```
\cs_new_protected:cpn { __color_model_devicen_parse_3:nn } #1#2
\cs_new:cpn { __color_parse_model_ #1 :w } ##1 , ##2 , ##3 , ##4 \s__color_stop
\__color_parse_number:n {##1} ~ \__color_parse_number:n {##2}
```

1443
\__color_parse_number:n {##3}  
\cs_new_eq:cN { __color_parse_mix_ #1 :nw } \__color_parse_mix_rgb:nw  
\cs_new_protected:cpn { __color_model_devicen_parse_4:nn } #1#2  
{  
\cs_new:cpn { __color_parse_model_ #1 :w } ##1 , ##2 , ##3 , ##4 , ##5 \s__color_stop  
{ #1  
 {  
 \__color_parse_number:n {##1} ~  
 \__color_parse_number:n {##2} ~  
 \__color_parse_number:n {##3} ~  
 \__color_parse_number:n {##4}  
}  
\cs_new_eq:cN { __color_parse_mix_ #1 :nw } \__color_parse_mix_cmyk:nw  
}  
\cs_new_protected:Npn \__color_model_devicen_parse_generic:nn #1#2  
{  
\cs_new:cpn { __color_parse_model_ #1 :w } ##1 , ##2 \s__color_stop  
{ #1  
 { \__color_model_devicen_parse:nw {#2} ##1 , ##2 , \q_nil , \s__color_stop }  
\cs_new:cpe { __color_parse_mix_ #1 :nw } ##1 ##2 \s__color_mark ##3 \s__color_stop  
{ \exp_not:N \__color_model_devicen_mix:nw {##1}  
##2 \c_space_tl \exp_not:N \q_nil \c_space_tl \exp_not:N \q_nil \c_space_tl \exp_not:N \q_nil \c_space_tl \exp_not:N \s__color_mark  
##3 \c_space_tl \exp_not:N \q_nil \c_space_tl \exp_not:N \q_nil \c_space_tl \exp_not:N \s__color_stop  
}  
\cs_new:Npn \__color_model_devicen_parse:nw #1#2 , #3 \s__color_stop  
{ \int_compare:nNnT {#1} > 0  
{  
 \quark_if_nil:nTF {#2}  
{ \prg_replicate:nn {#1} { 0 - } }  
{  
 \__color_parse_number:n {#2}  
 \int_compare:nNnT {#1} > 1 { - }  
 \exp_args:Nf \__color_model_devicen_parse:nw  
{ \int_eval:n { #1 - 1 } } #3 \s__color_stop  
}  
\exp_not:N \__color_model_devicen_mix:nw {##1}  
##2 \c_space_tl \exp_not:N \q_nil \c_space_tl \exp_not:N \q_nil \c_space_tl \exp_not:N \q_nil \c_space_tl \exp_not:N \s__color_mark  
##3 \c_space_tl \exp_not:N \q_nil \c_space_tl \exp_not:N \q_nil \c_space_tl \exp_not:N \q_nil \c_space_tl \exp_not:N \s__color_stop  
}  
\cs_new:Npn \__color_model_devicen_mix:nw #1#2 ~ #3 \s__color_mark #4 ~ #5 \s__color_stop  
{  
 \fp_eval:n { #2 * #1 + #4 * ( 1 - #1 ) }  
 \quark_if_nil:oF { \tl_head:w #3 \q_stop }  
{  
}  
}  
1444
To construct the tint transformation, we have to use PostScript. The aim is to have the final tint for each device colorant as

$$1 - \prod_{n}(1 - X_n D_{X_n})$$

where $X$ is a DeviceN colorant and $D$ is the amount of device colorant that the DeviceN colorant maps to. At the start of the process, the PostScript stack will contain the $X_n$ values, whilst we have the $D$ values on a per-DeviceN colorant basis. The more convenient approach for us is therefore to take each DeviceN colorant in turn and find the value $1 - X_n D_{X_n}$, multiplying as we go, and finalise with the subtraction. That contrasts to colorspace: it splits the process up by process color, which works better when you have a fixed list of colorants. (colorspace only supports up to 4 DeviceN colors, and only cmyk as the alternative space.) To set this up, we first need to know the number of values in the target color space: this is easily handled as there are a very small range of possibilities. Once we have that information, it’s relatively easy to build the required PostScript using some generic code.

As we always need to split the alternative values into parts, we use a shared auxiliary and only use a minimal difference between code paths. Construction of the tint transformation is as far as possible done using loops, which means there are some inefficiencies for device colors in the DeviceN space: we roll the stack one-at-a-time even if there is a potential shortcut. However, that way there is nothing to special-case. Once this is sorted, we can write the tint transform object, which will remain as the last object until we sort out the final step: the colorant list.
Here we need to set up conversion from the DeviceN space to the alternative at the \TeX\ level. This also means supplying methods for inter-converting to other parameter-based spaces. Essentially the approach is exactly the same as the PostScript, just expressed in \TeX\ terms.
\cs_new:Npn \__color_model_devicen_convert_aux:w \#1, \#2, \#3, \#4, \#5 \s__color_stop 
{
  \#1
\tl_if_blank:nF \#2 
  \{ 
   \#2
\tl_if_blank:nF \#3 
   \{ 
     \#3
\tl_if_blank:nF \#4 \{ \#4 \}
   \}
  \}
}\cs_new:Npn \__color_convert_devicen_cmyk:nnnnw
#1#2#3#4#5 \s__color_mark \#6 \s__color_stop
\{ \__color_convert_devicen_cmyk:nnnnnnnn {#5} \{#1} \{#2} \{#3} \{#4\} \#7 \#6 \s__color_mark \#8 \s__color_stop \}
\cs_new:Npn \__color_convert_devicen_cmyk:nnnnnnnn #1#2#3#4#5#6#7#8#9 
\{ \use:e \{ \exp_not:N \__color_convert_devicen_cmyk_aux:nnnnw #5 \s__color_mark {#1} \{#2\} \{#3\} \{#4\} \#6 \s__color_mark \#7 \s__color_stop \}
\cs_new:Npn \__color_convert_devicen_gray:nw
#1#2 \s__color_mark \#3 \s__color_stop
\{ \__color_convert_devicen_gray:nnn \{#1\} \{#2\} \#3 \s__color_mark \#4 \s__color_stop \}
\cs_new:Npn \__color_convert_devicen_gray:nnn #1#2#3 
\{ \use:e { \exp_not:N \__color_convert_devicen_gray_aux:nnnnw \#3 \s__color_mark {#1} \{#2\} \#4 \s__color_mark \#5 \s__color_stop \}
The signatures in the ICC file header indicating the underlying colorspace. We map it to three values: The number of components, the values corresponding to white, and the range.

\prop_const_from_keyval:Nn \c__color_icc_colorspace_signatures_prop
{ \prop_const_from_keyval:Nn \c__color_icc_colorspace_signatures_prop

\exp_not:N \__color_convert_devicen_rgb:nnnw
{ \fp_eval:n { #2 * (1 - (#1 * #5)) } }
{ \fp_eval:n { #3 * (1 - (#1 * #6)) } }
{ \fp_eval:n { #4 * (1 - (#1 * #7)) } }
}

\end_of_definition_for \__color_model_devicen:n and others.
\c__color_icc_colorspace_signatures_prop
For an ICC profile, we need a file name and a number of components. The file name is processed here so the backend can treat it as a string.

```latex
\cs_new_protected:Npn \__color_model_iccbased:n #1
\prop_get:NnNTF \l__color_internal_prop \file \l__color_internal_tl
\exp_args:NV \__color_model_iccbased:nnn \l__color_internal_tl {#1}
\msg_error:nnn { color } { ICCBased-requires-file } {#1}
```

Here, we can use the same internals as for DeviceN approach as we know the number of components. No conversion is possible, so there is no need to worry about that at all.

```latex
\cs_new_protected:Npn \__color_model_iccbased:nn #1#2
\prop_get:NnTF \l__color_internal_prop \file \l__color_internal_tl
\exp_args:NV \__color_model_iccbased:nnn \l__color_internal_tl {#1} {#2}
\msg_error:nnn { color } { ICCBased-unsupported-colorspace } {#2}
```

(End of definition for \c__color_icc_colorsparse_signatures_prop.)
93.13 Applying profiles

With a limited range of outcomes, this is largely about getting data to the backend.

```latex
\color_profile_apply:nn
\_color_profile_apply:nn
\_color_profile_apply_gray:n
\_color_profile_apply_rgb:n
\_color_profile_apply_cmyk:n
```

```
\cs_new_protected:Npn \color_profile_apply:nn #1#2
{\exp_args:Ne \_color_profile_apply:nn
 \{file_full_name:n \{#1\} \} \{#2\}}
```

```
\cs_new_protected:Npn \__color_profile_apply:nn #1#2
{\cs_if_exist_use:cF { __color_profile_apply_ \tl_to_str:n \{#2\} :n }
 \{\msg_error:nnn { color } { ICC-Device-unknown } {#2}
 \use_none:n \}
 \{#1\}}
```

```
\cs_new_protected:Npn \__color_profile_apply_gray:n #1
{\int_gincr:N \g__color_model_int
 \__color_backend_iccbased_device:nnn {#1} { Gray } { 1 }}
```

```
\cs_new_protected:Npn \__color_profile_apply_rgb:n #1
{\int_gincr:N \g__color_model_int
 \__color_backend_iccbased_device:nnn {#1} { RGB } { 3 }}
```

```
\cs_new_protected:Npn \__color_profile_apply_cmyk:n #1
{\int_gincr:N \g__color_model_int
 \__color_backend_iccbased_device:nnn {#1} { CMYK } { 4 }}
```

(End of definition for \color_profile_apply:nn and others. This function is documented on page 328.)

93.14 Diagnostics

```
\color_show:n
\color_log:n
\_color_show:Nn
\_color_show:n
```

Extract the information about a color and format for the user: the approach is similar to the keys module here.

```
\cs_new_protected:Npn \color_show:n
{\_color_show:Nn \msg_show:nneeee}
```

```
\cs_new_protected:Npn \color_log:n
{\_color_show:Nn \msg_log:nneeee}
```

```
\cs_new_protected:Npn \_color_show:Nn #1#2
{\color_if_exist:nT \{#2\}
 \{\exp_args:Nv \_color_show:n \{ l__color_named_ \_2 \_tl \}
 \prop_map_function:cN
 \{ l__color_named_ \_2 \_prop \}}
```

1451
93.15 Messages

(End of definition for \color_show:n and others. These functions are documented on page 324.)
LaTeX has been asked to apply an ICC profile but the device color space 
'\#1' is unknown.

\msg_new:nnnn { color } { ICCBased-unsupported-colorspace } 
{ ICCBased color space '#1' uses an unsupported data-colorspace. }

\msg_new:nnnn { color } { ICCBased-requires-file } 
{ ICCBased color space '#1' requires a file. }

\msg_new:nnnn { color } { model-already-defined } 
{ Color model '#1' already defined. }

\msg_new:nnnn { color } { separation-alternative-model } 
{ Separation color space '#1' requires an alternative model. }

\msg_new:nnnn { color } { separation-alternative-values } 
{ Separation color space '#1' requires values for the alternative. }

\msg_new:nnnn { color } { separation-requires-name }
LaTeX has been asked to create a separation color space, but no formal name was given with the correct information.

Unhandled color model in LaTeX2e value "#1": falling back on grayscale.

Unknown color model "#1": this has never been defined.

Unknown alternative model "#1": separation color space "#1" require a valid alternative space.

Unknown export format "#1": LaTeX has been asked to export a color in format "#1", but this has never been defined.

Unknown CIELAB illuminant "#1": LaTeX has been asked to create a color space using CIELAB illuminant "#1", but this does not exist.

Unknown model "#1": LaTeX has been asked to use a color model called "#1", but this model is not set up.

Unknown model type "#1": LaTeX has been asked to create a new color model called "#1", but this type of model was never set up.

Unknown color model "#1": this model is not set up.

Unknown color model type "#1": this type of color model was never set up.

LaTeX has been asked to use a color named "#1", but this has never been defined.

LaTeX has been asked to create a separation color space, but no formal name is required.

Unknown color model "#1": LaTeX has been asked to use a color model called "#1", but this model is not set up.
\prop_gput:Nnn \g_msg_module_type_prop { color } { }
\msg_new:nnn { color } { show }
{
\tl_if_empty:nTF {#2}
{ is-undefined. }
{ has-the-properties: #2 }
}
⟨/package⟩
Chapter 94

\texttt{l3pdf implementation}

\texttt{\s__pdf_stop} Internal scan marks.

\texttt{\g__pdf_init_bool} A boolean so we have some chance of avoiding setting things we are not allowed to. As we are potentially early in the format, we have to work a bit harder than ideal.

\texttt{\pdf_uncompress:} Simple to do.

\texttt{\pdf_uncompress:} This function is documented on page 331.
94.2 Objects

For returning object numbers.

Simple to do: all objects create a constant int so it is not a backend-specific name.

Object mappings are tracked in Lua for LuaTeX as this makes retrieving them much easier; as a result, there is a split in approaches. In Lua we store values in a table indexed by name. The Lua function here is set up to deal with both named and indexed objects: fits the Lua idiom well.

(End of definition for \pdf_object_new:n and others. These functions are documented on page 329.)
if index then
  return __pdf_objects_indexed[name][index] or 0
else
  return __pdf_objects_named[name] or 0
end
end

luacmd('__pdf_object_retrieve:n', function()
  local name = scan_string()
  return cprint(12, tostring(object_id(name)))
end, 'global')
ltx.pdf = ltx.pdf or {}
ltx.pdf.object_id = object_id

Whereas in \TeX{} we use integer constants.

\pdf_object_if_exist_p:n
\pdf_object_if_exist:nTF
\pdf_object_new_indexed:nn
\pdf_object_write_indexed:nnnn
\pdf_object_write_indexed:nnne
\pdf_object_ref_indexed:nn
\__kernel_pdf_object_id_indexed:nn

Again we split between the common code and the macro- or Lua-based implementation.
To make life easier for the Lua route, all of the potential expressions are expanded to braced numbers.
Again we split for Lua: the same idea as above but with nested tables. As we’ve arranged above that the \TeX code passes a braced number, we can use tonumber(scan_string()) rather than scan_int() for the index.

The non-Lua approach is to divide the range into blocks, and store in integer arrays that can simulate dynamic assignment.

The end of definition for \texttt{\_pdf\_object\_new:} and others. These functions are documented on page 330.
As we want blocks to start from one, and within the block for the top value to be “in” the block, we do a little bit of manipulation. By shifting down by one, we keep the values “in” the block, then we adjust the block/index number to get back on track.

\c__pdf_object_block_size_int
Sets the block size used for managing indexed objects.

\int_const:Nn \c__pdf_object_block_size_int { 10000 }

(End of definition for \c__pdf_object_block_size_int.)

Common variants.

\cs_generate_variant:Nn \__pdf_object_record:nnN { ne }
\cs_generate_variant:Nn \__pdf_object_retrieve:nn { ne }

(End of definition for \__pdf_object_record:nnN and others.)
94.3 Version

To compare version, we need to split the given value then deal with both major and minor version.

```latex
\texttt{\textbf{\textbackslash pdf\_version\_compare\_p:NN}} \texttt{\textbf{\textbackslash pdf\_version\_compare\:NnTF}} \texttt{\textbf{\textbackslash pdf\_version\_compare\_=:w}} \texttt{\textbf{\textbackslash pdf\_version\_compare\_<:w}} \texttt{\textbf{\textbackslash pdf\_version\_compare\_:w}}
```

```latex
\texttt{\textbf{\textbackslash prg\_new\_conditional:NNnn} \textbf{\textbackslash pdf\_version\_compare:NN} \#1\#2 \{ p, T, F, TF \}}
\texttt{\textbf{\textbackslash use:c \{ \textbf{\textbackslash pdf\_version\_compare\_=} w \} \#2 . \text{\_\_pdf\_\_stop} \}}
\texttt{\textbf{\textbackslash cs\_new\_cpn \{ \textbf{\textbackslash pdf\_version\_compare\_=} w \} \#1 . \#2 . \#3 \text{\_\_pdf\_\_stop} \}}

\texttt{\{ \textbf{\textbackslash bool\_lazy\_and:nnTF} \}}
\texttt{\{ \textbf{\textbackslash int\_compare\_p:NNn} \textbf{\textbackslash pdf\_backend\_version\_major:} = \{\#1\} \}}
\texttt{\{ \textbf{\textbackslash int\_compare\_p:NNn} \textbf{\textbackslash pdf\_backend\_version\_minor:} = \{\#2\} \}}
\texttt{\{ \textbf{\textbackslash prg\_return\_true:} \}}
\texttt{\{ \textbf{\textbackslash prg\_return\_false:} \}}

\texttt{\{ \textbf{\textbackslash bool\_lazy\_or:mmTF} \}}
\texttt{\{ \textbf{\textbackslash int\_compare\_p:NNn} \textbf{\textbackslash pdf\_backend\_version\_major:} < \{\#1\} \}}
\texttt{\{ \textbf{\textbackslash bool\_lazy\_and:p:nn} \}}
\texttt{\{ \textbf{\textbackslash int\_compare\_p:NNn} \textbf{\textbackslash pdf\_backend\_version\_major:} = \{\#1\} \}}
\texttt{\{ \textbf{\textbackslash int\_compare\_p:NNn} \textbf{\textbackslash pdf\_backend\_version\_minor:} < \{\#2\} \}}
\texttt{\{ \textbf{\textbackslash prg\_return\_true:} \}}
\texttt{\{ \textbf{\textbackslash prg\_return\_false:} \}}

\texttt{\{ \textbf{\textbackslash bool\_lazy\_or:mmTF} \}}
\texttt{\{ \textbf{\textbackslash int\_compare\_p:NNn} \textbf{\textbackslash pdf\_backend\_version\_major:} > \{\#1\} \}}
```
\bool_lazy_and_p:nn
\int_compare_p:nNn {__pdf_backend_version_major:} = {#1}
\int_compare_p:nNn {__pdf_backend_version_minor:} > {#2}
\prg_return_true:
\prg_return_false:

(End of definition for \pdf_version_compare:NnTF and others. This function is documented on page 331.)

\pdf_version_gset:n
\pdf_version_min_gset:n
\__pdf_version_gset:w
\cs_new_protected:Npn \pdf_version_gset:n #1
\cs_new_protected:Npn \pdf_version_min_gset:n #1
{ \pdf_version_compare:NnT < {#1} { \__pdf_version_gset:w #1 . . \s__pdf_stop } }
\cs_new_protected:Npn \__pdf_version_gset:w #1 . #2 . #3\s__pdf_stop
\bool_if:NF \g__pdf_init_bool
{ \__pdf_backend_version_major_gset:n {#1} \__pdf_backend_version_minor_gset:n {#2} }

(End of definition for \pdf_version_gset:n, \pdf_version_min_gset:n, and \__pdf_version_gset:w. These functions are documented on page 331.)

\pdf_version:
\pdf_version_major:
\pdf_version_minor:
\cs_new:Npn \pdf_version:
\cs_new:Npn \pdf_version_major:
\cs_new:Npn \pdf_version_minor:
{ \__pdf_backend_version_major: . \__pdf_backend_version_minor: }

(End of definition for \pdf_version:, \pdf_version_major:, and \pdf_version_minor:. These functions are documented on page 331.)

94.4 Page size

\pdf_pagesize_gset:nn
\cs_new_protected:Npn \pdf_pagesize_gset:nn #1#2
{ \__pdf_backend_pagesize_gset:nn {#1} {#2} }

(End of definition for \pdf_pagesize_gset:nn. This function is documented on page 331.)
94.5 Destinations

\pdf_destination:nn

\cs_new_protected:Npn \pdf_destination:nn #1#2
\{ \__pdf_backend_destination:nn {#1} {#2} \}

(End of definition for \pdf_destination:nn. This function is documented on page 332.)

\pdf_destination:nnnn

\cs_new_protected:Npn \pdf_destination:nnnn #1#2#3#4
\{ \hbox_to_zero:n \{ \__pdf_backend_destination:nnnn {#1} {#2} {#3} {#4} \}
\}

(End of definition for \pdf_destination:nnnn. This function is documented on page 332.)

94.6 PDF Page size (media box)

Everything here is delayed to the start of the document so that the backend will definitely be loaded.

\cs_if_exist:NT \@kernel@before@begindocument
\{ \tl_gput_right:Nn \@kernel@before@begindocument
\{ \bool_lazy_all:nT
\{ \cs_if_exist_p:N \stockheight \}
\{ \cs_if_exist_p:N \stockwidth \}
\{ \cs_if_exist_p:N \IfDocumentMetadataTF \}
\{ \IfDocumentMetadataTF \{ \c_true_bool \} \{ \c_false_bool \} \}
\{ \int_compare_p:nNn \tex_mag:D = \{ 1000 \} \}
\}
\}
\}
\bool_lazy_and:nnTF
\{ \dim_compare_p:nNn \stockheight > \{ \c_opt \} \}
\{ \dim_compare_p:nNn \stockwidth > \{ \c_opt \} \}
\{ \__pdf_backend_pagesize_gset:nn \stockwidth \stockheight \}
\}
\}
\bool_lazy_or:nnF
\{ \dim_compare_p:nNn \stockheight < \{ \c_opt \} \}
\{ \dim_compare_p:nNn \stockwidth < \{ \c_opt \} \}
\{ \bool_lazy_and:nnT
\{ \dim_compare_p:nNn \paperheight > \{ \c_opt \} \}
\{ \dim_compare_p:nNn \paperwidth > \{ \c_opt \} \}
\{ \__pdf_backend_pagesize_gset:nn \paperwidth \paperheight \}
\}

1463
}
}  
}  
}  
}  
}  
{/tex}  
{/package}
Chapter 95

13deprecation implementation

95.1 Patching definitions to deprecate

\__kernel_patch_deprecation:nnNNpn \{ (date) \} \{ (replacement) \} \{ (definition) \}
\{ (function) \} \{ (parameters) \} \{ (code) \}

defines the \{ (function) \} to produce an error and run its \{ (code) \}.

We make \debug_on:n { deprecation } turn the \{ (function) \} into an \outer error, and \debug_off:n { deprecation } restore whatever the behaviour was without \debug_on:n { deprecation }.

In the explanations below, \{ (definition) \} \{ (function) \} \{ (parameters) \} \{ (code) \} or assignments that only differ in the scope of the \{ (definition) \} will be called “the standard definition”.

\__kernel_patch_deprecation:nnNNpn \__deprecation_patch_aux:nnNNnn \__deprecation_warn_once:nnNnn \__deprecation_just_error:nnNN

(The parameter text is grabbed using \#5\#.) The arguments of \__kernel_deprecation_code:nn are run upon \debug_on:n { deprecation } and \debug_off:n { deprecation }, respectively. In both scenarios we the \{ (function) \} may be \outer so we undefine it with \text{ tex\_let:D } before redefining it, with \__kernel_deprecation_error:Nnn or with some code added shortly.

\cs_new_protected:Npn \__kernel_patch_deprecation:nnNNpn \#1\#2\#3\#4\#5\#6
\cs_new_protected:Npn \__deprecation_patch_aux:nnNNnn \#1\#2\#3\#4\#5\#6
\cs_new_protected:Npn \__deprecation_warn_once:nnNnn \#1\#2\#3\#4\#5\#6
\cs_new_protected:Npn \__deprecation_just_error:nnNN
\cs_if_eq:NNTF \#3 \cs_gset_protected:Npn
\__kernel_deprecation_code:nn
\{ \__deprecation_patch_aux:nnNNnn \#1\#2\#3\#4\#5\#6
\{ \scan_stop: \}
\__kernel_deprecation_error:Nnn \#4 \{ \#2 \} \{ \#1 \}
\scan_stop: \}
\{ \scan_stop: \}
\cs_if_eq:NNTF \#3 \cs_gset_protected:Npn
\{ \__deprecation_warn_once:nnNnn \#1\#2\#3\#4\#5\#6 \}
\{ \__deprecation_patch_aux:nn \#3 \{ \#4 \} \{ \#6 \} \}
\scan_stop: \}

1465
In case we want a warning, the \texttt{function} is defined to produce such a warning without grabbing any argument, then redefine itself to the standard definition that the \texttt{function} should have, with arguments, and call that definition. The e-type expansion and \texttt{exp_not:n} avoid needing to double the \texttt{#}, which we could not do anyways. We then deal with the code for \texttt{debug_off:n \{deprecation\}}: presumably someone doing that does not need the warning so we simply do the standard definition.

\begin{verbatim}
\cs_new_protected:Npn \__deprecation_warn_once:nnNnn #1#2#3#4#5
\plural
\__kernel_if_debug:TF
\plural
\exp_not:N \msg_warning:nneee
\plural
\{ \token_to_str:N #3 \}
\plural
\{ \tl_to_str:n \{#2\} \}
\plural
\exp_not:n \cs_gset_protected:Npn #3 \{#5\}
\plural
\exp_not:n \cs_gset_protected:Npn #3 \{#4 \}
\plural
\__kernel_deprecation_code:nn { }
\plural
\{ \cs_set_protected:Npn #3 \{#4 \} \}
\end{verbatim}

In case we want neither warning nor error, the \texttt{function} is given its standard definition. Here \texttt{#1} is \texttt{cs_new:Npn} or \texttt{cs_new_protected:Npn} and \texttt{#2} is \texttt{function} \texttt{parameters} \texttt{(code)}, so \texttt{#1#2} performs the assignment. For \texttt{debug_off:n \{deprecation\}} we want to use the same assignment but with a different scope, hence the \texttt{cs_if_eq:NNTF} test.

\begin{verbatim}
\cs_new_protected:Npn \__deprecation_patch_aux:Nn #1#2
\plural
\exp_not:N \cs_gset_protected:Npn \{ \cs_set_protected:Npn \{#2\} \}
\end{verbatim}

\texttt{End of definition for \__kernel_patch_deprecation:nnNNpn and others.}

The \texttt{outer} definition here ensures the command cannot appear in an argument.

\begin{verbatim}
\__kernel_deprecation_error:Nnn
\plural
\exp_not:N \msg_expandable_error:nmmm
\plural
\exp_not:N \msg_error:nneee
\end{verbatim}

\texttt{End of definition for \__kernel_deprecation_error:nnn and others.}
95.2 Deprecated l3basics functions

\cs_argument_spec:N

For the present, do not deprecate fully as L\TeX{}2ε will need to catch up: one for Fall 2022.

\cs_new:Npn \cs_argument_spec:N { \cs_parameter_spec:N } \par

95.3 Deprecated l3file functions

\iow_shipout_x:Nn \iow_shipout_x:Nx \iow_shipout_x:cn \iow_shipout_x:cx

Previously described as x-type, but the hash behaviour is really e-type. Currently not “live” as we need to have a transition.

\cs_new_protected:Npn \iow_shipout_x:Nn { \iow_shipout_e:Nn } \par

95.4 Deprecated l3keys functions

.\str_set_x:N \str_set_x:c .\str_gset_x:N .\str_gset_x:c

\cs_new_protected:cpn { \c__keys_props_root_str .\str_set_x:N } \par

(End of definition for .\str_set_x:N and .\str_gset_x:c.)

.\tl_set_x:N \tl_set_x:c .\tl_gset_x:N .\tl_gset_x:c

(End of definition for .\tl_set_x:N and .\tl_gset_x:c.)
We need a transition here so for the present this is commented out: only needed for \texttt{latex-lab} code so this should not last for too long.

(End of definition for \texttt{.tl_set_x:N} and \texttt{.tl_gset_x:N}.)

95.5 Deprecated \texttt{l3msg} functions

95.6 Deprecated \texttt{l3pdf} functions
\bool_gset_true:N \g__pdf_init_bool
}\cs_generate_variant:Nn \pdf_object_write:nn { nx }

(End of definition for \pdf_object_new:nn and \pdf_object_write:nn)

## 95.7 Deprecated l3prg functions

\bool_case_true:n
\bool_case_true:nT
\__kernel_patch_deprecation:nnNNpn { 2023-05-03 } { \bool_case:n }
\cs_new:Npn \bool_case_true:n { \bool_case:n }
\__kernel_patch_deprecation:nnNNpn { 2023-05-03 } { \bool_case:nT }
\cs_new:Npn \bool_case_true:nT { \bool_case:nT }
\__kernel_patch_deprecation:nnNNpn { 2023-05-03 } { \bool_case:nF }
\cs_new:Npn \bool_case_true:nF { \bool_case:nF }
\__kernel_patch_deprecation:nnNNpn { 2023-05-03 } { \bool_case:nTF }
\cs_new:Npn \bool_case_true:nTF { \bool_case:nTF }

(End of definition for \bool_case_true:nTF.)

## 95.8 Deprecated l3str functions

\str_lower_case:n\str_lower_case:f\str_upper_case:n\str_upper_case:f\str_fold_case:n\str_fold_case:V
\__kernel_patch_deprecation:nnNNpn { 2020-01-03 } { \str_lowercase:n }
\cs_new:Npn \str_lower_case:n { \str_lowercase:n }
\__kernel_patch_deprecation:nnNNpn { 2020-01-03 } { \str_lowercase:f }
\cs_new:Npn \str_lower_case:f { \str_lowercase:f }
\__kernel_patch_deprecation:nnNNpn { 2020-01-03 } { \str_uppercase:n }
\cs_new:Npn \str_upper_case:n { \str_uppercase:n }
\__kernel_patch_deprecation:nnNNpn { 2020-01-03 } { \str_uppercase:f }
\cs_new:Npn \str_upper_case:f { \str_uppercase:f }
\__kernel_patch_deprecation:nnNNpn { 2020-01-03 } { \str_casefold:n }
\cs_new:Npn \str_fold_case:n { \str_casefold:n }
\__kernel_patch_deprecation:nnNNpn { 2020-01-03 } { \str_casefold:V }
\cs_new:Npn \str_fold_case:V { \str_casefold:V }

(End of definition for \str_lower_case:n, \str_upper_case:n, and \str_fold_case:n.)

\str_foldcase:n\str_foldcase:V
\__kernel_patch_deprecation:nnNNpn { 2020-10-17 } { \str_casefold:n }
\cs_new:Npn \str_foldcase:n { \str_casefold:n }
\__kernel_patch_deprecation:nnNNpn { 2020-10-17 } { \str_casefold:V }
\cs_new:Npn \str_foldcase:V { \str_casefold:V }

(End of definition for \str_foldcase:n.)
This command was made internal, with one more argument. There is no easy way to compute a reasonable value for that extra argument so we take a value that is big enough to accommodate all of Unicode.

\__kernel_patch_deprecation:nnNNpn { 2020-08-20 } { }
\cs_new_protected:Npn \str_declare_eight_bit_encoding:nnn #1
{ \__str_declare_eight_bit_encoding:nnnn {#1} { 1114112 } }

(End of definition for \str_declare_eight_bit_encoding:nnn.)

95.9  Deprecated \l3seq functions

\seq_indexed_map_inline:Nn
\seq_indexed_map_function:NN
\__kernel_patch_deprecation:nnNNpn { 2020-06-18 } { \seq_map_indexed_inline:Nn }
\cs_new_protected:Npn \seq_indexed_map_inline:Nn { \seq_map_indexed_inline:Nn }
\__kernel_patch_deprecation:nnNNpn { 2020-06-18 } { \seq_map_indexed_function:NN }
\cs_new:Npn \seq_indexed_map_function:NN { \seq_map_indexed_function:NN }

(End of definition for \seq_indexed_map_inline:Nn and \seq_indexed_map_function:NN.)

\seq_mapthread_function:NNN
\__kernel_patch_deprecation:nnNNpn { 2023-05-10 } { \seq_map_pairwise_function:NNN }
\cs_new:Npn \seq_mapthread_function:NNN { \seq_map_pairwise_function:NNN }

(End of definition for \seq_mapthread_function:NNN.)

\seq_set_map_x:NNn
\seq_gset_map_x:NNn
\__kernel_patch_deprecation:nnNNpn { 2023-10-26 } { \seq_set_map_e:NNn }
\cs_new_protected:Npn \seq_set_map_x:NNn { \seq_set_map_e:NNn }
\__kernel_patch_deprecation:nnNNpn { 2023-10-26 } { \seq_gset_map_e:NNn }
\cs_new_protected:Npn \seq_gset_map_x:NNn { \seq_gset_map_e:NNn }

(End of definition for \seq_set_map_x:NNn and \seq_gset_map_x:NNn.)

95.10  Deprecated \l3sys functions

\sys_load_deprecation:
\__kernel_patch_deprecation:nnNNpn { 2021-01-11 } { (no-longer-required) }
\cs_new_protected:Npn \sys_load_deprecation: { }

(End of definition for \sys_load_deprecation.)
95.11 Deprecated \texttt{l3text} functions

\__kernel_patch_deprecation:nnNNpn { 2023-07-08 } \{ \texttt{text\_titlecase\_first:n} \}
\cs_new:Npn \texttt{text\_titlecase\_first:n} #1
\{ \texttt{text\_titlecase\_first:n} { \texttt{text\_lowercase:n} {#1} } \}
\__kernel_patch_deprecation:nnNNpn { 2023-07-08 } \{ \texttt{text\_titlecase\_first:nn} \}
\cs_new:Npn \texttt{text\_titlecase\_first:nn} #1#2
\{ \texttt{text\_titlecase\_first:nn} {#1} { \texttt{text\_lowercase:n} {#2} } \}

(End of definition for \texttt{text\_titlecase\_n} and \texttt{text\_titlecase\_nn}.)

95.12 Deprecated \texttt{l3tl} functions

\__kernel_patch_deprecation:nnNNpn { 2020-01-03 } \{ \texttt{\tl\_lowercase\_n} \}
\cs_new:Npn \texttt{\tl\_lowercase\_n} #1
\{ \texttt{\tl\_lowercase\_n} {#1} \}
\__kernel_patch_deprecation:nnNNpn { 2020-01-03 } \{ \texttt{\tl\_lowercase\_nn} \}
\cs_new:Npn \texttt{\tl\_lowercase\_nn} #1#2
\{ \texttt{\tl\_lowercase\_nn} {#1} {#2} \}
\__kernel_patch_deprecation:nnNNpn { 2020-01-03 } \{ \texttt{\tl\_uppercase\_n} \}
\cs_new:Npn \texttt{\tl\_uppercase\_n} #1
\{ \texttt{\tl\_uppercase\_n} {#1} \}
\__kernel_patch_deprecation:nnNNpn { 2020-01-03 } \{ \texttt{\tl\_uppercase\_nn} \}
\cs_new:Npn \texttt{\tl\_uppercase\_nn} #1#2
\{ \texttt{\tl\_uppercase\_nn} {#1} {#2} \}
\__kernel_patch_deprecation:nnNNpn { 2020-01-03 } \{ \texttt{\tl\_mixed\_case\_n} \}
\cs_new:Npn \texttt{\tl\_mixed\_case\_n} #1
\{ \texttt{\tl\_mixed\_case\_n} {#1} \}
\__kernel_patch_deprecation:nnNNpn { 2020-01-03 } \{ \texttt{\tl\_mixed\_case\_nn} \}
\cs_new:Npn \texttt{\tl\_mixed\_case\_nn} #1#2
\{ \texttt{\tl\_mixed\_case\_nn} {#1} {#2} \}

(End of definition for \texttt{\tl\_lowercase\_n} and others.)

