2* The present implementation has a long ancestry, beginning in the summer of 1977, when Michael F. Plass and Frank M. Liang designed and coded a prototype based on some specifications that the author had made in May of that year. This original protoTEX included macro definitions and elementary manipulations on boxes and glue, but it did not have line-breaking, page-breaking, mathematical formulas, alignment routines, error recovery, or the present semantic nest; furthermore, it used character lists instead of token lists, so that a control sequence like `\halign` was represented by a list of seven characters. A complete version of TEX was designed and coded by the author in late 1977 and early 1978; that program, like its prototype, was written in the SAIL language, for which an excellent debugging system was available. Preliminary plans to convert the SAIL code into a form somewhat like the present "web" were developed by Luis Trabb Pardo and the author at the beginning of 1979, and a complete implementation was created by Ignacio A. Zabala in 1979 and 1980. The TEX82 program, which was written by the author during the latter part of 1981 and the early part of 1982, also incorporates ideas from the 1979 implementation of TEX in MESA that was written by Leonidas Guibas, Robert Sedgewick, and Douglas Wyatt at the Xerox Palo Alto Research Center. Several hundred refinements were introduced into TEX82 based on the experiences gained with the original implementations, so that essentially every part of the system has been substantially improved. After the appearance of "Version 0" in September 1982, this program benefited greatly from the comments of many other people, notably David R. Fuchs and Howard W. Trickey. A final revision in September 1989 extended the input character set to eight-bit codes and introduced the ability to hyphenate words from different languages, based on some ideas of Michael J. Ferguson.

No doubt there still is plenty of room for improvement, but the author is firmly committed to keeping TEX82 "frozen" from now on; stability and reliability are to be its main virtues.

On the other hand, the `WEB` description can be extended without changing the core of TEX82 itself, and the program has been designed so that such extensions are not extremely difficult to make. The *banner* string defined here should be changed whenever TEX undergoes any modifications, so that it will be clear which version of TEX might be the guilty party when a problem arises.

This program contains code for various features extending TEX, therefore this program is called 'XeTeX' and not 'TEX'; the official name 'TEX' by itself is reserved for software systems that are fully compatible with each other. A special test suite called the "`TRIP` test" is available for helping to determine whether a particular implementation deserves to be known as 'TEX' [cf. Stanford Computer Science report CS1027, November 1984].

MLTEX will add new primitives changing the behaviour of TEX. The *banner* string has to be changed. We do not change the *banner* string, but will output an additional line to make clear that this is a modified TEX version.

A similar test suite called the "`e-TRIP` test" is available for helping to determine whether a particular implementation deserves to be known as 'ε-TEX'.

define *eTeX_version* = 2 { `\eTeXversion` }
define *eTeX_revision* ≡ `".6"` { `\eTeXrevision` }
define *eTeX_version_string* ≡ `´-2.6´` { current ε-TEX version }

define *XeTeX_version* = 0 { `\XeTeXversion` }
define *XeTeX_revision* ≡ `".999993"` { `\XeTeXrevision` }
define *XeTeX_version_string* ≡ `´-0.999993´` { current XeTeX version }

define *XeTeX_banner* ≡ `´This␣is␣XeTeX,␣Version␣3.141592653´`, *eTeX_version_string*,
XeTeX_version_string { printed when XeTeX starts }

define *TeX_banner_k* ≡ `´This␣is␣TeXk,␣Version␣3.141592653´` { printed when TEX starts }
define *TeX_banner* ≡ `´This␣is␣TeX,␣Version␣3.141592653´` { printed when TEX starts }

define *banner* ≡ *XeTeX_banner*
define *banner_k* ≡ *XeTeX_banner*

define *TEX* ≡ *XETEX* { change program name into *XETEX* }

define *TeXXeT_code* = 0 { the TEX--XET feature is optional }

define *XeTeX_dash_break_code* = 1 { non-zero to enable breaks after en- and em-dashes }
define *XeTeX_upwards_code* = 2 { non-zero if the main vertical list is being built upwards }
define *XeTeX_use_glyph_metrics_code* = 3 { non-zero to use exact glyph height/depth }
define *XeTeX_inter_char_tokens_code* = 4 { non-zero to enable `\XeTeXinterchartokens` insertion }
define *XeTeX_input_normalization_code* = 5 { normalization mode:, 1 for NFC, 2 for NFD, else none }
define *XeTeX_default_input_mode_code* = 6 { input mode for newly opened files }
define *XeTeX_input_mode_auto* = 0
define *XeTeX_input_mode_utf8* = 1
define *XeTeX_input_mode_utf16be* = 2
define *XeTeX_input_mode_utf16le* = 3
define *XeTeX_input_mode_raw* = 4
define *XeTeX_input_mode_icu_mapping* = 5
define *XeTeX_default_input_encoding_code* = 7 { *str_number* of encoding name if *mode* = ICU }
define *XeTeX_tracing_fonts_code* = 8 { non-zero to log native fonts used }
define *XeTeX_interword_space_shaping_code* = 9 { controls shaping of space chars in context when using native fonts; set to 1 for contextual adjustment of space width only, and 2 for full cross-space shaping (e.g. multi-word ligatures) }
define *XeTeX_generate_actual_text_code* = 10
{ controls output of /ActualText for native-word nodes }
define *XeTeX_hyphenatable_length_code* = 11 { sets maximum hyphenatable word length }
define *eTeX_states* = 12 { number of ε-TEX state variables in *eqtb* }

4* The program begins with a normal Pascal program heading, whose components will be filled in later, using the conventions of `WEB`. For example, the portion of the program called '⟨ Global variables 13 ⟩' below will be replaced by a sequence of variable declarations that starts in §13 of this documentation. In this way, we are able to define each individual global variable when we are prepared to understand what it means; we do not have to define all of the globals at once. Cross references in §13, where it says "See also sections 20, 26, ...," also make it possible to look at the set of all global variables, if desired. Similar remarks apply to the other portions of the program heading.

define *mtype* ≡ *t@&y@&p@&e* { this is a `WEB` coding trick: }
format *mtype* ≡ *type* { '**mtype**' will be equivalent to '**type**' }
format *type* ≡ *true* { but '*type*' will not be treated as a reserved word }

⟨ Compiler directives 9 ⟩
program *TEX*; { all file names are defined dynamically }
const ⟨ Constants in the outer block 11* ⟩
mtype ⟨ Types in the outer block 18 ⟩
var ⟨ Global variables 13 ⟩
procedure *initialize*; { this procedure gets things started properly }
var ⟨ Local variables for initialization 19* ⟩
begin ⟨ Initialize whatever TEX might access 8* ⟩
end;
⟨ Basic printing procedures 57 ⟩
⟨ Error handling procedures 82 ⟩

6* For Web2c, labels are not declared in the main program, but we still have to declare the symbolic names.

define *start_of_TEX* = 1 { go here when TEX's variables are initialized }
define *final_end* = 9999 { this label marks the ending of the program }

7* Some of the code below is intended to be used only when diagnosing the strange behavior that sometimes occurs when TeX is being installed or when system wizards are fooling around with TeX without quite knowing what they are doing. Such code will not normally be compiled; it is delimited by the codewords '**debug** ... **gubed**', with apologies to people who wish to preserve the purity of English.

Similarly, there is some conditional code delimited by '**stat** ... **tats**' that is intended for use when statistics are to be kept about TeX's memory usage. The **stat** ... **tats** code also implements diagnostic information for `\tracingparagraphs`, `\tracingpages`, and `\tracingrestores`.

```
define debug ≡ ifdef('TEXMF_DEBUG')
define gubed ≡ endif('TEXMF_DEBUG')
format debug ≡ begin
format gubed ≡ end

define stat ≡ ifdef('STAT')
define tats ≡ endif('STAT')
format stat ≡ begin
format tats ≡ end
```

8* This program has two important variations: (1) There is a long and slow version called `INITEX`, which does the extra calculations needed to initialize TeX's internal tables; and (2) there is a shorter and faster production version, which cuts the initialization to a bare minimum. Parts of the program that are needed in (1) but not in (2) are delimited by the codewords '**init** ... **tini**' for declarations and by the codewords '**Init** ... **Tini**' for executable code. This distinction is helpful for implementations where a run-time switch differentiates between the two versions of the program.

```
define init ≡ ifdef('INITEX')
define tini ≡ endif('INITEX')
define Init ≡
    init
    if ini_version then
      begin
define Tini ≡
    end ; tini
format Init ≡ begin
format Tini ≡ end
format init ≡ begin
format tini ≡ end
```

⟨ Initialize whatever TeX might access 8* ⟩ ≡
  ⟨ Set initial values of key variables 23* ⟩
  Init ⟨ Initialize table entries (done by `INITEX` only) 189 ⟩ **Tini**

See also section 1710*.

This code is used in section 4*.

11*. The following parameters can be changed at compile time to extend or reduce TeX's capacity. They may have different values in `INITEX` and in production versions of TeX.

define *file_name_size* ≡ *maxint*
define *ssup_error_line* = 255
define *ssup_max_strings* ≡ 2097151
{ Larger values than 65536 cause the arrays to consume much more memory. }
define *ssup_trie_opcode* ≡ 65535
define *ssup_trie_size* ≡ "3FFFFF
define *ssup_hyph_size* ≡ 65535 { Changing this requires changing (un)dumping! }
define *iinf_hyphen_size* ≡ 610 { Must be not less than *hyph_prime*! }
define *max_font_max* = 9000 { maximum number of internal fonts; this can be increased, but *hash_size* + *max_font_max* should not exceed 29000. }
define *font_base* = 0 { smallest internal font number; must be ≥ *min_quarterword*; do not change this without modifying the dynamic definition of the font arrays. }

⟨ Constants in the outer block 11* ⟩ ≡
hash_offset = 514; { smallest index in hash array, i.e., *hash_base* }
{ Use *hash_offset* = 0 for compilers which cannot decrement pointers. }
trie_op_size = 35111;
{ space for "opcodes" in the hyphenation patterns; best if relatively prime to 313, 361, and 1009. }
neg_trie_op_size = −35111; { for lower *trie_op_hash* array bound; must be equal to −*trie_op_size*. }
min_trie_op = 0; { first possible trie op code for any language }
max_trie_op = *ssup_trie_opcode*; { largest possible trie opcode for any language }
pool_name = *TEXMF_POOL_NAME*; { this is configurable, for the sake of ML-TeX }
{ string of length *file_name_size*; tells where the string pool appears }
engine_name = *TEXMF_ENGINE_NAME*; { the name of this engine }
inf_mem_bot = 0; *sup_mem_bot* = 1; *inf_main_memory* = 3000; *sup_main_memory* = 256000000;
inf_trie_size = 8000; *sup_trie_size* = *ssup_trie_size*; *inf_max_strings* = 3000;
sup_max_strings = *ssup_max_strings*; *inf_strings_free* = 100; *sup_strings_free* = *sup_max_strings*;
inf_buf_size = 500; *sup_buf_size* = 30000000; *inf_nest_size* = 40; *sup_nest_size* = 4000;
inf_max_in_open = 6; *sup_max_in_open* = 127; *inf_param_size* = 60; *sup_param_size* = 32767;
inf_save_size = 600; *sup_save_size* = 80000; *inf_stack_size* = 200; *sup_stack_size* = 30000;
inf_dvi_buf_size = 800; *sup_dvi_buf_size* = 65536; *inf_font_mem_size* = 20000;
sup_font_mem_size = 147483647; { *integer*-limited, so 2 could be prepended? }
sup_font_max = *max_font_max*; *inf_font_max* = 50; { could be smaller, but why? }
inf_pool_size = 32000; *sup_pool_size* = 40000000; *inf_pool_free* = 1000; *sup_pool_free* = *sup_pool_size*;
inf_string_vacancies = 8000; *sup_string_vacancies* = *sup_pool_size* − 23000;
sup_hash_extra = *sup_max_strings*; *inf_hash_extra* = 0; *sup_hyph_size* = *ssup_hyph_size*;
inf_hyph_size = *iinf_hyphen_size*; { Must be not less than *hyph_prime*! }
inf_expand_depth = 10; *sup_expand_depth* = 10000000;

This code is used in section 4*.