\__kernel_patch_deprecation:nnNNpn { 2022-05-23 } \{ \texttt{\tl\_case\_n} \}
\cs_new:Npn \texttt{\tl\_case\_n} { \texttt{\tl\_case\_n} }
\__kernel_patch_deprecation:nnNNpn { 2022-05-23 } \{ \texttt{\tl\_case\_n} \}
\cs_new:Npn \texttt{\tl\_case\_n} { \texttt{\tl\_case\_n} }
\__kernel_patch_deprecation:nnNNpn { 2022-05-23 } \{ \texttt{\tl\_case\_n} \}
\cs_new:Npn \texttt{\tl\_case\_n} { \texttt{\tl\_case\_n} }
\__kernel_patch_deprecation:nnNNpn { 2022-05-23 } \{ \texttt{\tl\_case\_n} \}
\cs_new:Npn \texttt{\tl\_case\_n} { \texttt{\tl\_case\_n} }
\__kernel_patch_deprecation:nnNNpn { 2022-05-23 } \{ \texttt{\tl\_case\_n} \}
\cs_new:Npn \texttt{\tl\_case\_n} { \texttt{\tl\_case\_n} }
\cs_generate_variant:Nn \tl\_case\_n { c }
\prg_generate_conditional_variant:Nnn \tl\_case\_n { c } { T , F , TF }

(End of definition for \texttt{\tl\_case\_n}.)
\_\_kernel\_\_patch\_\_deprecation:nnNNpn \{ 2023-10-18 \} \{ \tl_build\_begin:N \}
\cs\_\_new\_\_protected:Npn \tl_build\_clear:N \{ \tl_build\_begin:N \}
\_\_kernel\_\_patch\_\_deprecation:nnNNpn \{ 2023-10-18 \} \{ \tl_build\_gbegin:N \}
\cs\_\_new\_\_protected:Npn \tl_build\_gclear:N \{ \tl_build\_gbegin:N \}

(End of definition for \tl_build\_clear:N and \tl_build\_gclear:N.)

\_\_kernel\_\_patch\_\_deprecation:nnNNpn \{ 2023-10-25 \} \{ \tl_build\_get\_intermediate:NN \}
\cs\_\_new\_\_protected:Npn \tl_build\_get:NN \{ \tl_build\_get\_intermediate:NN \}

(End of definition for \tl_build\_get:NN.)

95.13 Deprecated l3token functions

\char\_\_to\_utfviii\_\_bytes:n
\_\_kernel\_\_patch\_\_deprecation:nnNNpn \{ 2022-10-09 \} \{ [ \codepoint\_\_generate:nn ] \}
\cs\_\_new:Npn \char\_\_to\_utfviii\_\_bytes:n \{ \_\_kernel\_\_codepoint\_\_to\_\_bytes:n \}

(End of definition for \char\_\_to\_utfviii\_\_bytes:n.)

\char\_\_to\_nfd:N
\char\_\_to\_nfd:n
\_\_kernel\_\_patch\_\_deprecation:nnNNpn \{ 2022-10-09 \} \{ \codepoint\_\_to\_nfd:n \}
\cs\_\_new:Npn \char\_\_to\_nfd:N \#1 \{ \codepoint\_\_to\_nfd:n \#1 \}
\_\_kernel\_\_patch\_\_deprecation:nnNNpn \{ 2022-10-09 \} \{ \codepoint\_\_to\_nfd:n \}
\cs\_\_new:Npn \char\_\_to\_nfd:n \{ \codepoint\_\_to\_nfd:n \}

(End of definition for \char\_\_to\_nfd:N and \char\_\_to\_nfd:n.)

\char\_\_lower\_\_case:N
\char\_\_upper\_\_case:N
\char\_\_fold\_\_case:Nn
\char\_\_str\_\_lower\_\_case:N
\char\_\_str\_\_upper\_\_case:N
\char\_\_str\_\_fold\_\_case:Nn
\_\_kernel\_\_patch\_\_deprecation:nnNNpn \{ 2020-01-03 \} \{ \text\_\_lowercase:n \}
\cs\_\_new:Npn \char\_\_lower\_\_case:N \{ \text\_\_lowercase:n \}
\_\_kernel\_\_patch\_\_deprecation:nnNNpn \{ 2020-01-03 \} \{ \text\_\_upppercase:n \}
\cs\_\_new:Npn \char\_\_upper\_\_case:N \{ \text\_\_upppercase:n \}
\_\_kernel\_\_patch\_\_deprecation:nnNNpn \{ 2020-01-03 \} \{ \text\_\_titlecase\_\_first:n \}
\cs\_\_new:Npn \char\_\_mixed\_\_case:N \{ \text\_\_titlecase\_\_first:n \}
\_\_kernel\_\_patch\_\_deprecation:nnNNpn \{ 2020-01-03 \} \{ \str\_\_casefold:n \}
\cs\_\_new:Npn \char\_\_fold\_\_case:N \{ \str\_\_casefold:n \}
\_\_kernel\_\_patch\_\_deprecation:nnNNpn \{ 2020-01-03 \} \{ \str\_\_lowercase:n \}
\cs\_\_new:Npn \char\_\_str\_\_lower\_\_case:N \{ \str\_\_lowercase:n \}
\_\_kernel\_\_patch\_\_deprecation:nnNNpn \{ 2020-01-03 \} \{ \str\_\_upppercase:n \}
\cs\_\_new:Npn \char\_\_str\_\_upper\_\_case:N \{ \str\_\_upppercase:n \}
\_\_kernel\_\_patch\_\_deprecation:nnNNpn \{ 2020-01-03 \} \{ \str\_\_titlecase:n \}
\cs\_\_new:Npn \char\_\_str\_\_mixed\_\_case:N \{ \str\_\_titlecase:n \}
\_\_kernel\_\_patch\_\_deprecation:nnNNpn \{ 2020-01-03 \} \{ \str\_\_casefold:n \}
\cs\_\_new:Npn \char\_\_str\_\_fold\_\_case:N \{ \str\_\_casefold:n \}

(End of definition for \char\_\_lower\_\_case:N and others.)
A little extra fun here to deal with the expansion.

\tl_map_inline:nn
{
    \peek_catcode_remove:NTF \peek_charcode_remove:NTF \peek_meaning_remove:NTF
    {\use:e
        {\_kernel_patch_deprecation:nnNNppn { 2022-01-11 } \peek_remove_spaces:n
            {\cs_gset_protected:Npn \exp_not:c\{ peek_ #1 _ignore_spaces:NTF \} ##1##2##3
                \peek_remove_spaces:n
                {\exp_not:c\{ peek_ #1 :NTF \} ##1 {##2} {##3} }
            }
        \_kernel_patch_deprecation:nnNNppn { 2022-01-11 } \peek_remove_spaces:n
            {\cs_gset_protected:Npn \exp_not:c\{ peek_ #1 _ignore_spaces:NT \} ##1##2
                \peek_remove_spaces:n
                {\exp_not:c\{ peek_ #1 :NT \} ##1 {##2} }
            }
        \_kernel_patch_deprecation:nnNNppn { 2022-01-11 } \peek_remove_spaces:n
            {\cs_gset_protected:Npn \exp_not:c\{ peek_ #1 _ignore_spaces:NF \} ##1##2
                \peek_remove_spaces:n
                {\exp_not:c\{ peek_ #1 :NF \} ##1 {##2} }
            }
    }
}\_kernel_patch_deprecation:nnNNppn { 2022-10-17 } \text_lowercase:n
\_kernel_patch_deprecation:nnNNppn { 2022-10-17 } \text_uppercase:n
\_kernel_patch_deprecation:nnNNppn { 2022-10-17 } \text_titlecase_first:n
\_kernel_patch_deprecation:nnNNppn { 2022-10-17 } \str_casefold:n
\_kernel_patch_deprecation:nnNNppn { 2022-10-17 } \str_lowercase:n
\_kernel_patch_deprecation:nnNNppn { 2022-10-17 } \str_titlecase_first:n
\_kernel_patch_deprecation:nnNNppn { 2022-10-17 } \str_uppercase:n
}\_kernel_patch_deprecation:nnNNppn { 2022-10-17 } \text_titlecase_first:n
\_kernel_patch_deprecation:nnNNppn { 2022-10-17 } \str_casefold:n
\_kernel_patch_deprecation:nnNNppn { 2022-10-17 } \str_lowercase:n
\_kernel_patch_deprecation:nnNNppn { 2022-10-17 } \str_titlecase_first:n
\_kernel_patch_deprecation:nnNNppn { 2022-10-17 } \str_uppercase:n
\_kernel_patch_deprecation:nnNNppn { 2022-10-17 } \str_casefold:n
\_kernel_patch_deprecation:nnNNppn { 2022-10-17 } \str_lowercase:n
\_kernel_patch_deprecation:nnNNppn { 2022-10-17 } \str_titlecase_first:n
\_kernel_patch_deprecation:nnNNppn { 2022-10-17 } \str_uppercase:n
\_kernel_patch_deprecation:nnNNppn { 2022-10-17 } \str_casefold:n

(End of definition for \char_lowercase:N and others.)

(End of definition for \peek_catcode_ignore_spaces:NTF and others.)
95.14 Deprecated \l3prop functions

\prop_put_if_new:Nnn
\prop_put_if_new:NVn
\prop_put_if_new:NnV
\prop_put_if_new:cnn
\prop_put_if_new:cVn
\prop_put_if_new:cnV
\prop_gput_if_new:Nnn
\prop_gput_if_new:NVn
\prop_gput_if_new:NnV
\prop_gput_if_new:cnn
\prop_gput_if_new:cVn
\prop_gput_if_new:cnV

(End of definition for \prop_put_if_new:Nnn and \prop_gput_if_new:Nnn.)

(End of definition for \prop_put_if_new:Nnn and \prop_gput_if_new:Nnn.)
Chapter 96

l3debug implementation

Internal kernel functions that are only defined here are listed in l3kernel-functions, see 41.1.

\s__debug_stop

Internal scan marks.

\__debug_use_i_delimit_by_s_stop:nw

Functions to gobble up to a scan mark.

\q__debug_recursion_tail
\q__debug_recursion_stop

Internal quarks.

\__debug_if_recursion_tail_stop:N

Functions to query recursion quarks.

\debug_on:n
\debug_off:n
\__debug_all_on:
\__debug_all_off:

\cs_gset_protected:Npn \debug_on:n #1
{ \exp_args:No \clist_map_inline:nn { \tl_to_str:n {#1} } }
\debug_suspend: \debug_resume: \__debug_suspended:T \l__debug_suspended_tl

Suspend and resume locally all debug-related errors and logging except deprecation errors. The \debug_suspend: and \debug_resume: pairs can be nested. We keep track of nesting in a token list containing a number of periods. At first begin with the “non-suspended” version of \__debug_suspended:T.

(End of definition for \debug_on:n and others. These functions are documented on page 30.)
When debugging is enabled these two functions set up functions that test their argument (when check-declarations is active)

- \__kernel_chk_var_exist:N and \__kernel_chk_cs_exist:N, two functions that test that their argument is defined;
- \__kernel_chk_var_local:N, \__kernel_chk_var_global:N and \__kernel_chk_var_scope:NN that checks that its argument #2 has scope #1.

\begin{verbatim}
\cs_new_protected:Npn \__kernel_chk_var_exist:N #1 { }
\cs_new_protected:Npn \__kernel_chk_cs_exist:N #1 { }
\cs_generate_variant:Nn \__kernel_chk_cs_exist:N { c }
\cs_new:Npn \__kernel_chk_flag_exist:NN { }
\cs_new_protected:Npn \__kernel_chk_var_local:N #1 { }
\cs_new_protected:Npn \__kernel_chk_var_global:N #1 { }
\cs_new_protected:Npn \__kernel_chk_var_scope:NN #1#2 { }
\cs_new_protected:cpn { __debug_check-declarations_on: }
\cs_set_protected:Npn \__kernel_chk_var_exist:N ##1
\__debug_suspended:T \use_none:nnn
\cs_if_exist:NF ##1
{ \msg_error:nne { debug } { non-declared-variable } }
{ \token_to_str:N ##1 }
\cs_set_protected:Npn \__kernel_chk_cs_exist:N ##1
\__debug_suspended:T \use_none:nnn
\cs_if_exist:NF ##1
{ \msg_error:nne { kernel } { command-not-defined } }
{ \token_to_str:N ##1 }
\cs_set:Npn \__kernel_chk_flag_exist:NN ##1##2
\__debug_suspended:T \use_iii:nnnn
\flag_if_exist:NTF ##2
{ ##1 ##2 }
{ \msg_expandable_error:nnn { kernel } { bad-variable } {##2} }
\cs_set_protected:Npn \__kernel_chk_var_scope:NN
\__debug_suspended:T \use_none:nnn
\__debug_chk_var_scope_aux:NN
\cs_set_protected:Npn \__kernel_chk_var_local:N #1
\__debug_suspended:T \use_none:nnn
\cs_set_protected:Npn \__kernel_chk_var_global:N #1
\__debug_suspended:T \use_none:nnn
\cs_set_protected:Npn \__kernel_chk_var_scope:NN
\end{verbatim}
\_debug_suspended:T \use_none:nnnnn
\_kernel_chk_var_exist:N \#1
\_debug_chk_var_scope_aux:NN \#1
}
\cs_set_protected:Npn \_kernel_chk_var_global:N \#1
{
\_debug_suspended:T \use_none:nnnnn
\_kernel_chk_var_exist:N \#1
\_debug_chk_var_scope_aux:NN g \#1
}
\cs_new_protected:cpn { __debug_check-declarations_off: }
{
\cs_set_protected:Npn \_kernel_chk_var_exist:N \#1 { }
\cs_set_protected:Npn \_kernel_chk_cs_exist:N \#1 { }
\cs_set:Npn \_kernel_chk_flag_exist:NN \#1 \#1 { }
\cs_set_protected:Npn \_kernel_chk_var_local:N \#1 { }
\cs_set_protected:Npn \_kernel_chk_var_global:N \#1 { }
\cs_set_protected:Npn \_kernel_chk_var_scope:NN \#1 \#2 { }
}
(End of definition for \_debug_check-declarations_on: and others.)
\_debug_chk_var_scope_aux:NN
\_debug_chk_var_scope_aux:N
\_debug_chk_var_scope_aux:Nn
First check whether the name of the variable \#2 starts with \langle letter\rangle. If it does then pass that letter, the \langle scope\rangle, and the variable name to \_debug_chk_var_scope_-aux:NNn. That function compares the two letters and triggers an error if they differ (the \scan_stop: case is not reachable here). If the second character was not \_ then pass the same data to the same auxiliary, except for its first argument which is now a control sequence. That control sequence is actually a token list (but to avoid triggering the checking code we manipulate it using \cs_set_nopar:Npn) containing a single letter \langle scope\rangle according to what the first assignment to the given variable was.
\cs_new_protected:Npn \_debug_chk_var_scope_aux:NN \#1 \#2
{
\exp_args:NNf \_debug_chk_var_scope_aux:Nn \#1 \#2 \cs_to_str:N \#2 }
\cs_new_protected:Npn \_debug_chk_var_scope_aux:NN \#1 \#2
{
\if:w _ \use_i:nn \_debug_use_i_delimit_by_s_stop:nw \#2 ? ? \s__debug_stop
\exp_after:wN \_debug_chk_var_scope_aux:NNn
\_debug_use_i_delimit_by_s_stop:nw \#2 ? \s__debug_stop
\#1 \#2 \fi:
}
\cs_new_protected:Npn \_debug_chk_var_scope_aux:NNn \#1 \#2 \#3
\if:w \#1 \#2 \#3
\else:
\exp_args:Nc \_debug_chk_var_scope_aux:NNn
{ \_debug_chk_/ \#2 }
\#1 \#2 \fi:
\cs_new_protected:Npn \_debug_chk_var_scope_aux:NNn \#1 \#2 \#3
\if:w \#1 \#2 
\else:
\exp_after:wN \_debug_chk_var_scope_aux:NNn
\_debug_use_i_delimit_by_s_stop:nw \#1 \#2 \s__debug_stop
\#1 \#2 
\else:
\msg_error:nneee { debug } { local-global } \#1 \#2 \{ \iow_char:N \#3 \}
1478
\__debug_log-functions_on: \_\_debug_log-functions_off: \_\_kernel__debug_log:e

These two functions (corresponding to the expl3 option log-functions) control whether \_\_kernel__debug_log:e writes to the log file or not. By default, logging is off.

\cs_new_protected:cpn { \_\_debug_log_functions_on: } { \cs_set_protected:Npn \_\_kernel__debug_log:e { \_\_debug_suspended:T \use_none:nn \iow_log:e } }
\cs_new_protected:cpn { \_\_debug_log_functions_off: } { \cs_set_protected:Npn \_\_kernel__debug_log:e { \use_none:n } }

(End of definition for \_\_debug_log-functions_on:, \_\_debug_log-functions_off:, and \_\_kernel__debug_log:e.)

\__debug_check-expressions_on: \_\_debug_check-expressions_off: \_\_kernel__chk_expr:nNn
\_\_kernel__chk_expr_aux:nNn

When debugging is enabled these two functions set \_\_kernel__chk_expr:nNn to test or not whether the given expression is valid. The idea is to evaluate the expression within a brace group (to catch trailing \use_none:nn or similar), then test that the result is what we expect. This is done by turning it to an integer and hitting that with \tex_romannumeral:D after replacing the first character by -0. If all goes well, that primitive finds a non-positive integer and gives an empty output. If the original expression evaluation stopped early it leaves a trailing \tex_relax:D, which stops the second evaluation (used to convert to integer) before it encounters the final \tex_relax:D. Since \tex_romannumeral:D does not absorb \tex_relax:D the output will be nonempty. Note that #3 is empty except for \mu expressions for which it is \tex_mutoglue:D to avoid an “incompatible glue units” error. Note also that if we had omitted the first \tex_relax:D then for instance 1+2\relax+3 would incorrectly be accepted as a valid integer expression.

\cs_new_protected:cpn { \_\_debug_check_expressions_on: } { \cs_set:Npn \_\_kernel__chk_expr:nNn #1#2#3#4 {#1} }
\cs_new:Npn \_\_kernel__chk_expr:nNn #1#2#3#4 {#1}
\cs_new:Npn \_\_debug_chk_expr_aux:nNn #1#2#3#4
\tl_if_empty:oF { \tex_romannumeral:D - 0 \tex_after:wN \use_none:n }
\tl_if_empty:oF { \relax:D #2 \text_stop: } #2

(End of definition for \_\_debug_check-expressions_on:, \_\_debug_check-expressions_off:, \_\_kernel__chk_expr:nNn, \_\_kernel__chk_expr_aux:nNn, and \_\_debug_chk_var_scope_aux:Nn.)
\int_value:w \ #3 \ #2 \ #1 \ \scan_stop:

{ }

\msg_expandable_error:nnnn
{ debug } \{ expr \} {#4} \{#1\}

\#1

\}

(End of definition for \__debug_check-expressions_on: and others.)

\__debug_deprecation_on:
\__debug_deprecation_off:

Make deprecated commands throw errors if the user requests it. This relies on two token lists, filled up in \l3deprecation by calls to \__kernel_deprecation_code:nn.

\cs_new_protected:Npn \__debug_deprecation_on:
{ \g__debug_deprecation_on_tl }
\cs_new_protected:Npn \__debug_deprecation_off:
{ \g__debug_deprecation_off_tl }

(End of definition for \__debug_deprecation_on: and \__debug_deprecation_off.)

\l__debug_internal_tl
\l__debug_tmpa_tl
\l__debug_tmpb_tl

For patching.

\__debug_generate_parameter_list:NNN
\__debug_build_parm_text:n
\__debug_build_arg_list:n
\__debug_arg_list_from_signature:nNN
\__debug_arg_check_invalid:N
\__debug_parm_terminate:w
\__debug_arg_if_braced:n
\__debug_get_base_form:N
\__debug_arg_return:N

\begin{itemize}
\item Some functions don’t take the arguments their signature indicates. For instance, \clist_concat:NNN doesn’t take (directly) any argument, so patching it with something that uses \#1, \#2, or \#3 results in “Illegal parameter number in definition of \clist_concat:NNN”.
\end{itemize}

Instead of changing the definition of the macros, we’ll create a copy of such macros, say, \__debug_clist_concat:NNN which will be defined as \debug code with \#1, \#2 and \#3<\clist_concat:NNN>.

For that we need to identify the signature of every function and build the appropriate parameter list.

\__debug_generate_parameter_list:NNN takes a function in \#1 and returns two parameter lists: \#2 contains the simple \#1#2#3\#3 as would be used in the (parameter text) of the definition and \#3 contains the same parameters but with braces where necessary.

With the current implementation the resulting \#3 is, for example for \some_function:NnNn, \#1{\#2}{\#3}{\#4}. While this is correct, it might be unnecessary. Bracing everything will usually have the same outcome (unless the function was misused in the first place). What should be done?

\cs_new_protected:Npn \__debug_generate_parameter_list:NNN \#1#2#3
{ \__kernel_tl_set:Nx \l__debug_internal_tl \#1 \l__debug_tmpa_tl \l__debug_tmpb_tl \l__debug_internal_tl \l__debug_tmpa_tl \l__debug_tmpb_tl

(End of definition for \l__debug_internal_tl, \l__debug_tmpa_tl, and \l__debug_tmpb_tl.)

\begin{itemize}
\item Some functions don’t take the arguments their signature indicates. For instance, \clist_concat:NNN doesn’t take (directly) any argument, so patching it with something that uses \#1, \#2, or \#3 results in “Illegal parameter number in definition of \clist_concat:NNN”.
\end{itemize}

Instead of changing the definition of the macros, we’ll create a copy of such macros, say, \__debug_clist_concat:NNN which will be defined as \debug code with \#1, \#2 and \#3<\clist_concat:NNN>.

For that we need to identify the signature of every function and build the appropriate parameter list.

\__debug_generate_parameter_list:NNN takes a function in \#1 and returns two parameter lists: \#2 contains the simple \#1#2#3\#3 as would be used in the (parameter text) of the definition and \#3 contains the same parameters but with braces where necessary.

With the current implementation the resulting \#3 is, for example for \some_function:NnNn, \#1{\#2}{\#3}{\#4}. While this is correct, it might be unnecessary. Bracing everything will usually have the same outcome (unless the function was misused in the first place). What should be done?
Argument types \(w, p, T,\) and \(F\) shouldn’t be included in the parameter lists, so we abort the loop if either is found.

The macro below gets the base form of an argument type given a variant. It serves only to differentiate arguments which should be braced from ones which shouldn’t. If all were to be braced this would be unnecessary. I moved the \(n\) and \(N\) variants to the beginning of the test as they are much more common here.
Simple patching by adding material at the start and end of (a collection of) functions is straightforward as we know the catcode setup. The approach is essentially that in etoolbox. Notice the need to worry about spaces: those are otherwise lost as normally in expl3 code they would be ~.

As discussed above, some functions don’t take arguments, so we can’t patch something that uses an argument in them. For these functions \_\_kernel_patch:nnn is used. It starts by creating a copy of the function (say, clist_concat:NNN) with a \_\_debug_ prefix in the name. This copy won’t be changed. The code redefines the original function to take the exact same arguments as advertised in its signature (see \_\_debuggenerate_parameter_list:NNN above). The redefined function also contains the debug code in the proper position. If a function with the same name and the \_\_debug_ prefix was already defined, then the macro patches that definition by adding more debug code to it.

(End of definition for \_\_debug_generate_parameter_list:NNN and others.)
Some functions, however, won’t work with the signature reading setup above because their signature contains weird arguments. These functions need to be patched using \__kernel_patch_weird:nnn, which won’t make a copy of the function, rather it will patch the debug code directly into it. This means that whatever argument the debug
code uses must be actually used by the patched function.

\cs_set_protected:Npn \__kernel_patch_weird:nnn
\group_begin:
\char_set_catcode_other:N \#
\__kernel_patch_weird_aux:nnn
\group_end:
\cs_set_protected:Npn \__kernel_patch_weird_aux:nnn \#1#2#3
{\char_set_catcode_parameter:N \#
\char_set_catcode_space:N \%
\tex_endlinechar:D -1 \scan_stop:
\tl_map_inline:nn {#3}
{ \__debug_patch_weird:Nnn ##1 {#1} {#2} }
\group_end:
\cs_set_protected:Npn \__debug_patch_weird:Nnn \#1\#2\#3
{\use:e
{\tex_endlinechar:D -1 \scan_stop:
\exp_not:N \tex_scantokens:D
{\tex_global:D \cs_prefix_spec:N \#1
\tex_def:D \exp_not:N \#1
\cs_parameter_spec:N \#1
{\tl_to_str:n {#2}
\cs_replacement_spec:N \#1
\tl_to_str:n {#3}
}
}
}
}

(End of definition for \__kernel_patch:nnn and others.)

Patching the second argument to ensure it exists. This happens before we alter #1 so the ordering is correct. For many variable types such as \texttt{int} a low-level error occurs when #2 is unknown, so adding a check is not needed.

\__kernel_patch:nnn
{ \__kernel_chk_var_exist:N \#2 }
{ }
{\bool_set_eq:NN
\bool_gset_eq:NN
\clist_set_eq:NN
\clist_gset_eq:NN
\fp_set_eq:NN
\fp_gset_eq:NN
\prop_set_eq:NN
\prop_gset_eq:NN
\seq_set_eq:NN
\seq_gset_eq:NN
\str_set_eq:NN
\str_gset_eq:NN

1484
Patching both second and third arguments.

\_\_kernel\_patch:nnn
\{\n\_\_kernel\_chk\_var\_exist:N \#2
\_\_kernel\_chk\_var\_exist:N \#3
\}\n\{\n\clist\_concat:NNN
\clist\_gconcat:NNN
\prop\_concat:NNN
\prop\_gconcat:NNN
\seq\_concat:NNN
\seq\_gconcat:NNN
\str\_concat:NNN
\str\_gconcat:NNN
\tl\_concat:NNN
\tl\_gconcat:NNN
\}
\cs\_gset\_protected:Npn \_\_kernel\_tl\_set:Nx \{ \cs\_set\_nopar:Npe \}
\cs\_gset\_protected:Npn \_\_kernel\_tl\_gset:Nx \{ \cs\_gset\_nopar:Npe \}

Patching where the first argument to a function needs scope-checking: either local or global (so two lists).

\_\_kernel\_patch:nnn
\{\_\_kernel\_chk\_var\_local:N \#1\}
\{\}
\{\n\bool\_set:Nn
\bool\_set\_eq:NN
\bool\_set\_true:N
\bool\_set\_false:N
\box\_set\_eq:NN
\box\_set\_eq\_drop:NN
\box\_set\_to\_last:N
\clist\_clear:N
\clist\_set\_eq:NN
\dim\_zero:N
\dim\_set:Nn
\dim\_set\_eq:NN
\dim\_add:Nn
\dim\_sub:Nn
\fp\_set\_eq:NN
\int\_zero:N
\int\_set\_eq:NN
\int\_add:Nn
\int\_sub:Nn
\int\_incr:N
\int\_decr:N

\"}
\_kernel_patch:nn
{ \_kernel_chk_var_global:N #1 }
{ }
{ }
\bool_gset:Nn
\bool_gset_eq:NN
\bool_gset_true:N
\bool_gset_false:N
\box_gset_eq:NN
\box_gset_eq_drop:NN
\box_gset_to_last:NN
\cctab_gset:NN
\clist_gclear:N
\clist_gset_eq:NN
\dim_gset_eq:NN
\dim_gzero:N
\dim_gset:Nn
\dim_gadd:Nn
\dim_gsub:Nn
\fp_gset_eq:NN
\int_gzero:N
\int_gset_eq:NN
\int_gadd:Nn
\int_gsub:Nn
\int_gincr:Nn
\int_gdecr:Nn
\int_gset:Nn
\hbox_gset:Nn
\hbox_gset_to_wd:Nnn
\hbox_gset:Nw
\hbox_gset_to_wd:Nnw
\muskip_gzero:N
\muskip_gset:Nn
\muskip_gadd:Nn
\muskip_gsub:Nn
\muskip_gset_eq:NN
\prop_gclear:N
\prop_gconcat:NNN
\prop_gpop:NnN
\prop_gpop:NnNT
\prop_gpop:NnNF
\prop_gpop:NnNTF
\prop_gput:Nnn
\prop_gput_if_not_in:Nnn
\prop_gmove:Nn
\prop_gset_eq:NN
\prop_gset_from_keyval:Nn
\seq_gset_eq:NN
\skip_gzero:N
\skip_gset:Nn
\skip_gset_eq:NN
\skip_gadd:Nn
Scoping for constants.

\__kernel_patch:nnn
\{ \__kernel_chk_var_scope:NN c #1 \}
\}
\{
\bool_const:Nn
\cctab_const:Nn
\dim_const:Nn
\int_const:Nn
\intarray_const_from_clist:Nn
\muskip_const:Nn
\prop_const_from_keyval:Nn
\prop_const_linked_from_keyval:Nn
\skip_const:Nn
\str_const:Nn
\tl_const:Nn
\}

Flag functions.

\__kernel_patch:nnn
\{ \__kernel_chk_flag_exist:NN \}
\}
\{
\flag_ensure_raised:N
\flag_height:N
\flag_if_raised:NT
\}
Various one-offs.

Patch various internal commands to log definitions of functions. First, a kernel internal. Then internals from the cs, keys and msg modules.
Variant\token_to_str:N \#4-%
already-defined;- not- changing- it- \msg_line_context: 
}
}
}
{ }\__cs_generate_variant:wwNN 
\langle \@@=keys\rangle
\__kernel_patch:nnn
{
\cs_if_exist:cF { \c__keys_code_root_str #1 }
{ \__kernel_debug_log:e { Defining-key-#1-\msg_line_context: } }
}
{ }\__keys_cmd_set_direct:nn
\langle \@@=msg\rangle
\__kernel_patch:nnn
{
\__kernel_debug_log:e
{ Defining-message- #1 / #2 -\msg_line_context: }
}
{ }\__msg_chk_free:nn
\langle \@@=prg\rangle
Internal functions from prg module.
\__kernel_patch_weird:nnn
{ \__kernel_chk_cs_exist:c { #5 _p : #6 } }
{ }
{ \__prg_set_eq_conditional_p_form:wNnnnn }
\__kernel_patch_weird:nnn
{ \__kernel_chk_cs_exist:c { #5 : #6 TF } }
{ }
{ \__prg_set_eq_conditional_TF_form:wNnnnn }
\__kernel_patch_weird:nnn
{ \__kernel_chk_cs_exist:c { #5 : #6 T } }
{ }
{ \__prg_set_eq_conditional_T_form:wNnnnn }
\__kernel_patch_weird:nnn
{ \__kernel_chk_cs_exist:c { #5 : #6 F } }
{ }
{ \__prg_set_eq_conditional_F_form:wNnnnn }
\langle \@@=regex\rangle
Internal functions from regex module.
\__kernel_patch:nnn
{ \__regex_trace_push:nnN { regex } { 1 } \__regex_escape_use:nnnn \group_begin:
\__kernel_t1_set:Nx \l__regex_internal_a_tl
{ \__regex_trace_pop:nnN { regex } { 1 } \__regex_escape_use:nnnn }
\use_none:nnn
{ }
}
\begin{verbatim}
{ \_regex_escape_use:nnn }
\_kernel_patch:nnn
{ \_regex_trace_push:nnN { regex } { 1 } \_regex_build:N }
{
{ \_regex_trace_states:n { 2 }
\_regex_trace_pop:nnN { regex } { 1 } \_regex_build:N }
}
{ \_regex_build:N }
\_kernel_patch:nnn
{ \_regex_trace_push:nnN { regex } { 1 } \_regex_build_for_cs:n }
{
{ \_regex_trace_states:n { 2 }
\_regex_trace_pop:nnN { regex } { 1 } \_regex_build_for_cs:n }
\_regex_build_for_cs:n
\_kernel_patch:nnn
{ \_regex_trace:nne { regex } { 2 }
{
regex-new-state-
L=\int_use:N \_\_regex_left_state_int \rightarrow-
R=\int_use:N \_\_regex_right_state_int \rightarrow-
M=\int_use:N \_\_regex_max_state_int \rightarrow-
\int_eval:n { \_\_regex_max_state_int + 1 }
}
}
{ \_regex_build_new_state: }
\_kernel_patch:nnn
{ \_regex_trace_push:nnN { regex } { 1 } \_regex_group_aux:nnnnN }
{ \_regex_trace_pop:nnN { regex } { 1 } \_regex_group_aux:nnnnN }
{ \_regex_group_aux:nnnnN }
\_kernel_patch:nnn
{ \_regex_trace_push:nnN { regex } { 1 } \_regex_branch:n }
{ \_regex_trace_pop:nnN { regex } { 1 } \_regex_branch:n }
{ \_regex_branch:n }
\_kernel_patch:nnn
{ \_regex_trace:nne { regex } { 1 } { initializing }
{ }
\_regex_match_init: }
\end{verbatim}
Patching arguments is a bit more involved: we do these one at a time. The basic idea is the same, using a \# token that is a string.

The functions here can get a bit repetitive, so we define a helper which can reuse the same patch code repeatedly. The main part of the patch is the same, so we just have to deal with the part which varies depending on the type of expression.
Patching expandable expressions, first the one-argument versions, then the two-argument ones.
\cs_set_protected:Npn \__kernel_patch_eval:nn #1\#2
\tl_map_inline:nn {#1}
\exp_args:NNe \__kernel_patch:Nn ##1

\__kernel_patch_eval:nn
\__kernel_patch_eval:nn
\__kernel_patch_eval:nn
\__kernel_patch_eval:nn
\__kernel_patch_eval:nn
\__kernel_patch_eval:nn
\__kernel_patch_eval:nn
\__kernel_patch_eval:nn

\__kernel_patch_eval:nn
\__kernel_patch_eval:nn
\__kernel_patch_eval:nn
\__kernel_patch_eval:nn
\__kernel_patch_eval:nn
\__kernel_patch_eval:nn
\__kernel_patch_eval:nn
\__kernel_patch_eval:nn
\__kernel_patch_eval:nn
\__kernel_patch_eval:nn
\__kernel_patch_eval:nn

1493
\exp_not:N \__kernel_chk_expr:nNn \{ \c_hash_str 2 \}
\exp_not:n \{#2\}
\exp_not:N \#1
\}
\}
\}
\}
⟨@@=dim⟩
\__kernel_patch_eval:nn
\{ \dim_max:nn
\dim_min:nn
\}
\{ \__dim_eval:w \{ \} \}
⟨@@=int⟩
\__kernel_patch_eval:nn
\{ \int_max:nn
\int_min:nn
\int_div_truncate:nn
\int_mod:nn
\}
\{ \__int_eval:w \{ \} \}

Conditionals: three argument ones then one argument ones
\cs_set_protected:Npn \__kernel_patch_cond:nn { dim_compare } \#1\#2
\clist_map_inline:nn { :nT , :nF , :nTF , _p:nNn }
\exp_args: Nce \__kernel_patch:Nn \{ \#1 \#1 \}
\}
\}
\}
\exp_not:N \__kernel_chk_expr:nNn \{ \c_hash_str 1 \}
\exp_not:n \{#2\}
\exp_not:c \{ \#1 \#1 \}
\}
\}
\}
\exp_not:N \__kernel_chk_expr:nNn \{ \c_hash_str 2 \}
\exp_not:N \{ \c_hash_str 3 \}
\exp_not:c \{ \#1 \#1 \}
\}
\}
\}
⟨@@=dim⟩
\__kernel_patch_cond:nn \{ dim_compare \} \{ \__dim_eval:w \{ \} \}
⟨@@=int⟩
\__kernel_patch_cond:nn \{ int_compare \} \{ \__int_eval:w \{ \} \}
\cs_set_protected:Npn \__kernel_patch_cond:nn \#1\#2
\clist_map_inline:nn { :nT , :nF , :nTF , _p:nNn }
\exp_args: Nce \__kernel_patch:Nn \{ \#1 \#1 \}
Step functions.