12:* Like the preceding parameters, the following quantities can be changed at compile time to extend or reduce TEX's capacity. But if they are changed, it is necessary to rerun the initialization program `INITEX` to generate new tables for the production TEX program. One can't simply make helter-skelter changes to the following constants, since certain rather complex initialization numbers are computed from them. They are defined here using `WEB` macros, instead of being put into Pascal's **const** list, in order to emphasize this distinction.

define *hash_size* = 15000 { maximum number of control sequences; it should be at most about (*mem_max* − *mem_min*)/10; see also *font_max* }
define *hash_prime* = 8501 { a prime number equal to about 85% of *hash_size* }
define *hyph_prime* = 607 { another prime for hashing `\hyphenation` exceptions; if you change this, you should also change *iinf_hyphen_size*. }
define *biggest_char* = 65535 { the largest allowed character number; must be ≤ *max_quarterword*, this refers to UTF16 codepoints that we store in strings, etc; actual character codes can exceed this range, up to *biggest_usv* }
define *too_big_char* = 65536 { *biggest_char* + 1 }
define *biggest_usv* = "`10FFFF` { the largest Unicode Scalar Value }
define *too_big_usv* = "`110000` { *biggest_usv* + 1 }
define *number_usvs* = "`110000` { *biggest_usv* + 1 }
define *special_char* = "`110001` { *biggest_usv* + 2 }
define *biggest_reg* = 255 { the largest allowed register number; must be ≤ *max_quarterword* }
define *number_regs* = 256 { *biggest_reg* + 1 }
define *font_biggest* = 255 { the real biggest font }
define *number_fonts* = *font_biggest* − *font_base* + 2
define *number_math_families* = 256
define *number_math_fonts* = *number_math_families* + *number_math_families* + *number_math_families*
define *math_font_biggest* = *number_math_fonts* − 1
define *text_size* = 0 { size code for the largest size in a family }
define *script_size* = *number_math_families* { size code for the medium size in a family }
define *script_script_size* = *number_math_families* + *number_math_families*
{ size code for the smallest size in a family }
define *biggest_lang* = 255 { the largest hyphenation language }
define *too_big_lang* = 256 { *biggest_lang* + 1 }
define *hyphenatable_length_limit* = 4095
{ hard limit for hyphenatable length; runtime value is *max_hyphenatable_length* }

16:* Here are some macros for common programming idioms.

define *negate*(`#`) ≡ `#` ← −`#` { change the sign of a variable }
define *loop* ≡ **while** *true* **do** { repeat over and over until a **goto** happens }
format *loop* ≡ *xclause* { `WEB`'s **xclause** acts like '**while** *true* **do**' }
define *do_nothing* ≡ { empty statement }
define *return* ≡ **goto** *exit* { terminate a procedure call }
format *return* ≡ *nil*
define *empty* = 0 { symbolic name for a null constant }

19* The original Pascal compiler was designed in the late 60s, when six-bit character sets were common, so it did not make provision for lowercase letters. Nowadays, of course, we need to deal with both capital and small letters in a convenient way, especially in a program for typesetting; so the present specification of TeX has been written under the assumption that the Pascal compiler and run-time system permit the use of text files with more than 64 distinguishable characters. More precisely, we assume that the character set contains at least the letters and symbols associated with ASCII codes *'40* through *'176*; all of these characters are now available on most computer terminals.

Since we are dealing with more characters than were present in the first Pascal compilers, we have to decide what to call the associated data type. Some Pascals use the original name *char* for the characters in text files, even though there now are more than 64 such characters, while other Pascals consider *char* to be a 64-element subrange of a larger data type that has some other name.

In order to accommodate this difference, we shall use the name *text_char* to stand for the data type of the characters that are converted to and from *ASCII_code* when they are input and output. We shall also assume that *text_char* consists of the elements *chr*(*first_text_char*) through *chr*(*last_text_char*), inclusive. The following definitions should be adjusted if necessary.

define *text_char* ≡ *ASCII_code* { the data type of characters in text files }
define *first_text_char* = 0 { ordinal number of the smallest element of *text_char* }
define *last_text_char* = *biggest_char* { ordinal number of the largest element of *text_char* }

⟨ Local variables for initialization 19* ⟩ ≡
i: *integer*;

See also sections 188 and 981.

This code is used in section 4*.

20* The TeX processor converts between ASCII code and the user's external character set by means of arrays *xord* and *xchr* that are analogous to Pascal's *ord* and *chr* functions.

⟨ Global variables 13 ⟩ +≡
xchr: ↑*text_char*; { dummy variable so tangle doesn't complain; not actually used }

23* The ASCII code is "standard" only to a certain extent, since many computer installations have found it advantageous to have ready access to more than 94 printing characters. Appendix C of *The TeXbook* gives a complete specification of the intended correspondence between characters and TeX's internal representation.

If TeX is being used on a garden-variety Pascal for which only standard ASCII codes will appear in the input and output files, it doesn't really matter what codes are specified in *xchr*[0 .. *'37*], but the safest policy is to blank everything out by using the code shown below.

However, other settings of *xchr* will make TeX more friendly on computers that have an extended character set, so that users can type things like '≠' instead of '`\ne`'. People with extended character sets can assign codes arbitrarily, giving an *xchr* equivalent to whatever characters the users of TeX are allowed to have in their input files. It is best to make the codes correspond to the intended interpretations as shown in Appendix C whenever possible; but this is not necessary. For example, in countries with an alphabet of more than 26 letters, it is usually best to map the additional letters into codes less than *'40*. To get the most "permissive" character set, change '␣' on the right of these assignment statements to *chr*(*i*).

⟨ Set initial values of key variables 23* ⟩ ≡

See also sections 24*, 62, 78*, 81, 84, 101, 122, 191, 241*, 280, 284*, 302, 317, 398, 417, 473, 516, 525, 586*, 591, 629, 632, 642, 687, 696, 704, 727, 819, 941, 982*, 1044, 1087, 1321, 1336, 1354, 1397, 1412, 1516, 1562, 1628, 1647, 1671, 1679*, 1688*, and 1692*.

This code is used in section 8*.

24* The following system-independent code makes the *xord* array contain a suitable inverse to the information in *xchr*. Note that if *xchr*[*i*] = *xchr*[*j*] where *i* < *j* < *'177*, the value of *xord*[*xchr*[*i*]] will turn out to be *j* or more; hence, standard ASCII code numbers will be used instead of codes below *'40* in case there is a coincidence.

⟨ Set initial values of key variables 23* ⟩ +≡

26:* Most of what we need to do with respect to input and output can be handled by the I/O facilities that are standard in Pascal, i.e., the routines called *get*, *put*, *eof*, and so on. But standard Pascal does not allow file variables to be associated with file names that are determined at run time, so it cannot be used to implement TeX; some sort of extension to Pascal's ordinary *reset* and *rewrite* is crucial for our purposes. We shall assume that *name_of_file* is a variable of an appropriate type such that the Pascal run-time system being used to implement TeX can open a file whose external name is specified by *name_of_file*.

```
⟨ Global variables 13 ⟩ +≡
name_of_file: ↑UTF8_code;  { we build filenames in utf8 to pass to the OS }
name_of_file16: ↑UTF16_code;  { but sometimes we need a UTF16 version of the name }
name_length: 0 .. file_name_size;
    { this many characters are actually relevant in name_of_file (the rest are blank) }
name_length16: 0 .. file_name_size;
```

27:* All of the file opening functions are defined in C.

28:* And all the file closing routines as well.

30:* Input from text files is read one line at a time, using a routine called *input_ln*. This function is defined in terms of global variables called *buffer*, *first*, and *last* that will be described in detail later; for now, it suffices for us to know that *buffer* is an array of *ASCII_code* values, and that *first* and *last* are indices into this array representing the beginning and ending of a line of text.

```
⟨ Global variables 13 ⟩ +≡
buffer: ↑UnicodeScalar;  { lines of characters being read }
first: 0 .. buf_size;  { the first unused position in buffer }
last: 0 .. buf_size;  { end of the line just input to buffer }
max_buf_stack: 0 .. buf_size;  { largest index used in buffer }
```

31* The *input_ln* function brings the next line of input from the specified file into available positions of the buffer array and returns the value *true*, unless the file has already been entirely read, in which case it returns *false* and sets $last \leftarrow first$. In general, the *ASCII_code* numbers that represent the next line of the file are input into $buffer[first]$, $buffer[first+1]$, ..., $buffer[last-1]$; and the global variable *last* is set equal to *first* plus the length of the line. Trailing blanks are removed from the line; thus, either $last = first$ (in which case the line was entirely blank) or $buffer[last-1] \neq$ "␣".

An overflow error is given, however, if the normal actions of *input_ln* would make $last \geq buf\_size$; this is done so that other parts of TeX can safely look at the contents of $buffer[last+1]$ without overstepping the bounds of the *buffer* array. Upon entry to *input_ln*, the condition $first < buf\_size$ will always hold, so that there is always room for an "empty" line.

The variable *max_buf_stack*, which is used to keep track of how large the *buf_size* parameter must be to accommodate the present job, is also kept up to date by *input_ln*.

If the *bypass_eoln* parameter is *true*, *input_ln* will do a *get* before looking at the first character of the line; this skips over an *eoln* that was in $f\uparrow$. The procedure does not do a *get* when it reaches the end of the line; therefore it can be used to acquire input from the user's terminal as well as from ordinary text files.

Standard Pascal says that a file should have *eoln* immediately before *eof*, but TeX needs only a weaker restriction: If *eof* occurs in the middle of a line, the system function *eoln* should return a *true* result (even though $f\uparrow$ will be undefined).

Since the inner loop of *input_ln* is part of TeX's "inner loop"—each character of input comes in at this place—it is wise to reduce system overhead by making use of special routines that read in an entire array of characters at once, if such routines are available. The following code uses standard Pascal to illustrate what needs to be done, but finer tuning is often possible at well-developed Pascal sites.

We define *input_ln* in C, for efficiency. Nevertheless we quote the module 'Report overflow of the input buffer, and abort' here in order to make `WEAVE` happy, since part of that module is needed by e-TeX.