Odds and ends

This one has catcode changes so must be done by hand.
Flip the switch for deprecated code.
\cs_set_protected:Npn \_kernel_if_debug:TF #1#2 (#1)

(End of definition for \_kernel_if_debug:TF.)

⟨/package⟩
Index

The italic numbers denote the pages where the corresponding entry is described, numbers underlined point to the definition, all others indicate the places where it is used.

Symbols

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Pages</th>
</tr>
</thead>
<tbody>
<tr>
<td>!</td>
<td>19942, 19958</td>
</tr>
<tr>
<td>`</td>
<td>19158, 19161, 32057, 34380, 34398, 34422, 34428, 34432, 34438, 34442, 34448, 34454, 34461, 34462, 34468, 34472, 34474, 34583</td>
</tr>
<tr>
<td>#</td>
<td>10674, 14143, 19158, 34317, 39301, 39306, 39363, 39368, 39787, 39792</td>
</tr>
<tr>
<td>$</td>
<td>5260, 14142, 19158, 19161, 34317</td>
</tr>
<tr>
<td>%</td>
<td>9378, 14135, 19158, 19161, 34317, 39306, 39363, 39368, 39787, 39792</td>
</tr>
<tr>
<td>&amp;</td>
<td>32057, 34380, 34392, 34419, 34426, 34430, 34435, 34440, 34443, 34445, 34452, 34457, 34458, 34465, 34470, 34473, 34481, 34482, 34529, 34530, 34537, 34538, 34549, 34550, 34555, 34556, 34584, 34585, 34607, 34608, 34611, 34612, 34613, 34614</td>
</tr>
<tr>
<td>(</td>
<td>31669</td>
</tr>
<tr>
<td>)</td>
<td>31669</td>
</tr>
<tr>
<td>*</td>
<td>9139, 9151, 13735, 13758, 19339, 19341, 19345, 19353</td>
</tr>
<tr>
<td>**</td>
<td>274</td>
</tr>
<tr>
<td>+</td>
<td>274</td>
</tr>
<tr>
<td>,</td>
<td>21471, 34329</td>
</tr>
<tr>
<td>-</td>
<td>150</td>
</tr>
<tr>
<td>.</td>
<td>274</td>
</tr>
<tr>
<td>.</td>
<td>32057, 34380, 34397, 34485, 34486, 34495, 34496, 34505, 34506, 34523, 34535, 34536, 34586, 34587, 34617, 34618, 34621, 34622</td>
</tr>
<tr>
<td>/</td>
<td>149, 4143</td>
</tr>
<tr>
<td></td>
<td>14141</td>
</tr>
<tr>
<td>:</td>
<td>43, 406, 425, 2392, 2393, 2394, 2395, 2396, 2397, 2398, 2400, 2402, 2403, 2404, 2411, 2414, 2417, 2423, 2579, 2581, 2586, 2591, 2593, 2598, 2650, 2651, 2652, 2653, 2654, 2665</td>
</tr>
<tr>
<td>:\</td>
<td>43, 2396, 2654</td>
</tr>
<tr>
<td>:v</td>
<td>43, 2417</td>
</tr>
<tr>
<td>:v_unbraced</td>
<td>43, 2578</td>
</tr>
<tr>
<td>:c</td>
<td>43, 2398</td>
</tr>
</tbody>
</table>
\begin{R}
\end{R}
\belowdisplayshortskip 166
\belowdisplayskip 34291
\binoppenalty \bfseries \belowdisplayskip \beginR
\__bitset_set_false:Nn \__bitset_set:NNnN \__bitset_set:NNn \bitset_to_bin:N \bitset_to_arabic:N \bitset_show:N \bitset_set_true:Nn \bitset_set_false:Nn \bitset_new:N \bitset_log_named_index:N \bitset_log:N \bitset_if_exist_p:N \bitset_if_exist:NTF \bitset_if_exist:N \bitset_gset_true:Nn \bitset_gset_inverse:N \bitset_gset:N \bitset_gclear:N \bitset_addto_named_index:Nn \bitset_gset_eq:NN \bitset_case_true:nTF \bitset_case_true:n \bitset_case:nTF \bitset_const:Nn \bitset_do_while:nn \bitset_do_until:Nn \bitset_do_until:nn \bitset_do_until:n \bitset_for:nn \bitset_for:n \bitset_for:Nn \bitset_if:n \bitset_if:NTF \bitset_if:N \bitset_bodydirection \bitset_bodydir 
\bool_if:NTF \bool_if:n \bool_if:N \bool_gset_true:N \bool_gset_inverse:N \bool_gset:N \bool_do_while:nn \bool_do_until:nn \bool_do_until:n \bool_for:nn \bool_for:n \bool_for:Nn \bool_if:n \bool_if:NTF \bool_if:N

\index
\begin{theindex}
\item \_bitset_addto_named_index:Nn ........... 286, 30180, 30180
\item \_bitset_clear:N ........... 287, 30270, 30270, 30278
\item \_bitset_gclear:N ........... 287, 30270, 30274, 30279
\item \_bitset_gset_false:Nn ........... 287, 30226, 30242, 30269
\item \_bitset_gset_true:N ........... 287, 30226, 30238, 30267
\item \_bitset_if_exist:N ........... 30186, 30188
\item \_bitset_if_exist:NTF ........... 287, 30185
\item \_bitset_if_exist:p:N ........... 287, 30185
\item \_bitset_item:Nn ........... 287, 30300, 30300, 30315
\item \_bitset_log:N ........... 289, 30316, 30318, 30319
\item \_bitset_log_named_index:N ........... 288, 30331, 30334, 30336
\item \_bitset_new:N ........... 286, 30163, 30163, 30178
\item \_bitset_new:Nn ........... 286, 30163, 30169, 30179
\item \_bitset_set_false:Nn ........... 287, 30226, 30240, 30268
\item \_bitset_set_true:N ........... 287, 30226, 30236, 30266
\item \_bitset_show:N ........... 287, 288, 30316, 30316, 30317
\item \_bitset_show_named_index:N ........... 288, 30331, 30334, 30336
\item \_bitset_to_arabic:N ........... 285, 288, 1254, 30280, 30280, 30298, 30297
\item \_bitset_to_bin:N ........... 286, 288, 30280, 30294, 30299, 30326
\item \_bitset_gset_false:Nn ........... 30189, 30195, 30243
\item \_bitset_gset_true:N ........... 30189, 30191, 30239
\item \_bitset_internal_int ........... 30220, 30224, 30228
\item \_bitset_set:NNn ........... 30223, 30229, 30242, 30243, 30244
\item \_bitset_set:NNnN ........... 30189, 30190, 30192, 30194, 30196, 30197
\item \_bitset_set_aux:Nn ........... 30236
\item \_bitset_set_false:N ........... 30189, 30193, 30241
\item \_bitset_set_true:N ........... 30189, 30189, 30237
\item \_bitset_test_digits:n ........... 30211
\item \_bitset_test_digits:NTF ........... 30221, 30223, 30235
\item \_bitset_test_digits_end ........... 30225, 30227, 30234, 30235
\item \_bitset_test_digits_end:n ........... 30221
\item \_bitset_to_int:nN ........... 30280, 30285, 30289, 30292
\item booldir ........... 790
\item bodydirection ........... 791
\item \_bitset_gset_inverse:N ........... 30221
\item \_bitset_test_digits:n ........... 30221
\item \_bitset_test_digits:nTF ........... 30857, 38860, 38862, 38864
\item \_bitset_const:Nn ........... 66, 8283, 8283, 8288, 39597
\item \_bitset_do_until:n ........... 70, 8506, 8508, 8509, 8511
\item \_bitset_do_until:nn ........... 70, 8512, 8533, 8536
\item \_bitset_do_while:N ........... 70, 8506, 8506, 8507, 8510
\item \_bitset_do_while:nn ........... 70, 8512, 8520, 8523
\item \_bitset_gset:N ........... 241, 22015
\item \_bitset_gset:Nn ........... 66, 8305, 8310, 8316, 39516
\item \_bitset_gset_eq:NN ........... 66, 4482, 6648, 8301, 8302, 39398, 39517
\item \_bitset_gset_false:N ........... 66, 6596, 8289, 8295, 8300, 8321, 14351, 14360, 39519
\item \_bitset_gset_inverse:N ........... 241, 22023
\item \_bitset_gset_inverse:Nn ........... 66, 8317, 8320, 8322
\item \_bitset_gset_true:N ........... 66, 6661, 8289, 8293, 8299, 8321, 8854, 14341, 38451, 38472, 38552, 38635, 38853, 39518
\item \_bitset_if:N ........... 8328, 8336
\item \_bitset_if:n ........... 8379
\item \_bitset_if:NTF ........... 67, 108, 2154, 5375, 5384, 5825, 5998, 6084, 6102, 6120, 6271, 6490, 6498, 6730, 7340, 7363, 7436, 7663, 7830, 7836, 7877, 8247, 8252, 8318, 8321, 8328, 8339, 8398, 8501, 8503, 8507, 8509, 8852,
\bool_if:nTF \bool_if_exist:n \bool_if:nTF \bool_if_exist:n \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nTF \bool_if:nFT
bool internal commands:
\__bool_!:Nw .................. 8409
\__bool_&:0 .................. 8421
\__bool_&:1 .................. 8421
\__bool_&:2 .................. 8421
\__bool_(:Nw ................ 8414
\__bool_):0 .................. 8421
\__bool_):1 .................. 8421
\__bool_):2 .................. 8421
\__bool_case:NnTF ............. 8540
\__bool_case:nTF .............. 8541, 8543, 8545, 8547, 8548
\__bool_case:w 8540, 8550, 8553, 8557
\__bool_case_end:nw ......... 8556, 8559
\__bool_choose:NNN ............ 8416, 8420, 8421, 8421
\__bool_get_next:NN .......... 589, 8395, 8399, 8399, 8411, 8417, 8432, 8433, 8434, 8435, 8436, 8437
\__bool_if_p:n ................. 8387, 8387, 8388
\__bool_if_p_aux:w ............. 588, 8387, 8390, 8397
\__bool_if_recursion_tail_stop_-do:nn ........ 8327, 8327, 8450, 8476
\__bool_lazy_all:n ............ 8438, 8439, 8448, 8453
\__bool_lazy_any:n ............ 8464, 8465, 8474, 8479
\__bool_p:Nw ................ 8419
\__bool_show:NN 8352, 8352, 8354, 8356
\__bool_use_i_delimit_by_q_-recursion_stop:nw ................. 8325, 8325, 8452, 8478
\__bool_l:0 .................. 8421
\__bool_l:1 .................. 8421
\__bool_l:2 .................. 8421
\botmark ............................. 169
\botmarks .......................... 475
\boundary .......................... 792
\box .................................. 170
box commands:
\box_autosize_to_wd_and_ht:Nnn ........ 311, 35244, 35244, 35246
\box_autosize_to_wd_and_ht_plus_-dp:Nnn 311, 35244, 35250, 35255
\box_clear:N 302, 303, 34642, 34642, 34646, 34649, 35483, 35570, 35647
\box_clear_new:N ................. 302, 303, 34642, 34648, 34652
\box_dp:N 304, 34649, 34686, 34698, 35118, 35233, 35252, 35258, 35332, 35339, 35344, 35684, 35685, 35795, 35800, 35825, 35842, 36013, 36291, 36312, 36615
\box_gautosize_to_wd_and_ht:Nnn .. 311, 35244, 35247, 35249
\box_gautosize_to_wd_and ht:plus_dp:Nnn 311, 35244, 35256, 35261
\box_gclear:N 302, 303, 34642 34644, 34647, 34651, 35492
\box_gclear_new:N .. 302, 303, 34648, 34650, 34653
\box_gresize_to_ht:NN .......... 311, 35127, 35140, 35142
\box_gresize_to_ht_plus_dp:NN .. 312, 35127, 35160, 35162
\box_gresize_to_wd:NN ........... 312, 35127, 35180, 35182
\box_gresize_to_wd_and_ht:NN ...... 312, 35127, 35197, 35199
\box_gresize_to_wd_and ht_plus_-dp:NN .... 318, 35088, 35094, 35099, 36127
\box_grotate:NN ................. 313, 34970, 34973, 34975, 35062
\box_gscale:NN ................ 313, 35215, 35218, 35220, 36169
\box_gset_clipped:N ........... 313, 35312, 35315, 35317
\box_gset_dp:NN ................ 304, 34679, 34685, 34687
\box_gset_eq:NN ............... 303, 34645, 34654, 34656, 34659, 35322, 35373, 35669, 39520
\box_gset_eq:drop:NN ............ 310, 34660, 34662, 34665, 39521
\box_gset_h:NN .......................... 304, 34679, 34694, 34696
\box_gset_to_last:N .......... 305, 34734, 34735, 34738, 39522
\box_gset_trim:NNNN ...... 313, 35318, 35321, 35323
\box_gset_viewport:NNNN .... 313, 35369, 35372, 35374
\box_gset_wd:NN 304, 34670, 34703, 34705
\box_h:NN ....................... 304, 34721, 34670, 34673, 34677
Index

bp .......................... 280
\breakafterdimode .......................... 795
\brokenpenalty .......................... 172

C
\c ................................... 32057,
34380, 34403, 34424, 34450, 34507,
34508, 34527, 34528, 34531, 34532,
34539, 34540, 34551, 34552, 34559,
34560, 34563, 34564, 34619, 34620
\catcode 66, 85, 86, 88, 89, 90, 91, 92,
96, 97, 98, 99, 100, 101, 102, 103, 173
\catcodedateable ........................... 796
cc .................................. 280

cctab commands:
\cctab_begin:N ........................ 290, 1257, 1258,
1260, 1263–1265, 30519, 30519, 30532
\cctab_const:N ........................ ...
290, 1257, 30525, 30562, 30652,
30662, 30664, 30671, 30715, 30958
\cctab_end: .............................. 290,
1257, 1258, 1261–1265, 30533, 30533
\cctab_gsave_current:N
................................. 289, 30455, 30455, 30460
\cctab_gset:N ............................ 289,
290, 1257, 30443, 30443, 30454, 30523
\cctab_if_exist:N ........................ 30607, 30609
\cctab_if_exist:NTF ........................ 290, 30607, 30614
\cctab_if_exist:Np: ........................ 290, 30607
\cctab_item:N .......................... 290, 30591, 30591, 30606
\cctab_new:N ............................ 289, 1257,
1258, 1267, 30574, 30376, 30396,
30415, 30663, 30726, 30727, 30928
\cctab_select:N .......................... 125, 289,
290, 14424, 30448, 30471, 30471,
30473, 30657, 30666, 30673, 30717
\c_code_cctab .......................... 290, 14424, 30676
\c_document_cctab ........................ 290, 1260,
30676
\c_inited_cctab
................................. ...
291, 30448, 30657, 30663
\c_other_cctab .......................... 291, 30663
\g_tampa_cctab .......................... 291, 30726
\g_temp_cctab .......................... 291, 30726

cctab internal commands:
\g__cctab_allocate_int
............................. 30370, 30512, 30514, 30516
\g__cctab_begin_aux: ........................ ...
1262, 30500, 30502, 30510, 30524
\g__cctab_chk_group_begin:n
............................. 1263, 30525, 30544, 30544, 30550
\g__cctab_chk_group_end:n
............................. 1263, 30538, 30544, 30551
\g__cctab_chk_if_valid:N
............................. 30611
\g__cctab_endlinechar_prop
............................. 1259, 30373, 30424, 30426, 30479
\g__cctab_group_seq
............................. 30369, 30546, 30553
\g__cctab_gset:n ........................ 30416, 30418, 30432,
30450, 30458, 30528, 30569, 30713
\g__cctab_gset_aux:n
............................. 30416, 30419, 30420
\g__cctab_gstore:Nmm
............................. 30374, 30394, 30403, 30404,
30405, 30406, 30408, 30409, 30411, 30412
\l__cctab_internal_a_tl
............................. 1262, 1263, 30371, 30479,
30480, 30505, 30515, 30523, 30526,
30528, 30535, 30537, 30539, 30540
\l__cctab_internal_b_tl
............................. 30371, 30553, 30557, 30564
\l__cctab_internal_cctab
............................. 30461
\l__cctab_internal_cctab_name:
..........................
30461, 30464, 30482, 30483, 30484,
30485
\l__cctab_item:N . 30592, 30595, 30599
\l__cctab_nesting_number:N
............................. 30526, 30539, 30569,
30570, 30572
\l__cctab_nesting_number:w
............................. 30569, 30574, 30579
\l__cctab_new:N 1258, 1263, 30374,
30379, 30381, 30388, 30399, 30463,
30483, 30504, 30513, 30655, 30680
\l__cctab_next_cctab
............................. 30500
\l__cctab_select:N 1261, 30471
\l__cctab_stack_seq
............................. 1257, 30367, 30527,
30535, 30587
\l__cctab_used_seq
............................. 1257, 1262, 1263,
30367, 30523, 30537
cell .................................. 276
\char ................................ 174, 19515
chare commands:
\l__char_active_seq ........................ 90, 200, 19156
\char_fold_case:N ........................ 38940, 38956
\char_foldcase:N ........................ 38965, 38972
\char_generate:nn
............................. 124, 196,
446, 467, 553, 702, 765, 896, 3544,
3545, 3546, 3547, 3549, 3550, 3551,
4108, 4182, 4198, 4210, 4630, 5668,
6042, 12405, 12421, 14265, 14522,
14538, 19183, 19183, 19282, 19958,
30782, 30790, 30809, 30812, 30815,
Index

clist internal commands:
\_\_clist_concat:NNN  ..........  18477, 18478, 18480, 18481
\_\_clist_count:n  .  18853, 18858, 18862
\_\_clist_count:w  ..........  18853, 18870, 18874, 18878
\_\_clist_get:w  ..........  18534, 18539, 18542, 18576
\_\_clist_if_empty:n:w  ........................................  18685, 18687, 18692, 18695
\_\_clist_if_empty:n:wW  ........................................  18685, 18696, 18698
\_\_clist_if_in_return:nnW  ........................................  18699, 18701, 18706, 18709
\_\_clist_if_wrap:n  ........................................  18402
\_\_clist_if_wrap:nTP .  870, 18402, 18427, 18469, 18616, 18630, 18711
\_\_clist_if_wrap:w  ........................................  870, 18402, 18406, 18425
\_\_clist_internal_clist  ........................................  873, 18376, 18510, 18511, 18523, 18524, 18705, 18706, 18707, 18794, 18795, 18805, 18806
\_\_clist_internal_remove_clist  ........................................  18600, 18608, 18611, 18613, 18615, 18620
\_\_clist_internal_remove_seq  ........................................  18600, 18632, 18633, 18634
\_\_clist_item:nnN  ........................................  18949, 18951, 18957, 18972, 18982
\_\_clist_item:nw  18980, 18986, 18989
\_\_clist_item:n_end:n  ........................................  18980, 18997, 19005
\_\_clist_item:n_loop:nnw  ........................................  18949, 18955, 18973, 18977
\_\_clist_item:n_loop:nnw  ........................................  18980, 18990, 18991, 18994, 18999
\_\_clist_item:n_strip:n  ........................................  18980, 19006, 19007
\_\_clist_item:n_strip:w  ........................................  18980, 19007, 19008
\_\_clist_map_function:Nw  ........................................  880, 18736, 18740, 18746, 18751, 18783
\_\_clist_map_function:end:w  ........................................  880, 18736, 18749, 18753, 18757
\_\_clist_map_function:nN  ........................................  881, 18760, 18762, 18768, 18772
\_\_clist_map_tokens:nw  ........................................  18808, 18812, 18818, 18824
\_\_clist_map_tokens:end:w  ........................................  18808, 18821, 18826, 18830
\_\_clist_map_tokens:nw  ........................................  18833, 18835, 18839, 18847

\_\_clist_map_unbrace:wn  ..............  881, 18760, 18771, 18775, 18845, 18933
\_\_clist_map_variable:Nn  ..............  882, 18798, 18799, 18801
\_\_clist_pop:nN  ..............  18545, 18546, 18548, 18549
\_\_clist_pop:wN  .  18545, 18562, 18568
\_\_clist_pop:wwNNN  ..............  875, 18545, 18554, 18557, 18590
\_\_clist_pop_TP:NNN  ..............  18571, 18582, 18584, 18585
\_\_clist_pop_left:NNNn  ..............  18504, 18505, 18507, 18508
\_\_clist_pop_right:NNNn  ..............  18517, 18518, 18520, 18521
\_\_clist_rand_item:nn  ..............  19009, 19010, 19011
\_\_clist_remove_all:  ........................................  18602, 18603, 18605, 18606
\_\_clist_reverse:wwNNw  ..............  875, 18669, 18671, 18672, 18676
\_\_clist_reverse_end:ww  ..............  875, 18669, 18673, 18679
\_\_clist_sanitize:n  ..............  18389, 18399, 18431, 18497, 18499
\_\_clist_sanitize:no  ..............  870, 18389, 18391, 18395, 18399
\_\_clist_set_from_seq:n  ..............  18450, 18462, 18466
\_\_clist_set_from_seq:NNN  ..............  18450, 18451, 18453, 18454
\_\_clist_show:NN  ..............  19022, 19024, 19024, 19026
\_\_clist_show:Nn  ..............  19022, 19024, 19024, 19026
\_\_clist_show:tmpr  ..............  18382, 18382, 18637, 18659, 18713, 18722, 18726, 18728, 18863, 18881
\_\_clist_trim_next:w  ..............  870, 881, 18383, 18383, 18386, 18392, 18400, 18763, 18773
\_\_clist_use:Nw  885, 18916, 18918, 18919, 18920, 18926, 18929, 18945
\_\_clist_use:nwwNw  18833, 18897, 18911
\_\_clist_use:nwwnw  ..............  884, 18883, 18894, 18896, 18908
\_\_clist_use:wN  ..............  18883, 18890, 18891, 18907
\_clist_use_end:w .............. 855, 18916, 18920, 18939, 18945
\_clist_use_i_delimit_by_s-_stop:nw ...... 18379, 18381, 18976
\_clist_use_more:w .................. 855, 18916, 18921, 18942, 18945
\_clist_use_none_delimit_by_s-_mark:w ...... 18379, 18379, 18931
\_clist_use_none_delimit_by_s-_stop:w .......... 877, 18379, 18380, 18397, 18640, 18748, 18755, 18770, 18820, 18828, 18843, 18876, 18918, 18962, 18967
\_clist_use_one:w 18916, 18919, 18937
\_clist_wrap_item:w .................. 870, 18398, 18426, 18426
\closein .................................. 177
\closeout ................................ 178
\clubpenalties .......................... 476
\clubpenalty ........................... 179
cm ..................................... 280
code commands:
  \_code:n .................................. 242, 22043
codepoint commands:
  \codepoint_generate:nn 294, 30774, 30784, 30821, 30974, 32472, 32488, 32489, 32567, 32571, 32575, 32653, 32660, 32699, 32842, 32846, 32906, 32906, 33213, 33320, 33322, 33336, 33338, 33403, 33405, 33407, 33420, 33457, 33483, 33564, 33579, 33603, 33611, 33623, 33435, 34387, 34414, 38943
  \codepoint_str_generate:n ................ 294, 14124, 14127, 14129, 30774, 30778, 30795, 31025, 31065, 31252, 31263, 31291, 31315, 31337, 32622, 32638
  \codepoint_to_category:n ................ 295, 30941, 30941, 32499
  \codepoint_to_nfd:n ..................... 295, 30950, 30950, 32688, 33244, 38945, 38946, 38947, 38948
codepoint internal commands:
  \_codepoint_add:nn ..................... 31059, 31060, 31061, 31071
  \c_codepoint_block_size_int ............. 30981, 30991, 31075, 31105, 31116, 31119, 31124, 31127, 31130, 31178, 31192, 31224, 31229, 31241
  \_codepoint_case:nn ................... 31312, 31326, 31327, 31328, 31329, 31330
  \_codepoint_case:nnn ................. 31312, 31314, 31317
  \_codepoint_casefold:n 31312, 31329
  \_codepoint_category_Ch_tl . 31083
  \_codepoint_data:nnn .................. 31218, 31220, 31238
  \_codepoint_data_auxi:w .............. 30995, 31000, 31002, 31012, 31213, 31245, 31273, 31278, 31308
  \_codepoint_data_auzip:nw .......... 30918, 31022, 31257, 31261, 31281, 31282, 31284, 31286
  \_codepoint_data_auzip:w ............ 31020, 31031
  \_codepoint_data_auxiv:nnn .......... 31036, 31052
  \_codepoint_data_auxv:mmmw ......... 31056, 31078
  \_codepoint_data_category:n ........ 31038, 31044
  \g_codepoint_data_ior ................ 30982, 31206, 31208, 31244, 31270, 31276, 31277, 31299, 31310
  \_codepoint_data_offset:mm ........... 31039, 31040, 31046, 31062
  \_codepoint_finalise_blocks:blocks: .... 31100, 31111
  \_codepoint_matche_range:blocks: ... 31170, 31173
  \_codepoint_finalise_blocks:nn ........ 31181, 31189
  \_codepoint_finalise_blocks:mmmw ... 31191, 31196, 31202
  \_codepoint_generate:n ............. 20774, 30846, 30847, 30850, 30852, 30857
  \_codepoint_generate:nnn .......... 30774, 30834, 30840
  \_codepoint_lowercase:n 31312, 31327
  \l_codepoint_matched_block_tl .. ... 30994, 31135, 31140, 31143, 31161
  \l_codepoint_next_codepoint-_fint_tl . 30993, 31054, 31068, 31096
  \_codepoint_nfd:n 30968, 31336, 31336
  \_codepoint_nfd:nn ................. 31336, 31337, 31338
  \_codepoint_range:nn ............... 31082, 31084, 31085, 31088, 31089, 31090, 31093, 31169
  \_codepoint_range:nnn .............. 31100, 31111
  \_codepoint_range_aux:nnn ........... 31095, 31098
  \_codepoint_save_blocks:nn ........ 31076, 31117, 31126, 31133
  \_codepoint_str_generate:nnn ...... 30774, 30802, 30807
  \_codepoint_templa_tl .............. 31208, 31210, 31213
Index

coffin internal commands:

\_\_coffin_align:NnNnNnN 36237, 36288, 36309, 36316, 36316, 36412
\_\_coffin_aligned_coffin ............... 35677, 36238, 36239, 36243, 36249, 36252, 36255, 36271, 36272, 36289, 36290, 36291, 36292, 36293, 36296, 36300, 36304, 36305, 36310, 36311, 36312, 36313, 36314, 36347, 36363, 36413, 36414, 36605, 36612, 36614, 36616, 36618
\_\_coffin_aligned_internal_-.\_\_coffin ....... 35677, 36326, 36333
\_\_coffin_attach:NnNnNnN ............... 36274, 36276, 36282, 36286
\_\_coffin_attach_mark:NnNnNnN ....... 36274, 36307, 36482, 36498, 36514
\_\_coffin_bottom_corner_dim ........... 35954, 35968, 35990, 36069, 36080, 36081, 36101, 36109
\_\_coffin_bounding_prop ................. 35950, 35977, 36006, 36008, 36014, 36016, 36025, 36088
\_\_coffin_bounding_shift_dim ........... 35953, 35985, 36007, 36093, 36094
\_\_coffin_calculate_intersection:Nnn . . 35847, 35847, 36318, 36321, 36324
\_\_coffin_calculate_intersection:nmmnnn . . . 35847, 35911, 35919
\_\_coffin_calculate_intersection:nmmnnnn . 35847, 35853, 35862, 35849
\_\_coffin_corners_prop .................. 35427, 35504, 35703, 35710
\_\_coffin_corners_prop .................. 35451, 35968, 35972, 35995, 36000, 36031, 36071, 36098, 36145, 36149, 36155, 36161, 36196, 36210
\_\_coffin_cos_fp . . . . . . . . . . . . . 1394, 1397, 35948, 35967, 36052, 36061
\_\_coffin_display_attach:Nnnnn ........ 36520, 36554, 36571, 36590, 36596
\_\_coffin_display_coffin ................. 36417, 36536, 36542, 36607, 36608, 36613, 36615, 36617, 36618
\_\_coffin_display_coord_coffin .......... 36417, 36484, 36499, 36515, 36557, 36572, 36591
\_\_coffin_display_font_tl ............... 36462, 36487, 36560
\_\_coffin_display_handles_-.\_\_coffin ........ 36520, 36577, 36582, 36588
\_\_coffin_display_handles_-.\_\_coffin ........ 36520, 36540, 36544
\_\_coffin_display_handles_prop ....... 36420, 36490, 36494, 36563, 36567
\_\_coffin_display_offset_dim ........... 36457, 36516, 36517, 36592, 36593
\_\_coffin_display_poly_coffin .......... 36417, 36477, 36483, 36522, 36555
\_\_coffin_display_poly_coffin .......... 36417, 36477, 36483, 36522, 36555
\_\_coffin_display_poly_pole ........... 36541, 36561, 36562, 36563, 36564, 36565, 36566, 36567, 36568, 36569, 36578, 36583
\_\_coffin_display_poles_prop ........... 36461, 36527, 36532, 36535, 36537, 36539, 36546
\_\_coffin_display_x_dim ............... 36459, 36552, 36602
\_\_coffin_display_y_dim ............... 36459, 36553, 36604
\_\_coffin_empty_coffin ................. 36407, 36412
\_\_coffin_error_bool ................... 35448, 35851, 35855, 35869, 35891, 35922, 36548, 36550
\_\_coffin_find_bounding_shift: ........ 35980, 36085, 36085
\_\_coffin_find_bounding_shift_-.\_\_aux:nn 36685, 36689, 36691
\_\_coffin_find_corner_maxima:N ........ 35979, 36065, 36065
\_\_coffin_find_corner_maxima_-.\_\_aux:nn 36605, 36672, 36674
\_\_coffin_get_pole:Nnn ................. 35990, 35949, 35950, 36374, 36375, 36378, 36379, 36529, 36530, 36533
\_\_coffin_greset_structure:N .......... 35493, 35700, 35707, 35768
\_\_coffin_gset_pole:N ................... 35545, 35622, 35714, 35755
\_\_coffin_gupdate_corners:N ............ 35769, 35772, 35774
\_\_coffin_gupdate_poles:N ............... 35770, 35803, 35805
\_\_coffin_if_exist:NTF ................. 35470, 35470, 35481, 35490, 35512, 35525, 35550, 35585, 35598, 35627, 35655, 35667, 35722, 35740, 36028, 36059
\_\_coffin_internal_box .................. 35424, 35559, 35565, 35570, 35636, 35642, 35647, 35982, 35989, 35991, 35992, 35994
\_\_coffin_internal_dim ................. 35424, 36013, 36015, 36019, 36076, 36179, 36244, 36246, 36247
\_\_coffin_internal_tl ................. 35424, 36019, 36046, 36484, 36499, 36515, 36557, 36572, 36591
\_\_coffin_internal_tl ................. 35424, 36019, 36046, 36484, 36499, 36515, 36557, 36572, 36591
\_\_coffin_join:NnNnNnN ................ 36223, 36225, 36231, 36235
Index

1519

\_color_parse_model_HSB:w .................. 37010, 37054
\_color_parse_model_Hsb:w .................. 37010, 37014
\_color_parse_model_hsb:w .................. 37010, 37012
\_color_parse_model_hsb:0:nnnn ............. 37010
\_color_parse_model_hsb:1:nnnn ............. 37010
\_color_parse_model_hsb:2:nnnn ............. 37010
\_color_parse_model_hsb:3:nnnn ............. 37010
\_color_parse_model_hsb:4:nnnn ............. 37010
\_color_parse_model_hsb:5:nnnn ............. 37010
\_color_parse_model_hsb:aux:nnnn .......... 37010, 37029, 37033
\_color_parse_model_hsb:-aux:nnnn .......... 37037, 37045
\_color_parse_model_HSB_HTML:w .......... 37010, 37061
\_color_parse_model_HSB_Aux:w .......... 37062, 37063
\_color_parse_model_RGB:w .................. 37010, 37072
\_color_parse_model_Rgb:w .................. 36985, 36987
\_color_parse_model_HSB:n .................. 37151, 37154, 37156
\_color_parse_model_HSB:nn .................. 37151, 37158, 37169, 37173
\_color_parse_model_HSB:w .................. 37151, 37151
\_color_parse_model_HSB:n .................. 37100, 37101
\_color_parse_model_RGB:n .................. 37879, 37885, 37899, 37900, 37901, 37913, 37914, 37915, 37916, 37943
\_color_parse_number:w .................. 36085, 37007, 37008
\_color_parse_set_eq:Nn .................. 36820, 36824, 36855
\_color_parse_set_eq:Nn .................. 36827, 36828, 36831
\_color_parse_std:n .................. 36807, 36919, 36942
\_color_profile_apply:nn .................. 38230, 38232, 38235
\_color_profile_apply_cmyk:n .................. 38230, 38254
\_color_profile_apply_gray:n .................. 38230, 38244
\_color_profile_apply_rgb:n .................. 38230, 38249
\_color_select:N .................. 36092, 36694, 36694, 37214, 37221, 37348
\_color_select:nn .................. 36694, 36696, 36700, 36701
\_color_select:nnN .................. 37210, 37226, 37236, 37243, 37464
\_color_select_loop:Nw .................. 37210, 37230, 37232, 37240
\_color_select_main:Nw .................. 37210, 37218, 37223, 37270, 37370, 37376, 37518
\_color_select_math:N .................. 36694, 36699, 37278
\_color_select_swap:Nnn .................. 37210, 37239, 37252
\_color_set:nn .................. 37286, 37394, 37397
\_color_set:nn .................. 37286, 37388, 37391
\_color_set:mm .................. 37286, 37408, 37411
\_color_set_aux:nn .................. 37286, 37438, 37442
\_color_set_colom:mm .................. 37286, 37444, 37449
\_color_set_loop:nn ................. 37286, 37455, 37456, 37459, 37470
\_color_show:n .................. 38259, 38270, 38279
\_color_show:Nn .................. 38259, 38260, 38262, 38263
\_color_tmp:w .................. 37551, 37559, 37560, 37561, 37562, 37563
\_color_value_tl .................. 36803, 36835, 36836, 36840, 36842, 36846, 36866, 36869, 36907, 36921, 36924, 36929, 36938, 37190, 37193, 37199, 37254, 37256, 37981, 37983
Index