`@{`⟨ Report overflow of the input buffer, and abort 35* ⟩`@}`

32* The user's terminal acts essentially like other files of text, except that it is used both for input and for output. When the terminal is considered an input file, the file variable is called *term_in*, and when it is considered an output file the file variable is *term_out*.

define *term_out* ≡ *stdout* { the terminal as an output file }

⟨ Global variables 13 ⟩ +≡
init *ini_version*: *boolean*; { are we INITEX? }
dump_option: *boolean*; { was the dump name option used? }
dump_line: *boolean*; { was a %&format line seen? }
tini
dump_name: *const_cstring*; { format name for terminal display }
term_in: *unicode_file*;
bound_default: *integer*; { temporary for setup }
bound_name: *const_cstring*; { temporary for setup }
mem_bot: *integer*;
{ smallest index in the *mem* array dumped by INITEX; must not be less than *mem_min* }
main_memory: *integer*; { total memory words allocated in initex }
extra_mem_bot: *integer*; { *mem_min* ← *mem_bot* − *extra_mem_bot* except in INITEX }
mem_min: *integer*; { smallest index in TeX's internal *mem* array; must be *min_halfword* or more; must be equal to *mem_bot* in INITEX, otherwise ≤ *mem_bot* }
mem_top: *integer*; { largest index in the *mem* array dumped by INITEX; must be substantially larger than *mem_bot*, equal to *mem_max* in INITEX, else not greater than *mem_max* }
extra_mem_top: *integer*; { *mem_max* ← *mem_top* + *extra_mem_top* except in INITEX }
mem_max: *integer*; { greatest index in TeX's internal *mem* array; must be strictly less than *max_halfword*; must be equal to *mem_top* in INITEX, otherwise ≥ *mem_top* }
error_line: *integer*; { width of context lines on terminal error messages }
half_error_line: *integer*; { width of first lines of contexts in terminal error messages; should be between 30 and *error_line* − 15 }
max_print_line: *integer*; { width of longest text lines output; should be at least 60 }
max_strings: *integer*; { maximum number of strings; must not exceed *max_halfword* }
strings_free: *integer*; { strings available after format loaded }
string_vacancies: *integer*; { the minimum number of characters that should be available for the user's control sequences and font names, after TeX's own error messages are stored }
pool_size: *integer*; { maximum number of characters in strings, including all error messages and help texts, and the names of all fonts and control sequences; must exceed *string_vacancies* by the total length of TeX's own strings, which is currently about 23000 }
pool_free: *integer*; { pool space free after format loaded }
font_mem_size: *integer*; { number of words of *font_info* for all fonts }
font_max: *integer*; { maximum internal font number; ok to exceed *max_quarterword* and must be at most *font_base*+*max_font_max* }
font_k: *integer*; { loop variable for initialization }
hyph_size: *integer*; { maximum number of hyphen exceptions }
trie_size: *integer*; { space for hyphenation patterns; should be larger for INITEX than it is in production versions of TeX. 50000 is needed for English, German, and Portuguese. }
buf_size: *integer*; { maximum number of characters simultaneously present in current lines of open files and in control sequences between \csname and \endcsname; must not exceed *max_halfword* }
stack_size: *integer*; { maximum number of simultaneous input sources }
max_in_open: *integer*;
{ maximum number of input files and error insertions that can be going on simultaneously }
param_size: *integer*; { maximum number of simultaneous macro parameters }
nest_size: *integer*; { maximum number of semantic levels simultaneously active }
save_size: *integer*; { space for saving values outside of current group; must be at most *max_halfword* }

dvi_buf_size: *integer*; { size of the output buffer; must be a multiple of 8 }
expand_depth: *integer*; { limits recursive calls to the *expand* procedure }
parse_first_line_p: *cinttype*; { parse the first line for options }
file_line_error_style_p: *cinttype*; { format messages as file:line:error }
eight_bit_p: *cinttype*; { make all characters printable by default }
halt_on_error_p: *cinttype*; { stop at first error }
quoted_filename: *boolean*; { current filename is quoted }
{ Variables for source specials }
src_specials_p: *boolean*; { Whether *src_specials* are enabled at all }
insert_src_special_auto: *boolean*;
insert_src_special_every_par: *boolean*;
insert_src_special_every_parend: *boolean*;
insert_src_special_every_cr: *boolean*;
insert_src_special_every_math: *boolean*;
insert_src_special_every_hbox: *boolean*;
insert_src_special_every_vbox: *boolean*;
insert_src_special_every_display: *boolean*;

33* Here is how to open the terminal files. *t_open_out* does nothing. *t_open_in*, on the other hand, does the work of "rescanning," or getting any command line arguments the user has provided. It's defined in C.

define *t_open_out* ≡ { output already open for text output }

34* Sometimes it is necessary to synchronize the input/output mixture that happens on the user's terminal, and three system-dependent procedures are used for this purpose. The first of these, *update_terminal*, is called when we want to make sure that everything we have output to the terminal so far has actually left the computer's internal buffers and been sent. The second, *clear_terminal*, is called when we wish to cancel any input that the user may have typed ahead (since we are about to issue an unexpected error message). The third, *wake_up_terminal*, is supposed to revive the terminal if the user has disabled it by some instruction to the operating system. The following macros show how these operations can be specified with UNIX. *update_terminal* does an *fflush*. *clear_terminal* is redefined to do nothing, since the user should control the terminal.

define *update_terminal* ≡ *fflush*(*term_out*)
define *clear_terminal* ≡ *do_nothing*
define *wake_up_terminal* ≡ *do_nothing* { cancel the user's cancellation of output }

35* We need a special routine to read the first line of TEX input from the user's terminal. This line is different because it is read before we have opened the transcript file; there is sort of a "chicken and egg" problem here. If the user types '`\input paper`' on the first line, or if some macro invoked by that line does such an `\input`, the transcript file will be named '`paper.log`'; but if no `\input` commands are performed during the first line of terminal input, the transcript file will acquire its default name '`texput.log`'. (The transcript file will not contain error messages generated by the first line before the first `\input` command.)

The first line is even more special if we are lucky enough to have an operating system that treats TEX differently from a run-of-the-mill Pascal object program. It's nice to let the user start running a TEX job by typing a command line like '`tex paper`'; in such a case, TEX will operate as if the first line of input were '`paper`', i.e., the first line will consist of the remainder of the command line, after the part that invoked TEX.

The first line is special also because it may be read before TEX has input a format file. In such cases, normal error messages cannot yet be given. The following code uses concepts that will be explained later. (If the Pascal compiler does not support non-local **goto**, the statement '**goto** *final_end*' should be replaced by something that quietly terminates the program.)

Routine is implemented in C; part of module is, however, needed for e-TeX.

```
⟨ Report overflow of the input buffer, and abort 35* ⟩ ≡
  begin cur_input.loc_field ← first; cur_input.limit_field ← last − 1; overflow("buffer␣size", buf_size);
  end
```

This code is used in sections 31* and 1567.

37* The following program does the required initialization. Iff anything has been specified on the command line, then *t_open_in* will return with *last* > *first*.

```
function init_terminal: boolean; { gets the terminal input started }
  label exit;
  begin t_open_in;
  if last > first then
    begin loc ← first;
    while (loc < last) ∧ (buffer[loc] = ´␣´) do incr(loc);
    if loc < last then
      begin init_terminal ← true; goto exit;
      end;
    end;
  loop begin wake_up_terminal; write(term_out, ´**´); update_terminal;
    if ¬input_ln(term_in, true) then { this shouldn't happen }
      begin write_ln(term_out); write_ln(term_out, ´!␣End␣of␣file␣on␣the␣terminal...␣why?´);
      init_terminal ← false; return;
      end;
    loc ← first;
    while (loc < last) ∧ (buffer[loc] = "␣") do incr(loc);
    if loc < last then
      begin init_terminal ← true; return; { return unless the line was all blank }
      end;
    write_ln(term_out, ´Please␣type␣the␣name␣of␣your␣input␣file.´);
    end;
exit: end;
```

38* String handling. Control sequence names and diagnostic messages are variable-length strings of eight-bit characters. Since Pascal does not have a well-developed string mechanism, TEX does all of its string processing by homegrown methods.

Elaborate facilities for dynamic strings are not needed, so all of the necessary operations can be handled with a simple data structure. The array *str_pool* contains all of the (eight-bit) ASCII codes in all of the strings, and the array *str_start* contains indices of the starting points of each string. Strings are referred to by integer numbers, so that string number s comprises the characters $str\_pool[j]$ for $str\_start\_macro[s] \le j < str\_start\_macro[s+1]$. Additional integer variables *pool_ptr* and *str_ptr* indicate the number of entries used so far in *str_pool* and *str_start*, respectively; locations $str\_pool[pool\_ptr]$ and $str\_start\_macro[str\_ptr]$ are ready for the next string to be allocated.

String numbers 0 to 255 are reserved for strings that correspond to single ASCII characters. This is in accordance with the conventions of WEB, which converts single-character strings into the ASCII code number of the single character involved, while it converts other strings into integers and builds a string pool file. Thus, when the string constant `"."` appears in the program below, WEB converts it into the integer 46, which is the ASCII code for a period, while WEB will convert a string like `"hello"` into some integer greater than 255. String number 46 will presumably be the single character '.'; but some ASCII codes have no standard visible representation, and TEX sometimes needs to be able to print an arbitrary ASCII character, so the first 256 strings are used to specify exactly what should be printed for each of the 256 possibilities.

Elements of the *str_pool* array must be ASCII codes that can actually be printed; i.e., they must have an *xchr* equivalent in the local character set. (This restriction applies only to preloaded strings, not to those generated dynamically by the user.)

Some Pascal compilers won't pack integers into a single byte unless the integers lie in the range $-128 \ldots 127$. To accommodate such systems we access the string pool only via macros that can easily be redefined.

```
define si(#) ≡ #   { convert from ASCII_code to packed_ASCII_code }
define so(#) ≡ #   { convert from packed_ASCII_code to ASCII_code }
define str_start_macro(#) ≡ str_start[(#) − too_big_char]
⟨ Types in the outer block 18 ⟩ +≡
  pool_pointer = integer;   { for variables that point into str_pool }
  str_number = 0 .. ssup_max_strings;   { for variables that point into str_start }
  packed_ASCII_code = 0 .. biggest_char;   { elements of str_pool array }
```

39* ⟨ Global variables 13 ⟩ +≡

```
str_pool: ↑packed_ASCII_code;   { the characters }
str_start: ↑pool_pointer;   { the starting pointers }
pool_ptr: pool_pointer;   { first unused position in str_pool }
str_ptr: str_number;   { number of the current string being created }
init_pool_ptr: pool_pointer;   { the starting value of pool_ptr }
init_str_ptr: str_number;   { the starting value of str_ptr }
```

47* The initial values of *str_pool*, *str_start*, *pool_ptr*, and *str_ptr* are computed by the INITEX program, based in part on the information that WEB has output while processing TEX.

```
⟨ Declare additional routines for string recycling 1685* ⟩
  init function get_strings_started: boolean;
      { initializes the string pool, but returns false if something goes wrong }
  label done, exit;
  var g: str_number;   { garbage }
  begin pool_ptr ← 0; str_ptr ← 0; str_start[0] ← 0; ⟨ Make the first 256 strings 48 ⟩;
  ⟨ Read the other strings from the TEX.POOL file and return true, or give an error message and return
      false 51* ⟩;
exit: end;
  tini
```

49* The first 128 strings will contain 95 standard ASCII characters, and the other 33 characters will be printed in three-symbol form like '^^A' unless a system-dependent change is made here. Installations that have an extended character set, where for example $xchr[\'32] = \'\#\'$, would like string *'32* to be printed as the single character *'32* instead of the three characters *'136*, *'136*, *'132* (^^Z). On the other hand, even people with an extended character set will want to represent string *'15* by ^^M, since *'15* is *carriage_return*; the idea is to produce visible strings instead of tabs or line-feeds or carriage-returns or bell-rings or characters that are treated anomalously in text files.

Unprintable characters of codes 128–255 are, similarly, rendered ^^80–^^ff.