11414, 11420, 11427, 11428, 11433,
11453, 11468, 11474, 11546, 11551,
11558, 11559, 11560, 11587, 11600,
11612, 11622, 11624, 11626, 11635,
11643, 11767, 11768, 11773, 12109,
12111, 12113, 12115, 12117, 12122,
12132, 12138, 12145, 12147, 12151,
12153, 12157, 12159, 12163, 12171,
12189, 12191, 12193, 12195, 12205,
12210, 12215, 12220, 12228, 12236,
12241, 12246, 12251, 12259, 12279,
12281, 12286, 12291, 12299, 12307,
12309, 12314, 12319, 12327, 12355,
12362, 12364, 12366, 12368, 12380,
12399, 12414, 12432, 12454, 12456,
12458, 12460, 12467, 12500, 12506,
12534, 12536, 12540, 12542, 12626,
12627, 12628, 12710, 12723, 12750,
12752, 12754, 12826, 12828, 13100,
13102, 13234, 13236, 13288, 13254,
13257, 13270, 13273, 13306, 13308,
13310, 13312, 13321, 13327, 13333,
13350, 13351, 13353, 13356, 13359,
13370, 13375, 13380, 13388, 13390,
13440, 13444, 13449, 13454, 13459,
13464, 13476, 13478, 13480, 13482,
13488, 13503, 13516, 13518, 13522,
15254, 13671, 13686, 13695, 13705,
13707, 14154, 14161, 14170, 14302,
14318, 14334, 14430, 14434, 14436,
14366, 14386, 14394, 14404, 14413,
14497, 14505, 14507, 14509, 14519,
14525, 14549, 14560, 14566, 14569,
14571, 14576, 14602, 14608, 14614,
14642, 14716, 14764, 14817, 14900,
14906, 14929, 14959, 14980, 15055,
15151, 15156, 15158, 15160, 15236,
15238, 15240, 15245, 15255, 15334,
15539, 15541, 15594, 15596, 15398,
15403, 15413, 16347, 16449, 16451,
16462, 16469, 16475, 16493, 16502,
16507, 16512, 16517, 16522, 16528,
16533, 16540, 16546, 16555, 16565,
16567, 16569, 16574, 16579, 16591,
16636, 16675, 16681, 16684, 16687,
16690, 16701, 16706, 16711, 16716,
16727, 16733, 16735, 16737, 16739,
16741, 16778, 16780, 16782, 16788,
16790, 16798, 16806, 16819, 16821,
16829, 16831, 16833, 16845, 16847,
16849, 16874, 16876, 16886, 16892,
16905, 16914, 16939, 16941, 16943,
16968, 16969, 16970, 16995, 17027,
17035, 17045, 17056, 17058, 17060,
Index

\currentcjtktoken .......... 1136, 1200
\currentgrouplevel ........ 477
\currentgroupotype ........ 478
\currentifbrance ........... 479
\currentiflevel ............ 480
\currentifotype ............ 481
\currentspacingmode ........ 1137
\currentxspacingmode ....... 1138

D
\d ....................... 32057, 34380, 34402
\date ..................... 366
\day ........................ 186, 1290, 9010
dd .......................... 280
\deadcycles ................ 187

debug commands:
\debug off:n ............... 30, 366, 1465, 1466, 1509, 1575, 1579, 39031, 39039, 39059
\debug on:n ................ 30, 366, 694, 1465, 1509, 1569, 1573, 39031, 39049
\debug resume: ............. 30, 1528, 1383, 1476, 1581, 1582, 30392, 35507, 39067, 39073
\debug suspend: ............. 30, 1528, 1383, 1476, 1581, 1581, 30390, 35500, 39067, 39068

debug internal commands:
_\debug_add_to_debug_code:Nnn ........ 39297, 39312, 39335
_\debug_all_off: .............. 39031, 39057
_\debug_all_on: .............. 1572, 1578, 39031, 39047
_\debug_arg_check_invalid:N ....... 39216, 39238, 39244
_\debug_arg_if_braced:N ........ 39250
_\debug_arg_if_braced:n ....... 39216, 39260, 39261
_\debug_arg_if_braced:NTF ....... 39216, 39239
_\debug_arg_list_from_signature:nNN ........ 39216, 39227, 39232, 39241
_\debug_arg_return:N .......... 39216, 39274, 39275, 39276, 39277, 39279, 39280, 39281, 39282, 39283, 39295
_\debug_build_arg_list:n ...... 39216, 39223, 39230
_\debug_build_parm_text:n .... 39216, 39221, 39225
_\debug_check-declarations_off: .......... 39083
_\debug_check-declarations_on: .... 39083
_\debug_check-expressions_off: .... 39182
_\debug_check-expressions_on: .... 39182
_\debug_check_declarations_on: .... 39182
_\debug_check_declarations_off: .... 39182
_\debug_chk_expr_aux:NNN ....... 39182, 39187, 39195
_\debug_chk_var_scope_aux:NNN ....... 39124, 39129, 39135, 39147, 39147
_\debug_chk_var_scope_aux:NNN ....... 39147, 39148, 39149
_\debug_chk_var_scope_aux:NNN ....... 39178, 39179, 39152, 39156, 39161
_\debug_deprecation_off: ......... 39209, 39211
_\debug_deprecation_off_tl ....... 39209, 39209
_\debug_deprecation_on: .......... 1583, 39212
_\debug_deprecation_on_tl ....... 39209, 39209
_\debug_generate_parameter_list:NNN ....... 1459, 1482, 39216, 39261, 39321
_\debug_get_base_form:N .......... 39216, 39260, 39272
_\debug_if_recursion_tail-stop:N .......... 39229, 39300, 39237
_\debug_insert_debug_code:Nnn .... 39297
_\debug_internal_tl .......... 39213, 39218, 39221, 39223
_\debug_log-functions_off: ...... 39174
_\debug_log-functions_on: ........ 39174
_\debug_parm_terminate:w ....... 39216, 39246, 39247, 39248, 39249, 39257
_\debug_patch_weird:N ............ 39297, 39372, 39375
_\debug_setup_debug_code:Nnn .... 39297, 39313, 39318
_\debug_suspended:TF .......... 1476, 39067, 39071, 39079, 39082, 39094, 39103, 39112, 39122, 39127, 39133, 39177, 39186
_\debug_suspended_tl .......... 39067
_\debug_tmp:w ................ 39339, 39358
_\debug_tmpa_tl .............. 39213, 39321, 39326
_\debug_tmpb_tl .............. 39213, 39321, 39330
_\debug_use_i_delimit_by_a-stop:nw .... 39025, 39025, 39151, 39153
_\debug_use_none_delimit_by_q-recursion_stop:w .......... 39028, 39258
\def ........................ 36, 37, 38, 60, 62, 67, 68, 70, 84, 106, 143, 188
default commands:
\default:n .................. 242, 22069
\defaultlyphenchar .......... 189
\defaultskewchar ............ 190
\deferred ................... 806
deg .......................... 279
\delcode .......................... 191
Index

\delimiter ........................................ 192
\delimiterfactor ................................ 193
\delimitershortfall ............................. 194
\deprecation internal commands:
\_\_deprecation_j ust_error:nnNN 38744
\_\_deprecation_patch_aux:nn ... 38744, 38756, 38777
\_\_deprecation_patch_aux:nnNNnn .
\_\_deprecation_patch_aux:nnNNnnn
\_\_deprecation_warn_once:nnNNnn ..... 38744, 38755, 38758
\detokenize ................................ 30, 94, 482
\DH ........................................... 32063, 33647, 34343
dim commands:
\dim_abs:n 224, 971, 20809, 20809, 39887
\dim_add:Nn ................................ 224, 20791, 20791, 20798, 39448, 39822
\dim_case:n ....... 227, 20898, 20904
\dim_case:nnFF ........... 227, 20899, 20894, 20899
\dim_compare:n .............. 20849
\dim_compare:nNn ............ 20844
\dim_compare:nNnFF ........... 225–228, 263, 20844
\dim_compare:nTF .............. 20913, 20919, 20957, 20966, 20972,
\dim_compare:nTF .............. 20984, 20987, 20998, 21152, 35332,
\dim_compare:nTF .............. 35349, 35383, 35397, 35407, 35865,
\dim_compare:nTF .............. 35868, 35873, 35887, 35900, 35985,
\dim_compare:nTF .............. 36241, 36246, 36256, 36385, 36397
\dim_compare:nTF .............. 225, 226,
\dim_compare:nTF .............. 228, 20849, 20921, 20929, 20938, 20944
\dim_compare:p:n .............. 226, 20849
\dim_compare:p:nNn ............ 225, 20844, 38717,
\dim_compare:p:nNn ............ 38718, 38725, 38726, 38729, 38730
\dim_compare:nNn ............ 223, 944, 958, 20758, 20758, 20763,
\dim_compare:nNn ............ 21173, 21174, 22998, 35992, 39826
\dim_compare:nNn ............ 228, 20919, 20941, 20945
\dim_compare:nNn ............ 227, 20947, 20969, 20973
\dim_compare:nNn ............ 228, 20919, 20935, 20939
\dim_compare:nNn ............ 227, 20947, 20963, 20967
\dim_eval:n ................ 225,
\dim_eval:n ................ 226, 229, 944, 1358, 1359, 20761,
\dim_eval:n ................ 20892, 20897, 20902, 20907, 21002,
\dim_eval:n ................ 21030, 21037, 21168, 21172, 35664,
\dim_eval:n ................ 35641, 35727, 35745, 35786, 35790,
\dim_eval:n ................ 35791, 35795, 35799, 35800, 35817,
\dim_eval:n ................ 35822, 35828, 35835, 35842, 36007,
\dim_eval:n ................ 36010, 36011, 36018, 36100, 36101,
\dim_eval:n ................ 36108, 36109, 36212, 36219, 36368,
\dim_eval:n ................ 36369, 36633, 36634, 36635, 39884
\dim_gadd:Nn ................ 224, 20791, 20793, 20799, 39529, 39823
\dim_gset:N ..................... 242, 22079
\dim_gset:N ..................... 224,
\dim_gset_eq:NN ............... 224, 20791, 20781, 20784, 39528, 39821
\dim_gset_eq:NN ............... 224, 20785, 20788, 20790, 39526
\dim_gsub:N ..................... 224, 20791, 20802, 20808, 39530, 39825
\dim_gzero:N ............... 223, 20764,
\dim_gzero:N ............... 20765, 20768, 20772, 21194, 39527
\dim_gzero_new:N ............. 223, 20760, 20771, 20774
\dim_if_exist:N ............... 20775, 20777
\dim_if_exist:NTF ............. 224, 20770, 20772, 20775
\dim_if_exist_p:N ............. 224, 20775
\dim_log:N ...................... 231, 21169, 21169, 21170
\dim_log:n ...................... 232, 21169, 21171
\dim_max:n ...................... 224, 20809, 20816, 36079, 36083, 39933
\dim_min:n ...................... 224, 20629, 20677, 36081, 36094, 39934
\dim_new:N ...................... 222,
\dim_new:N ...................... 20752, 20752, 20757, 20760, 20770,
\dim_new:N ...................... 20772, 21175, 21176, 21177, 21178,
\dim_new:N ...................... 34961, 34962, 34963, 34964, 34965,
\dim_new:N ...................... 34966, 34967, 34968, 35425, 35449,
\dim_new:N ...................... 35450, 35453, 35454, 35455, 35456,
\dim_new:N ...................... 35953, 35954, 35955, 35956, 35957,
\dim_new:N ...................... 36115, 36116, 36457, 36459, 36460
\dim_ratio:n .................... 225, 20840, 20840
\dim_set:N ..................... 224, 22079, 20779, 20783, 34988, 34989,
\dim_set:N ..................... 34990, 35022, 35033, 35117, 35118,
\dim_set:N ..................... 35119, 35134, 35232, 35233, 35234,
\dim_set:N ..................... 35236, 35238, 35240, 35554, 35630,
\dim_set:N ..................... 35867, 35871, 35889, 35893, 35924,
\dim_set:N ..................... 35938, 36013, 36048, 36056, 36067,
\dim_set:N ..................... 36068, 36069, 36070, 36076, 36078,
\dim_set:N ..................... 36080, 36082, 36087, 36093, 36176,
\dim_set:N ..................... 36178, 36180, 36188, 36190, 36244,
\dim_set:N ..................... 36319, 36320, 36322, 36324, 36342,
\dim_set:N ..................... 36334, 36458, 36552, 36553, 36599,
\dim_set:N ..................... 36600, 36601, 36603, 39446, 39820
\dim_set_eq:NN ............... 224, 20785
\dim_set_eq:NN ............... 20785, 20787, 35579, 35580, 39447
\dim_show:N ............... 231, 21165, 21165, 21166
\dim_show:n ............... 231, 21167, 21167
\dim_sign:n ............... 229, 21032, 21032, 39888
\dim_step_function:nN ............ 228, 950, 20975, 20975, 201027, 39988, 39992, 39996, 40000
\dim_step_inline:nnn ............ 228, 21005, 21005
\dim_step_variable:nnnN ............ 229, 21005, 21012
\dim_sub:N ......................... 224, 20791, 20800, 20807, 39449, 39824
\dim_to_decimal:n ............ 229, 954, 21052, 21052, 21088, 21116, 21153, 21162, 39885
\dim_to_decimal_in_bp:n ........ 230, 21069
\dim_to_decimal_in_cc:n ........ 230, 21069
\dim_to_decimal_in_cm:n ........ 230, 21069
\dim_to_decimal_in_dd:n ........ 230, 21069
\dim_to_decimal_in_in:n ........ 230, 21069
\dim_to_decimal_in_mm:n ........ 230, 21069
\dim_to_decimal_in_pc:n ........ 230, 21069
\dim_to_decimal_in_sp:n ........ 231, 1066, 21067, 21067, 24216, 24253, 24847, 39886
\dim_to_decimal_in_unit:nn ........ 231, 21095, 21095
\dim_to_fp:n ............ 231, 1063, 1082, 21067, 21012, 29102, 35026, 35027, 35037, 35038, 35106, 35109, 35110, 35135, 35150, 35151, 35170, 35171, 35189, 35206, 35209, 35210, 35266, 35268, 35287, 35387, 35389, 35390, 35402, 35403, 35404, 35605, 35605, 36060, 36061, 36134, 36137, 36138, 36189, 36191, 40019
\dim_until_do:nn ............ 228, 20919, 20927, 20932
\dim_until_do:NNnn ............ 227, 20947, 20955, 20960
\dim_use:N ............ 229, 1558, 20812, 20814, 20819, 20820, 20826, 20827, 20828, 20852, 20871, 21031, 21035, 21050, 21050, 21051, 21055, 21248, 21249, 21324, 21325, 36015, 36019, 36026, 36032, 36041, 36042, 36043, 36197, 36204, 36350
\dim_while_do:nn ............ 228, 20919, 20919, 20924
\dim_while_do:NNnn ............ 228, 20947, 20947, 20952
\dim_zero:N ............ 223, 20764, 20764, 20767, 20770, 21193, 34991, 35120, 35235, 35858, 35859, 39445
\dim_zero_new:N ............ 223, 20769, 20769, 20773
\c_max_dim ............ 230, 232, 235, 1012, 21173, 21267, 23025, 23067, 23075, 36067, 36068, 36069, 36070, 36087
\g_tempa_dim ............ 232, 21175
\l_tempa_dim ............ 232, 21175
\l_tempb_dim ............ 232, 21175
\l_tmbp_dim ............ 232, 21175
\c_zero_dim ............ 232, 20984, 20987, 21040, 21173, 21266, 23092, 34847, 34871, 35336, 35347, 35353, 35365, 35383, 35387, 35395, 35397, 35401, 35407, 35417, 35865, 35868, 35873, 35887, 35890, 35895, 36241, 36246, 36256
dim internal commands:
\__dim_abs:N ............ 20809, 20811, 20814
\__dim_branch_unit:w ....... 954, 21105, 21110, 21110
\__dim_case:nTF ............ 20889, 20892, 20897, 20902, 20907, 20099
\__dim_case:nw ............ 20889, 20910, 20911, 20915
\__dim_case_end:nw ........ 20889, 20914, 20917
\__dim_chk_unit:w ........ 954, 21097, 21100, 21100
\__dim_compare:.w ........ 2049, 20851, 20854
\__dim_compare:wN ........ 946, 20849, 20857, 20860, 20870
\__dim_compare!:w ........ 2049
\__dim_compare:<:w ........ 2049
\__dim_compare:=:w ........ 2049
\__dim_compare>:w ........ 2049
\__dim_compare_end:w ........ 20857, 20881
\__dim_compare_error: ........ 946, 20849, 20852, 20854, 20883, 20887
\__dim_convert_remainder:w ....... 955, 21140, 21144, 21144
\__dim_eval:w ........ 20746, 20747, 20780, 20782, 20792, 20796, 20801, 20805, 20812, 20818, 20819, 20820, 20826, 20827, 20828, 20843, 20846, 20852, 20871, 20876, 20978, 20979, 20980, 21031, 21035, 21055, 21068, 21075, 21098, 21106, 21153, 39828, 39890, 39936, 39967, 39991, 39995, 39999
\__dim_eval_end: ........ 20746, 20748, 20780, 20782, 20792, 20796, 20801, 20805, 20812, 20822, 20830, 20843, 20846, 21031, 21035, 21055, 21068, 21075, 21098, 21153
\__dim_get_quotient:w ....... 955, 21117, 21120, 21120
\__dim_get_remainder:w ....... 955, 21127, 21132, 21138, 21138
Index

16963, 17016, 17033, 17043, 17097, 17102, 17476, 17477, 17486, 17509, 17526, 17527, 17529, 17546, 17547, 17591, 17644, 17652, 17679, 17687, 17693, 17716, 17724, 17762, 17770, 17855, 18072, 18165, 18153, 18158, 18274, 18300, 18327, 18336, 18540, 18555, 18578, 18592, 19192, 19193, 19194, 19195, 19200, 19201, 19209, 19210, 19211, 19240, 19246, 19264, 19273, 19275, 19321, 19322, 19324, 19325, 19326, 19327, 19328, 19329, 19330, 19359, 19364, 19369, 19374, 19381, 19388, 19393, 19398, 19403, 19408, 19413, 19420, 19425, 19447, 19458, 19459, 19509, 19510, 19632, 19641, 19650, 19658, 19735, 19764, 19765, 19773, 19811, 19825, 19834, 19844, 19866, 19867, 19868, 19978, 19985, 19987, 20212, 20219, 20263, 20311, 20320, 20468, 20507, 20516, 20547, 20548, 20549, 20550, 20559, 20560, 20561, 20562, 20585, 20586, 20587, 20588, 20597, 20598, 20599, 20600, 20815, 20838, 20847, 20864, 20868, 20882, 20885, 21047, 21048, 21091, 21096, 21106, 21107, 21613, 21668, 21680, 23033, 23034, 23069, 23077, 23141, 23160, 23174, 23260, 23276, 23280, 23292, 23302, 23397, 23450, 23453, 23454, 23459, 23473, 23511, 23512, 23513, 23514, 23515, 23516, 23517, 23518, 23519, 23520, 23521, 23522, 23535, 23537, 23548, 23551, 23556, 23570, 23703, 23794, 23795, 23804, 23805, 23816, 23817, 23818, 23829, 23830, 23831, 23838, 23849, 23850, 23851, 23861, 23862, 23866, 23867, 23875, 23878, 23879, 23887, 23988, 23918, 23941, 23976, 23993, 24012, 24013, 24022, 24028, 24048, 24049, 24077, 24086, 24103, 24110, 24118, 24119, 24220, 24221, 24222, 24225, 24228, 24283, 24309, 24310, 24317, 24325, 24354, 24355, 24358, 24360, 24361, 24366, 24376, 24379, 24381, 24386, 24417, 24430, 24435, 24441, 24444, 24445, 24479, 24480, 24507, 24508, 24521, 24524, 24535, 24558, 24577, 24587, 24603, 24612, 24618, 24624, 24628, 24633, 24639, 24654, 24665, 24682, 24690, 24692, 24698, 24719, 24747, 24770, 24803, 24805, 24928,
file commands:

```
\file_compare_timestamp:nNn ....... 11364, 11372
\file_compare_timestamp:nNnTF ....... 102, 11364
\file_compare_timestamp:pNn ....... 102, 11364
\g_file_curr_dir_str ...... 99, 10928, 11457, 11463, 11476
\g_file_curr_ext_str ...... 99, 10928, 11459, 11465, 11478
\g_file_curr_name_str ...... 99, 10928, 11458, 11464, 11477
\file_full_name:n ....... 103, 11054, 11054, 11059, 11060, 11071
\file_full_name:nNTF ....... 103, 11054, 11056
\file_get:nnN ....... 103, 11071
\file_get:nnNTF ....... 102, 11205, 11205, 11210, 11211, 11219
\file_get_full_name:n ....... 103, 11205, 11205, 11210, 11211, 11219
\file_get_full_name:nNTF ....... 102, 11220, 11062, 11205, 11207, 11416, 11142, 11435
\file_get_hex_dump:n ....... 100, 11310, 11310, 11312, 11312, 11322, 11324
\file_get_hex_dump:nnnN ....... 100, 11346, 11346, 11351, 11352, 11361
\file_get_hex_dump:nnnNTF ....... 100, 11346, 11348
\file_get_mdfive_hash:n ....... 100, 11310, 11311
\file_get_mdfive_hash:nnnN ....... 101, 11310, 11313, 11315, 11326, 11328
\file_get_mdfive_hash:nnnNTF ....... 101, 11310, 11314
```
flag commands:
\flag_clear:N \ldots \ 182, 5625, 7513, 7514, 14499, 14527, 14621, 14650, 14719, 14767, 14768, 14820, 14821, 15057, 15058, 15059, 15060, 15061, 15162, 15257, 15258, 15259, 15260, 15415, 15416, 15417, 18290, 18295, 18306, 18349, 29781, 39639
\flag_clear:n \ldots \ 18348, 18349
\flag_clear_new:N \ldots \ 182, 778, 15002, 15003, 15004, 15005, 15185, 15186, 15187, 15365, 15366, 18205, 18305, 18307, 18350
\flag_clear_new:n \ldots \ 18348, 18350
\flag_ensure_raised:N \ldots \ 182, 5625, 5674, 18345, 18346, 18347, 18361, 23633, 23642, 23650, 23667, 23676, 23707, 29780, 39613
\flag_ensure_raised:n \ldots \ 18348, 18361
\flag_height:N \ldots \ 182, 7523, 7525, 14343, 18315, 18331, 18341, 18344, 18350, 39614
\flag_height:n \ldots \ 18348, 18359, 18369
\flag_if_exist:N \ldots \ 18317, 18319
\flag_if_exist:NTF \ldots \ 182, 18306, 18317, 18352, 18353, 18355, 39113
\flag_if_exist:nTF \ldots \ 18348, 18352, 18353, 18354
\flag_if_exist_p:N \ldots \ 182, 364, 18317, 18351
\flag_if_exist_p:n \ldots \ 182, 1848
\flag_if_raised:N \ldots \ 18321, 18329
\flag_if_raised:NTF \ldots \ 182, 5633, 14336, 14341, 14344, 15029, 15035, 15040, 15047, 15219, 15224, 15229, 15380, 15387, 18321, 18356, 18357, 18358, 29783, 39015, 39016, 39017
\flag_if_raised:nTF \ldots \ 18348, 18356, 18357, 18358
\flag_if_raised:p:N \ldots \ 182, 18321, 18355, 39018
\flag_if_raised:p:n \ldots \ 182, 18348
\flag_log:N \ldots \ 182, 18308, 18310, 18311
\flag_log:n \ldots \ 18362, 18363
\flag_new:N \ldots \ 181, 182, 778, 5615, 7373, 7374, 14205, 14206, 18285, 18287, 18288, 18289, 18306, 18348, 23599, 23600, 23601, 23602, 29767, 39632
\flag_new:n \ldots \ 18348, 18348
\flag_raise:N \ldots \ 182, 7566, 7574, 14485, 14535, 14635, 14668, 14739, 14752, 14790, 14795, 14876, 15078, 15079, 15102, 15103, 15116, 15117, 15136, 15137, 15143, 15144, 15174, 15324, 15325, 15434, 15435, 15439, 15440, 15454, 15455, 18342, 18342, 18344, 18360, 39619
\flag_raise:n \ldots \ 18348, 18360
\flag_show:N \ldots \ 182, 18308, 18308, 18309
\flag_show:n \ldots \ 18362, 18362
\l_tmpa_flag \ldots \ 183, 364, 18288, 39117
\l_tmpb_flag \ldots \ 183, 18288
flag internal commands:
\__flag_clear:WN \ldots \ 18290, 18292, 18296, 18302
\__flag_height_end:WN \ldots \ 18311, 18331, 18335, 18340
\__flag_height_loop:WN \ldots \ 18311, 18331, 18332, 18337
\__flag_show:NN \ldots \ 18308, 18308, 18310, 18312
\__flag_show:Nn \ldots \ 18362, 18362, 18363, 18364
\floatingpenalty \ldots \ 234
floor \ldots \ 276
fmtname \ldots \ 8687, 8690, 8691, 8703, 8713, 31674, 31935, 31936, 35576, 35577, 36464, 36465, 38445
\font \ldots \ 235
\fontcha dp \ldots \ 491
\fontcharht \ldots \ 492
\fontcharic \ldots \ 493
\fontcharwd \ldots \ 494
\fontdimen \ldots \ 1004, 236
\fontencoding \ldots \ 34259
\fontfamily \ldots \ 34260
\fontid \ldots \ 826
\fontname \ldots \ 237
\fontseries \ldots \ 34261
\fontshape \ldots \ 34262
\fontsize \ldots \ 34265
\footnotesize \ldots \ 34301
\forcecjktoken \ldots \ 1203
\formatname \ldots \ 827
fp commands:
c_e_fp \ldots \ 269, 272, 25532
fp_abs:n \ldots \ 274, 280, 1218, 29142, 29142, 35135, 35237, 35239, 35241, 36179, 36181
fp_add:Nn \ldots \ 269, 1217, 1218, 25443, 25443, 25449
fp_clear_function:n \ldots \ 268, 29065, 29965
fp_clear_variable:n \ldots \ 268, 29733, 29733, 29904
fp_compare:n \ldots \ 25557
fp_compare:Nn \ldots \ 25573
Index

\fp_to_dim:n ........................................ 261, 270, 29067, 29070, 35024, 35035, 35135, 35876, 35988, 35926, 35940, 36050, 36058, 36189, 36191
\fp_to_int:N .................................. 261, 29083, 29083, 29084
\fp_to_int:n .................................. 261, 29083, 29085, 37588
\fp_to_scientific:N ................................ 262, 28890, 28890, 28921, 28928
\fp_to_scientific:n ................................ 262, 28890, 28893
\fp_to_tl:n ........................................ 262, 284, 23591, 25461, 29023, 29023, 29024, 29788
\fp_to_tl:n ........................................ 262, 23217, 23632, 23641, 23666, 23675, 23704, 23501, 23516, 25529, 25531, 25767, 25768, 25787, 25798, 29023, 29025
\fp_trap:nn .................................. 270, 1637, 23603, 23718, 23719, 23720, 23721
\l_fm_underflow_flag ............ 270, 23599
\fp_until_do:nn ......................... 265, 25704, 25706, 25711
\fp_until_do:nnn ......................... 265, 25772, 25734, 25739
\fp_use:N .................................. 262, 284, 29137, 29137, 29138
\fp_while_do:nn ......................... 265, 25704, 25714, 25719
\fp_while_do:nnn ................. 265, 25772, 25742, 25747
\fp_zero:N ................................ 260, 25433, 25435, 25435, 25438
\fp_zero_new:N .................. 260, 25437, 25437, 25441
\c_inf_fp .................................. 269, 279, 23231, 24813, 26303, 26385, 26723, 27484, 27507, 27709, 27712, 27716, 27739, 27941, 28104, 30145
\c_minus_inf_fp ................. 269, 279, 23231, 26304, 26388, 26721, 27259, 28105, 30146
\c_minus_zero_fp .................. 269, 23231, 26300, 28824, 30144
\c_nan_fp .................................. 269, 279, 1040, 1065, 23231, 23643, 23651, 23723, 23929, 23948, 23954, 23977, 24144, 24152, 24160, 24238, 24295, 24334, 24725, 24802, 24814, 25303, 25318, 25791, 27683, 29201, 29780, 29959, 30058, 30117, 30143
\c_one_degree_fp .................. 269, 279, 24816, 25534
\c_one_fp .................................. 269, 1093, 1592, 24817, 25246, 25267, 25532, 25891, 26744, 27478, 27678, 27729, 27914, 28028, 28058, 28607, 29217
\c_pi_fp .................. 269, 279, 1075, 24815, 25534
\g_tampa_fp .................. 269, 25536
\l_tampa_fp .................. 269, 25536
\g_tmpb_fp .................. 269, 25536
\l_tmpb_fp .................. 266, 267, 269, 25536
\c_zero_fp .................. 269, 1097, 1111, 1246, 23231, 23285, 24188, 25258, 25270, 25418, 25433, 25434, 25893, 25896, 26132, 26299, 27487, 27508, 27706, 27742, 28822, 28928, 29112, 30142, 34992, 34994, 34999, 35284, 35293, 36153, 36092, 38033
\g_flux ............... 25900, 25007, 25916, 25917
\g_flux_o:ww ............... 1100, 1109, 25897
\g_flux_o:ww ............... 29549
\g_flux_o:ww ............... 25897
\g_flux_o:ww ............... 26261
\g_flux_o:ww ............... 26274
\g_flux_o:ww ............... 1112, 1113, 1114, 25895
\g_flux_o:ww ............... 29549
\g_flux_o:ww ............... 1112, 1113, 25890
\g_flux_o:ww ............... 29549
\g_flux_o:ww ............... 1112, 1113, 1114, 25895
\g_flux_o:ww ............... 29549
\g_flux_o:ww ............... 1112, 1113, 1114, 25895
\g_flux_o:ww ............... 29549
\g_flux_o:ww ............... 1112, 1113, 1114, 25895
\g_flux_o:ww ............... 29549
\g_flux_o:ww ............... 1112, 1113, 1114, 25895
\g_flux_o:ww ............... 29549
\g_acos_o:ww ............... 1206, 1209, 28763, 28763
\g_acos_o:ww ............... 28903, 28905, 28955, 28961
\g_acos_o:ww ............... 28805, 28628
\g_acos_o:ww ............... 28805, 28844
\g_acos_o:ww ............... 28815, 28815
\g_add:NNN ............... 25443, 25443, 25445, 25446, 25447
\g_add_big_i:NNN ............... 1115
\g_add_big_i:NNN ............... 1112, 1115, 26052, 26059, 26059
\g_add_big_i:NNN ............... 1115
\g_add_big_i:NNN ............... 26055, 26059, 26067
\g_add:NNN ............... 26001, 26021, 26021
\g_add:NNN ............... 26001, 26000, 26036, 26036
\g_add:NNN ............... 26003, 26009, 26099, 26099
\g_add:NNN ............... 26003, 26009, 26099, 26099
\g_add:NNN ............... 26003, 26009, 26099, 26099
\_fp_add_significand\_o:NnNw
  1115, 1116, 26062, 26070, 26075, 26075
\_fp_add_significant\_pack:NN\_NwNw
  26075, 26079, 26082
\_fp_add_significant\_test\_o:N  ...
  26075, 26077, 26089
\_fp_add_zeros\_o:Nw
  25999, 26011, 26011
\_fp_and\_return\_wNw
  25897, 25903, 25910, 25922
\_fp_array\_bounds\:NN\_TF
  30014, 30014, 30045, 30115
\_fp_array\_bounds\_error\:NN
  30014, 30017, 30021, 30028
\_fp_array\_count\:n  23344, 23344, 23913, 25651, 25652, 26784, 28863
\_fp_array\_get\:NN\_Nw
  30033, 30036, 30043
\_fp_array\_get\:w  30033, 30049, 30060
\_fp_array\_get\_normal\:w
  30033, 30064, 30070
\_fp_array\_get\_recover\:Nw
  30033, 30050, 30055
\_fp_array\_get\_special\:nn\_NN
  30014
\_fp_array\_if\_all\_fp\_nTF
  1246
\_fp_array\_if\_all\_fp\_wT
  23446, 23446, 25296
\_fp_array\_loop\_int
  23446, 23448, 23515, 23545
\_fp_array\_new\:NN
  29979, 29986, 29988, 30000
\_fp_array\_item\:N  300097, 30121, 30126
\_fp_array\_item\:NN\_Nw
  30037, 30116, 30119
\_fp_array\_item\:Nw
  30097, 30099, 30107, 30113
\_fp_array\_item\:w  30097, 30129, 30131
\_fp_array\_item\_normal\:w
  30097, 30133, 30149
\_fp_array\_item\_special\:w
  30097, 30128, 30137
\_fp_array\_loop\_int
  29980, 30086, 30089, 30092
\_fp_array\_new\:NN\_NN
  29981
\_fp_array\_new\:NN\_NN
  29990, 29994
\_fp_array\_to\_clist\:n
  23981, 29148, 29148, 29214, 29651
\_fp_array\_to\_clist\_loop\:Nw
  29148, 29154, 29159, 29164
\_fp_anec\_o\:w
  28828, 28828
\_fp_asin\_aux\_o\:NnNw
  1207, 1209, 28793, 28796, 28796, 28855
\_fp_asin\_isqrt\:wn
  28796, 28799, 28806
\_fp_asin\_normal\_o\:NwNnNw
  28754, 28770, 28781, 28781
\_fp_asin\_o\:w
  28748, 28748
\_fp_atan\_aux\_i\:w
  1204, 28673, 28687, 28687
\_fp_atan\_aux\_i\:w
  28687, 28688, 28689
\_fp_atan\_combine\_aux\_w
  28714, 28728, 28735
\_fp_atan\_combine\_o\:Nwww
  1202, 28632, 28649, 28714
\_fp_atan\_default\:w
  1093, 1202, 28595, 28599, 28605, 28607
\_fp_atan\_div\:www
  1202, 28660, 28662, 28662
\_fp_atan\_inf\_o\:NNw
  1202, 28620, 28621, 28622, 28630, 28630, 28766, 28839
\_fp_atan\_near\:www
  28007, 28009, 28595, 28595
\_fp_atan\_Taylor\_break\:w
  28598, 28701, 28711
\_fp_atan\_Taylor\_loop\:www
  1204, 28693, 28698, 28706
\_fp_atan\_test\_o\:NwNwww
  1202, 28660, 28662, 28662
\_fp_atan\_o\:Nw
  28007, 28009, 28595, 28595
\_fp_atan\_Taylor\_break\:w
  28598, 28701, 28711
\_fp_atan\_Taylor\_loop\:www
  1204, 28693, 28698, 28706
\_fp_atan\_test\_o\:NwNwww
  1202, 28660, 28662, 28662
\_fp_atan\_o\:Nw
  28007, 28009, 28595, 28595
\_fp_atan\_Taylor\_break\:w
  28598, 28701, 28711
\_fp_atan\_Taylor\_loop\:www
  1204, 28693, 28698, 28706
\_fp_atan\_test\_o\:NwNwww
  1202, 28660, 28662, 28662
\_fp_atan\_o\:Nw
  28007, 28009, 28595, 28595
\_fp_atan\_Taylor\_break\:w
  28598, 28701, 28711
\_fp_atan\_Taylor\_loop\:www
  1204, 28693, 28698, 28706
\_fp_atan\_test\_o\:NwNwww
  1202, 28660, 28662, 28662
\_fp_atan\_o\:Nw
  28007, 28009, 28595, 28595
\_fp_atan\_Taylor\_break\:w
  28598, 28701, 28711
\_fp_atan\_Taylor\_loop\:www
  1204, 28693, 28698, 28706
\_fp_atan\_test\_o\:NwNwww
  1202, 28660, 28662, 28662
\_fp_atan\_o\:Nw
  28007, 28009, 28595, 28595
\_fp_atan\_Taylor\_break\:w
  28598, 28701, 28711
\_fp_atan\_Taylor\_loop\:www
  1204, 28693, 28698, 28706
\_fp_atan\_test\_o\:NwNwww
  1202, 28660, 28662, 28662
\_fp_atan\_o\:Nw
\_fp_mul_significand_keep:NNNNNw