The boolean expression defined here should be *true* unless TeX internal code number k corresponds to a non-troublesome visible symbol in the local character set. An appropriate formula for the extended character set recommended in *The TeXbook* would, for example, be '$k \in [0, \'10 .. \'12, \'14, \'15, \'33, \'177 .. \'377]$'. If character k cannot be printed, and $k < \'200$, then character $k + \'100$ or $k - \'100$ must be printable; moreover, ASCII codes $[\'41 .. \'46, \'60 .. \'71, \'136, \'141 .. \'146, \'160 .. \'171]$ must be printable. Thus, at least 80 printable characters are needed.

51* ⟨ Read the other strings from the TEX.POOL file and return *true*, or give an error message and return *false* 51* ⟩ ≡

```
g ← loadpoolstrings((pool_size − string_vacancies));
if g = 0 then
  begin wake_up_terminal; write_ln(term_out, '!␣You␣have␣to␣increase␣POOLSIZE.');
  get_strings_started ← false; return;
  end;
get_strings_started ← true;
```

This code is used in section 47*.

52* Empty module

53* Empty module

54* On-line and off-line printing. Messages that are sent to a user's terminal and to the transcript-log file are produced by several '*print*' procedures. These procedures will direct their output to a variety of places, based on the setting of the global variable *selector*, which has the following possible values:

term_and_log, the normal setting, prints on the terminal and on the transcript file.
log_only, prints only on the transcript file.
term_only, prints only on the terminal.
no_print, doesn't print at all. This is used only in rare cases before the transcript file is open.
pseudo, puts output into a cyclic buffer that is used by the *show_context* routine; when we get to that routine we shall discuss the reasoning behind this curious mode.
new_string, appends the output to the current string in the string pool.
0 to 15, prints on one of the sixteen files for `\write` output.

The symbolic names '*term_and_log*', etc., have been assigned numeric codes that satisfy the convenient relations $no\_print + 1 = term\_only$, $no\_print + 2 = log\_only$, $term\_only + 2 = log\_only + 1 = term\_and\_log$.

Three additional global variables, *tally* and *term_offset* and *file_offset*, record the number of characters that have been printed since they were most recently cleared to zero. We use *tally* to record the length of (possibly very long) stretches of printing; *term_offset* and *file_offset*, on the other hand, keep track of how many characters have appeared so far on the current line that has been output to the terminal or to the transcript file, respectively.

```
define no_print = 16 { selector setting that makes data disappear }
define term_only = 17 { printing is destined for the terminal only }
define log_only = 18 { printing is destined for the transcript file only }
define term_and_log = 19 { normal selector setting }
define pseudo = 20 { special selector setting for show_context }
define new_string = 21 { printing is deflected to the string pool }
define max_selector = 21 { highest selector setting }
```

```
⟨ Global variables 13 ⟩ +≡
log_file: alpha_file; { transcript of TeX session }
selector: 0 .. max_selector; { where to print a message }
dig: array [0 .. 22] of 0 .. 15; { digits in a number being output }
tally: integer; { the number of characters recently printed }
term_offset: 0 .. max_print_line; { the number of characters on the current terminal line }
file_offset: 0 .. max_print_line; { the number of characters on the current file line }
trick_buf: array [0 .. ssup_error_line] of ASCII_code; { circular buffer for pseudoprinting }
trick_count: integer; { threshold for pseudoprinting, explained later }
first_count: integer; { another variable for pseudoprinting }
```

65* Here is the very first thing that TeX prints: a headline that identifies the version number and format package. The *term_offset* variable is temporarily incorrect, but the discrepancy is not serious since we assume that this part of the program is system dependent.

```
⟨ Initialize the output routines 55 ⟩ +≡
  if src_specials_p ∨ file_line_error_style_p ∨ parse_first_line_p then wterm(banner_k)
  else wterm(banner);
  wterm(version_string);
  if format_ident = 0 then wterm_ln(´␣(preloaded␣format=´, dump_name, ´)´)
  else begin slow_print(format_ident); print_ln;
    end;
  if shellenabledp then
    begin wterm(´␣´);
    if restrictedshell then
      begin wterm(´restricted␣´);
      end;
    wterm_ln(´\write18␣enabled.´);
    end;
  if src_specials_p then
    begin wterm_ln(´␣Source␣specials␣enabled.´)
    end;
  if translate_filename then
    begin wterm(´␣(WARNING:␣translate-file␣"´); fputs(translate_filename, stdout);
    wterm_ln(´"␣ignored)´);
    end;
  update_terminal;
```

77* The global variable *interaction* has four settings, representing increasing amounts of user interaction:

```
define batch_mode = 0 { omits all stops and omits terminal output }
define nonstop_mode = 1 { omits all stops }
define scroll_mode = 2 { omits error stops }
define error_stop_mode = 3 { stops at every opportunity to interact }
define unspecified_mode = 4 { extra value for command-line switch }
define print_err(#) ≡
    begin if interaction = error_stop_mode then wake_up_terminal;
    if file_line_error_style_p then print_file_line
    else print_nl("! ");
    print(#);
    end
```

⟨ Global variables 13 ⟩ +≡
interaction: *batch_mode* .. *error_stop_mode*; { current level of interaction }
interaction_option: *batch_mode* .. *unspecified_mode*; { set from command line }

78* ⟨ Set initial values of key variables 23* ⟩ +≡

```
if interaction_option = unspecified_mode then interaction ← error_stop_mode
else interaction ← interaction_option;
```

80* A global variable *deletions_allowed* is set *false* if the *get_next* routine is active when *error* is called; this ensures that *get_next* and related routines like *get_token* will never be called recursively. A similar interlock is provided by *set_box_allowed*.

The global variable *history* records the worst level of error that has been detected. It has five possible values: *spotless*, *warning_issued*, *error_message_issued*, *fatal_error_stop*, and *output_failure*.

Another global variable, *error_count*, is increased by one when an *error* occurs without an interactive dialog, and it is reset to zero at the end of every paragraph. If *error_count* reaches 100, TeX decides that there is no point in continuing further.

```
define spotless = 0 { history value when nothing has been amiss yet }
define warning_issued = 1 { history value when begin_diagnostic has been called }
define error_message_issued = 2 { history value when error has been called }
define fatal_error_stop = 3 { history value when termination was premature }
define output_failure = 4 { history value when output driver returned an error }
```

⟨ Global variables 13 ⟩ +≡
deletions_allowed: *boolean*; { is it safe for *error* to call *get_token*? }
set_box_allowed: *boolean*; { is it safe to do a `\setbox` assignment? }
history: *spotless* .. *output_failure*; { has the source input been clean so far? }
error_count: −1 .. 100; { the number of scrolled errors since the last paragraph ended }

85* The *jump_out* procedure just cuts across all active procedure levels. The body of *jump_out* simply calls '*close_files_and_terminate*;' followed by a call on some system procedure that quietly terminates the program.

```
format noreturn ≡ procedure
define do_final_end ≡
        begin update_terminal; ready_already ← 0;
        if (history ≠ spotless) ∧ (history ≠ warning_issued) then uexit(1)
        else uexit(0);
        end
```

```
⟨ Error handling procedures 82 ⟩ +≡
noreturn procedure jump_out;
  begin close_files_and_terminate; do_final_end;
  end;
```

86* Here now is the general *error* routine.

```
⟨ Error handling procedures 82 ⟩ +≡
procedure error; { completes the job of error reporting }
  label continue, exit;
  var c: ASCII_code; { what the user types }
    s1, s2, s3, s4: integer; { used to save global variables when deleting tokens }
  begin if history < error_message_issued then history ← error_message_issued;
  print_char("."); show_context;
  if (halt_on_error_p) then
    begin history ← fatal_error_stop; jump_out;
    end;
  if interaction = error_stop_mode then ⟨ Get user's advice and return 87 ⟩;
  incr(error_count);
  if error_count = 100 then
    begin print_nl("(That␣makes␣100␣errors;␣please␣try␣again.)"); history ← fatal_error_stop;
    jump_out;
    end;
  ⟨ Put help message on the transcript file 94 ⟩;
exit: end;
```

88* It is desirable to provide an '`E`' option here that gives the user an easy way to return from TeX to the system editor, with the offending line ready to be edited. We do this by calling the external procedure *call_edit* with a pointer to the filename, its length, and the line number. However, here we just set up the variables that will be used as arguments, since we don't want to do the switch-to-editor until after TeX has closed its files.

There is a secret '`D`' option available when the debugging routines haven't been commented out.

```
define edit_file ≡ input_stack[base_ptr]
```

```
⟨ Interpret code c and return if done 88* ⟩ ≡
  case c of
  "0","1","2","3","4","5","6","7","8","9": if deletions_allowed then
      ⟨ Delete c − "0" tokens and goto continue 92 ⟩;
 debug "D": begin debug_help; goto continue; end; gubed
  "E": if base_ptr > 0 then
      if input_stack[base_ptr].name_field ≥ 256 then
        begin edit_name_start ← str_start_macro(edit_file.name_field);
        edit_name_length ← str_start_macro(edit_file.name_field + 1) − str_start_macro(edit_file.name_field);
        edit_line ← line; jump_out;
        end;
  "H": ⟨ Print the help information and goto continue 93 ⟩;
  "I": ⟨ Introduce new material from the terminal and return 91 ⟩;
  "Q","R","S": ⟨ Change the interaction level and return 90 ⟩;
  "X": begin interaction ← scroll_mode; jump_out;
    end;
  othercases do_nothing
  endcases;
  ⟨ Print the menu of available options 89 ⟩
```

This code is used in section 87.

97* The following procedure prints TeX's last words before dying.

```
define succumb ≡
        begin if interaction = error_stop_mode then interaction ← scroll_mode;
              { no more interaction }
        if log_opened then error;
        debug if interaction > batch_mode then debug_help;
        gubed
        history ← fatal_error_stop; jump_out; { irrecoverable error }
        end
```

```
⟨ Error handling procedures 82 ⟩ +≡
noreturn procedure fatal_error(s : str_number); { prints s, and that's it }
    begin normalize_selector;
    print_err("Emergency␣stop"); help1(s); succumb;
    end;
```

98* Here is the most dreaded error message.

```
⟨ Error handling procedures 82 ⟩ +≡
noreturn procedure overflow(s : str_number; n : integer); { stop due to finiteness }
    begin normalize_selector; print_err("TeX␣capacity␣exceeded,␣sorry␣["); print(s);
    print_char("="); print_int(n); print_char("]");
    help2("If␣you␣really␣absolutely␣need␣more␣capacity,")
    ("you␣can␣ask␣a␣wizard␣to␣enlarge␣me."); succumb;
    end;
```

99* The program might sometime run completely amok, at which point there is no choice but to stop. If no previous error has been detected, that's bad news; a message is printed that is really intended for the TeX maintenance person instead of the user (unless the user has been particularly diabolical). The index entries for 'this can't happen' may help to pinpoint the problem.

```
⟨ Error handling procedures 82 ⟩ +≡
noreturn procedure confusion(s : str_number); { consistency check violated; s tells where }
  begin normalize_selector;
  if history < error_message_issued then
    begin print_err("This␣can't␣happen␣("); print(s); print_char(")");
    help1("I'm␣broken.␣Please␣show␣this␣to␣someone␣who␣can␣fix␣can␣fix");
    end
  else begin print_err("I␣can't␣go␣on␣meeting␣you␣like␣this");
    help2("One␣of␣your␣faux␣pas␣seems␣to␣have␣wounded␣me␣deeply...")
    ("in␣fact,␣I'm␣barely␣conscious.␣Please␣fix␣it␣and␣try␣again.");
    end;
  succumb;
  end;
```

108* Physical sizes that a TEX user specifies for portions of documents are represented internally as scaled points. Thus, if we define an 'sp' (scaled point) as a unit equal to 2^{-16} printer's points, every dimension inside of TEX is an integer number of sp. There are exactly 4,736,286.72 sp per inch. Users are not allowed to specify dimensions larger than $2^{30} - 1$ sp, which is a distance of about 18.892 feet (5.7583 meters); two such quantities can be added without overflow on a 32-bit computer.

The present implementation of TEX does not check for overflow when dimensions are added or subtracted. This could be done by inserting a few dozen tests of the form '**if** $x \geq$ *'10000000000* **then** *report_overflow*', but the chance of overflow is so remote that such tests do not seem worthwhile.

TEX needs to do only a few arithmetic operations on scaled quantities, other than addition and subtraction, and the following subroutines do most of the work. A single computation might use several subroutine calls, and it is desirable to avoid producing multiple error messages in case of arithmetic overflow; so the routines set the global variable *arith_error* to *true* instead of reporting errors directly to the user. Another global variable, *remainder*, holds the remainder after a division.

define *remainder* ≡ *tex_remainder*

⟨ Global variables 13 ⟩ +≡
arith_error: *boolean*; { has arithmetic overflow occurred recently? }
remainder: *scaled*; { amount subtracted to get an exact division }

113* When TEX "packages" a list into a box, it needs to calculate the proportionality ratio by which the glue inside the box should stretch or shrink. This calculation does not affect TEX's decision making, so the precise details of rounding, etc., in the glue calculation are not of critical importance for the consistency of results on different computers.

We shall use the type *glue_ratio* for such proportionality ratios. A glue ratio should take the same amount of memory as an *integer* (usually 32 bits) if it is to blend smoothly with TEX's other data structures. Thus *glue_ratio* should be equivalent to *short_real* in some implementations of Pascal. Alternatively, it is possible to deal with glue ratios using nothing but fixed-point arithmetic; see *TUGboat* **3**,1 (March 1982), 10–27. (But the routines cited there must be modified to allow negative glue ratios.)

define *set_glue_ratio_zero*(#) ≡ # ← 0.0 { store the representation of zero ratio }
define *set_glue_ratio_one*(#) ≡ # ← 1.0 { store the representation of unit ratio }
define *float*(#) ≡ # { convert from *glue_ratio* to type *real* }
define *unfloat*(#) ≡ # { convert from *real* to type *glue_ratio* }
define *float_constant*(#) ≡ #.0 { convert *integer* constant to *real* }

⟨ Types in the outer block 18 ⟩ +≡

132* Packed data. In order to make efficient use of storage space, TeX bases its major data structures on a *memory_word*, which contains either a (signed) integer, possibly scaled, or a (signed) *glue_ratio*, or a small number of fields that are one half or one quarter of the size used for storing integers.

If x is a variable of type *memory_word*, it contains up to four fields that can be referred to as follows:

$x.int$	(an *integer*)
$x.sc$	(a *scaled* integer)
$x.gr$	(a *glue_ratio*)
$x.hh.lh$, $x.hh.rh$	(two halfword fields)
$x.hh.b0$, $x.hh.b1$, $x.hh.rh$	(two quarterword fields, one halfword field)
$x.qqqq.b0$, $x.qqqq.b1$, $x.qqqq.b2$, $x.qqqq.b3$	(four quarterword fields)

This is somewhat cumbersome to write, and not very readable either, but macros will be used to make the notation shorter and more transparent. The Pascal code below gives a formal definition of *memory_word* and its subsidiary types, using packed variant records. TeX makes no assumptions about the relative positions of the fields within a word.

Since we are assuming 32-bit integers, a halfword must contain at least 16 bits, and a quarterword must contain at least 8 bits. But it doesn't hurt to have more bits; for example, with enough 36-bit words you might be able to have *mem_max* as large as 262142, which is eight times as much memory as anybody had during the first four years of TeX's existence.

N.B.: Valuable memory space will be dreadfully wasted unless TeX is compiled by a Pascal that packs all of the *memory_word* variants into the space of a single integer. This means, for example, that *glue_ratio* words should be *short_real* instead of *real* on some computers. Some Pascal compilers will pack an integer whose subrange is '0 .. 255' into an eight-bit field, but others insist on allocating space for an additional sign bit; on such systems you can get 256 values into a quarterword only if the subrange is '−128 .. 127'.

The present implementation tries to accommodate as many variations as possible, so it makes few assumptions. If integers having the subrange '*min_quarterword* .. *max_quarterword*' can be packed into a quarterword, and if integers having the subrange '*min_halfword* .. *max_halfword*' can be packed into a halfword, everything should work satisfactorily.

It is usually most efficient to have *min_quarterword* = *min_halfword* = 0, so one should try to achieve this unless it causes a severe problem. The values defined here are recommended for most 32-bit computers.

```
define min_quarterword = 0 { smallest allowable value in a quarterword }
define max_quarterword = "FFFF { largest allowable value in a quarterword }
define min_halfword ≡ -"FFFFFFF { smallest allowable value in a halfword }
define max_halfword ≡ "3FFFFFFF { largest allowable value in a halfword }
```

133* Here are the inequalities that the quarterword and halfword values must satisfy (or rather, the inequalities that they mustn't satisfy):

```
⟨ Check the "constant" values for consistency 14 ⟩ +≡
  init if (mem_min ≠ mem_bot) ∨ (mem_max ≠ mem_top) then bad ← 10;
  tini
  if (mem_min > mem_bot) ∨ (mem_max < mem_top) then bad ← 10;
  if (min_quarterword > 0) ∨ (max_quarterword < "7FFF) then bad ← 11;
  if (min_halfword > 0) ∨ (max_halfword < "3FFFFFFF) then bad ← 12;
  if (min_quarterword < min_halfword) ∨ (max_quarterword > max_halfword) then bad ← 13;
  if (mem_bot − sup_main_memory < min_halfword) ∨ (mem_top + sup_main_memory ≥ max_halfword)
      then bad ← 14;
  if (max_font_max < min_halfword) ∨ (max_font_max > max_halfword) then bad ← 15;
  if font_max > font_base + max_font_max then bad ← 16;
  if (save_size > max_halfword) ∨ (max_strings > max_halfword) then bad ← 17;
  if buf_size > max_halfword then bad ← 18;
  if max_quarterword − min_quarterword < "FFFF then bad ← 19;
```

134* The operation of adding or subtracting *min_quarterword* occurs quite frequently in TeX, so it is convenient to abbreviate this operation by using the macros *qi* and *qo* for input and output to and from quarterword format.

The inner loop of TeX will run faster with respect to compilers that don't optimize expressions like '$x+0$' and '$x-0$', if these macros are simplified in the obvious way when *min_quarterword* = 0. So they have been simplified here in the obvious way.

The WEB source for TeX defines $hi(\#) \equiv \# + min\_halfword$ which can be simplified when *min_halfword* = 0. The Web2C implementation of TeX can use $hi(\#) \equiv \#$ together with *min_halfword* < 0 as long as *max_halfword* is sufficiently large.

define $qi(\#) \equiv \#$ { to put an *eight_bits* item into a quarterword }
define $qo(\#) \equiv \#$ { to take an *eight_bits* item from a quarterword }
define $hi(\#) \equiv \#$ { to put a sixteen-bit item into a halfword }
define $ho(\#) \equiv \#$ { to take a sixteen-bit item from a halfword }

135* The reader should study the following definitions closely:

define $sc \equiv int$ { *scaled* data is equivalent to *integer* }

⟨ Types in the outer block 18 ⟩ +≡
quarterword = *min_quarterword* .. *max_quarterword*; *halfword* = *min_halfword* .. *max_halfword*;
two_choices = 1 .. 2; { used when there are two variants in a record }
four_choices = 1 .. 4; { used when there are four variants in a record }
`#include␣"texmfmem.h";` *word_file* = *gzFile*;

138* The *mem* array is divided into two regions that are allocated separately, but the dividing line between these two regions is not fixed; they grow together until finding their "natural" size in a particular job. Locations less than or equal to *lo_mem_max* are used for storing variable-length records consisting of two or more words each. This region is maintained using an algorithm similar to the one described in exercise 2.5–19 of *The Art of Computer Programming*. However, no size field appears in the allocated nodes; the program is responsible for knowing the relevant size when a node is freed. Locations greater than or equal to *hi_mem_min* are used for storing one-word records; a conventional `AVAIL` stack is used for allocation in this region.

Locations of *mem* between *mem_bot* and *mem_top* may be dumped as part of preloaded format files, by the `INITEX` preprocessor. Production versions of TeX may extend the memory at both ends in order to provide more space; locations between *mem_min* and *mem_bot* are always used for variable-size nodes, and locations between *mem_top* and *mem_max* are always used for single-word nodes.

The key pointers that govern *mem* allocation have a prescribed order:

$$null \le mem\_min \le mem\_bot < lo\_mem\_max < hi\_mem\_min < mem\_top \le mem\_end \le mem\_max.$$

Empirical tests show that the present implementation of TeX tends to spend about 9% of its running time allocating nodes, and about 6% deallocating them after their use.

```
⟨ Global variables 13 ⟩ +≡
yzmem: ↑memory_word;  { the big dynamic storage area }
zmem: ↑memory_word;  { the big dynamic storage area }
lo_mem_max: pointer;  { the largest location of variable-size memory in use }
hi_mem_min: pointer;  { the smallest location of one-word memory in use }
```

147* A call to *get_node* with argument *s* returns a pointer to a new node of size *s*, which must be 2 or more. The *link* field of the first word of this new node is set to null. An overflow stop occurs if no suitable space exists.