........... 26319, 26324, 26326, 26342
\_fp_mul_significand_large_f:NNNNNw ... 26449, 26535, 26533
\_fp_mul_significand_o:nnnNnnn

........... 1123, 26317, 26319, 26319
\_fp_mul_significand_small_f:NNWWN ... 26347, 26364, 26364
\_fp_mul_significand_test_f:NNN

........... 1124, 26321, 26344, 26344
\c\_fp_myrriad_int 23329
26802, 26833, 26834, 26911, 26972
\_fp_neg_sign:N ........... 1113, 23262, 23262, 25983, 26136
\_fp_new_function:n

........... 29820, 29821, 29822
\_fp_new_variable:n

........... 29748, 29750, 29752
\_fp_not_o:w 1100, 24752, 25888, 25888
\c\_fp_one_fixed_t1 ..... 26700, 27416, 27629, 27826, 27853, 28636,
28703, 28808, 29259, 29269, 29310
\_fp_overflow:w 1023, 1037, 1039,
23277, 23711, 23716, 27507, 27940
\c\_fp_overflowing_fp

........... 23243, 29821, 29875
\_fp_pack:NNNNNw

........... 24146, 24149, 26808,
26818, 26821, 26824, 26827, 26830,
26891, 26894, 26897, 26900, 26903,
27791, 27979, 27979, 27980
\c\_fp_pack_big:NNNNNw

........... 23430, 23433,
26583, 26586, 26589, 26592, 26595,
26598, 26601, 26605, 26918, 26928,
26938, 26948, 26951, 26954, 26961
\_fp_pack_Bigg:NNNNNNw

........... 23435, 23438, 26432,
26435, 26438, 26450, 26453, 26456
\_fp_pack_eight:wNNNNNNN

........... 1029, 1119, 23442, 23442,
26245, 26554, 26983, 28132, 28133
\_fp_pack_twice_four:wNNNNNNN

........... 1029, 23440, 23440, 24029, 24030,
2618, 26188, 26984, 26985, 26986,
27018, 27019, 27020, 27209, 27209,
27544, 27545, 27546, 28134, 28135,
28249, 28340, 28341, 28342, 29125
\_fp_parse:n

........... 1051, 1063, 1075, 1083,
1096, 1097, 1105, 1217, 1218, 1216,
24000, 24211, 24809, 24809, 25421,
25423, 25425, 25448, 25551, 25560,
25577, 25587, 25755, 25805, 26730,
28896, 28950, 29028, 29073, 29088,
29141, 29143, 29145, 29147, 30040
\_fp_parse_after:w

........... 24860, 24872, 24880, 24885
\_fp_parse_apply_binary:NWWN

........... 1055, 1056, 1059, 1087, 24907, 24907, 25098
\_fp_parse_apply_binary_chk:NN

.... 24907, 24912, 24924, 24938, 24951
\_fp_parse_apply_binary_error:NNN.... 24907, 24927, 24931
\_fp_parse_apply_comma:NWWN

........... 1087, 25057, 25068, 25083
\_fp_parse_apply_compare:NNNNWW

........... 25245, 25254
\_fp_parse_apply_compare_aux:NNWW .... 25266, 25269, 25274
\_fp_parse_apply_function:NNWW

........... 1078, 24701, 24701, 24862
\_fp_parse_apply_unary:NNWW

........... 24706, 24706, 24738, 24853
\_fp_parse_apply_unary_chk:NNNNWWN

........... 24717, 24718, 24721
\_fp_parse_apply_unary_chk:NNNNW

........... 24706
\_fp_parse_apply_unary_chk:NNW

........... 24706, 24708, 24713
\_fp_parse_apply_unary_error:NNW

........... 24706, 24729, 24732, 26768
\_fp_parse_apply_unary_type:NN

........... 24706, 24709, 24727
\_fp_parse_caseless_inf:N .... 24819, 24819
\_fp_parse_caseless_infinity:N

........... 24819, 24820
\_fp_parse_caseless_nan:N

........... 24819, 24821
\_fp_parse_compare:NNNNNN

........... 25186, 25187, 25189,
25191, 25194, 25207, 25215, 25276
\_fp_parse_compare_aux:NNNNNN

........... 25186, 25210, 25218, 25232
\_fp_parse_compare_auxi:NNNNNN

........... 25186,
25223, 25224, 25225, 25226, 25230
\_fp_parse_compare_end:NNNN

........... 25186, 25227, 25241
\_fp_parse_continue:NW

........... 1055, 1056, 1058, 24896, 24899, 24900, 24909,
25085, 25234, 25945, 25953, 25963
\_fp_parse_continue_compare:NNNN

........... 25277, 25292
Index

1572

36865, 37419, 37508, 37550, 39297, 39300, 39362, 39710, 39783, 39786
\c_group_begin_token 114, 200, 208, 718, 898, 3652, 4195, 12916, 12956, 19311, 19334, 19358, 31469, 34809, 34815, 34829, 34835, 34913, 34919, 34934, 34940, 34975, 37435, 37437, 37353
\group_end: ................. 13, 14, 462, 571, 829, 1260, 1264, 1413, 1421, 1423, 2303, 2306, 2312, 2696, 2885, 3062, 3228, 3270, 3479, 3553, 3588, 3609, 3633, 3694, 3876, 4025, 4125, 4484, 4578, 4924, 4932, 5413, 5738, 6035, 6083, 6090, 6998, 6507, 6508, 6915, 7222, 7421, 7449, 7537, 7682, 7729, 7811, 7812, 8625, 8684, 8906, 9103, 9142, 9154, 9368, 9381, 9400, 9600, 10385, 10628, 10699, 10955, 11093, 11597, 12187, 12392, 12442, 12602, 12616, 14331, 13739, 13762, 14268, 14383, 14426, 14731, 14778, 14837, 14899, 15329, 15461, 16090, 17019, 17024, 19170, 19177, 19276, 19281, 19351, 19355, 19383, 19479, 19528, 19884, 21085, 21474, 21500, 21559, 22624, 25021, 29940, 30451, 30660, 30674, 31216, 31217, 31311, 31381, 31423, 31701, 33660, 33693, 34324, 34627, 34772, 36690, 36870, 37423, 37513, 37565, 39316, 39373, 39781, 39800, 40020
\c_group_end_token ............ 200, 898, 3655, 19314, 19334, 19363, 31470, 34823, 34928, 37439, 37357
\group_insert_after:N .......... 14, 1427, 1427, 4480, 36697, 37349, 37350, 37383, 37667
\group_log_list: ............... 14, 2315, 2317
\group_show_list: .............. 14, 2315, 2315
groups commands:
\.\groups:n ................... 243, 22095
\gtoksapp ........................ 836
\gtokspre ........................ 837
H
\H ................................ 65, 32057, 34380, 34400, 34547, 34548, 34575, 34576
\halign ................................ 242
\hangafter ................................ 243
\hangindent ................................ 244
\hbadness ................................ 245
\hbox ................................ 246
hbox commands:
\hbox:n ................................ 302, 306, 34780, 34780, 35007, 35304, 36473
\hbox_gset:N ................. 306, 34782, 34787, 34793, 34974, 35097, 35141, 35161, 35181, 35198, 35219, 35248, 35259, 35316, 35527, 35962, 39539
\hbox_gset:Nw ............ 307, 34800, 34812, 34819, 35600, 39541
\hbox_gset_end: ...................... 307, 34806, 34825, 35603
\hbox_gset_to:wn ............... 307, 34794, 34799, 34805, 39540
\hbox_gset_to:Wn ............... 307, 34826, 34832, 34839, 39542
\hbox_overlap_center:n .......... 307, 34850, 34850
\hbox_overlap_left:n ........... 307, 34850, 34852
\hbox_overlap_right:n ........... 307, 34850, 34854
\hbox_set:N .......................... 302, 306, 307, 321, 34782, 34782, 34792, 34971, 35003, 35004, 35091, 35138, 35158, 35178, 35195, 35216, 35245, 35253, 35277, 35313, 35326, 35334, 35342, 35351, 35360, 35377, 35385, 35393, 35399, 35412, 35514, 35959, 35982, 36229, 36326, 36605, 39458
\hbox_set:Nw .......................... 307, 34806, 34806, 34818, 35587, 39460
\hbox_set_end: ...................... 307, 34806, 34820, 34825, 35590
\hbox_set_to:W:nn ................. 307, 34794, 34794, 34804, 39459
\hbox_set_to:Wn ................. 307, 34826, 34826, 34838, 39461
\hbox_to:wn ......................... 306, 34840, 34846, 34848, 32959
\hbox_to:zero:n ................. 306, 34840, 34845, 34851, 34853, 34855, 36780
\hbox_unpack:N ..................... 307, 34856, 34856, 34858, 36243
\hbox_unpack_drop:N .............. 310, 34856, 34857, 34859
hcoffin commands:
\hcoffin_gset:N ................. 316, 35510, 35523, 35535
\hcoffin_gset:Nw ................. 316, 35583, 35596, 35608
\hcoffin_gset_end: ............... 316, 35583, 35601, 35610
\hcoffin_set:N ................. 316, 317, 35510, 35510, 35522, 36477, 36484, 36522, 36557
\hcoffin_set:Nw .................. 316, 35583, 35583, 35595
Index

\hcoffin_set_end: ................. 316, 35583, 35588, 35609
\hfi .................................. 1144
\hfii ................................ 247
\hfill ................................ 248
\hfilneg ................................ 249
\hfuzz ................................ 250
\hjcode ................................ 828
\hoffset ................................ 251
\holdinginserts ...................... 252

hook commands:
\hook_gput_code:nnn ................... 30583, 30585
\hookpack ................................ 829
\hookrule ................................ 253
\hoysize ................................ 254
\hookip ................................ 255
\has .................................. 256
\ht .................................. 257
\huge .................................. 34298
\huge .................................. 34302
\hyphenation ......................... 258
\hyphenationbounds ................... 830
\hyphenationmin ....................... 831
\hyphenchar ................................ 259
\hyphenpenalty ...................... 260
\hyphenpenaltymode .................. 832

I
\i ...................................... 33658,
34355, 34456, 34458, 34460, 34462,
34513, 34516, 34519, 34522, 34593

if commands:
\if:w .................................. 28, 29, 195, 387, 388, 420,
488, 692, 707, 708, 711, 720, 721,
728, 1392, 1398, 1804, 2173, 2174,
2774, 2777, 2778, 2779, 2780, 2795,
2796, 2797, 2798, 2799, 2800, 2801,
2802, 2803, 2867, 2868, 2870, 4083,
4719, 4748, 4802, 11118, 12560,
12570, 12663, 12970, 12990, 13005,
13553, 13560, 13565, 18153, 19645,
21091, 21106, 21107, 23972, 24345,
24349, 24371, 24464, 24496, 24515,
24581, 24595, 24612, 24633, 24672,
24684, 24970, 25012, 25132, 25147,
25551, 27725, 27775, 29236, 31247,
31250, 31272, 31453, 39151, 39163,
39165, 39246, 39247, 39248, 39249,
39263, 39264, 39274, 39275, 39276,
39277, 39278, 39279, 39280, 39281,
39282, 40049, 40051, 40055, 40057
\ifbool:N ................................ 72, 584, 1402, 8279, 8330

if commands:
\if_box_empty:N ...................... 314, 34718, 34720, 34730
\if_case:w ............................. 179, 745,
774, 786, 857, 1032, 1125, 1180,
1224, 2080, 3663, 3862, 4049, 4431,
4703, 5517, 5546, 5603, 6275, 6328,
6947, 6994, 7092, 7409, 7920, 7931,
10784, 13832, 13906, 14244, 15275,
17540, 17454, 18044, 18077, 19267,
23273, 23526, 23541, 23912, 23941,
25220, 25261, 25988, 26123, 26198,
26223, 26275, 26719, 26755, 26752,
27030, 27257, 27284, 27442, 27477,
27635, 27680, 27731, 27857, 27880,
27913, 27972, 28012, 28027, 28042,
28057, 28072, 28087, 28013, 28666,
28750, 28765, 28817, 28830, 28914,
28968, 29042, 29233, 30062, 30141
\if_catcode:w .......................... 29, 718, 719,
897, 912, 1432, 1400, 2914, 3652,
3655, 3810, 3812, 3814, 3816, 3818,
3820, 3822, 4050, 4051, 4195, 12911,
12954, 19290, 19293, 19296, 19299,
19302, 19305, 19308, 19311, 19314,
19317, 19358, 19363, 19368, 19373,
19380, 19387, 19392, 19397, 19402,
19407, 19412, 19419, 19446, 19755,
19814, 19819, 19866, 19867, 23172,
24100, 24305, 24623, 24670, 24969,
25011, 31411, 31412, 31454, 31469,
31470, 31471, 31472, 31473, 31474,
31475, 31476, 31477, 31500, 31503,
31506, 31509, 31512, 31515, 31518
\if_charcode:w ........................ 29, 195, 455, 718, 719, 749,
912, 1392, 1399, 3722, 3746, 3795,
4091, 4128, 4130, 4640, 4650, 5161,
5716, 6884, 6887, 11378, 11387,
12897, 12947, 13990, 14224, 14888,
19424, 19816, 23529, 25564, 25927
\if_cs_exist:N ...................... 29, 1407, 1407,
1832, 1870, 2690, 19218, 19454, 19654
\if_cs_exist:w ........................ 29, 1407, 1408,
1435, 1842, 1855, 1884, 1899, 1924,
1935, 2068, 4236, 11110, 18298,
18323, 18334, 20258, 20305, 21610
\if_dim:w ................................ 238, 20746,
20746, 20834, 20846, 20869, 21040
\if_eof:w ................................ 99,
644, 10322, 10322, 10327, 10412, 10430
\iffalse: ................................ 28, 65, 208, 446, 467, 533, 565,
566, 569, 596, 665, 701, 796, 710,
717, 831, 848, 895, 936, 1392, 1393,
Index

7387, 7504, 7617, 8404, 8426, 10720, 11015, 12550, 12603, 12617, 12938, 13340, 13382, 13395, 13665, 13743, 13766, 13927, 13965, 14396, 15123, 15272, 15287, 15314, 15429, 16362, 16368, 16394, 16406, 16414, 16446, 16453, 16477, 16481, 16559, 16595, 16613, 16927, 16959, 17014, 17029, 17037, 17473, 17476, 17486, 17521, 17526, 17527, 17679, 18536, 18551, 18573, 18587, 19451, 19490, 19493, 19628, 19730, 19758, 19769, 19807, 19865, 20219, 20316, 20464, 20952, 20512, 20815, 20862, 21043, 23255, 23276, 23288, 23298, 23393, 23448, 23457, 23548, 23563, 23565, 23703, 23791, 23801, 23813, 23826, 23827, 23846, 23847, 23861, 23862, 23873, 23874, 23940, 23987, 24022, 24025, 24041, 24048, 24101, 24104, 24220, 24221, 24222, 24223, 24226, 24322, 24435, 24441, 24671, 24715, 24926, 24997, 25258, 25276, 25326, 25336, 25622, 25623, 25624, 25625, 25626, 25627, 25849, 25861, 25862, 25890, 25902, 25909, 25926, 25989, 26024, 26038, 26084, 26091, 26167, 26179, 26279, 26282, 26293, 26346, 26419, 26489, 26492, 26499, 26532, 26533, 26536, 26757, 27003, 27195, 27205, 27254, 27353, 27433, 27482, 27496, 27643, 27677, 27689, 27702, 27705, 27708, 27711, 27736, 27837, 27841, 27900, 27917, 27923, 28507, 28560, 28611, 28612, 28614, 28615, 28635, 28652, 28719, 28817, 28893, 28967, 29041, 29111, 29116, 29222, 29254, 29265, 29372, 29946, 30075, 30128, 30134, 31210, 31426, 32040, 32904 \if_mode_horizontal: 29, 1403, 1404, 8613
\if_mode_inner: 29, 1403, 1404, 8615
\if_mode_math: 29, 1403, 1404, 8617
\if_mode_vertical: 29, 1403, 1405, 2385, 8611
\if_predicate:w 63, 65, 72, 8279, 8279, 8381, 8442, 8457, 8458, 8483, 8494
\if_true: 28, 66, 1392, 1392, 1887, 1893, 1903, 1909, 12710, 12714, 12996, 13002
\if_vbox:N 314, 34718, 34719, 34724 936
\ifabsdim 937
\ifabsnum 937
\ifcase 262
\ifcat 263
\ifcondition 838
\ifsname 366, 663, 501
\ifdbox 1145
\ifdir 1146
\ifdefined 502
\ifdim 264
\ifDocumentMetadataTF 38711, 38712
\ifeof 265
\iffalse 266
\IfFileExists 666
\iffontchar 503
\ifbox 267
\ifmode 268
\ifinsname 672
\ifinner 269
\iffont 1147
\ifbox 1148
\ifdir 1149
\ifmode 270
\ifnum 10, 22, 52, 56, 271
\ifodd 272
\ifpdfabsdim 624
\ifpdfabsnum 625
\ifpdfprimitive 626
\ifprimitive 775
\iftbox 1150
\iftdir 1152
\iftfont 1151
\ifttrue 273
\ifvbox 274
\ifvmode 275
\ifvoid 276
\fix 4, 8, 13, 17, 53, 54, 61, 277
\ifybox 1153
\ifydir 1154
\ignorereligaturesinfont 938
\ignorespaces 278
\IJ 32065, 33649, 34345
\ij 32065, 33649, 34357
\immediate 68, 279
\immediateassigned 839
\immediateassignment 840
in 280
\indent 280
\inf 279
\infty 24223, 24224
\inherit commands: .\inherit:n 243, 22097
\inhbitglue 1155
\inhbitxspcode 1156
\initcatcodetable 841
Index

174  73, 174, 853, 1105, 7697, 7705, 17828, 17829, 17862, 17864, 18295, 40004, 40008, 40012, 40016

\int_step_function:nnN ............ 174, 1013, 16985, 17865, 17866, 17869, 23006, 30400, 30434, 30491, 31136, 31179

\int_step_inline:nn ............ 174, 2326, 6597, 8269, 9120, 9122, 9124, 10197, 10450, 14579, 14588,
                              17865, 17867, 30407, 30410, 30668

\int_step_inline:nnn ............ 174, 1107, 17865, 17866, 17868, 17869

\int_step_variable:nnN ............ 174, 17865, 17876

\int_step_variable:nnnn ............ 174, 17865, 17878

\int_step_variable:nnnnn ............ 174, 17865, 17879, 17880

\int_to_symbols:nnnn ............ 170, 4415, 5117, 6458, 6466, 6475, 9350, 10887,
                                17610, 17612, 17620, 39454, 39836

\int_to_Alph:n ............ 175, 176, 17921, 17953

\int_to_Alph:n ............ 175, 176, 17921, 17921

\int_to_arabic:n ............ 175, 17898, 17899

\int_to_Base:n ............ 176

\int_to_Base:n ............ 176

\int_to_Base:n ............ 176, 177, 17985, 17987, 18112

\int_to_base:n ............ 176, 177, 17985, 18108, 18110, 18114

\int_to_base:n ............ 176, 1867107, 18107

\int_to_Hex:n ............ 176, 177, 4626, 18107, 18111, 31267, 37588

\int_to_hex:n ............ 176, 177, 17985, 18108, 18110, 18114

\int_to_oct:n ............ 176, 177, 17985, 18107, 18107

\int_to_Roman:n ............ 176, 177, 18107, 18113

\int_to_Roman:n ............ 176, 177, 18115, 18125

\int_to_Roman:n ............ 176, 177, 18115, 18115

\int_to_symbols:n ............ 175, 17900, 17900, 17916, 17923, 17955

\int_until:do:n ............ 173, 17772, 17780, 17785

\int_until:do:n ............ 173, 17800, 17808, 17813

\int_use:n \165, 167, 170, 1045, 1051, 2329, 2331, 2333, 3894, 4022, 4042,
            4913, 5000, 5078, 5098, 5102, 5113, 5114, 5120, 5121, 5127, 5128,
            5285, 6114, 6204, 6209, 6230, 6232, 6337, 6350, 6351, 6751, 6803, 6901,
            6912, 7067, 7450, 7534, 7535, 7726, 7727, 8263, 8734, 8736, 8739, 8755,
            8757, 8761, 8764, 8769, 9310, 10009,
            1578

\int_range:nn ............ 177, 177, 1231, 13132, 17170, 18204, 19014, 19019, 20320, 20320

\int_rand:n ............ \177, 22928, 23112, 230414, 29414

\int_set_eq:NN ............ 1222

\int_set:Nn ............ 169, 365, 2310, 2324, 2325, 2330, 2332, 2333, 3134,
            3156, 3157, 3172, 3180, 3181, 3193, 3194, 3212, 3215, 3637, 3700, 4044,
            4140, 4143, 4273, 5594, 6132, 6196, 6199, 6237, 6239, 6308, 6359, 6360,
            6370, 6381, 6405, 6423, 6472, 6604, 6606, 6630, 6675, 6674, 6715, 6750,
            7450, 7522, 7524, 7668, 7694, 7717, 7719, 10358, 10360, 10581, 10583,
            10602, 10612, 10625, 10672, 10678, 10690, 10695, 12374, 12409, 14720,
            14769, 14822, 16998, 17634, 17634,
            17638, 21775, 23167, 30480, 30481,
            30496, 30667, 30682, 30718, 34767,
            34768, 34769, 34770, 39457, 39832

\int_set_eq:NN ............ 169, 3173, 3203, 4411, 4881, 4885, 4894, 4896,
            4939, 5006, 5302, 5402, 5514, 5514,
            6152, 6172, 6189, 6194, 6214, 6248,
            6249, 6299, 6402, 6403, 6455, 6504,
            6582, 6605, 6609, 6610, 6642, 6628,
            6631, 6670, 6680, 6811, 6812, 7401,
            7618, 7911, 8801, 11076, 12372,
            12375, 17602, 17602, 17603, 39452

\int_show:N ............ 177, 18256, 18256, 18257

\int_show:n ............ 178, 628, 863, 18258, 18258

\int_sign:n ............ 169, 951, 17465, 17465, 30563, 39896

\int_step_function:nn ............ 174, 17828, 17861, 30686, 30687

\int_step_function:nnn ............ 174, 6681, 7518, 17828, 17863, 19260,
                                   30688, 30689, 30690, 30691, 30712
int internal commands:
\_\_int_abs:N ........ 17479, 17481, 17485
\_\_int_case:nnN ........ 17726, 17729, 17734, 17739, 17744, 17746
\_\_int_case:nn ........ 17726, 17747, 17748, 17752
\_\_int_case_end:nn .... 17726, 17751, 17754
\_\_int_compare:nn .... 849, 17657, 17689, 17697, 17699, 17701, 17703, 17705, 17707, 17709
\_\_int_compare:NN .... 848, 17657, 17669, 17673
\_\_int_compare:NN .... 848, 17657, 17665, 17667, 17669
\_\_int_compare:w .... 847, 17657, 17659, 17662
\_\_int_compare:=:NNw ... 17657
\_\_int_compare:<:NNw ... 17657
\_\_int_compare:>:NNw ... 17657
\_\_int_compare;<=:NNw ... 17657
\_\_int_compare;=:NNw ... 17657
\_\_int_compare;!=:NNw ... 17657
\_\_int_compare:w .... 849, 17657
\_\_int_compare_error....... 17511, 17514, 17519, 17542
\_\_int_compare_error:NNw ........ 847, 848, 17642, 17674, 17682
\_\_int_const:nN .......... 17556, 17557, 17559, 17579
\_\_int_constdef:Nw . 17556, 17574, 17585, 17586, 17587, 17589, 17590
\_\_int_div_truncate:NwNW .... 17511, 17514, 17519, 17542
\_\_int_eval:w ........ 365, 842, 843, 848, 17450, 17451, 17463, 17464, 17468, 17482, 17490, 17491, 17498, 17499, 17513, 17515, 17516, 17533, 17536, 17537, 17538, 17545, 17577, 17611, 17613, 17615, 17617, 17635, 17637, 17660, 17694, 17712, 17720, 17758, 17766, 17831, 17832, 17833, 17859, 18044, 18071, 18077, 18104, 39840, 39898, 39945, 39969, 39985, 39986, 40007, 40011, 40015
\_\_int_eval_end: ........ 17450, 17452, 17463, 17468, 17482, 17517, 17533, 17539, 17548, 17577, 17611, 17613, 17615, 17617, 17635, 17637, 17712, 17758, 17766, 18044, 18071, 18077, 18104
\_\_int_from_alph:N ........ 869, 18161, 18174, 18176

4893, 5115, 5551, 6078, 6151, 6182, 6603, 6880, 7395, 7396, 7444, 7631, 7906, 10732, 16097, 17592, 17594, 17597, 21767, 22933, 23088, 23117, 29896, 30086, 37976, 39451
\_\_int_zero_new:N .......... 169, 17596, 17596, 17600
\c_max_char_int ............. 178, 4623, 18266, 19201, 19942
\c_max_int ... 178, 255, 421, 1222, 4050, 4051, 4052, 18265, 29367, 34745, 34751, 36646, 36649
\c_max_register_int .......... 178, 434, 4453, 3132, 3157, 3194, 10009, 10009, 16972, 17450
\c_one_int ................. 178, 3740, 4054, 4700, 4823, 5386, 5892, 5900, 5943, 6088, 6194, 6309, 6382, 6393, 6404, 6407, 6424, 6446, 6473, 6598, 6601, 6609, 6630, 6683, 6700, 6701, 6771, 6773, 6774, 6848, 7018, 7031, 7053, 7519, 7700, 7708, 8271, 17623, 17625, 17627, 17629, 18264, 18303, 18398, 19860, 23144, 23163, 26616, 26620, 26624, 26678, 27270, 27413, 27554, 27860, 28700, 28784, 29226, 29229, 29236, 29421, 29949
\g_tmapa_int ............ 178, 18276
\l_tmapa_int ........ 4, 178, 18276
\l_tapb_int ........ 178, 18276
\l_tampb_int ........ 4, 178, 18276
\l_zero_int 178, 376, 388, 715, 1444, 1807, 1809, 3763, 3843, 2944, 4056, 4316, 4458, 4470, 4911, 5409, 5664, 5805, 5841, 5899, 6081, 6180, 6202, 6293, 6319, 6380, 6420, 6471, 6533, 6684, 6781, 6989, 6970, 7038, 7109, 7118, 7129, 7161, 7170, 7187, 7210, 7480, 7496, 7530, 7563, 7571, 7622, 7624, 7696, 7719, 7720, 7723, 8891, 9061, 11076, 11398, 12372, 13170, 13219, 13543, 13548, 13871, 13906, 14634, 15433, 17546, 17547, 17561, 17592, 17593, 17644, 17652, 17720, 17837, 17840, 18264, 19200, 19866, 19867, 19868, 20885, 21063, 23007, 23131, 23566, 23792, 23796, 23798, 23802, 23806, 23819, 23832, 23839, 23852, 23864, 23876, 23885, 23888, 23892, 24005, 24010, 25036, 25558, 26262, 26267, 26679, 26680, 26681, 27447, 27514, 27737, 27755, 27778, 27810, 28536, 29016, 29208, 29237, 29337, 29401, 29922, 30228
\__int_from_alph:nN  
\__int_from_base:N  
\__int_from_base:nN  
\__int_from_roman:NN  
\__int_to_Base:nnN  
\__int_to_base:nn  
\__int_step:wwwN  
\__int_sign:Nw  
\__int_show:nN  
\__int_pass_signs_end:wn  
\__int_pass_signs:wn  
\__int_mod:ww  
\__int_maxmin:wwN  
\__int_from_roman_error:w  
\__int_from_base:nnN  
\__int_from_base:N  
\__int_from_alph:nN  
\intarray:NN  
\intarray_item:Nn  
\intarray_if_exist_p:N  
\intarray_if_exist:NTF  
\intarray_if_exist:N  
\intarray_gset:Nnn  
\intarray_const_from_clist:Nn  
\intarray_gzero:N  
\intarray_count:N  
\intarray_gzero:w  
\intarray_item:Nn  
\intarray_if_exist_p:N  
\intarray_if_exist:NTF  
\intarray_if_exist:N  
\intarray_from_roman_x:w  
\intarray_from_roman_X:w  
\intarray_from_roman_v:w  
\intarray_from_roman_V:w  
\intarray_from_roman_m:w  
\intarray_from_roman_l:w  
\intarray_from_roman_L:w  
\intarray_from_roman_i:w  
\intarray_from_roman_d:w  
\intarray_from_roman_c:w  
\intarray_from_roman_aux:N  
\intarray_from_roman_error:w  
\intarray_from_roman:w  
\intarray_from_roman:N  
\intarray_from_letter:n  
\intarray_from_Letter:n  
\intarray_from_base:nnN  
\intarray_from_base:N  
\intarray_from_alph:nN  
\intarray_commands:  
\intarray_count:N  
\intarray_gzero:w  
\intarray_item:Nn  
\intarray_if_exist_p:N  
\intarray_if_exist:NTF  
\intarray_if_exist:N  
\intarray_from_roman_x:w  
\intarray_from_roman_X:w  
\intarray_from_roman_v:w  
\intarray_from_roman_V:w  
\intarray_from_roman_m:w  
\intarray_from_roman_l:w  
\intarray_from_roman_L:w  
\intarray_from_roman_i:w  
\intarray_from_roman_d:w  
\intarray_from_roman_c:w  
\intarray_from_roman_aux:N  
\intarray_from_roman_error:w  
\intarray_from_roman:w  
\intarray_from_roman:N  
\intarray_from_letter:n  
\intarray_from_Letter:n  
\intarray_from_base:nnN  
\intarray_from_base:N  
\intarray_from_alph:nN  
\intarray_commands:
\ior_shell_open:Nn .......................... 91, 367, 10260, 10260, 11645
\ior_show:N .......................... 92, 10294, 10294, 10295
\ior_show_list: .................. 92, 10310, 10310
\ior_str_get:NN ...................... 92, 93, 96, 10352, 10352, 10363, 10421
\ior_str_get:NNTF .................. 93, 10352, 10353
\ior_str_get_term:nN .......... 96, 10373, 10375
\ior_str_map_inline:Nn ............ 94, 10392, 10394, 31270, 31299, 31372
\ior_str_map_variable:NNNN ...... 94, 10418, 10420
\g_tampa_iorn .................. 99, 10212
\g_tamp_iorn .................. 99, 10212
ior internal commands:
  \l__ior_file_name_tl ............... 10217, 10220, 10222
  \_ior_get:NN ...................... 10339, 10341, 10348, 10374, 10393
  \_ior_get_term:NN ................ 10373, 10374, 10376, 10377
  \l__ior_internal_tl ............... 10192, 10302, 10305, 10411, 10415
  \_ior_list:N ...................... 10310, 10310, 10311, 10312
  \_ior_map_inline:NNN ............ 10392, 10393, 10395, 10396
  \_ior_map_inline:NNNn ............ 10372, 10399, 10402
  \_ior_map_inline_loop:NNNN ...... 10392, 10405, 10409, 10416
  \_ior_map_variable:NNNNn ...... 10418, 10419, 10421, 10422
  \_ior_map_variable_loop:NNNNn ... 10418, 10424, 10427, 10434
  \_ior_new:N ..................... 640, 10228, 10228, 10232, 10233, 10245
  \_ior_new_aux:N ................ 10232, 10236
  \l__ior_open_stream:Nn ........... 10239, 10243, 10247, 10251
  \_ior_show:nN .................. 10260, 10263, 10266, 10275
  \_ior_str_get:NN .................. 10294, 10294, 10296, 10298
  \_ior_str_get:NN ................. 10352, 10354, 10368, 10376, 10395
  \l__ior_stream_tl ............... 10195, 10242, 10246, 10253
  \g__ior_streanm_prop .......... 644, 10196, 10254, 10287, 10302, 10317
  \g__ior_streanm_seq .......... 10194, 10242, 10288, 10289
  \c__ior_term_iorn ............. 10193, 10210, 10284, 10290, 10326, 10383
  \c__ior_term_noprompt_iorn ... .......................... 10372, 10382
iow commands:
  \iow_char:N .................. 83, 97, 3491, 3494, 3495, 3519, 3520, 3527, 3528,
                            4597, 4598, 4605, 4607, 4609, 4611,
                            4613, 4615, 5259, 5260, 5994, 6001,
                            6002, 6003, 6127, 7953, 7956, 7957,
                            7962, 7996, 8005, 8009, 8014, 8034,
                            8036, 8037, 8039, 8042, 8044, 8049,
                            8051, 8053, 8058, 8062, 8065, 8066,
                            8069, 8071, 8075, 8077, 8083, 8085,
                            8089, 8091, 8095, 8100, 8102, 8144,
                            8146, 8151, 8153, 8159, 8164, 8169,
                            8173, 8183, 8186, 8190, 8191, 8195,
                            8203, 8274, 9855, 9858, 9859, 9891,
                            9919, 10077, 10060, 10060, 11175,
                            11177, 11178, 11179, 14698, 27674,
                            30735, 30737, 30738, 30741, 30743,
                            30744, 30747, 30749, 30750, 30751,
                            30755, 30762, 39169, 40039, 40040
  \iow_close:N .................. 91, 92, 10492, 10526, 10526, 10537
  \iow_indent:n ................ 98,
                            653, 654, 8230, 9812, 9953, 10024,
                            10032, 10048, 10050, 10050, 10053,
                            10065, 10082, 10087, 14696, 15021,
                            15209, 23735, 23747, 38288, 38317,
                            38339, 38362, 38371, 38380, 38401
  \l__ior_line_count_int ........ 98, 99, 461, 654, 3993, 3997, 9350, 10601,
                            10601, 10609, 10696, 10734
  \iow_log:N .... 92, 10538, 10540, 10541
  \iow_log:n ..................... 96,
                            365, 1975, 1975, 9550, 9557, 10563,
                            10593, 10594, 10595, 13271, 39177
  \iow_log_list: ................. 92, 10554, 10555
  \iow_new:N ........................ 91, 10467, 10467, 10468, 10469, 10470
  \iow_newline: ................. 83, 96–98, 368, 618,
                            650, 728, 3943, 6049, 9372, 10599,
                            10599, 10679, 10688, 10694, 11174,
                            25467, 25468, 30633, 36634, 36635
  \iow_now:N ....... 96, 97, 8938, 10586, 10586, 10589, 10591,
                            10592, 10593, 10594, 10596, 10597
  \iow_open:N .... 91, 10483, 10483, 10489
  \iow_shell_open:Nn ............ 91, 367, 10510, 10510
  \iow_shipout:N ................ 96, 97,
                            650, 8970, 10569, 10569, 10571, 10572
  \iow_shipout_e:N ............ 96,
                            97, 10566, 10566, 10568, 38805, 38806

Index
1583
iow internal commands:

\_iow\_shipout\_x:Nn ............... 650, 38805, 38806, 38807
\_iow\_show:N .................. 92, 10538, 10538, 10539
\_iow\_show\_list: ........... 92, 10554, 10554
\_iow\_term:n ................. 96, 1975, 1977, 8264, 9384, 9450, 9455, 9563, 9589, 10593, 10596, 10597, 10598
\_iow\_wrap:nnN 96–99, 628, 654, 728, 9348, 9351, 9363, 9521, 9555, 9561, 9568, 10642, 10648, 10653, 10665, 10668, 10686, 10702, 13255, 13271
\_iow\_wrap\_allow\_break: . 98, 10639, 10639, 10642, 10648, 10681, 10686
\_iow\_wrap\_break:n ........ 653
\_c\_log\_iow .................... 65, 65, 65, 10438, 10528, 10593, 10594
\_c\_term\_iow ........ 99, 645, 646, 10438, 10467, 10528, 10534, 10596, 10597
\_g\_tpma\_iow .................. 99, 10469
\_g\_tpmb\_iow .................. 99, 10469

\_l\_iow\_file\_name\_tl ............... 10482, 10485, 10487
\_\_iow\_indent:n ............... 653, 10560, 10566, 10682
\_\_iow\_indent\_error:n ........ 653, 10560, 10662, 10687
\_l\_iow\_indent\_int ........... 10618, 10732, 10750, 10862, 10879, 10887
\_l\_iow\_indent\_tl ........... 10618, 10733, 10749, 10861, 10880, 10888, 10897
\_l\_iow\_internal\_tl ............. 10437, 10546, 10549
\_l\_iow\_line\_break\_bool ........ 10622, 10728, 10856, 10870, 10878, 10886, 10894, 10896, 10901, 10903
\_l\_iow\_line\_part\_tl ............. 656, 657, 659, 10620, 10730, 10742, 10763, 10821, 10824, 10855, 10869, 10871, 10877, 10885, 10908
\_l\_iow\_line\_target\_int ............ 659, 10604, 10690, 10692, 10695, 10857, 10862, 10897
\_l\_iow\_line\_tl ........... 10620, 10729, 10746, 10836, 10852, 10868, 10869, 10877, 10885, 10907, 10908, 10913, 10915
\_\_iow\_list:N .................. 10554, 10554, 10555, 10556
\_\_iow\_new:N ................. 10471, 10471, 10475, 10476, 10496
\_\_iow\_new\_aux:N .......... 10475, 10479
\_l\_iow\_newline\_tl ........... 10603, 10608, 10689, 10691, 10694, 10912
\_l\_iow\_one\_indent\_int ............... 10605, 10879, 10887
\_l\_iow\_one\_indent\_tl ............... 10605, 10695, 10880
\_\_iow\_open\_stream:Nn ............... 10483, 10494, 10498, 10502, 10509
\_\_iow\_set\_indent:n ............... 651, 10605, 10608, 10617
\_\_iow\_shell\_open:nN ............... 10510, 10513, 10516, 10525
\_\_iow\_show:N .................. 10538, 10538, 10540, 10542
\_l\_iow\_stream\_tl ............ 10448, 10493, 10497, 10504
\_g\_iow\_streams\_prop ........ 649, 10449, 10505, 10531, 10546, 10561
\_g\_iow\_streams\_seq .......... 10447, 10493, 10532, 10533
\_\_iow\_tmp:w .................. 657, 10736, 10760, 10817, 10849, 10917, 10925
\_\_iow\_unindent:w .............. 651, 10605, 10607, 10615, 10889
\_\_iow\_use\_i\_delimit\_by\_s\_stop:nw ........ 10456, 10465, 10572
\_\_iow\_with:nN ................. 10573, 10577, 10579, 10585
\_\_iow\_wrap\_allow\_break: ......... 653, 10639, 10644, 10681
\_\_iow\_wrap\_allow\_break:n ......... 10866, 10866
\_\_iow\_wrap\_allow\_break\_error: ........ 653, 10639, 10645, 10686
\_c\_iow\_wrap\_allow\_break\_marker\_tl ......... 10624, 10644
\_\_iow\_wrap\_break:w .......... 10803, 10817, 10819
\_\_iow\_wrap\_end\_w .......... 657, 10817, 10826, 10846
\_\_iow\_wrap\_first\_w .......... 10817, 10823, 10829
\_\_iow\_wrap\_loop:w .......... 10817, 10832, 10840, 10844
\_\_iow\_wrap\_break\_none:w ........ 10817, 10831, 10834
\_\_iow\_wrap\_chunk:nw .............. 10734, 10736, 10738, 10872, 10873, 10881, 10890, 10897
\_\_iow\_wrap\_do: .......... 10698, 10703, 10703
\_\_iow\_wrap\_end:n .......... 10892, 10899
\_\_iow\_wrap\_end\_chunk:w ........ 655, 10754, 10761, 10811, 10853
\_c\_iow\_wrap\_end\_marker\_tl ......... 10624, 10708

Index

1584
\_kernel chk var global: N ................................. 364, 1477, 39083, 39088, 39131, 39144, 39513
\_kernel chk var local: N ................................. 364, 1477, 39083, 39087, 39125, 39143, 39433
\_kernel chk var scope: NN ................................. 364, 1477, 39083, 39089, 39120, 39145, 39294, 39262, 39630, 39635, 39641, 39645, 39649
\_kernel codepoint case: nn ......................... 371, 14110, 31312, 31312, 32544
\_kernel codepoint data: nn ......................... 371, 30946, 31218, 31218, 31249, 31332
\_kernel codepoint_to_bytes: n ......................... 365, 15480, 30803, 30835, 30863, 30863, 38944
l kernel color_stack_int ......................... 373
\_kernel cs parm from arg count: nnTF ......................... 365, 1641, 2075, 2075, 2122
\_kernel debug_log: n ................................. 365, 1479, 39174, 39176, 39180, 39181, 39565, 39664, 39767, 39685
\_kernel dependency_version check: Nn ............... 366, 11624, 11624
\_kernel dependency_version check: nn ............... 366, 11624, 11625, 11626
\_kernel deprecation_code: n ......................... 366, 1465, 1480, 1583, 1585, 38748, 38774, 38781, 38782
\_kernel deprecation_error: Nn .......................... 1465, 38751, 38784, 38784
\_kernel exp not: w ................................. 366, 415, 446, 463, 715, 735, 938, 2670, 2670, 2672, 2674, 2676, 2679, 2684, 4073, 12141, 12167, 12168, 12175, 12176, 12190, 12192, 12194, 12196, 12208, 12213, 12218, 12224, 12225, 12232, 12233, 12239, 12244, 12249, 12255, 12256, 12263, 12264, 12280, 12284, 12289, 12295, 12296, 12303, 12304, 12308, 12312, 12317, 12233, 12321, 12332, 12510, 12820, 12874, 12879, 12887, 13083, 13246, 13324, 13330, 13345, 13363, 13401, 13452, 13457, 13462, 17675, 20620, 20635, 21353, 30789, 30828, 30842, 30859, 31715, 32999, 34060
\l kernel expl bool ................................. 105, 108, 122, 135, 1391
\c kernel expl_date_tl ................................. 680, 1391, 11628, 11631, 11667, 11671
\_kernel file input pop: ................................. 366, 14433, 11473
\_kernel file input push: n ................................. 366, 14433, 11467
\_kernel file missing: n ................................. 366, 10215, 11428, 11428, 11437
\_kernel file name_quote: n ................................. 366, 10215, 11428, 11428, 11437
\_kernel file name sanitize: n ................................. 366, 675, 10486, 10971, 10971, 11100, 11431, 11499, 11508
\_kernel group show: NN ................................. 2315, 2316, 2318, 2319
\_kernel if debug: TF ................................. 1568, 1568, 38762, 40073, 40073
\_kernel int add: nnn ................................. 366, 17543, 17543, 29367
\_kernel int_array_gset: Nnn ................................. 367, 1007, 1009, 1012, 6600, 6706, 6709, 7407, 7479, 7481, 7487, 7495, 7497, 7500, 7621, 7623, 7627, 7629, 7641, 7644, 22866, 22867, 22937, 23007, 23019, 23044, 23044, 23059, 23127, 23175, 30073, 30074, 30076, 30080, 30081, 30082, 30401, 30402, 30436, 30439, 31158
\_kernel int_array_gset range from clist: Nnn ................................. 367, 6769, 22964, 22965, 23165, 23165
\_kernel int_array_item: Nn ................................. 367, 1008, 1013, 1195, 4319, 6717, 6744, 6688, 6688, 6695, 6929, 6948, 7484, 7675, 7894, 22890, 22915, 23006, 23096, 23108, 23142, 23161, 27654, 27660, 27663, 27666, 28399, 28402, 28405, 28408, 28411, 28414, 28417, 28420, 28423, 30122, 30123, 30124, 30126, 30129, 30149, 30197
\_kernel int_array_range to clist: Nnn ................................. 367, 6698, 22994, 22950, 23146, 23146
\_kernel ior open: Nn ................................. 367, 641, 10222, 10239, 10250, 10273
\_kernel iow open: Nn ................................. 367, 10483, 10487, 10490, 10501, 10523
\_kernel iow with: Nnn ................................. 368, 618, 650, 728, 9385, 9387, 9591, 9593, 10573, 10573, 10588, 13260, 13262
\_kernel kern: N ................................. 368, 1391, 13943, 34634, 35006, 35297, 35306, 35528, 35330, 35379, 35381, 35984, 36242, 36247, 36329, 36330, 36608, 36609

Index
1586
\_1\_kernel_keyeval\_allow\_blank:NN
\_keys\_bool:
20165, 20171, 20173, 21447, 21508
\_kernel\_msg\_error:nnn:9762, 9768
\_kernel\_msg\_error:nnn:9762, 9770
\_kernel\_msg\_error:nnnn:9762, 9772
\_kernel\_msg\_expandable\_error:nnn:9774, 9774, 9776
\_kernel\_msg\_expandable\_error:nnnn:9774, 9774, 9778
\_kernel\_msg\_info:nnnn:9762, 9762
\_kernel\_msg\_log\_eval:NN
368, 8851, 9753, 9755, 18263, 21172, 21265, 21333, 25531
\_kernel\_msg\_new:nnn:9758, 9760, 9986
\_kernel\_msg\_new:nnnn:9758, 9758
\_kernel\_msg\_show\_eval:NN
368, 8349, 9753, 9753, 18259, 21168, 21261, 21292, 25529
\_kernel\_msg\_warning:nnn:9762, 9764
\_kernel\_msg\_warning:nnnn:9762, 9766
\_kernel\_patch:NN
39784, 39866, 39867, 39915, 39950, 39974, 39988, 40004, 40019
\_kernel\_patch:nnnn
1483, 39297, 9360, 39660, 39690, 39694, 39698, 39702
\_kernel\_patch\_weird\_aux:nnnn
39297, 39364, 39366
\_kernel\_pdf\_object\_id:nn
368, 38463, 38480
\_kernel\_pdf\_object\_id\_indexed:nn
368, 38542, 38560
\_kernel\_prefix\_arg\_replacement:VN
2336, 2338, 2346, 2355, 2364
\_g\_kernel\_prg\_map\_int
368, 458, 712, 853, 950, 1391, 3892, 3894, 3903, 4020, 4022, 4030, 4042, 8626, 10398, 10400, 10407, 12712, 12714, 12716, 12721, 13673, 13675, 13679, 13684, 17213, 17214, 17220, 17221, 17271, 17273, 17275, 17277, 17871, 17874, 17882, 17885, 17896, 18780, 18782, 18784, 18789, 20569, 20570, 20575, 20577, 21007, 21010, 21017, 21028, 25812, 25815, 25819, 25822, 25833, 34046, 34048, 34050, 34052
\_kernel\_primitive:NN
\keys_if_exist_p:nnn \keys_key_str 249, 2172, 21841, 21842, 22344, 22345, 22441, 22445, 22470, 22473, 22474, 22522, 22579
\keys_key_str ........................................ 21573, 21841, 21842, 22345
\keys_log:nn ........................................ 251, 22659, 22661
\keys_path_str ........................................ 249, 976, 21577, 21645, 21663, 21682, 21699, 21717, 21719, 21721, 21724, 21736, 21739, 21743, 21751, 21753, 21754, 21757, 21772, 21788, 21801, 21805, 21816, 21819, 21827, 21833, 21837, 21840, 21844, 21855, 21857, 21859, 21862, 21873, 21882, 21887, 21889, 21904, 21914, 21920, 21924, 21939, 21948, 21990, 22001, 22044, 22335, 22343, 22381, 22384, 22420, 22424, 22429, 22438, 22452, 22454, 22455, 22459, 22467, 22502, 22532, 22555, 22567, 22576
\keys_path_ttl 21578, 21682, 21743
\keys_precompile:nnn 251, 22315, 22315
\keys_set:nn ........ 246, 242, 249, 248–251, 22217, 22217, 22247, 22257, 22319
\keys_set_exclude_groups:nnn .................. 250, 22275, 22289, 22294, 38825, 38826
\keys_set_exclude_groups:nnnn ................. 250, 22275, 22286, 22288, 38828, 38829
\keys_set_exclude_groups:nnnn ................. 250, 22275, 22285, 22287, 22291, 38831, 38832
\keys_set_filter:nnn ......................... 38825, 38826, 38827
\keys_set_filter:nnnn 38825, 38829, 38830
\keys_set_filter:nnnN .................................. 38825, 38832, 38833
\keys_set_groups:nnn ......................... 251, 22275, 22309, 22314
\keys_set_groups:nnnN 251, 22275, 22306, 22308
\keys_set_groups:nnnnN 251, 22275, 22295, 22305, 22307, 22311
\keys_set_known:nn ........ 249, 22258, 22272, 22274
\keys_set_known:nnn .......... 249, 22258, 22269, 22271
\keys_set_known:nnnn ........ 249, 22258, 22258, 22268, 22270, 22273
\keys_show:nn ..................................... 251, 22659, 22669
\l_keys_usage_load_prop ........ 248, 21594, 21958, 21965, 21972
\l_keys_usage_preamble_prop ........ 248, 21594, 21960, 21967, 21974
\keys_value_tl ........... 249, 21588, 21843, 21844, 21939, 22423, 22427, 22433, 22444, 22455, 22474, 22494, 22498, 22524, 22534, 22562
\keys_internal_commands:\keys_bool_set:NN ........................................ 21707, 21707, 21709, 21728, 22016, 22018, 22020, 22022
\keys_bool_set:NNNN ........................................ 21707, 21708, 21711, 21713
\keys_bool_set_inverse:NN ........................................ 21707, 21710, 21712, 22024, 22026, 22028, 22030
\keys_check_forbidden:NN ........................................ 21707, 21934
\keys_check_groups: ........................................ 22385, 22393
\keys_check_required:NN ........................................ 21907, 21943
\c_keys_check_root_str ........................................ 21561, 21914, 21920, 21924, 22454, 22473
\keys_choice_find:n ........................................ 21730, 22573, 22573, 22589
\keys_choice_find:n ........................................ 22573, 22576, 22578, 22582
\keys_choice_make: ........................................ 21716, 21729, 21729, 21761, 21854, 22032
\keys_choice_make:N ........................................ 21729, 21730, 21732, 21733
\keys_choice_make_aux: ........................................ 21729, 21745, 21747, 21749
\keys_choice_make:n ........................................ 21760, 21760, 22034, 22036, 22038, 22040, 22042
\\keys_choices_make:NN ........................................ 21760, 21761, 21763, 21764
\keys_angles_str ........................................ 21717, 21719, 21771, 21781, 21781, 21855, 21857, 21859, 21889, 22001, 22044
\keys_angles_str ........................................ 21721, 21753, 21754, 21751, 21782, 21784, 21873, 21881, 39680
\c_keys_code_root_str ........................................ 997, 21561, 21785, 21788, 21837, 22452, 22470, 22486, 22512, 22584, 22664, 22655, 22676, 39676
\keys_functions:NN ........................................ 21786, 21786, 21795, 22054, 22056, 22058, 22060, 22062, 22064, 22066, 22068
\keys_functions:NN ........................................ 21608, 21608, 21800, 21815, 21903, 21923
\keys_base_inherit ........................................ 22416, 22430, 22435
\c_keys_base_root_str ........................................ 21561, 21801, 21805, 22420,
Index

\_keys_default_set:n 22424, 22441, 22445, 22491, 22495
\_keys_inherit_root_str 21726, 21796, 21796, 21864, 22070, 22072, 22074, 22076, 22078
\_keys_define:n 21622, 21626, 21626
\_keys_define:nn 21622, 21626, 21631
\_keys_define:nnn 21616, 21617, 21619, 21625
\_keys_define_aux:nn 21626, 21629, 21634, 21636
\_keys_define_code:n 21640, 21690, 21690
\_keys_define_code:w 21600, 21604, 21705
\l\_keys_exclude Bool 21583, 22225, 22253, 22264, 22281, 22301, 22388, 22406, 22411
\_keys_execute: 22349, 22389, 22408, 22412, 22450, 22450
\_keys_execute:nn 21844, 22450, 22455, 22474, 22498, 22506, 22507, 22508, 22585, 22586
\_keys_execute_inherit: 21834, 22450, 22460, 22464
\_keys_execute_unknown: 996, 22450, 22461, 22478, 22480
\_keys_find_key_module:WN 21839, 21887, 22223, 22343, 22353
\_keys_find_key_module_auxi:Nw 22223, 22355, 22458, 22366, 22371
\_keys_find_key_module_auxii:Nw 22223, 22345, 22362, 22363
\_keys_find_key_module_auxii:Nn 22223
\_keys_find_key_module_auxii:Nn 22223, 22366, 22373
\_keys_groups_clist 995, 21570, 21812, 21813, 21820, 22383, 22398
\c\_keys_groups_root_str 21561, 21816, 21819, 22381, 22384
\_keys_groups_set:n 21810, 21810, 22096
\_keys_inherit:n 21823, 21823, 22098
\_keys_inherit_clist 21571, 21826, 21828
\c\_keys_inherit_root_str 21561, 21827, 21833, 22429, 22438, 22459, 22467
\l\_keys_inherit_str 21579, 21836, 22442, 22472, 22575, 22579
\_keys_initialise:n 21830, 21830, 22100, 22102, 22104, 22106, 22108
\_keys_legacy_if_inverse:nn 21848
\_keys_legacy_if_inverse:nnnn 21848
\_keys_legacy_if_inverse:nnn 21848
\_keys_legacy_if_inverse:n 21848, 22118, 22120
\_keys_legacy_if_set:nnn 21849, 21851, 21852
\_keys_legacy_if_set_inverse:nn 21850, 22122, 22124
\_keys_meta_make:n 21871, 21871, 21872, 21873, 21876, 21879, 22128
\_keys_meta_make:nn 21871, 21871, 21872, 21873, 21876, 21879, 22128
\_keys_module_str 21574, 21617, 21621, 21623, 21665, 21876, 21987, 21994, 22240, 22243, 22245, 22236, 22331, 22341, 22344, 22350, 22486, 22491, 22495, 22498, 22502
\_keys_multichoice_find:n 21732, 22573, 22588
\_keys_multichoice_make: 21720, 21731, 21763, 21810, 21848
\_keys_multichoice_make:nn 21760, 21762, 22132, 22134, 22136, 22138, 22140
\_keys_no_value_bool 21575, 21628, 21633, 21692, 21936, 21945, 22325, 22330, 22418, 22488, 22523, 22533, 22561
\_keys_only_known_bool 21576, 22224, 22252, 22263, 22280, 22300, 22482
\_keys_parent:n 21736, 21739, 21743, 21833, 22429, 22438, 22459, 22467, 22590, 22590
\_keys_parent_auxi:w 22590, 22592, 22595, 22601, 22605
\_keys_parent_auxii:w 22590, 22592, 22595, 22601, 22605
\_keys_property_find:n 21638, 21640, 21649
Index

\_\_keys\_value\_or\_default:n ........ 22446, 22416, 22416
\_\_keys\_value\_requirement:nn ........ 21807, 21907, 21907, 22012, 22202, 22204
\_\_keys\_variable\_set:NnnN ........ 21998, 21998, 22008, 22011, 22046, 22048, 22050, 22052, 22166, 22168, 22170, 22172, 22174, 22176, 22178, 22180, 22182, 22184, 22186, 22188, 22190, 22192, 22194, 22196, 38810, 38812, 38814, 38816, 38818, 38820, 38822, 38824
\_\_keys\_variable\_set\_required:NnnN ........ 21998, 21998, 22008, 22009, 22014, 22080, 22082, 22084, 22086, 22088, 22090, 22092, 22094, 22110, 22112, 22114, 22116, 22112, 22114, 22142, 22144, 22146, 22148, 22158, 22160, 22162, 22164
\_keyval\_commands:
\_keyval\_parse:Nnn ........ 254, 971, 21351, 21361, 21475, 21622, 22244
\_keyval\_parse:nnn ........ 253, 254, 924, 965, 970, 20172, 21351, 21351, 21361, 21476
\_keyval\_internal\_commands:
\_\_keyval\_blank\_key\_error:w ........ 21482, 21491, 21504, 21506
\_\_keyval\_blank\_true:w ........ 21436, 21504, 21504
\_\_keyval\_clean\_up\_active:w ........ 968, 21378, 21391, 21412, 21412, 21444, 21464
\_\_keyval\_clean\_up\_other:w ........ 968, 21417, 21422, 21433, 21433
\_\_keyval\_end\_loop\_active:w ........ 21365, 21458, 21466
\_\_keyval\_end\_loop\_other:w ........ 969, 21375, 21458, 21458
\_\_keyval\_if\_blank:w ........ 21436, 21482, 21491, 21501, 21502
\_\_keyval\_if\_empty:w ........ 21501, 21501, 21502
\_\_keyval\_if\_recursion\_tail:w ........ 21364, 21374, 21501, 21503
\_\_keyval\_key:nn ........ 968, 21438, 21477, 21489, 21504
\_\_keyval\_loop\_active:nnn ........ 21356, 21362, 21362, 21465
\_\_keyval\_loop\_other:nnn ........ 966, 21366, 21372, 21372, 21448, 21456, 21468, 21487, 21496, 21505, 21506, 21511
\_\_keyval\_misplaced\_equal\_after_-active_error:w ........ 967, 21385, 21389, 21440, 21440
\_\_keyval\_misplaced\_equal\_in_-split_error:w ........ 21396, 21401, 21405, 21409, 21425, 21430, 21440, 21450
\_\_keyval\_pair:nnnn ........ 967, 968, 21411, 21432, 21477, 21480
\_\_keyval\_split\_active:w ........ 966, 967, 21368, 21370, 21376, 21395, 21460
\_\_keyval\_split\_active\_auxi:w ........ 21377, 21382, 21382, 21413, 21463
\_\_keyval\_split\_active\_auxii:w ........ 967, 21382, 21386, 21388, 21442
\_\_keyval\_split\_active\_auxiii:w ........ 967, 21382, 21392, 21393
\_\_keyval\_split\_active\_auxiv:w ........ 967, 21382, 21397, 21400
\_\_keyval\_split\_other\_auxiii:w ........ 21382, 21406, 21408
\_\_keyval\_split\_other\_w ........ 21368, 21384, 21404, 21415, 21424
\_\_keyval\_split\_other\_auxi:w ........ 968, 21416, 21419, 21419, 21434
\_\_keyval\_split\_other\_auxii:w ........ 21419, 21420, 21421
\_\_keyval\_split\_other\_auxiii:w ........ 968, 21419, 21426, 21429
\_\_keyval\_tmp:w ........ 969, 21349, 21473, 21478, 21499, 21520, 21558
\_\_keyval\_trim:n ........ 21392, 21411, 21420, 21432, 21438, 21504, 21519, 21522
\_\_keyval\_trim\_auxi:w ........ 21519, 21524, 21533, 21536, 21541
\_\_keyval\_trim\_auxii:w ........ 21519, 21528, 21541
\_\_keyval\_trim\_auxiii:w ........ 21519, 21519, 21529, 21543, 21546, 21550, 21554
\_\_keyval\_trim\_auxiv:w ........ 21519, 21531, 21552
\knacode ........ 673
\knbcode ........ 674
\knbscode ........ 675
\kuten ........ 1165, 1206
\L ........ 32066, 33650, 34346
\l ........ 32066, 33650, 34358
\label ........ 31662, 31672, 34319
\language ........ 288
\LARGE ........ 34299
\Large ........ 34300
\large \ldots 34303
\lastallocatedtoks \ldots 3187
\lastbox \ldots 289
\lastkern \ldots 290
\lastlinefit \ldots 506
\lastnamedcs \ldots 842
\lastnodechar \ldots 1166
\lastnodesubtype \ldots 1168
\lastnodetype \ldots 507
\lastpenalty \ldots 291
\lastsavedboxresourceindex \ldots 940
\lastsavedimagere-sourceindex \ldots 942
\lastsavedimagere-sourcepages \ldots 944
\lastskip \ldots 292
\lastxpos \ldots 946
\lastypos \ldots 947
\latelu \ldots 843
\latelu-function \ldots 844
\lccode \ldots 64, 65, 293
\leaders \ldots 294
\left \ldots 295
\leftghost \ldots 845
\lefthyphenmin \ldots 296
\lefmarginkern \ldots 676
\leftskip \ldots 297
\legacy commands:
\legacy_if:n \ldots 12100
\legacy_if:nTF \ldots 108, 12100
\legacy_if:get:n \ldots 243, 22117
\legacy_if:get:nn \ldots 108, 12117, 12122
\legacy_if:get=false:n \ldots 108, 12109, 12115, 12124
\legacy_if:get:inverse:n \ldots 243, 22117
\legacy_if:get=true:n \ldots 108, 12109, 12113, 12124
\legacy_if:p:n \ldots 108, 12100
\legacy_if:set:n \ldots 243, 22117
\legacy_if:set:nn \ldots 108, 12117, 12117
\legacy_if:set=false:n \ldots 108, 12109, 12111, 12119
\legacy_if:set:inverse:n \ldots 243, 22117
\legacy_if:set=true:n \ldots 108, 12109, 12109, 12119
\legacy_if \ldots 298
\letcharcode \ldots 5, 140, 141, 299
\letterspacefont \ldots 846
\limits \ldots 11760
\LineBreak \ldots 11762
\lineskip \ldots 11767
\lineskiplimit \ldots 301
\linewidth \ldots 302
\ln \ldots 303
\ln \ldots 27779, 27782
\localbrokenpenalty \ldots 275
\localinterlinepenalty \ldots 849
\localletbox \ldots 850
\localleftbox \ldots 855
\localrightbox \ldots 856
\loccount \ldots 10204, 10457
\loctoks \ldots 3159, 3160, 3186
\logb \ldots 275
\long \ldots 143, 304, 19517, 19521
\LongText \ldots 305
\looseness \ldots 306
\lowercase \ldots 67, 307
\lcode \ldots 678
\ltx pdf.object commands:
\ltx pdf.object_id \ldots 38486
\ltx utils \ldots 106, 11826
\ltx.utils.filedump \ldots 106, 11899
\ltx.utils.filemd5sum \ldots 106, 11920
\ltx.utils.filemodeate \ldots 106, 11929
\ltx.utils.filesize \ldots 107, 11982
\lua commands:
\lua \ldots 106, 11763, 11765, 11770, 11771, 11785
\lua load module:n \ldots 106, 11772, 11773, 11796
\lua now:n \ldots 106, 11774, 11775, 11776
\lua_shipout:n \ldots 105, 11765, 11766, 11768, 11776, 30596
\lua_shipout e:n \ldots 105, 11765, 11767, 11769, 11796
\lua internal commands:
\_lua_err_msg_str \ldots 11772, 11778
\_lua_escapen str \ldots 11760, 11761, 11767, 11770
\_lua_load_module_p:n \ldots 11775, 12042
\_lua now:n \ldots 11760, 11761, 11765
\_lua_shipout:n \ldots 11760, 11762, 11767
\_lua_escapestring \ldots 857
\_luafunction \ldots 858
\_luafunctioncall \ldots 859
\_la texbanner \ldots 860
\_latexrevision \ldots 861
\_latexversion \ldots 10, 56, 862
Index

\muskip_zero_new:N   .................................. 236, 21290, 21290, 21294
\g_tampa_muskip      .................................. 238, 21336
\l_tampa_muskip      .................................. 238, 21336
\g_tampb_muskip      .................................. 238, 21336
\l_tampb_muskip      .................................. 238, 21336
\c_zero_muskip      238, 21285, 21287, 21334
\muskipdef          336
\mutoglue           .................................. 511

\n ........................................ 8919, 8921, 8923, 12040
\nan                    279
\nc                      280
\ncd                    .................................. 280
\newbox                .................................. 844
\newcatcodetable      ................. 30383
\newcount              .................................. 844
\newdimen              .................................. 844
\newlinechar            59, 337
\newlubytecode         .................................. 19
\next                  36, 73, 81
\NG                     32067, 33651, 34347
\NG                    32067, 33651, 34359
\noalign                 338
\noautosize            1169
\noautoxspace           1170
\nobox                   ................. 339
\nobreakspace           .......... 34273
\noexpand               60, 75, 78, 84, 340
\nohrule                .................................. 886
\noindent               341
\no kern               887
\no lig              888
\no limits             342
\nonscript              .................................. 343
\nostopmode             ................... 344
\normaldeviate         948
\normalend              1314, 1315
\normaleveryjob        1316
\normalexpanded        1325
\normalfont             ................. 34287
\normalhoffset         1328
\normalinput            1317
\normalitaliccrcorrection  1327, 1329
\normallanguage        1318
\normalleft             1335, 1336
\normalmathop            1319
\normalmiddle           1337
\normalmonth              1320
\normalouter            1321
\normalover              1322
\normalright             1338
\normalshadetokens      1331
\normalize              .................................. 34304
\normalunexpanded       .................................. 1324
\normalcenter          .................................. 1323
\normalvoffset         .................................. 1330
\nospaces                889
\notexpended: (token) .................................. 210
\nowrule                .................................. 890
\nulldelimiterspace     .................................. 345
\nullfont               .................................. 346
\num                    .................................. 259
\number                  .................................. 347
\numexpr                 .................................. 512

O

\O     .................................. 32068, 33652, 34348, 34613
\o     .................................. 32068, 33652, 34360, 34614
\odelcode            .................................. 1209
\odelimiter          .................................. 1210
\OE    .................................. 32069, 33653, 34349
\oe     .................................. 32069, 33653, 34361
\omathaccent          .................................. 1211
\omathchar            .................................. 1212
\omathchardef         .................................. 1213
\omathcode            .................................. 1214
\omit                  .................................. 348
\openin                .................................. 349
\openout               .................................. 350
\or     .................................. 351

or commands:

\or: .................................. 179, 745, 747, 895, 1032, 1392, 1394, 2082, 2083, 2084,
      2085, 2086, 2087, 2088, 2089, 2090, 3065, 3066, 3067, 3068, 3069, 3070,
      3071, 3072, 3073, 3074, 3075, 3076, 3077, 3078, 3079, 3080, 3081,
      3082, 3083, 3084, 3085, 3086, 3087, 3088, 3089, 3090, 3091, 3092,
      3093, 3094, 3095, 3096, 3097, 3098, 3099, 3100, 3101, 3102, 3103,
      3104, 3105, 3106, 3107, 3108, 3109, 3110, 3111, 3112, 3113, 3114,
      3115, 3116, 3117, 3118, 3119, 3120, 3121, 3122, 3123, 3124, 3125,
Index

pdfimagecolor .......................... 559  \pdfrefximage .......................... 598
pdfimageresolution ..................... 560  \pdfresettimer ......................... 661
dpfincludechars ....................... 561  \pdfrestore ................................ 599
dpfincusioncopyfonts ................... 562  \pdfsetval ............................... 600
dpfincursionerrorlevel ................. 563  \pdfrunninglinkoff .................... 601
dpfinfo ................................ 565  \pdfrunninglinkon ...................... 602
dpfinfoomtitdate ....................... 566  \pdfsave ................................ 603
dpfinsertth ............................. 644  \pdfsavepos ................................ 662
dpfinterwordspaceoff .................. 567  \pdfsetmatrix ................................ 604
dpfinterwordspaceon ................... 568  \pdfsetrandomseed ....................... 663
dpflastannot ........................... 569  \pdfshellescape ......................... 664
dpflastinedepth ....................... 645  \pdfstartlink ........................... 605
dpflastlink ................................ 570  \pdfstartthread ......................... 606
dpflastmatch ............................ 646  \pdfstrcmp ................................ 133, 5, 699

dpflastobj ................................ 571  \pdfsuppresstexpinfo .................... 607
dpflastxform ............................ 572  \pdfsuppresstwarnidupdest ............ 608
dpflastximage ........................... 573  \pdfsuppresstwarnidupmap ............... 610
dpflastximagecolordepth .............. 574  \pdfsuppresstwarnidpagegroup ........... 612
dpflastximpagestes ..................... 576  \pdftext .................................. 668
dpflastxpos ................................ 647  \pdftextrevision .......................... 669

dpflastyp ............................... 648  \pdfthread ................................ 614

dpflinkmargin ........................... 577  \pdfthreadmargin ......................... 615

dpdfliteral ............................. 578  \pdfthreadmargins ......................... 615

dpdfmajorversion ...................... 581  \pdftracingfonts ..................... 665, 1266, 1267

dpdfmapfile ............................. 579  \pdftrailer ................................ 616

dpdfmapline .............................. 580  \pdftrailerid ............................. 617

dpdfmatch ................................ 649  \pdfunescapehex ......................... 666

dpdfmdfivesum ......................... 701  \pdfuniformdeviate ....................... 667

dpdfminorversion ...................... 582  \pdfuniqueqername ....................... 618

dpdfnames ................................ 583  \pdfvariable .............................. 902

dpdfnobluntintunicode ................. 584  \pdfvorigin .............................. 619

dpdfnoligatures ......................... 650  \pdfxf .................................. 620

dpdfnormaldeviate ...................... 651  \pdfxfname .............................. 621

dpdfobj ................................ 585  \pdfximage .............................. 622

dpdfobjcompresslevel ................... 586  \pdfximagebbox ........................... 623

dpdfonitcharget .......................... 587

\peek commands:
\peek_after:Nw ................ 73, 205, 4045,
\peek_analysis_map_break: .. ...
\peek_analysis_map_break:n ... 208, 4005, 4006, 4008, 4028
\peek_analysis_map INLINE:n ... 208, 4005, 4007, 7825
\peek_catcode:NTF ........... 206, 19846, 37287
\peek_catcode:ignore_spaces:NTF
\peek_catcode:remove:NTF ...... 206, 19846, 37353
\peek_catcode:remove:ignore: -
\peek_charcode:NTF ....... 206, 209, 210, 19846
\peek_charcode:ignore_spaces:NTF
\peek_charcode:ignore_spaces:NTF