If *get_node* is called with $s = 2^{30}$, it simply merges adjacent free areas and returns the value *max_halfword*.

```
function get_node(s : integer): pointer;  { variable-size node allocation }
  label found, exit, restart;
  var p: pointer;  { the node currently under inspection }
    q: pointer;  { the node physically after node p }
    r: integer;  { the newly allocated node, or a candidate for this honor }
    t: integer;  { temporary register }
  begin restart: p ← rover;  { start at some free node in the ring }
  repeat ⟨ Try to allocate within node p and its physical successors, and goto found if allocation was
      possible 149 ⟩;
    p ← rlink(p);  { move to the next node in the ring }
  until p = rover;  { repeat until the whole list has been traversed }
  if s = ´10000000000 then
    begin get_node ← max_halfword; return;
    end;
  if lo_mem_max + 2 < hi_mem_min then
    if lo_mem_max + 2 ≤ mem_bot + max_halfword then
      ⟨ Grow more variable-size memory and goto restart 148 ⟩;
  overflow("main memory size", mem_max + 1 − mem_min);  { sorry, nothing satisfactory is left }
found: link(r) ← null;  { this node is now nonempty }
  stat var_used ← var_used + s;  { maintain usage statistics }
  tats
  ⟨ Initialize bigger nodes with SyncTeX information 1714* ⟩;
  get_node ← r;
exit: end;
```

157*. An *hlist_node* stands for a box that was made from a horizontal list. Each *hlist_node* is seven words long, and contains the following fields (in addition to the mandatory *type* and *link*, which we shall not mention explicitly when discussing the other node types): The *height* and *width* and *depth* are scaled integers denoting the dimensions of the box. There is also a *shift_amount* field, a scaled integer indicating how much this box should be lowered (if it appears in a horizontal list), or how much it should be moved to the right (if it appears in a vertical list). There is a *list_ptr* field, which points to the beginning of the list from which this box was fabricated; if *list_ptr* is *null*, the box is empty. Finally, there are three fields that represent the setting of the glue: *glue_set*(*p*) is a word of type *glue_ratio* that represents the proportionality constant for glue setting; *glue_sign*(*p*) is *stretching* or *shrinking* or *normal* depending on whether or not the glue should stretch or shrink or remain rigid; and *glue_order*(*p*) specifies the order of infinity to which glue setting applies (*normal*, *fil*, *fill*, or *filll*). The *subtype* field is not used in TeX. In ε-TeX the *subtype* field records the box direction mode *box_lr*.

define *synctex_field_size* = 1 { Declare the SyncTeX field size to store the SyncTeX information: we will put file tag and line into lh and rh fields of one word }
define *sync_tag*(#) ≡ *mem*[# − *synctex_field_size*].*hh*.*lh* { The tag subfield }
define *sync_line*(#) ≡ *mem*[# − *synctex_field_size*].*hh*.*rh* { The line subfield }

define *hlist_node* = 0 { *type* of hlist nodes }
define *box_node_size* = 7 + *synctex_field_size* { number of words to allocate for a box node }
define *width_offset* = 1 { position of *width* field in a box node }
define *depth_offset* = 2 { position of *depth* field in a box node }
define *height_offset* = 3 { position of *height* field in a box node }
define *width*(#) ≡ *mem*[# + *width_offset*].*sc* { width of the box, in sp }
define *depth*(#) ≡ *mem*[# + *depth_offset*].*sc* { depth of the box, in sp }
define *height*(#) ≡ *mem*[# + *height_offset*].*sc* { height of the box, in sp }
define *shift_amount*(#) ≡ *mem*[# + 4].*sc* { repositioning distance, in sp }
define *list_offset* = 5 { position of *list_ptr* field in a box node }
define *list_ptr*(#) ≡ *link*(# + *list_offset*) { beginning of the list inside the box }
define *glue_order*(#) ≡ *subtype*(# + *list_offset*) { applicable order of infinity }
define *glue_sign*(#) ≡ *type*(# + *list_offset*) { stretching or shrinking }
define *normal* = 0 { the most common case when several cases are named }
define *stretching* = 1 { glue setting applies to the stretch components }
define *shrinking* = 2 { glue setting applies to the shrink components }
define *glue_offset* = 6 { position of *glue_set* in a box node }
define *glue_set*(#) ≡ *mem*[# + *glue_offset*].*gr* { a word of type *glue_ratio* for glue setting }

160*. A *rule_node* stands for a solid black rectangle; it has *width*, *depth*, and *height* fields just as in an *hlist_node*. However, if any of these dimensions is -2^{30}, the actual value will be determined by running the rule up to the boundary of the innermost enclosing box. This is called a "running dimension." The *width* is never running in an hlist; the *height* and *depth* are never running in a vlist.

define *rule_node* = 2 { *type* of rule nodes }
define *rule_node_size* = 4 + *synctex_field_size* { number of words to allocate for a rule node }
define *null_flag* ≡ −'10000000000 { -2^{30}, signifies a missing item }
define *is_running*(#) ≡ (# = *null_flag*) { tests for a running dimension }

163* A *mark_node* has a *mark_ptr* field that points to the reference count of a token list that contains the user's `\mark` text. In addition there is a *mark_class* field that contains the mark class.

```
define mark_node = 4 { type of a mark node }
define small_node_size = 2 { number of words to allocate for most node types }
define medium_node_size = small_node_size + synctex_field_size { number of words to allocate for
    synchronized node types like math, kern, glue and penalty nodes }
define mark_ptr(#) ≡ link(# + 1) { head of the token list for a mark }
define mark_class(#) ≡ info(# + 1) { the mark class }
```

166* The *new_ligature* function creates a ligature node having given contents of the *font*, *character*, and *lig_ptr* fields. We also have a *new_lig_item* function, which returns a two-word node having a given *character* field. Such nodes are used for temporary processing as ligatures are being created.

```
function new_ligature(f : internal_font_number; c : quarterword; q : pointer): pointer;
  var p: pointer; { the new node }
  begin p ← get_node(small_node_size); type(p) ← ligature_node; font(lig_char(p)) ← f;
  character(lig_char(p)) ← c; lig_ptr(p) ← q; subtype(p) ← 0; new_ligature ← p;
  end;
function new_lig_item(c : quarterword): pointer;
  var p: pointer; { the new node }
  begin p ← get_node(small_node_size); character(p) ← c; lig_ptr(p) ← null; new_lig_item ← p;
  end;
```

171* A *math_node*, which occurs only in horizontal lists, appears before and after mathematical formulas. The *subtype* field is *before* before the formula and *after* after it. There is a *width* field, which represents the amount of surrounding space inserted by `\mathsurround`.

In addition a *math_node* with *subtype* > *after* and *width* = 0 will be (ab)used to record a regular *math_node* reinserted after being discarded at a line break or one of the text direction primitives (`\beginL`, `\endL`, `\beginR`, and `\endR`).

```
define math_node = 9 { type of a math node }
define before = 0 { subtype for math node that introduces a formula }
define after = 1 { subtype for math node that winds up a formula }
define M_code = 2
define begin_M_code = M_code + before { subtype for \beginM node }
define end_M_code = M_code + after { subtype for \endM node }
define L_code = 4
define begin_L_code = L_code + begin_M_code { subtype for \beginL node }
define end_L_code = L_code + end_M_code { subtype for \endL node }
define R_code = L_code + L_code
define begin_R_code = R_code + begin_M_code { subtype for \beginR node }
define end_R_code = R_code + end_M_code { subtype for \endR node }
define end_LR(#) ≡ odd(subtype(#))
define end_LR_type(#) ≡ (L_code * (subtype(#) div L_code) + end_M_code)
define begin_LR_type(#) ≡ (# - after + before)
function new_math(w : scaled; s : small_number): pointer;
  var p: pointer; { the new node }
  begin p ← get_node(medium_node_size); type(p) ← math_node; subtype(p) ← s; width(p) ← w;
  new_math ← p;
  end;
```

176*. And here's a function that creates a glue node for a given parameter identified by its code number; for example, *new_param_glue*(*line_skip_code*) returns a pointer to a glue node for the current `\lineskip`.

```
function new_param_glue(n : small_number): pointer;
  var p: pointer;  { the new node }
    q: pointer;  { the glue specification }
  begin p ← get_node(medium_node_size); type(p) ← glue_node; subtype(p) ← n+1; leader_ptr(p) ← null;
  q ← ⟨ Current mem equivalent of glue parameter number n 250 ⟩; glue_ptr(p) ← q;
  incr(glue_ref_count(q)); new_param_glue ← p;
  end;
```

177*. Glue nodes that are more or less anonymous are created by *new_glue*, whose argument points to a glue specification.

```
function new_glue(q : pointer): pointer;
  var p: pointer;  { the new node }
  begin p ← get_node(medium_node_size); type(p) ← glue_node; subtype(p) ← normal;
  leader_ptr(p) ← null; glue_ptr(p) ← q; incr(glue_ref_count(q)); new_glue ← p;
  end;
```

180*. The *new_kern* function creates a kern node having a given width.

```
function new_kern(w : scaled): pointer;
  var p: pointer;  { the new node }
  begin p ← get_node(medium_node_size); type(p) ← kern_node; subtype(p) ← normal; width(p) ← w;
  new_kern ← p;
  end;
```

183*. Anyone who has been reading the last few sections of the program will be able to guess what comes next.

```
function new_penalty(m : integer): pointer;
  var p: pointer;  { the new node }
  begin p ← get_node(medium_node_size); type(p) ← penalty_node; subtype(p) ← 0;
      { the subtype is not used }
  penalty(p) ← m; new_penalty ← p;
  end;
```

190*. If TeX is extended improperly, the *mem* array might get screwed up. For example, some pointers might be wrong, or some "dead" nodes might not have been freed when the last reference to them disappeared. Procedures *check_mem* and *search_mem* are available to help diagnose such problems. These procedures make use of two arrays called *free* and *was_free* that are present only if TeX's debugging routines have been included. (You may want to decrease the size of *mem* while you are debugging.)

```
define free ≡ free_arr
⟨ Global variables 13 ⟩ +≡
  { The debug memory arrays have not been mallocated yet. }
  debug free: packed array [0 .. 9] of boolean; { free cells }
  was_free: packed array [0 .. 9] of boolean; { previously free cells }
  was_mem_end, was_lo_max, was_hi_min: pointer; { previous mem_end, lo_mem_max, and hi_mem_min }
  panicking: boolean; { do we want to check memory constantly? }
  gubed
```

200* Boxes, rules, inserts, whatsits, marks, and things in general that are sort of "complicated" are indicated only by printing '[]'.

```
procedure short_display(p : integer); { prints highlights of list p }
  var n: integer; { for replacement counts }
  begin while p > mem_min do
    begin if is_char_node(p) then
      begin if p ≤ mem_end then
        begin if font(p) ≠ font_in_short_display then
          begin if (font(p) > font_max) then print_char("*")
          else ⟨ Print the font identifier for font(p) 297 ⟩;
          print_char("␣"); font_in_short_display ← font(p);
          end;
        print_ASCII(qo(character(p)));
        end;
      end
    else ⟨ Print a short indication of the contents of node p 201 ⟩;
    p ← link(p);
    end;
  end;
```

202* The *show_node_list* routine requires some auxiliary subroutines: one to print a font-and-character combination, one to print a token list without its reference count, and one to print a rule dimension.

```
procedure print_font_and_char(p : integer); { prints char_node data }
  begin if p > mem_end then print_esc("CLOBBERED.")
  else begin if (font(p) > font_max) then print_char("*")
    else ⟨ Print the font identifier for font(p) 297 ⟩;
    print_char("␣"); print_ASCII(qo(character(p)));
    end;
  end;
procedure print_mark(p : integer); { prints token list data in braces }
  begin print_char("{");
  if (p < hi_mem_min) ∨ (p > mem_end) then print_esc("CLOBBERED.")
  else show_token_list(link(p), null, max_print_line − 10);
  print_char("}");
  end;
procedure print_rule_dimen(d : scaled); { prints dimension in rule node }
  begin if is_running(d) then print_char("*")
  else print_scaled(d);
  end;
```

212*. The code will have to change in this place if *glue_ratio* is a structured type instead of an ordinary *real*. Note that this routine should avoid arithmetic errors even if the *glue_set* field holds an arbitrary random value. The following code assumes that a properly formed nonzero *real* number has absolute value 2^{20} or more when it is regarded as an integer; this precaution was adequate to prevent floating point underflow on the author's computer.

```
⟨ Display the value of glue_set(p) 212* ⟩ ≡
  g ← float(glue_set(p));
  if (g ≠ float_constant(0)) ∧ (glue_sign(p) ≠ normal) then
    begin print(", glue set ");
    if glue_sign(p) = shrinking then print("- ");  { The Unix pc folks removed this restriction with a
        remark that invalid bit patterns were vanishingly improbable, so we follow their example without
        really understanding it. if abs(mem[p + glue_offset].int) < ´4000000 then print(´?.?´) else }
    if fabs(g) > float_constant(20000) then
      begin if g > float_constant(0) then print_char(">")
      else print("< -");
      print_glue(20000 * unity, glue_order(p), 0);
      end
    else print_glue(round(unity * g), glue_order(p), 0);
    end
```

This code is used in section 210.