\peekafter:Nw ................ 73, 205, 4045,
\peek_analysis_map:break: ..... ....
\peek_analysis_map:break:n .... 208, 4005, 4006, 4008, 4028
\peek_analysis_map:inline:n .... 208, 4005, 4007, 7825
\peek_catcode:ntf ............ 206, 19846, 37287
\peek_catcode:ignore_spaces:ntf
\peek_catcode:remove:ntf ...... 206, 19846, 37353
\peek_catcode:remove:ignore:-
\peek_charcode:ntf ....... 206, 209, 210, 19846
\peek_charcode:ignore_spaces:ntf
\peek_charcode:ignore_spaces:ntf

\peekafter:nw ............. 73, 205, 4045,
\__peek_execute_branches_-
    charcode: ...... 913, 19813, 19815
\__peek_execute_branches_-
    meaning: ...... 913, 19805, 19805
\__peek_execute_branches_N_type:
    \_ 19860, 19863, 19895, 19898, 19900
\__peek_false:w
    \_ 19860, 19870, 19880
\__peek_N_type:w . 19860, 19870, 19880
\__peek_N_type_aux:nw
    \_ 19860, 19872, 19885
\__peek_remove_filler:
    \_ 19737, 19750, 19753
\__peek_remove_filler:w
    \_ 19737, 19739, 19746, 19748, 19772
\__peek_remove_filler_exp:w
    \_ 19737, 19763, 19767
\__peek_remove_spaces:
    \_ 19721, 19725, 19728
\l\_peek_search_tl
    909, 913, 19704, 19779, 19830, 19840
\l\_peek_search_token
    \_ 19795, 19796, 19803, 19883
\l\_peek_token_generic:NNTF
    913, 914, 19793, 19793, 19739, 19796, 19796, 19797, 19798, 19894, 19898, 19900
\l\_peek_token_generic_aux:NNNTF
    \_ 19775, 19775, 19794, 19800
\l\_peek_token_generic:NNNF
    \_ 19795, 19796, 19803, 19804
\l\_peek_true:w
    \_ 913, 914, 19705, 19705, 19785, 19808, 19831, 19841, 19875, 19889, 19890
\l\_peek_true:aux:w
    \_ 19705, 19705, 19706, 19718, 19725, 19726, 19739, 19780, 19794
\l\_peek_true_remove:w
    \_ 19705, 19705, 19706, 19718, 19725, 19726, 19739, 19780, 19794
\l\_peek_use_none_delimit_by_s_-_stop:w
    \_ 913, 19711, 19711, 19873
\penalty
    \_ 374
\pi
    \_ 24226, 24227
\pi
    \_ 279
\postbreakpenalty
    \_ 1172
\postdisplaypenalty
    \_ 375
\postexhyphenchar
    \_ 903
\posthyphenchar
    \_ 904
\prebinoppenalty
    \_ 905
\postexhyphenchar
    \_ 903
\posthyphenchar
    \_ 904
\prebinoppenalty
    \_ 905
\postexhyphenchar
    \_ 903
\posthyphenchar
    \_ 904
\prebinoppenalty
    \_ 905
\texttt{\textbackslash prg\_break}: \ldots 73, 433, 440, 822, 2380, 2381, 2382, 3231, 3273, 3397, 3404, 3777, 3938, 3974, 4327, 4575, 5856, 5905, 6354, 6686, 6838, 7633, 7639, 8627, 13112, 14211, 14219, 14309, 14325, 14445, 14477, 14583, 14586, 14727, 14774, 14827, 14833, 15066, 15266, 15423, 17021, 17151, 17185, 17236, 17287, 17303, 17354, 17361, 17849, 18203, 18393, 18871, 23134, 23153, 23168, 23340, 23349, 23549, 23570, 23568, 25764, 29156, 29180, 29412, 29465
\texttt{\textbackslash prg\_break\_point}: \ldots 73, 73, 149, 194, 462, 473, 833, 853, 950, 2371, 2371, 2372, 3902, 4028, 6578, 6592, 6637, 7824, 8627, 10406, 10425, 12691, 12720, 12731, 13657, 13683, 13703, 13725, 17186, 17227, 17237, 17260, 17267, 17276, 17896, 18743, 18765, 18788, 18815, 18877, 20551, 20573, 20589, 21028, 25833, 33712, 33731, 34051
\texttt{\textbackslash prg\_do\_nothing}: \ldots 13, 73, 492, 546, 566, 675, 700, 761, 815, 821, 870, 887, 1039, 1217, 2369, 2369, 2380, 2789, 2816, 2915, 2916, 2917, 3319, 3477, 3478, 3748, 3797, 4077, 4093, 4423, 4908, 4951, 4952, 4959, 4960, 6898, 7126, 7549, 7553, 7605, 8986, 10785, 11082, 11502, 11541, 11543, 12386, 13006, 13477, 13479, 14374, 15316, 16577, 16613, 16614, 16751, 16758, 17116, 17118, 18386, 18392, 18400, 18562, 18763, 18773, 18836, 18847, 18923, 18935, 18990, 18994, 19001, 22507, 23621, 23655, 23681, 23689, 25233, 29135, 29534, 29688, 30548, 32167
\texttt{\textbackslash prg\_generate\_conditional\_-\_variant\_Nn}: \ldots 33, 65, 2905, 2905, 2959, 2727, 2783, 7313, 7315, 8356, 10227, 11071, 11219, 11324, 11328, 11332, 11336, 11361, 11372, 11413, 12556, 12566, 12590, 12593, 12609, 12623, 12629, 12640, 12660, 12907, 12925, 12934, 13556, 13568, 13576, 13607, 13642, 16420, 16442, 16884, 16985, 16965, 17025, 17119, 17121, 17135, 17137, 17139, 17141, 18329, 18580, 18594, 18595, 18732, 18734, 20422, 20424, 20426, 20383, 20385, 20518, 20541, 22650, 34725, 34727, 34731, 35468, 38934
\texttt{\textbackslash prg\_gset\_conditional\_Nn}: \ldots 63, 1618, 1620
\texttt{\textbackslash prg\_gset\_conditional\_Npnn}: \ldots 63, 1597, 1599, 1826, 1840, 1853, 1868, 1882, 1897
\texttt{\textbackslash prg\_gset\_eq\_conditional\_Nn}: \ldots 65, 1741, 1743
\texttt{\textbackslash prg\_gset\_protected\_conditional\_Nn}: \ldots 64, 1618, 1626
\texttt{\textbackslash prg\_gset\_protected\_conditional\_Npnn}: \ldots 64, 1597, 1605
\texttt{\textbackslash prg\_map\_break\_Nn}: \ldots 72, 73, 494, 462, 473, 833, 853, 950, 2371, 2372, 2372, 4006, 4008, 4322, 8627, 10389, 10391, 12758, 12760, 13716, 13718, 17174, 17176, 18850, 18852, 20605, 20607, 34041, 34043
\texttt{\textbackslash prg\_new\_conditional\_Nn}: \ldots 63, 1618, 1622, 8280
\texttt{\textbackslash prg\_new\_conditional\_Npnn}: \ldots 63, 65, 73, 79, 899, 913, 1597, 1601, 2256, 4717, 4746, 4829, 4858, 8289, 8328, 8379, 8440, 8455, 8466
Index

14355, 14364, 14917, 14923, 14955, 14970, 14976, 16407, 16415, 16426,
16436, 16585, 16596, 16923, 16960,
17024, 17042, 17067, 1713, 17721,
17759, 17767, 18324, 18414, 18417,
18423, 18577, 18591, 18689, 18719,
18729, 18935, 19351, 19369, 19471,
19381, 19388, 19393, 19398, 19403,
19408, 19413, 19420, 19425, 19441,
19447, 19455, 19505, 19629, 19655,
20240, 20373, 20381, 20503, 20513,
20533, 20539, 20847, 20882, 21232,
21241, 22647, 22656, 23257, 23532,
24676, 24967, 25552, 25570, 25579,
29558, 29661, 29765, 30229, 30617,
31541, 31599, 31603, 32905, 32909,
32991, 33006, 33009, 33012, 33015,
33018, 33021, 33041, 33047, 33050,
33053, 33056, 33059, 33062, 33065,
33068, 33071, 33074, 33077, 33080,
33083, 33086, 33089, 33118, 33121,
33136, 33139, 33153, 33156, 33159,
33162, 33178, 33181, 33184, 34722,
34724, 34730, 35463, 36715, 38540,
38618, 38660, 39672, 39263, 40031

\_prg_set_conditiona\_Nnnn ......... 63, 1618, 1618, 8280
\_prg_set_conditiona\_Npnn ......... 63-65, 389, 1597, 1597, 8280, 40024
\_prg_set_eq_conditiona\_NNNN ....... 65, 1741, 1741, 8280
\_prg_set_protected_conditiona\_Nnnn ......... 64, 1618, 1624, 8280
\_prg_set_protected_conditiona\_Npnn ......... 64, 1597, 1603, 8280

prg internal commands:
\_\_prg_break_point:aN ......... 404
\_\_prg_F\_true:aw ......... 1694, 1727, 1739
\_\_prg_generate_conditiona\_nnNNNNNN ......... 1613, 1642, 1651, 1651
\_\_prg_generate_conditiona\_NNNNNNNN ......... 1651, 1660, 1677, 1692
\_\_prg_generate_conditiona\_\_count:NNNN ......... 1618, 1619, 1621, 1623, 1625, 1627, 1629, 1630
\_\_prg_generate_conditiona\_\_count:nnNNNN ......... 1618, 1634, 1639
\_\_prg_generate_conditiona\_\_fast:nw ......... 383, 384, 1651, 1664, 1675
\_\_prg_generate_conditiona\_\_parm:NNNN ......... 1597, 1602, 1604, 1606, 1608, 1609
\_\_prg_generate_conditiona\_\_test:w ......... 1651, 1662, 1672

\_\_\_prg_generate_F_form:wNNnnnnN ......... 1694, 1721
\_\_\_prg_generate_F_form:wNNnnnnN ......... 1694, 1694
\_\_\_prg_generate_F_form:wNNnnnnN ......... 1694, 1713
\_\_\_prg_generate_TF_form:wNNnnnnN ......... 1694, 1729
\_\_\_prg_p_true:w ......... 1694, 1706, 1737
\_\_\_prg_replicate:N ......... 8562, 8569, 8570, 8572
\_\_\_prg_replicate_0:n ......... 8562
\_\_\_prg_replicate_1:n ......... 8562
\_\_\_prg_replicate_2:n ......... 8562
\_\_\_prg_replicate_3:n ......... 8562
\_\_\_prg_replicate_4:n ......... 8562
\_\_\_prg_replicate_5:n ......... 8562
\_\_\_prg_replicate_6:n ......... 8562
\_\_\_prg_replicate_7:n ......... 8562
\_\_\_prg_replicate_8:n ......... 8562
\_\_\_prg_replicate_9:n ......... 8562
\_\_\_prg_replicate_first:N ......... 8562, 8565, 8571
\_\_\_prg_replicate_first_0:n ......... 8562
\_\_\_prg_replicate_first_1:n ......... 8562
\_\_\_prg_replicate_first_2:n ......... 8562
\_\_\_prg_replicate_first_3:n ......... 8562
\_\_\_prg_replicate_first_4:n ......... 8562
\_\_\_prg_replicate_first_5:n ......... 8562
\_\_\_prg_replicate_first_6:n ......... 8562
\_\_\_prg_replicate_first_7:n ......... 8562
\_\_\_prg_replicate_first_8:n ......... 8562
\_\_\_prg_replicate_first_9:n ......... 8562
\_\_\_prg_set_eq_conditiona\_NNNNNN ......... 1741, 1742, 1744, 1746, 1747
\_\_\_prg_set_eq_conditiona\_nnNNNN ......... 1751, 1759, 1759
\_\_\_prg_set_eq_conditiona\_F\_\_form:NNN ......... 1759
\_\_\_prg_set_eq_conditiona\_F\_\_form:NNN ......... 1796, 39705
\_\_\_prg_set_eq_conditiona\_\_loop:NNNN ......... 1759, 1771, 1773, 1788
\_\_\_prg_set_eq_conditiona\_\_form:NNN ......... 1790, 39693
\_\_\_prg_set_eq_conditiona\_\_form:NNN ......... 1759
\_\_\_prg_set_eq_conditiona\_\_form:NNN ......... 1794, 39701
\__prop_split:NnTFn \ 926, 927, 933, 935, 20201, 20201, 20250, 20290, 20444
\__prop_split_aux:nTFn \ 20201, 20203, 20206
\__prop_split_fl:w \ 926, 927, 20201, 20208, 20216, 20219
\__prop_split_linked:w \ 20201, 20208, 20220, 20222
\__prop_split_test:wn \ 926, 927, 20201, 20211, 20217
\__prop_split_wrong:Nn \ 926, 927, 20201, 20222
\__prop_tmp:w \ 940, 941, 19905, 19905, 20151, 20155, 20688, 20707
\__prop_to_keyval:nn \ 19943, 19946, 19949, 19953, 19955
prop (prefix) internal commands:
\__prop (prefix) \ 915, 916
\protect \ 1298, 10084, 24161, 31964, 31931, 31946, 31955, 31969, 31971, 34225
\protected \ 82, 84, 106, 518, 19519, 19521
\protrudechars \ 952
\protrusionboundary \ 910
\ProvidesExplClass \ 9
\ProvidesExplFile \ 9, 39023
\ProvidesExplPackage \ 9
pt \ 280
\texfontname \ 1174
\texlineendmode \ 1175
\texminormode \ 1176
\textversion \ 1177
\textractingfonts \ 1178
\textversion \ 1179
\pxdimen \ 953

Q
quark commands:
quark \q_mark \ 147, 446, 16352, 32355, 33257, 33264, 33267, 33277
quark \q_nil \ 26, 27, 126, 147, 379, 808, 810, 812, 1551, 1554, 11148, 11150, 16352, 16406, 16425, 16431, 16446, 16447, 16453, 16477, 16481, 22592, 22593, 22595, 22597, 22599, 22601, 22607, 37926, 37932, 37933
quark \q_no_value \ 77, 93, 100–103, 146, 147, 153, 154, 161, 192, 217, 808, 810, 829, 830, 875, 931, 8872, 10340, 10353, 11057, 11208, 11311, 11314, 11317, 11320, 11349, 16352, 16414, 16435, 16441, 17030, 17038, 17050, 17076, 18537, 18552, 20232, 20359, 20365
quark \q_recursion_stop \ 147, 446, 148, 149, 379, 808, 1553, 1557, 5845, 5964, 16356, 17428, 30286, 31011, 31194, 33657, 33692, 33953, 33962, 33986, 34365, 34408, 34626, 37167
quark \q_recursion_tail \ 148, 149, 808, 809, 5845, 5963, 16356, 16362, 16368, 16377, 16384, 16389, 16394, 16401, 17428,
regex internal commands:
\__regex_A_test: \ 486, 5237, 5259, 5875, 5878, 5884, 6002, 6483, 6516
\__regex_action_cost:n \ 522, 526, 6272, 6273, 6281, 6730, 6756, 6756
\__regex_action_free:n \ 522, 534, 6295, 6301, 6302, 6313, 6371, 6375, 6400, 6425, 6429, 6432, 6460, 6468, 6478, 6492, 6535, 6728, 6732, 6732
\__regex_action_free_aux:n \ .......... 6732, 6733, 6735, 6736
\__regex_action_free_group:n \ 522, 534, 6321, 6440, 6443, 6732, 6734
\__regex_action_startWildcard:N \ 522, 6156, 6176, 6725, 6725
\__regex_action_submatch:n \ ......... 522, 6180, 6202, 6394, 6395, 6533, 6781, 6783, 6783
\__regex_action_submatch_aux:w \ .. 6783, 6785, 6788
\__regex_action_submatch_auxii:w \ .......... 6783, 6794, 6799
\__regex_action_submatch_auxii:w \ 6783, 6795, 6800, 6801, 6802
\__regex_action_submatch_auxiv:w \ .......... 6783, 6784, 6799
\__regex_action_success: \ ........ 522, 6159, 6205, 6223, 6804, 6804
\__regex_action_wildcard: \ 539
\__regex_add_begin_int \ 758, 7592, 7530, 7534, 7588, 7717, 7722, 7726, 7737, 7752
\__regex_add_end_int \ 7583, 7524, 7530, 7535, 7589, 7719, 7722, 7727, 7739, 7753
\__regex_all_catcodes_int \ 4783, 4895, 4999, 5595
\__regex_ascii_lower_int \ 4354, 4415, 4420
\__regex_ascii_max_control_int \ 4351, 4531
\__regex_ascii_max_int \ 4351, 4524, 4532, 4723
\__regex_ascii_min_int \ 4351, 4523, 4530
\__regex_assertion:Nn \ 486, 500, 532, 5233, 5255, 5564, 5995, 6483, 6483
\__regex_b_test: \ 486, 532, 5245, 5247, 5881, 6000, 6483, 6501
\__regex_balance_int \ 474, 546, 569, 4350, 6880, 6912, 7171, 7188, 7395, 7408, 7410, 7411, 7668, 7694, 7718, 7720
\__regex_balance_intarray: \ 471, 560, 6599, 6866, 7382, 7407
\__regex_balance_tll \ 546, 6822, 6881, 6911, 6937, 6954, 6964, 7039
\__regex_begin_flag: \ 7374, 7513, 7523, 7566
\__regex_branch:n \ 486, 504, 528, 4347, 4900, 4975, 5405, 5458, 5643, 5753, 5761, 5845, 5847, 5850, 5977, 6366, 6360, 39749, 39750, 39751
\__regex_break_point:TF \ .......... 475, 499, 526, 4363, 4364, 4365, 4369, 6272, 6273, 6489, 6506
\__regex_break_true:w \ .......... 475, 476, 4363, 4363, 4369, 4374, 4381, 4388, 4392, 4399, 4405, 4452, 4464, 4480, 5208, 6513, 6519, 6525
\__regex_build:N \ 558, 6139, 6141, 7280, 7287, 7307, 7311, 39718, 39721, 39723
\__regex_build:n \ 523, 558, 6139, 6139, 7274, 7285, 7306, 7309
\__regex_build_aux:NN \ 570, 6129, 6142, 6146, 6148, 7774, 7793, 7863
\__regex_build_aux:Nn \ 570, 6139, 6140, 6143, 7765, 7783, 7855
\__regex_build_for_cs:n \ 4475, 6212, 6212, 39725, 39728, 39730
\__regex_build_new_state: \ 6153, 6154, 6173, 6174, 6178, 6215, 6216, 6245, 6245, 6254, 6286, 6320, 6324, 6368, 6383, 6388, 6427, 6446, 6481, 6485, 6530, 39743
\__regex_build_tll \ 504, 575, 4344, 4892, 4899, 4917, 4922, 4925, 4926, 4929, 4930, 4933, 4993, 4996, 5036, 5050, 5054, 5177, 5191, 5232, 5254, 5267, 5299, 5312, 5316, 5398, 5401, 5404, 5410, 5411, 5414, 5457, 5574, 5751, 5758, 5764, 5785, 5801, 5819, 5976, 6033, 6036, 6047, 6077, 6092, 6096, 6099, 6105, 6879, 6902, 6913, 6916, 6967, 7036, 7093, 7096, 7110, 7178, 7921, 7924, 7932, 7935
\__regex_build_transition_-_left:NN \ 6241, 6241, 6429, 6443, 6460
Index

1617

... 4808
... 4808, 5067, 5082

... 4808, 4818, 4817

... 4808, 4820, 4820

... 4808, 4818, 4817

... 4808, 4818, 4817

... 4808, 4818, 4817

... 4808, 4818, 4817
\_\_regex_replacement_normal:n ........... 6890, 6896, 6968, 6968, 7021, 7051, 7078, 7113, 7121, 7136
\_\_regex_replacement_normal:_aux:N ........... 6968, 6974, 6988
\_\_regex_replacement_put:n ........... 6966, 6966, 6971, 7162, 7214, 7881
\_\_regex_replacement_put:_submatch:n . 7019, 7025, 7025, 7066
\_\_regex_replacement_put:_submatch_aux:n ........... 7025, 7028, 7034, 7882
\_\_regex_replacement_rbrace:N ........... 6885, 7065, 7107, 7107
\_\_regex_replacement_set:n ........... 6874, 6875, 6919, 6956
\_\_regex_replacement_tl ........... 7759, 7872, 7887
\_\_regex_replacement_u:w ........... 7160, 7167, 7881
\_\_regex_replacement_rbrace:N ........... 6885, 7025, 7034, 7882
\_\_regex_return:................... 558, 7275, 7281, 7309, 7311, 7385
\_\_regex_right_state_int ........... 6134, 6158, 6199, 6200, 6220, 6232, 6239, 6248, 6249, 6288, 6295, 6301, 6314, 6322, 6372, 6376, 6387, 6401, 6410, 6422, 6426, 6430, 6433, 6438, 6441, 6444, 6452, 6466, 6469, 6472, 6475, 6479, 6495, 6538, 39737
\_\_regex_right_state_seq ........... 6134, 6198, 6208, 6231, 6238, 6374
\_\_regex_saved_success_bool ...
\_\_regex_saved_success_bool ........... 536, 4476, 4483, 6568
\_\_regex_show:N ................... 556, 5973, 5973, 7254, 7266
\_\_regex_show:NN 7249, 7259, 7260, 7261
\_\_regex_show:Nn 7249, 7249, 7250, 7521
\_\_regex_show(Char:n ........... 6005, 6012, 6016, 6025, 6038, 6038
\_\_regex_show_class:NNNN ........... 5992, 6074, 6074
\_\_regex_show_group_aux:NNNN ........... 5987, 5989, 5991, 6065, 6065
\_\_regex_show_item_catcode:NNnF ........... 6019, 6021, 6110, 6110
\_\_regex_show_item_exact_cs:n ........... 5981, 5994, 5997, 6005, 6008, 6012, 6015, 6025, 6029, 6041, 6044, 6060, 6067, 6071, 6084, 6100, 6128
\_\_regex_show_lines_int ........... 4798, 6046, 6078, 6081, 6088
\_\_regex_show_one:n ........... 5981, 5994, 5997, 6005, 6008, 6012, 6015, 6025, 6029, 6041, 6044, 6060, 6067, 6071, 6084, 6100, 6128
\_\_regex_show_pop: ........... 6054, 6056, 6093, 6070
\_\_regex_show_prefix_seq ........... 4797, 5979, 5982, 6030, 6050, 6055, 6057
\_\_regex_show_push:n ........... 6051, 6054, 6054, 6051, 6061, 6068, 6079
\_\_regex_show_scope:nm ........... 6023, 6028, 6054, 6058, 6115
\_\_regex_single_match: ........... 536, 4473, 6444, 6644, 7418, 7455, 7660, 7813
\_\_regex_split:nn ........... 7325, 7471, 7471
\_\_regex_standard_escapechar: ........... 4722, 4722, 4569, 4938, 6150, 6171
\_\_regex_start_pos_int ........... 6524, 6543, 6619, 6624, 6630, 7477, 7489, 7502, 7505, 7628, 7711
\_\_regex_state_active_intarray ........... 471, 524, 535, 537, 6560, 6601, 6706, 6709, 6717, 6744
\_\_regex_step_int 471, 6557, 6603, 6668, 6707, 6711, 6719, 6733, 6735
\_\_regex_store_int ........... 535, 6629, 6758, 6761, 6761
\_\_regex_store_submatches: ........... 6701
\_\_regex_store_submatches:n ........... 6780
\_\_regex_store_submatches:nn ........... 6763, 6767
\_\_regex_submatch_balance:n ........... 6824, 6856, 6856, 6938, 7040, 7559
\_\_regex_submatch_begin:_intarray ........... 471, 544, 567, 6829, 6853, 6869, 6930, 7378, 7484, 7487, 7500, 7641
\_\_regex_submatch_case_intarray ........... 6949, 7378, 7623, 7629
\_\_regex_submatch_end_intarray ........... 471, 567, 6854, 6862, 7378, 7481, 7497, 7644, 7675, 7894
\_\_regex_submatch_int ........... 471, 560, 564, 564, 6610, 7475, 7496, 7498, 7501, 7503, 7506, 7519, 7618, 7622, 7624, 7625, 7700, 7708
\_\_regex_submatch_prev_intarray ........... 471, 560, 566, 6828, 6926, 7378, 7479, 7495, 7621, 7627
\_\_regex_success_bool ........... 536, 4477, 4479, 4482, 6568, 6596, 6649, 6661, 7340, 7363, 7387, 7436, 7517, 7603, 7830, 7836, 7877
\_\_regex_success_pos_int ........... 6543, 6565, 6624, 6811, 7477
\_\_regex_success_submatches_tl ........... 535, 597, 6555, 6814, 7632
\_\_regex_tests_action_cost:n ........... 6206, 6208, 6281, 6287, 6296, 6314

Index
\seq_remove_all:Nn \ 152, 156, 162, 163, 1645, 16845, 16869, 16870, 18633
\seq_remove_duplicates:N \ 156, 162, 163, 11569, 16829, 16829, 16830, 16843
\seq_reverse:N \ 157, 826, 16939, 16939, 16955
\seq_set_eq:NN \ 151, 163, 3242, 16682, 16693, 16694, 16695, 16696, 16820, 16968, 39405, 39479
\seq_set_filter:NNn \ 158, 836, 6113, 16778, 16778
\seq_set_from_clist:NN \ 152, 16701, 16701, 16721, 16722, 18632
\seq_set_map:NNn \ 156, 16564, 16574, 16576, 16580, 16584
\seq_set_map:n:Nn \ 156, 16574, 16874, 16878, 16880, 16884
\seq_set_map:nNn \ 159, 17329, 17329
\seq_set_map:e:NNn \ 160, 837, 6126, 17319, 17319, 38892, 38893
\seq_set_map:xx:NNn \ 160, 837, 6126, 17319, 17319, 38892, 38893
\seq_set_split:NNn \ 152, 6112, 6125, 9160, 16733, 16733, 16772, 16773, 19157, 19160
\seq_set_split_keep_spaces:NNn \ 152, 16733, 16737, 16767
\seq_show:N \ 164, 623, 728, 17418, 17418, 17419
\seq_shuffle:N \ 157, 16967, 16968, 17004
\seq_sort:Nn \ 156, 16874, 16878, 16880, 16884
\seq_split:Nn \ 164, 6129, 17363, 17363, 17397, 17399
\seq_sort:Nn \ 164, 6129, 17363, 17363, 17397, 17399
\seq_sort:Nnnn \ 160, 17363, 17363, 17385, 17385, 17398
\g__seq_internal_seq \ 164, 17443
\l__seq_internal_b_tl \ 164, 17443
\l__seq_internal_b_int \ 164, 17443
\l__seq_internal_a_tl \ 164, 17443
\l__seq_internal_a_int \ 164, 17443
\__seq_int_eval:w \ 16873, 16873, 16920, 16930
\__seq_item:n \ 156, 157, 3241, 3241, 3243, 3243, 16957
\__seq_item:nNn \ 164, 17443
\__seq_item:n \ 164, 17443
\__seq_pop:NNNN \ 17049, 17049, 17053
\__seq_pop:NNNN \ 17049, 17049, 17053
\__seq_pop:NNNN \ 17049, 17049, 17053
\__seq_pop_left:NNN \ 17057, 17057, 17059, 17059, 17060, 17125, 17128
str commands:
\c_ampsersand_str ........ 141, 14135
\c_atsign_str ........ 141, 14135
\c_backslash_str ........ 141, 14135
14840, 14842, 14865, 14894, 14896, 14928, 14937, 14941, 39342, 39352
\c_circumflex_str .... 141, 14135
\c_colon_str .... 141, 14135
19614, 19620, 21705, 37445, 37450
\c_dollar_str .... 141, 14148
\c_hash_str .... 141, 14135
14808, 14911, 15486, 15487, 15490, 15493, 31272, 31301, 31305, 31374, 39240, 39808, 39810, 39870, 39918, 39923, 39953, 39957, 39977
\c_left_brace_str . 141, 488, 4650, 5005, 5069, 5089, 5102, 5126, 5598, 5609, 5613, 5692, 5716, 6887, 14135
\c_percent_str . 141, 14135
141, 14145, 14180, 14964
\c_right_brace_str ........ 141, 4686, 5075, 5095, 5108, 5616, 5629, 5713, 6884, 14135, 22709
\str_case:Nn .... 132, 13584, 13609
\str_case:nn ........ 132, 13584, 13606, 13607, 13609, 37963, 38001
\str_case:NSnF .... 132, 13584, 13610, 13611, 13612
\str_case:NNNF .... 132, 593, 850, 947, 5940, 8795, 8830, 9214, 9927, 13584, 13589, 13594, 13599, 13610, 13611, 13612, 21909, 21954, 25480, 29606, 29620
\str_case:e:nn .... 133, 13584, 13619, 13641, 13642
\str_case:NNNF .... 133, 5073, 13584, 13624, 13629, 13634, 14863
\str_casefold:n .... 139, 146, 297, 14000, 14000, 14003, 24291, 37632, 38874, 38875, 38876, 38877, 38878, 38879, 38880, 38881, 38955, 38956, 38963, 38964, 38971, 38972, 38981, 38982
\c_str_cctab .... 291, 1261, 30663
\str_clear:N .... 130, 13413, 21867, 21836, 22241, 22242, 39485
\str_clear_new:N .... 130, 13413
\str_compare:nNn .... 13539, 13546
\str_compare:nNNF .... 133, 13539
\str_compare:p:nNn .... 133, 13539
\str_concat:NNN .... 130, 13413, 13436, 13438, 39425
\str_const:Nn ........ 130, 8648, 8667, 8685, 8717, 8786, 9017, 9176, 11735, 11742, 11746, 11750, 13440, 13444, 13471, 13415, 13416, 13417, 13418, 13419, 13410, 13411, 13412, 14143, 14144, 14145, 14146, 14147, 14904, 14905, 14927, 21561, 21562, 21563, 21564, 21565, 21566, 21567, 30999, 39606
\str_convert_pdfname:n .... 144, 15462, 15462, 37729
\str_count:N .... 135, 3998, 9356, 9357, 9530, 9531, 10613, 10691, 13932, 13933, 30201, 30206, 30282
\str_count:n .... 135, 3992, 13932, 13932, 13934
\str_count_ignore_spaces:n .... 135, 748, 3580, 13932, 13947
\str_count_spaces:N .... 135, 748, 13912, 13912, 13914
\str_count_spaces:n .... 135, 748, 13912, 13913, 13915, 13938
\str_declare_eight_bit_encoding:nnn .... 38882, 38883
\str_fold_case:n .... 38866, 38875, 38877
\str_fold_case:n .... 38878, 38879, 38881
\str_gclear:N .... 136, 13413, 39566
\str_gclear_new:N .... 130, 13413
\str_gconcat:NNN .... 130, 13413, 13439, 39426
\str_gput_left:N .... 131, 13440, 13454, 13473, 39568
\str_gput_right:N .... 131, 13440, 13464, 13475, 39569
\str_gremove_all:N .... 138, 13522, 13524, 13527
\str_gremove_once:N .... 138, 13516, 13518, 13521
\str_greplace_all:N .... 138, 13476, 13482, 13487, 13525
\str_greplace_once:N .... 138, 13476, 13478, 13485, 13519
.str_gset:N .... 244, 22165
\str_gset:N .... 131, 11476, 11477, 11478, 13440, 13442, 13470, 30192, 30196
\str_gset_convert:NNnn .... 144, 14344, 14346, 14358
\str_gset_convert:NNnTF .... 144, 14344
\str_gset:e:N .... 244, 22165
\str_gset_eq:NN .... 130, 11463, 11464, 11465, 13413, 13433, 13435, 30276, 39408, 39567
\str_gset:x:N .... 38809
Index

13469, 13710, 21621, 21623, 22243, 22245, 22250, 22472, 30190, 30194
\str_set_convert:Nnn : 144, 145, 759, 770, 14344, 14344, 14349
\str_set_convert:NnnTF : 144, 759, 14344
\str_set_e:N : 244, 22165
\str_set_eq:NN : 130, 13413, 13432, 13434, 30272, 39407, 39486
\str_set_x:N : 38809
\str(show):N : 140, 14153, 14154, 14159
\str(show):n : 140, 14153, 14153
\str:tail:N : 136, 13895, 13985, 13986
\str:tail:n : 136, 451, 13985, 13985, 13987, 31454
\str:tail:ignore_spaces:n : 136, 13985, 13994
\str:titlecase:n : 38961, 38962
\str:upper_case:n : 38866, 38871, 38873
\str:upper:case:n : 139, 297, 14000, 14002, 14005, 38870, 38871, 38872, 38873, 38959, 38960, 38979, 38980
\str:use:N : 135, 13413
\c:tilde_str : 141, 14135
\g:tap_str : 141, 14149
\l:tap_str : 138, 14149
\g:tab_str : 141, 14149
\l:tab_str : 141, 14149
\c: underscore_str : 141, 14135, 29902
\c:zero_str : 141, 14135, 30166, 30172, 30272, 30276

str internal commands:
\g: str_alias:prop : 762, 14177, 14147
\c: str:byte: distance : 14258
\c: str:byte:0 : 14258
\c: str:byte:1 : 14258
\c: str:byte:255 : 14258
\c: str:byte: (number) : 757
\l: str:byte: flag : 764, 14205, 14485, 14499, 14502, 14767, 14776, 14820, 14835
\str:case:nn : 13584, 13587, 13592, 13597, 13602, 13604
\str:case:nTF : 13584, 13604, 13605, 13613, 13617
\str:case:e:nn : 13584, 13622, 13627, 13632, 13637, 13639
\str:case:e:nTF : 13584, 13640, 13643, 13647
\str:case:end:nn : 13584, 13616, 13646, 13649
\str:change:case:nn : \__str_change_case_aux:nn : 14000, 14008, 14011
\str:change:case:char:n : 14000, 14025, 14034
\str:change:case:char:nn : 14000, 14045, 14074, 14077, 14083, 14092, 14105, 14114
\str:change:case:char:nnn : 14000, 14117, 14119
\str:change:case:char:aux:nn : 14000, 14099, 14115
\str:change:case:char:auxi:nn : 14000, 14052, 14056, 14062
\str:change:case:codepoint:n : 14000, 14038, 14044, 14047
\str:change:case:codepoint:nn : 14000, 14065, 14075
\str:change:case:codepoint:nnn : 14000, 14068, 14081
\str:change:case:codepoint:nnn : 14000, 14090
\str:change:case:codepoint:nNNNN : 14069
\str:change:case:end:nn : 14000, 14009
\str:change:case:end:nw : 14019, 14037
\str:change:case:loop:nn : 14000, 14013, 14021, 14032, 14112
\str:change:case:output:nn : 14000, 14016, 14018, 14031, 14107
\str:change:case:result:n : 14000, 14014, 14016, 14017, 14019
\str:change:case:space:n : 14000, 14024, 14029
\str:collect:delimit:by:q:stop:w : 13863, 13886, 13886
\str:collect:end:nnnnnnnn : 747, 13886, 13905, 13910
\str:collect:end:nn : 13886, 13893, 13903
\str:collect:loop:nn : 13886, 13888, 13899
\str:collect:loop:nnnnnnnnn : 13886, 13891, 13897
\str:convert:nn : 761, 762, 14389, 14390, 14404, 14404
\str:convert:nnnn : 762, 14404, 14408, 14413
\str:convert:NNNNN : 14386, 14391, 14394
\str:convert:NNNNNN : 14344, 14345, 14347, 14352, 14361, 14366
\_\_str_decode_utf_viii_aux:WN
                      .............. 15055, 15098, 15110
\_\_str_decode_utf_viii_contiuation:WN
                      .............. 15055, 15083, 15090, 15126
\_\_str_decode_utf_viii_end: ...
                      .............. 15055, 15065, 15140
\_\_str_decode_utf_viii_overflow:w
                      .............. 15055, 15124, 15133
\_\_str_decode_utf_viii_start:N ...
                      15055, 15064, 15070, 15088, 15091, 15108, 15111, 15131
\_\_str_decode_utf_vxi:Nw ...
                      785, 15236, 15237, 15239, 15248, 15251, 15252, 15255
\_\_str_decode_utf_vxi_bom:NN ...
                      15236, 15242, 15245
\_\_str_decode_utf_vxi_error:NN ...
                      .............. 15270, 15288, 15307, 15316, 15321, 15322
\_\_str_decode_utf_vxi_extra:NNw
                      .............. 15270, 15278, 15320
\_\_str_decode_utf_vxi_pair:NN ...
                      785, 786, 15264, 15270, 15279, 15282, 15285, 15309
\_\_str_decode_utf_vxi_pair_end:Nw ...
                      15270, 15273, 15289, 15311
\_\_str_decode_utf_xxxii:WN ...
                      .............. 789, 15394, 15395, 15397, 15406, 15409, 15410, 15413
\_\_str_decode_utf_xxxii_bom:NNNN ...
                      15394, 15400, 15403
\_\_str_decode_utf_xxxii_end:w ...
                      15394, 15430, 15450
\_\_str_decode_utf_xxxii_loop:NNNN ...
                      15394, 15421, 15427, 15448
\_\_str_encode_clist_char:n ...
                      14360, 14562, 14565
\_\_str_encode_eight_bit_aux:NN ...
                      .............. 14641, 14646, 14654
\_\_str_encode_eight_bit_aux:nn ...
                      .............. 14641, 14656, 14664
\_\_str_encode_native_char:n ...
                      .............. 14513, 14520, 14521, 14528, 14532
\_\_str_encode_utf_vii_loop:ww
                      .............. 778
\_\_str_encode_utf_vii_char:n ...
                      .............. 14980, 14981, 14982
\_\_str_encode_utf_vii_loop:ww ...
                      .............. 14980, 14984, 14991, 14997
\_\_str_encode_utf_vxi:NN ...
                      .............. 15151, 15153, 15157, 15159, 15160
\_\_str_encode_utf_vxi_be:nn ...
                      .............. 783
\_\_str_encode_utf_xvi:NN ...
                      .............. 15151, 15164, 15167
\_\_str_encode_utf_xvi_be:n ...
                      .............. 15334, 15336, 15340, 15343
\_\_str_decode_utf_xvi:WW ...
                      .............. 15334, 15345, 15348
\_\_str_decode_utf_xvi_bom:NN ...
                      .............. 15344, 15342, 15354
\_\_str_decode_utf_xvi_be:n ...
                      .............. 15344, 15356, 15359
\_\_str_end_flag .............. 15185, 15365
\_\_str_name ...
                      .............. 15187, 15202, 15229, 15260, 15366, 15373, 15387, 15416, 15454
\_\_str_error Bool .............. 14204, 14341, 14351, 14355, 14360, 14364
\_\_str_error_flag .............. 14205, 14257, 14529, 14535, 14621, 14623, 14635, 14650, 14652, 14668, 14719, 14730, 14739, 14752, 14768, 14777, 14790, 14795, 14821, 14836, 14876, 15057, 15068, 15079, 15103, 15117, 15137, 15144, 15162, 15165, 15174, 15257, 15268, 15324, 15417, 15425, 15435, 15440, 15455
\_\_str_escape_hex_char:N ...
                      .............. 14900, 14901, 14902
\_\_str_escape_name_char:n ...
                      .............. 14904, 14907, 14908, 15475, 15498
\_\_str_escape_name_not_str ...
                      .............. 773, 14904
\_\_str_escape_string_char:N ...
                      .............. 14927, 14930, 14931
\_\_str_escape_string_str ...
                      .............. 14927
\_\_str_escape_url_char:n ...
                      .............. 14959, 14960, 14961
\_\_str_extra .............. 15002, 15185
\_\_str_flag .......... 15003, 15012, 15035, 15059, 15078, 15186, 15201, 15224, 15259
\_\_str_filter_bytes:n ...
                      .............. 14467, 14473, 14490, 14501, 14782, 14844
\_\_str_filter_bytes_aux:N ...
                      .............. 14467, 14473, 14479, 14487
\_\_str_head:w 749, 13970, 13974, 13978
\_\_str_hexdecimal_use:N ...
                      .............. 14239
\_\_str_hexdecimal_use:NTF ...
                      .............. 772, 14239, 14736, 14746, 14785, 14787
\_\_str_if_contains_char:N ...
                      .............. 14207, 14916, 14922, 14935
\@changed@cmd ......... 31989, 34238
\@classoptionslist .... 9208, 9210, 9212
\@current@cmd ........... 31988, 34237
\@current@namestack ..... 661, 10950, 10952, 10953
\@expandfinalise@setup@... 8856, 8858, 30158, 30160, 30676, 30677, 31675, 31677, 33661, 33663, 38446, 38448
\@expandlua@data@bytecode .... 19
\@filelist ............ 104, 661, 674, 677, 11442, 11563, 11566, 11575, 11580
\@firstofone ........... 24
\@firstoftwo .......... 378
\@gobble .............. 26
\@gobbletwo .......... 26
\@Kernel@after@begindocument ... 8860, 31679, 33665, 38450
\@Kernel@before@begindocument .... 38703, 38705
\@protected@protect@testopt ... 1298, 31976
\@secondoftwo ......... 378
\@optoken ............. 200
\@tempa ............... 1243, 1257, 1260
\@ttfor ............... 360, 1243
\@ucclist ............ 1335, 33690
\@unexpandable@protect .... 1062
\@unusedoptionlist .... 9227
\@active@prefix ......... 31837
\@afterassignment ...... 565
\@aftergroup ........... 14
\@AtBeginDocument ........ 360
\@begingroup ............ 13
\@begin@group .......... 200
\@botmark .............. 904
\@box .................. 310
\@catcodetable ...... 1258, 1262, 1264
\@char .................. 212
\@chordef .............. 203, 204, 585, 588, 845, 1287
\@conditionally@traceoff .... 653, 9631, 10671
\@conditionally@traceon .... 9649
\@copy ................. 303
\@count ................ 211, 431
\@cr .................. 596
\@CR@shipout .......... 1244
\@csname ............... 21, 366, 663, 664
\@csstring ............ 388
\@current@group@level ... 401, 628, 1261
\@current@group@type .... 401, 628
\@day .................. 74
\@declare@file@substitution .. 30161
\@def .................. 211
\@detokenize ............ 115
\@development@branch@name 11651, 11652
\@dimen .................. 903
\@dimendef .............. 903
\@dimexpr .......... 1359
\@directlua ............ 105
\@dp .................. 304, 1063, 1064
\@dup@shipout .... 1245
\@e@alloc@e@codetable@count ... 30634
\@e@alloc@e@top .......... 431, 3173
\@edef .................. 3, 6, 694
\@egroup ............... 200
\@else .................. 28
\@end ............... 359, 621
\@endcsname ......... 21
\@endgroup ............. 13
\@endinput ............... 84
\@endlinechar ........ 93, 125, 289–
\@endtemplate .......... 73, 596
\@errhelp ............... 617
\@errshow .............. 617, 618
\@errorcontextlines .. 368, 618, 728, 1362
\@escapechar .......... 115, 387, 402, 471, 652
\@every@eof ............. 702
\@every@job ............. 606
\@every@par .......... 30, 207, 404
\@expand@after .......... 39, 41
\@expanded ............. 3, 6, 26, 34, 335, 407, 410, 423, 701
\@fi ................. 28, 210
\@first@mark ............ 417, 904
\@fmtname ............... 75
\@font .................. 210, 902
\@fontdimen ............ 61, 257, 1011–1014
\@frozen@every@display .... 1230
\@frozen@every@math .... 1231
\@future@let ..............
\@global ...... 446, 449, 451, 462, 596, 909, 912
\@hp@J@Torg@shipout ...... 1246
\@halign .............. 73, 103, 405, 596, 895
\@hskip ................. 235
\@ht .................. 304, 1063, 1064
\@hyphen ............... 904
\@hyphenchar .......... 1011
\@if .................... 29
\@ifcase .......... 179
\@ifcat ................. 29
\@ifc@name ............ 29
\@ifdef@ined .......... 29
\@ifdim ................. 238
\@iffalse ............... 99
\@ifh@box .............. 28, 66, 692
\@ifh@mode ............... 314
\@ifhmode ............... 29
\text{Index} 1643

\text{\textunderscore cstring}:D ........ 185, 1409
\text{\textunderscore currentcjktoken}:D ....... 1136, 1200
\text{\textunderscore currentgrouplevel}:D ...... 1263, 477, 20094, 20112, 30467, 30547, 30557, 30564, 37666
\text{\textunderscore currentgroupertype}:D .. 478
\text{\textunderscore currentifbranch}:D ....... 479
\text{\textunderscore currentiflevel}:D ........ 480
\text{\textunderscore currentiftype}:D ........ 481
\text{\textunderscore currentspacingmode}:D ... 1137
\text{\textunderscore currentspacingmode}:D .. 1138
\text{\textunderscore day}:D .................. 186, 1289, 1293
\text{\textunderscore deadcycles}:D ............ 187
\text{\textunderscore def}:D .................... 188, 688, 689, 690, 1472, 1476, 1481, 1486, 39325, 39349, 39383, 39796
\text{\textunderscore defaulthyperchar}:D ...... 189
\text{\textunderscore defaultskewchar}:D ........ 190
\text{\textunderscore deferred}:D ............... 806
\text{\textunderscore delcode}:D ................. 191
\text{\textunderscore delimiter}:D ............... 192
\text{\textunderscore delimitertfactor}:D ....... 193
\text{\textunderscore delimitershortfall}:D ..... 194
\text{\textunderscore detokenize}:D ............. 482, 1418, 1420
\text{\textunderscore dimen}:D .................. 195
\text{\textunderscore dimenlead}:D .............. 196
\text{\textunderscore dimexpr}:D ................. 483, 20747, 34631
\text{\textunderscore directlua}:D ............... 809, 1260, 1270, 8666, 9019, 9020, 9073, 11736, 11761
\text{\textunderscore disablecjktoken}:D .......... 1201
\text{\textunderscore discretionary}:D .......... 197
\text{\textunderscore discretionaryligaturemode}:D 808
\text{\textunderscore dishibitgldue}:D ........... 1139
\text{\textunderscore displayindent}:D .......... 198
\text{\textunderscore displaylimits}:D .......... 199, 37311
\text{\textunderscore displaystyle}:D ........... 200
\text{\textunderscore displaywidowpenalties}:D .. 484
\text{\textunderscore displaywidth}:D ............ 201
\text{\textunderscore displaywidth}:D ............ 202
\text{\textunderscore divide}:D .................. 203, 3150, 5270
\text{\textunderscore doublelyphendemerits}:D .. 204
\text{\textunderscore dp}:D ....................... 205, 34671
\text{\textunderscore draftmode}:D ............... 633, 934
\text{\textunderscore dtou}:D ..................... 1140
\text{\textunderscore dump}:D .................... 206
\text{\textunderscore dviextension}:D ............ 810
\text{\textunderscore dvifeedback}:D .......... 811
\text{\textunderscore dvivariable}:D ............ 812
\text{\textunderscore eachlinedepth}:D .......... 634
\text{\textunderscore eachlineheight}:D .......... 635
\text{\textunderscore edef}:D .................... 207, 1239, 1240, 1256, 1283, 1284, 1289, 1290, 1295, 1296, 1301, 1302, 1473, 1474, 1478, 1483, 1488, 10763, 10821, 38786
\text{\textunderscore efcode}:D ................... 671
\text{\textunderscore elapsedtime}:D ............. 1143
\text{\textunderscoreemergencystretch}:D ........ 209
\text{\textunderscore enablecjktoken}:D ........ 1202, 8654
\text{\textunderscore end}:D ..................... 210, 1226, 1315, 1967
\text{\textunderscore endcsname}:D ............... 211, 1410
\text{\textunderscore endgroup}:D ............... 212, 1224, 1264, 1307, 1423
\text{\textunderscore endinput}:D ............... 213, 9504, 11427, 11661
\text{\textunderscore endl}:D .................... 485
\text{\textunderscore endlinechar}:D ............ 214, 2325, 2332, 2333, 9387, 9406, 9593, 13262, 34770
\text{\textunderscore errorcontextlines}:D ....... 218, 2325, 2332, 2333, 9387, 9406, 9593, 13262, 34770
\text{\textunderscore errorstopmode}:D .......... 219
\text{\textunderscore escapechar}:D .............. 663, 220, 2310, 3637, 3699, 3700, 4044, 4142, 4143, 4146, 4174, 4273, 10380, 10625, 10672, 10678, 14720, 14769, 14822
\text{\textunderscore escapehex}:D ............... 637
\text{\textunderscore escapename}:D ............. 638
\text{\textunderscore escapestring}:D ........... 639
\text{\textunderscore eTeXglueshrinkorder}:D ..... 813
\text{\textunderscore eTeXgluestratchorder}:D ..... 814
\text{\textunderscore eTeXrevision}:D ............ 487
\text{\textunderscore eTeXversion}:D ............. 488
\text{\textunderscore etoksapp}:D ................ 816, 4295, 4296
\text{\textunderscore etokspre}:D ............... 817, 4289, 4290
\text{\textunderscore euc}:D ..................... 1143
\text{\textunderscore everycr}:D ................ 221
\text{\textunderscore everydisplay}:D ............ 222, 1230
\text{\textunderscore everyeof}:D ................ 226
\text{\textunderscore everyjob}:D ............... 224, 1309, 1316
\text{\textunderscore everymath}:D .............. 223
\text{\textunderscore everyvar}:D ............... 225, 1231
\text{\textunderscore everyvar}:D ............... 226
Index

\text{pdffakespace:D} ............. 549  \text{pdfsetmatrix:D} ............. 604
\text{pdffeedback:D} ............. 901  \text{pdfstartlink:D} ............. 605
\text{pdffontattr:D} ............. 550  \text{pdfstartthread:D} ............. 606
\text{pdffontname:D} ............. 551  \text{pdfsuppresstpxinfo:D} ............. 607
\text{pdffontobjnum:D} ............. 552  \text{pdfsuppresswarningdupdest:D} ............. 609
\text{pdfgamma:D} ............. 553  \text{pdfsuppresswarningdupmap:D} ............. 611
\text{pdfgentounicode:D} ............. 554  \text{pdfsuppresswarningpagegroup:D} ............. 613
\text{pdfglyphtounicode:D} ............. 555
\text{pdfhorigin:D} ............. 556  \text{pdfximage:D} ............. 557  \text{pdfximagegamma:D} ............. 558  \text{pdfximageapplygamma:D} ............. 559
\text{pdfimagecolor:D} ............. 560  \text{pdfthread:D} ............. 614
\text{pdfimagesresolution:D} ............. 561  \text{pdfthreadmargin:D} ............. 615
\text{pdfincludechars:D} ............. 561  \text{pdftrailer:D} ............. 616
\text{pdfincluisioncopyfonts:D} ............. 562  \text{pdftrailerid:D} ............. 617
\text{pdfinclusionerrorlevel:D} ............. 564
\text{pdfinfo:D} ............. 565  \text{pdfuniqueresname:D} ............. 618
\text{pdfinfoommitdate:D} ............. 566  \text{pdfvariable:D} ............. 902
\text{pdfinterwordspace:D} ............. 567  \text{pdfvorigin:D} ............. 619  \text{pdfxform:D} ............. 620, 958
\text{pdflastannot:D} ............. 568  \text{pdfxmlname:D} ............. 621
\text{pdflastlink:D} ............. 570  \text{pdfximage:D} ............. 622, 959
\text{pdflastobj:D} ............. 571  \text{pdfximagebbox:D} ............. 623
\text{pdflastform:D} ............. 572, 941  \text{pdfxform:D} ............. 624  \text{pdfxformname:D} ............. 625
\text{pdflastxform:D} ............. 573, 943  \text{pdfximage:D} ............. 626
\text{pdflastximage:D} ............. 574, 945  \text{pdfxform:D} ............. 627
\text{pdflastximagecolordepth:D} ............. 575
\text{pdflastximagepages:D} ............. 576, 945  \text{pdfxmlname:D} ............. 628
\text{pdflinkmargin:D} ............. 577  \text{pdfxmlns:D} ............. 629
\text{pdflinkmargin:D} ............. 577  \text{pdfxmlns:D} ............. 630
\text{pdfliteral:D} ............. 578  \text{pdfxmlns:D} ............. 631
\text{pdfmajorversion:D} ............. 581  \text{pdfxmlns:D} ............. 632
\text{pdfmapfile:D} ............. 579, 1279  \text{pdfxmlns:D} ............. 633
\text{pdfmapline:D} ............. 580, 1280  \text{pdfxmlns:D} ............. 634
\text{pdfminorversion:D} ............. 581
\text{pdfnames:D} ............. 582  \text{pdfxmlns:D} ............. 635
\text{pdfnobuiltintouname:D} ............. 584  \text{pdfxmlns:D} ............. 636
\text{pdfobj:D} ............. 585  \text{pdfxmlns:D} ............. 637
\text{pdfobjcompresslevel:D} ............. 586  \text{pdfxmlns:D} ............. 638
\text{pdfomitcharset:D} ............. 587  \text{pdfxmlns:D} ............. 639
\text{pdfomitcharset:D} ............. 587  \text{pdfxmlns:D} ............. 640
\text{pdfoutline:D} ............. 588  \text{pdfxmlns:D} ............. 641
\text{pdfoutput:D} ............. 599
\text{pdfoutput:D} ............. 599, 549, 589, 949, 1312, 8694, 8700, 8708, 9179
\text{pdfpageattr:D} ............. 590  \text{pdfxmlns:D} ............. 642
\text{pdfpagebox:D} ............. 592  \text{pdfxmlns:D} ............. 643
\text{pdfpageresources:D} ............. 594  \text{pdfxmlns:D} ............. 644
\text{pdfpagesattr:D} ............. 591, 595  \text{pdfxmlns:D} ............. 645
\text{pdfpagesattr:D} ............. 591, 595
\text{pdfrefobj:D} ............. 596  \text{pdfxmlns:D} ............. 646
\text{pdfrefxform:D} ............. 597, 955  \text{pdfxmlns:D} ............. 647
\text{pdfrefximage:D} ............. 598, 956  \text{pdfxmlns:D} ............. 648
\text{pdfrefximage:D} ............. 598, 956  \text{pdfxmlns:D} ............. 649
\text{pdfrestore:D} ............. 599  \text{pdfxmlns:D} ............. 650
\text{pdfrestore:D} ............. 599  \text{pdfxmlns:D} ............. 651
\text{pdfreval:D} ............. 600  \text{pdfxmlns:D} ............. 652
\text{pdfreval:D} ............. 600  \text{pdfxmlns:D} ............. 653
\text{pdfrunninglinkoff:D} ............. 601  \text{pdfxmlns:D} ............. 654
\text{pdfrunninglinkoff:D} ............. 601
\text{pdfrunninglink:D} ............. 602
\text{pdfsave:D} ............. 603  \text{pdfxmlns:D} ............. 655
\text{pdfxmlns:D} ............. 656
Index

1653

text commands:
\text_titlecase_all:nn ............. 31657, 31892
\text_titlecase:nn ............. 296, 297, 299, 31659, 31856, 32295
\text_case_switch:nnn ............ 298, 31849, 32345, 32646, 32646
\text_declare_case_equivalent:Nn ............. 298, 32608, 32608
\text_declare_expand_equivalent:Nn 296, 32050, 32055, 32058, 32073
\text_declare_lovercase_mapping:nn ............. 298, 32613, 32613
\text_declare_lovercase_mapping:nnn ............. 298, 32613, 32629
\text_declare_purify_equivalent:Nn 299, 34251, 34251, 34256, 34264, 34265, 34266, 34267, 34284, 34309, 34310, 34312, 34313, 34315, 34318, 34319, 34325, 34327, 34328, 34329, 34333, 34366, 34381
\text_declare_titlecase_mapping:nn ............. 298, 32613, 32615
\text_declare_titlecase_mapping:nnn ............. 298, 32613, 32631
\text_declare_uppercase_mapping:nnn ............. 298, 32613, 32633
\text_expand:n ................. 296, 297, 299, 300, 31713, 31713, 31713, 32103, 33070, 34064
\text_expand_exclude_tl ............. 296, 299, 31670, 31855
\text_letterlike_tl ............. 31657, 31912
\text_lowercase:n ............. 139, 198, 297, 32079, 32079, 38902, 38905, 38907, 38909, 38949, 38950, 38965, 38966, 38966
\text_lovercase:nn ............. 297, 32079, 32087, 38910, 38912
\text_map_break: ............. 300, 33700, 33712, 33721, 33731, 33747, 33788, 33936, 34040, 34041, 34043, 34051
\text_map_break:n ............. 300, 33706, 34042
\text_map_function:nN .......... 300, 33705, 33706, 33706, 34049
\text_map_inline:n ............. 300, 34044, 34044
\text_math_arg_tl .............. 296, 299, 31666, 31854, 32294, 34180
\text_math_delims_tl .............. 296, 299, 31668, 31772, 32224, 34109
\text_purify:n ............. 299, 31058, 34058
\text_titlecase:n ............. 38900, 38901
\text_titlecase:nn ............. 38900, 38904
\text_titlecase_all:nn ............. 139, 297, 32079, 32083
\text_uppercase:n ............. 139, 199, 297, 32079, 32081, 38913, 38915, 38951, 38952, 38967, 38968
\text_uppercase:n ............. 297, 32079, 32089, 38916, 38918

text internal commands:
\text_case_switch_marker: ............. 32133, 32136, 32181, 32195, 32325
\text_change_case:nnnn ............. 32133, 32136, 32181, 32195, 32325
\text_change_case:nnnn ............. 32133, 32136, 32181, 32195, 32325
\text_change_case_auxi:nnnn ............. 32133, 32136, 32181, 32195, 32325
\text_change_case_auxii:nnnn ............. 32133, 32136, 32181, 32195, 32325
\text_change_case_BCP:nnnn ............. 32133, 32136, 32181, 32195, 32325
\text_change_case_BCP:nnnn ............. 32133, 32136, 32181, 32195, 32325
\text_change_case_uppercase-boundary-upper_el-x-iota:nnnn ............. 33191
\text_change_case_uppercase-boundary-upper_el:nnn ............. 33191
\text_change_case_uppercase-boundary-upper_el:nnn ............. 33191
\text_change_case_uppercase-boundary-upper_el:nnn ............. 33191
\text_change_case_uppercase-boundary-upper_el:nnn ............. 33191
Index

\_\_tl_tmp:w .......................... 710, 716, 12635, 12636, 12642, 12655, 12832, 12871, 13019, 13034, 13302
\_\_tl_trim_mark: ........................ 715, 716, 12819, 12824, 12832, 12839, 12840, 12847, 12851, 12853, 12856, 12869
\_\_tl_trim_spaces:nn .............................. 971, 12818, 12824, 12832, 12834
\_\_tl_trim_spaces_aux:w ................... 716, 12832, 12836, 12847, 12850, 12856
\_\_tl_trim_spaces_auxi:w .................... 716, 12832, 12840, 12855
\_\_tl_trim_spaces_auxii:w .................... 716, 12832, 12841, 12858, 12861, 12865
\_\_tl_use_none_delimit_by_q_act:-stop:w .......................... 13019
\_\_tl_use_none_delimit_by_s_act:-stop:w .......................... 13053, 13058
\_\_tl_use_none_delimit_by_s:-stop:w .......................... 712, 12686, 12698, 12705, 12709, 12738, 12746
\_tojis .......................... 1192

token commands:
\c_alignment_token ............................. 200, 899, 1429, 3816, 19299, 19334, 19373, 31503, 37287
\c_catcode_letter_token .......................... 200, 900, 3812, 19290, 19334, 19402, 31515
\c_catcode_other_token .......................... 200, 900, 3810, 19293, 19334, 19407, 31518
\c_group_begin_token .......................... 200
\c_group_end_token .......................... 200
\c_math_subscript_token .............................. 200, 900, 3820, 19308, 19334, 19392, 31474, 31509, 37302
\c_math_superscript_token .............................. 200, 900, 3818, 19305, 19334, 19387, 31506, 37303
\c_math_toggle_token .......................... 200, 899, 3814, 19296, 19334, 19368, 31471, 31500
\c_parameter_token ................................... 200, 544, 899, 19334, 19377, 19380
\c_space_token ................................... 40, 115, 126, 200, 207, 718, 900, 3649, 3679, 3822, 4012, 4052, 4192, 4640, 4681, 6990, 10801, 12917, 12957, 19317, 19334, 19397, 19730, 19755, 19868, 31475, 31512
\token_case_catcode:Nn .............................. 205, 19661, 19666, 19665, 19667, 37300
\token_case_charcode:Nn .............................. 205, 19661, 19669
\token_case_charcode:NnTF .............................. 205, 19661, 19667, 38925, 38926
\token_case_meaning:Nn .............................. 205, 19661, 19671, 19673, 19675
\token_if_active:N .............................. 19410
\token_if_active:NTF .............................. 202, 19410
\token_if_active:p:N .............................. 202, 19410, 31828, 32001, 32485, 32528, 34202
\token_if_alignment:N .............................. 19371
\token_if_alignment:NTF .............................. 202, 19371
\token_if_alignment:p:N .............................. 202, 19371
\token_if_chardef:N .............................. 203, 19667, 19490
\token_if_chardef:p:N .............................. 203, 19480, 31437
\token_if_cs:N .............................. 19444
\token_if_cs:NTF .............................. 202, 19320, 19444, 31824, 32282, 32727, 32755, 32865, 32953, 33203, 33749, 33940, 34030, 34216
\token_if_catcode:N .............................. 19444, 32000, 33444, 33507, 33542, 33597, 33791, 34201
\token_if_dim_register:NTF .............................. 204, 3969, 19480
\token_if_dim_register:p:N .............................. 204, 19480
\token_if_eq_catcode:NN .............................. 19417
\token_if_eq_catcode:NNTF .............................. 202, 19444, 19664, 19666, 19668
\token_if_eq_catcode:p:NN .............................. 202, 19444, 206, 19417, 19662, 19664, 19666, 19668
\token_if_eq_catcode:NN .............................. 202, 19417
\token_if_eq_catcode:NNTF .............................. 203, 19422, 206, 19417, 19662, 19664, 19666, 19668
\token_if_eq_catcode:p:NN .............................. 203, 19422
\token_if_eq_charcode:NN .............................. 203, 206, 19422
\token_if_eq_charcode:NNTF .............................. 203, 19422
\token_if_eq_charcode:p:NN .............................. 203, 19422
\token_case_catcode:Nn .............................. 205, 19661, 19666, 19665, 19667, 37300
\token_case_charcode:Nn .............................. 205, 19661, 19669
\token_case_charcode:NnTF .............................. 205, 19661, 19671, 19673, 19675
\token_case_meaning:Nn .............................. 205, 19661, 19677, 38925, 38926
\token_case_meaning:NTF .............................. 205, 19661, 19680, 19682.
\texttt{\textbackslash止め} commands: use

\texttt{\textbackslash明るく} commands: undefine:

\texttt{\textbackslash明るく} commands:

\texttt{\textbackslash使用} commands:

\texttt{\textbackslash使用～～} commands:
Index

\vbox_set_end: ........................................... 308, 309, 34910, 34924, 34930, 35615
\vbox_set_split_to_ht:NNn ............................. 309, 34949, 34949, 34951, 39510
\vbox_set_to_ht:NNn .......................... 308, 309, 34918, 34988, 34908, 39507
\vbox_set_to_ht:Nw ...................... 309, 34931, 34931, 34943, 39509
\vbox_set_top:NN ................ 308, 34886, 34886, 34896, 35559, 35636, 39506
\vbox_to_h:nn .................. 308, 34864, 34864
\vbox_to_zero:n .................. 308, 34860, 34862
\vbox_unpack:N .................. 309, 34945, 34945, 34947, 35559, 35636
\vbox_unpack_drop:N ............... 310, 34945, 34946, 34948
\vcenter ........................................... 454
\vcoffin commands: ..................................... 478
\vcoffin_set:Nnw .................................. 706
\vcoffin_set:Nnn .................................. 707
\vcoffin_gset:Nnw .................................. 708
\vcoffin_gset:Nnn .................................. 709
\vcoffin_gset_end: .................................. 710
\vcoffin_set_top:Nn .................................. 711
\vcoffin_set_to_ht:Nnw .................................. 712
\vcoffin_set_to_ht:Nnn .................................. 713
\vcoffin_set_split_to_ht:NNn .................................. 714
\vcoffin_set:NN .................................. 715
\vcoffin_set:Nw .................................. 716
\vcoffin_set:N .................................. 717
\vcoffin_set:end: .................................. 718
\vcoffin_gset_end: .................................. 719
\vcoffin_gset:NNn .................................. 720
\vcenter ........................................... 721
\vcoffin_set:end: .................................. 722
\vcoffin_set:N .................................. 723
\vcoffin_gset:N .................................. 724
\vcoffin_gset:Nw .................................. 725
\vcenter ........................................... 726
\vcoffin_gset:N .................................. 727
\vcoffin_set:N .................................. 728
\vcoffin_set:Nn .................................. 729
\vcenter ........................................... 730
\vcoffin_set:N .................................. 731
\vcenter ........................................... 732
\vcoffin_set:N .................................. 733
\vcoffin_set:N .................................. 734
\vcenter ........................................... 735
\vcoffin_set:N .................................. 736
\vcenter ........................................... 737
\vcoffin_set:N .................................. 738
\vcenter ........................................... 739
\vcoffin_set:N .................................. 740
\vcenter ........................................... 741
\vcoffin_set:N .................................. 742
\vcenter ........................................... 743
\vcoffin_set:N .................................. 744
\vcenter ........................................... 745
\vcoffin_set:N .................................. 746
\vcenter ........................................... 747
\vcoffin_set:N .................................. 748
\vcenter ........................................... 749
\vcoffin_set:N .................................. 750
\vcenter ........................................... 751
\vcoffin_set:N .................................. 752
\vcenter ........................................... 753
\vcoffin_set:N .................................. 754
\vcenter ........................................... 755
\vcoffin_set:N .................................. 756
\vcenter ........................................... 757
\vcoffin_set:N .................................. 758
\vcenter ........................................... 759
\vcoffin_set:N .................................. 760
\vcenter ........................................... 761
\vcoffin_set:N .................................. 762
\vcenter ........................................... 763
\vcoffin_set:N .................................. 764
\vcenter ........................................... 765
\vcoffin_set:N .................................. 766
\vcenter ........................................... 767
\vcoffin_set:N .................................. 768
\vcenter ........................................... 769
\vcoffin_set:N .................................. 770
\vcenter ........................................... 771
\vcoffin_set:N .................................. 772
\vcenter ........................................... 773
\vcoffin_set:N .................................. 774
\vcenter ........................................... 775
\vcoffin_set:N .................................. 776
\vcenter ........................................... 777
\vcoffin_set:N .................................. 778
\vcenter ........................................... 779
\vcoffin_set:N .................................. 780
\vcenter ........................................... 781
\vcoffin_set:N .................................. 782
\vcenter ........................................... 783
\vcoffin_set:N .................................. 784
\vcenter ........................................... 785
\vcoffin_set:N .................................. 786
\vcenter ........................................... 787
\vcoffin_set:N .................................. 788
\vcenter ........................................... 789
\vcoffin_set:N .................................. 790
\vcenter ........................................... 791
<table>
<thead>
<tr>
<th>Command</th>
<th>Page</th>
<th>Command</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>\kanjiskip</td>
<td>1195</td>
<td>\xtokspre</td>
<td>931</td>
</tr>
<tr>
<td>\leaders</td>
<td>470</td>
<td></td>
<td></td>
</tr>
<tr>
<td>\spaceskip</td>
<td>471</td>
<td>\ybaselineshift</td>
<td>1197</td>
</tr>
<tr>
<td>\spcode</td>
<td>1196</td>
<td>\year</td>
<td>472, 1302, 9012</td>
</tr>
<tr>
<td>\xtoksapp</td>
<td>930</td>
<td>\yoko</td>
<td>1198</td>
</tr>
</tbody>
</table>