228* Now we are ready to delete any node list, recursively. In practice, the nodes deleted are usually charnodes (about 2/3 of the time), and they are glue nodes in about half of the remaining cases.

```
procedure flush_node_list(p : pointer); { erase list of nodes starting at p }
  label done; { go here when node p has been freed }
  var q: pointer; { successor to node p }
  begin while p ≠ null do
    begin q ← link(p);
    if is_char_node(p) then free_avail(p)
    else begin case type(p) of
      hlist_node, vlist_node, unset_node: begin flush_node_list(list_ptr(p)); free_node(p, box_node_size);
        goto done;
        end;
      rule_node: begin free_node(p, rule_node_size); goto done;
        end;
      ins_node: begin flush_node_list(ins_ptr(p)); delete_glue_ref(split_top_ptr(p));
        free_node(p, ins_node_size); goto done;
        end;
      whatsit_node: ⟨ Wipe out the whatsit node p and goto done 1418 ⟩;
      glue_node: begin fast_delete_glue_ref(glue_ptr(p));
        if leader_ptr(p) ≠ null then flush_node_list(leader_ptr(p));
        free_node(p, medium_node_size); goto done;
        end;
      kern_node, math_node, penalty_node: begin free_node(p, medium_node_size); goto done;
        end;
      margin_kern_node: begin free_node(p, margin_kern_node_size); goto done;
        end;
      ligature_node: flush_node_list(lig_ptr(p));
      mark_node: delete_token_ref(mark_ptr(p));
      disc_node: begin flush_node_list(pre_break(p)); flush_node_list(post_break(p));
        end;
      adjust_node: flush_node_list(adjust_ptr(p));
      ⟨ Cases of flush_node_list that arise in mlists only 740 ⟩
      othercases confusion("flushing")
      endcases;
      free_node(p, small_node_size);
    done: end;
    p ← q;
    end;
  end;
```

232* ⟨ Case statement to copy different types and set *words* to the number of initial words not yet copied 232* ⟩ ≡

```
case type(p) of
hlist_node, vlist_node, unset_node: begin r ← get_node(box_node_size);
  ⟨ Copy the box SyncTeX information 1733* ⟩;
  mem[r + 6] ← mem[p + 6]; mem[r + 5] ← mem[p + 5]; { copy the last two words }
  list_ptr(r) ← copy_node_list(list_ptr(p)); { this affects mem[r + 5] }
  words ← 5;
  end;
rule_node: begin r ← get_node(rule_node_size); words ← rule_node_size − synctex_field_size;
    { SyncTeX: do not let TeX copy the SyncTeX information }
  ⟨ Copy the rule SyncTeX information 1734* ⟩;
  end;
ins_node: begin r ← get_node(ins_node_size); mem[r + 4] ← mem[p + 4]; add_glue_ref(split_top_ptr(p));
  ins_ptr(r) ← copy_node_list(ins_ptr(p)); { this affects mem[r + 4] }
  words ← ins_node_size − 1;
  end;
whatsit_node: ⟨ Make a partial copy of the whatsit node p and make r point to it; set words to the
    number of initial words not yet copied 1417 ⟩;
glue_node: begin r ← get_node(medium_node_size); add_glue_ref(glue_ptr(p));
  ⟨ Copy the medium sized node SyncTeX information 1735* ⟩;
  glue_ptr(r) ← glue_ptr(p); leader_ptr(r) ← copy_node_list(leader_ptr(p));
  end;
kern_node, math_node, penalty_node: begin r ← get_node(medium_node_size);
  words ← medium_node_size;
  end;
margin_kern_node: begin r ← get_node(margin_kern_node_size); words ← margin_kern_node_size;
  end;
ligature_node: begin r ← get_node(small_node_size); mem[lig_char(r)] ← mem[lig_char(p)];
    { copy font and character }
  lig_ptr(r) ← copy_node_list(lig_ptr(p));
  end;
disc_node: begin r ← get_node(small_node_size); pre_break(r) ← copy_node_list(pre_break(p));
  post_break(r) ← copy_node_list(post_break(p));
  end;
mark_node: begin r ← get_node(small_node_size); add_token_ref(mark_ptr(p));
  words ← small_node_size;
  end;
adjust_node: begin r ← get_node(small_node_size); adjust_ptr(r) ← copy_node_list(adjust_ptr(p));
  end; { words = 1 = small_node_size − 1 }
othercases confusion("copying")
endcases
```

This code is used in section 231.

235* The next codes are special; they all relate to mode-independent assignment of values to TeX's internal registers or tables. Codes that are *max_internal* or less represent internal quantities that might be expanded by '`\the`'.

define *toks_register* = 72 { token list register (`\toks`) }
define *assign_toks* = 73 { special token list (`\output`, `\everypar`, etc.) }
define *assign_int* = 74 { user-defined integer (`\tolerance`, `\day`, etc.) }
define *assign_dimen* = 75 { user-defined length (`\hsize`, etc.) }
define *assign_glue* = 76 { user-defined glue (`\baselineskip`, etc.) }
define *assign_mu_glue* = 77 { user-defined muglue (`\thinmuskip`, etc.) }
define *assign_font_dimen* = 78 { user-defined font dimension (`\fontdimen`) }
define *assign_font_int* = 79 { user-defined font integer (`\hyphenchar`, `\skewchar`) }
define *set_aux* = 80 { specify state info (`\spacefactor`, `\prevdepth`) }
define *set_prev_graf* = 81 { specify state info (`\prevgraf`) }
define *set_page_dimen* = 82 { specify state info (`\pagegoal`, etc.) }
define *set_page_int* = 83 { specify state info (`\deadcycles`, `\insertpenalties`) }
{ (or `\interactionmode`) }
define *set_box_dimen* = 84 { change dimension of box (`\wd`, `\ht`, `\dp`) }
define *set_shape* = 85 { specify fancy paragraph shape (`\parshape`) }
{ (or `\interlinepenalties`, etc.) }
define *def_code* = 86 { define a character code (`\catcode`, etc.) }
define *XeTeX_def_code* = 87 { `\Umathcode`, `\Udelcode` }
define *def_family* = 88 { declare math fonts (`\textfont`, etc.) }
define *set_font* = 89 { set current font (font identifiers) }
define *def_font* = 90 { define a font file (`\font`) }
define *register* = 91 { internal register (`\count`, `\dimen`, etc.) }
define *max_internal* = 91 { the largest code that can follow `\the` }
define *advance* = 92 { advance a register or parameter (`\advance`) }
define *multiply* = 93 { multiply a register or parameter (`\multiply`) }
define *divide* = 94 { divide a register or parameter (`\divide`) }
define *prefix* = 95 { qualify a definition (`\global`, `\long`, `\outer`) }
{ (or `\protected`) }
define *let* = 96 { assign a command code (`\let`, `\futurelet`) }
define *shorthand_def* = 97 { code definition (`\chardef`, `\countdef`, etc.) }
{ or `\charsubdef` }
define *read_to_cs* = 98 { read into a control sequence (`\read`) }
{ (or `\readline`) }
define *def* = 99 { macro definition (`\def`, `\gdef`, `\xdef`, `\edef`) }
define *set_box* = 100 { set a box (`\setbox`) }
define *hyph_data* = 101 { hyphenation data (`\hyphenation`, `\patterns`) }
define *set_interaction* = 102 { define level of interaction (`\batchmode`, etc.) }
define *max_command* = 102 { the largest command code seen at *big_switch* }

237* The semantic nest. TeX is typically in the midst of building many lists at once. For example, when a math formula is being processed, TeX is in math mode and working on an mlist; this formula has temporarily interrupted TeX from being in horizontal mode and building the hlist of a paragraph; and this paragraph has temporarily interrupted TeX from being in vertical mode and building the vlist for the next page of a document. Similarly, when a `\vbox` occurs inside of an `\hbox`, TeX is temporarily interrupted from working in restricted horizontal mode, and it enters internal vertical mode. The "semantic nest" is a stack that keeps track of what lists and modes are currently suspended.

At each level of processing we are in one of six modes:

vmode stands for vertical mode (the page builder);
hmode stands for horizontal mode (the paragraph builder);
mmode stands for displayed formula mode;
$-$*vmode* stands for internal vertical mode (e.g., in a `\vbox`);
$-$*hmode* stands for restricted horizontal mode (e.g., in an `\hbox`);
$-$*mmode* stands for math formula mode (not displayed).

The mode is temporarily set to zero while processing `\write` texts.

Numeric values are assigned to *vmode*, *hmode*, and *mmode* so that TeX's "big semantic switch" can select the appropriate thing to do by computing the value $abs(mode) + cur\_cmd$, where *mode* is the current mode and *cur_cmd* is the current command code.

```
define vmode = 1 { vertical mode }
define hmode = vmode + max_command + 1 { horizontal mode }
define mmode = hmode + max_command + 1 { math mode }
procedure print_mode(m : integer); { prints the mode represented by m }
  begin if m > 0 then
    case m div (max_command + 1) of
    0: print("vertical␣mode");
    1: print("horizontal␣mode");
    2: print("display␣math␣mode");
    end
  else if m = 0 then print("no␣mode")
    else case (−m) div (max_command + 1) of
      0: print("internal␣vertical␣mode");
      1: print("restricted␣horizontal␣mode");
      2: print("math␣mode");
      end;
  end;
procedure print_in_mode(m : integer); { prints the mode represented by m }
  begin if m > 0 then
    case m div (max_command + 1) of
    0: print("´␣in␣vertical␣mode");
    1: print("´␣in␣horizontal␣mode");
    2: print("´␣in␣display␣math␣mode");
    end
  else if m = 0 then print("´␣in␣no␣mode")
    else case (−m) div (max_command + 1) of
      0: print("´␣in␣internal␣vertical␣mode");
      1: print("´␣in␣restricted␣horizontal␣mode");
      2: print("´␣in␣math␣mode");
      end;
  end;
```

239* **define** *mode* ≡ *cur_list.mode_field* { current mode }
define *head* ≡ *cur_list.head_field* { header node of current list }
define *tail* ≡ *cur_list.tail_field* { final node on current list }
define *eTeX_aux* ≡ *cur_list.eTeX_aux_field* { auxiliary data for ε-TeX }
define *LR_save* ≡ *eTeX_aux* { LR stack when a paragraph is interrupted }
define *LR_box* ≡ *eTeX_aux* { prototype box for display }
define *delim_ptr* ≡ *eTeX_aux* { most recent left or right noad of a math left group }
define *prev_graf* ≡ *cur_list.pg_field* { number of paragraph lines accumulated }
define *aux* ≡ *cur_list.aux_field* { auxiliary data about the current list }
define *prev_depth* ≡ *aux.sc* { the name of *aux* in vertical mode }
define *space_factor* ≡ *aux.hh.lh* { part of *aux* in horizontal mode }
define *clang* ≡ *aux.hh.rh* { the other part of *aux* in horizontal mode }
define *incompleat_noad* ≡ *aux.int* { the name of *aux* in math mode }
define *mode_line* ≡ *cur_list.ml_field* { source file line number at beginning of list }

⟨ Global variables 13 ⟩ +≡
nest: ↑*list_state_record*;
nest_ptr: 0 .. *nest_size*; { first unused location of *nest* }
max_nest_stack: 0 .. *nest_size*; { maximum of *nest_ptr* when pushing }
cur_list: *list_state_record*; { the "top" semantic state }
shown_mode: −*mmode* .. *mmode*; { most recent mode shown by `\tracingcommands` }

241* We will see later that the vertical list at the bottom semantic level is split into two parts; the "current page" runs from *page_head* to *page_tail*, and the "contribution list" runs from *contrib_head* to *tail* of semantic level zero. The idea is that contributions are first formed in vertical mode, then "contributed" to the current page (during which time the page-breaking decisions are made). For now, we don't need to know any more details about the page-building process.

⟨ Set initial values of key variables 23* ⟩ +≡
nest_ptr ← 0; *max_nest_stack* ← 0; *mode* ← *vmode*; *head* ← *contrib_head*; *tail* ← *contrib_head*;
eTeX_aux ← *null*; *prev_depth* ← *ignore_depth*; *mode_line* ← 0; *prev_graf* ← 0; *shown_mode* ← 0;
{ The following piece of code is a copy of module 991: }
page_contents ← *empty*; *page_tail* ← *page_head*; { *link*(*page_head*) ← *null*; }
last_glue ← *max_halfword*; *last_penalty* ← 0; *last_kern* ← 0; *last_node_type* ← −1; *page_depth* ← 0;
page_max_depth ← 0;

245* ⟨ Show the auxiliary field, a 245* ⟩ ≡

```
case abs(m) div (max_command + 1) of
0: begin print_nl("prevdepth␣");
  if a.sc ≤ ignore_depth then print("ignored")
  else print_scaled(a.sc);
  if nest[p].pg_field ≠ 0 then
    begin print(",␣prevgraf␣"); print_int(nest[p].pg_field);
    if nest[p].pg_field ≠ 1 then print("␣lines")
    else print("␣line");
    end;
  end;
1: begin print_nl("spacefactor␣"); print_int(a.hh.lh);
  if m > 0 then if a.hh.rh > 0 then
      begin print(",␣current␣language␣"); print_int(a.hh.rh); end;
  end;
2: if a.int ≠ null then
    begin print("this␣will␣begin␣denominator␣of:"); show_box(a.int); end;
end { there are no other cases }
```

This code is used in section 244.

246* The table of equivalents. Now that we have studied the data structures for TEX's semantic routines, we ought to consider the data structures used by its syntactic routines. In other words, our next concern will be the tables that TEX looks at when it is scanning what the user has written.

The biggest and most important such table is called *eqtb*. It holds the current "equivalents" of things; i.e., it explains what things mean or what their current values are, for all quantities that are subject to the nesting structure provided by TEX's grouping mechanism. There are six parts to *eqtb*:

1) *eqtb*[*active_base* .. (*hash_base* − 1)] holds the current equivalents of single-character control sequences.

2) *eqtb*[*hash_base* .. (*glue_base* − 1)] holds the current equivalents of multiletter control sequences.

3) *eqtb*[*glue_base* .. (*local_base* − 1)] holds the current equivalents of glue parameters like the current baselineskip.

4) *eqtb*[*local_base* .. (*int_base* − 1)] holds the current equivalents of local halfword quantities like the current box registers, the current "catcodes," the current font, and a pointer to the current paragraph shape. Additionally region 4 contains the table with MLTEX's character substitution definitions.

5) *eqtb*[*int_base* .. (*dimen_base* − 1)] holds the current equivalents of fullword integer parameters like the current hyphenation penalty.

6) *eqtb*[*dimen_base* .. *eqtb_size*] holds the current equivalents of fullword dimension parameters like the current hsize or amount of hanging indentation.

Note that, for example, the current amount of baselineskip glue is determined by the setting of a particular location in region 3 of *eqtb*, while the current meaning of the control sequence '`\baselineskip`' (which might have been changed by `\def` or `\let`) appears in region 2.

248*. Many locations in *eqtb* have symbolic names. The purpose of the next paragraphs is to define these names, and to set up the initial values of the equivalents.

In the first region we have *number_usvs* equivalents for "active characters" that act as control sequences, followed by *number_usvs* equivalents for single-character control sequences.

Then comes region 2, which corresponds to the hash table that we will define later. The maximum address in this region is used for a dummy control sequence that is perpetually undefined. There also are several locations for control sequences that are perpetually defined (since they are used in error recovery).

define *active_base* = 1 { beginning of region 1, for active character equivalents }
define *single_base* = *active_base* + *number_usvs* { equivalents of one-character control sequences }
define *null_cs* = *single_base* + *number_usvs* { equivalent of `\csname\endcsname` }
define *hash_base* = *null_cs* + 1 { beginning of region 2, for the hash table }
define *frozen_control_sequence* = *hash_base* + *hash_size* { for error recovery }
define *frozen_protection* = *frozen_control_sequence* { inaccessible but definable }
define *frozen_cr* = *frozen_control_sequence* + 1 { permanent '`\cr`' }
define *frozen_end_group* = *frozen_control_sequence* + 2 { permanent '`\endgroup`' }
define *frozen_right* = *frozen_control_sequence* + 3 { permanent '`\right`' }
define *frozen_fi* = *frozen_control_sequence* + 4 { permanent '`\fi`' }
define *frozen_end_template* = *frozen_control_sequence* + 5 { permanent '`\endtemplate`' }
define *frozen_endv* = *frozen_control_sequence* + 6 { second permanent '`\endtemplate`' }
define *frozen_relax* = *frozen_control_sequence* + 7 { permanent '`\relax`' }
define *end_write* = *frozen_control_sequence* + 8 { permanent '`\endwrite`' }
define *frozen_dont_expand* = *frozen_control_sequence* + 9 { permanent '`\notexpanded:`' }
define *prim_size* = 500 { maximum number of primitives }
define *frozen_special* = *frozen_control_sequence* + 10 { permanent '`\special`' }
define *frozen_null_font* = *frozen_control_sequence* + 12 + *prim_size* { permanent '`\nullfont`' }
define *frozen_primitive* = *frozen_control_sequence* + 11 { permanent '`\pdfprimitive`' }
define *prim_eqtb_base* = *frozen_primitive* + 1
define *font_id_base* = *frozen_null_font* − *font_base* { begins table of 257 permanent font identifiers }
define *undefined_control_sequence* = *frozen_null_font* + *max_font_max* + 1 { dummy location }
define *glue_base* = *undefined_control_sequence* + 1 { beginning of region 3 }

⟨ Initialize table entries (done by `INITEX` only) 189 ⟩ +≡
eq_type(*undefined_control_sequence*) ← *undefined_cs*; *equiv*(*undefined_control_sequence*) ← *null*;
eq_level(*undefined_control_sequence*) ← *level_zero*;
for *k* ← *active_base* **to** *eqtb_top* **do** *eqtb*[*k*] ← *eqtb*[*undefined_control_sequence*];

256*. Region 4 of *eqtb* contains the local quantities defined here. The bulk of this region is taken up by five tables that are indexed by eight-bit characters; these tables are important to both the syntactic and semantic portions of TEX. There are also a bunch of special things like font and token parameters, as well as the tables of `\toks` and `\box` registers.

define *par_shape_loc* = *local_base* { specifies paragraph shape }
define *output_routine_loc* = *local_base* + 1 { points to token list for `\output` }
define *every_par_loc* = *local_base* + 2 { points to token list for `\everypar` }
define *every_math_loc* = *local_base* + 3 { points to token list for `\everymath` }
define *every_display_loc* = *local_base* + 4 { points to token list for `\everydisplay` }
define *every_hbox_loc* = *local_base* + 5 { points to token list for `\everyhbox` }
define *every_vbox_loc* = *local_base* + 6 { points to token list for `\everyvbox` }
define *every_job_loc* = *local_base* + 7 { points to token list for `\everyjob` }
define *every_cr_loc* = *local_base* + 8 { points to token list for `\everycr` }
define *err_help_loc* = *local_base* + 9 { points to token list for `\errhelp` }
define *tex_toks* = *local_base* + 10 { end of TEX's token list parameters }

define *etex_toks_base* = *tex_toks* { base for ε-TEX's token list parameters }
define *every_eof_loc* = *etex_toks_base* { points to token list for `\everyeof` }
define *XeTeX_inter_char_loc* = *every_eof_loc* + 1 { not really used, but serves as a flag }
define *etex_toks* = *XeTeX_inter_char_loc* + 1 { end of ε-TEX's token list parameters }

define *toks_base* = *etex_toks* { table of *number_regs* token list registers }

define *etex_pen_base* = *toks_base* + *number_regs* { start of table of ε-TEX's penalties }
define *inter_line_penalties_loc* = *etex_pen_base* { additional penalties between lines }
define *club_penalties_loc* = *etex_pen_base* + 1 { penalties for creating club lines }
define *widow_penalties_loc* = *etex_pen_base* + 2 { penalties for creating widow lines }
define *display_widow_penalties_loc* = *etex_pen_base* + 3 { ditto, just before a display }
define *etex_pens* = *etex_pen_base* + 4 { end of table of ε-TEX's penalties }

define *box_base* = *etex_pens* { table of *number_regs* box registers }
define *cur_font_loc* = *box_base* + *number_regs* { internal font number outside math mode }
define *math_font_base* = *cur_font_loc* + 1 { table of *number_math_fonts* math font numbers }
define *cat_code_base* = *math_font_base* + *number_math_fonts*
{ table of *number_usvs* command codes (the "catcodes") }
define *lc_code_base* = *cat_code_base* + *number_usvs* { table of *number_usvs* lowercase mappings }
define *uc_code_base* = *lc_code_base* + *number_usvs* { table of *number_usvs* uppercase mappings }
define *sf_code_base* = *uc_code_base* + *number_usvs* { table of *number_usvs* spacefactor mappings }
define *math_code_base* = *sf_code_base* + *number_usvs* { table of *number_usvs* math mode mappings }
define *char_sub_code_base* = *math_code_base* + *number_usvs* { table of character substitutions }
define *int_base* = *char_sub_code_base* + *number_usvs* { beginning of region 5 }

define *par_shape_ptr* ≡ *equiv*(*par_shape_loc*)
define *output_routine* ≡ *equiv*(*output_routine_loc*)
define *every_par* ≡ *equiv*(*every_par_loc*)
define *every_math* ≡ *equiv*(*every_math_loc*)
define *every_display* ≡ *equiv*(*every_display_loc*)
define *every_hbox* ≡ *equiv*(*every_hbox_loc*)
define *every_vbox* ≡ *equiv*(*every_vbox_loc*)
define *every_job* ≡ *equiv*(*every_job_loc*)
define *every_cr* ≡ *equiv*(*every_cr_loc*)
define *err_help* ≡ *equiv*(*err_help_loc*)
define *toks*(#) ≡ *equiv*(*toks_base* + #)
define *box*(#) ≡ *equiv*(*box_base* + #)
define *cur_font* ≡ *equiv*(*cur_font_loc*)
define *fam_fnt*(#) ≡ *equiv*(*math_font_base* + #)

define *cat_code*(#) ≡ *equiv*(*cat_code_base* + #)
define *lc_code*(#) ≡ *equiv*(*lc_code_base* + #)
define *uc_code*(#) ≡ *equiv*(*uc_code_base* + #)
define *sf_code*(#) ≡ *equiv*(*sf_code_base* + #)
define *math_code*(#) ≡ *equiv*(*math_code_base* + #)
{ Note: *math_code*(*c*) is the true math code plus *min_halfword* }
define *char_sub_code*(#) ≡ *equiv*(*char_sub_code_base* + #)
{ Note: *char_sub_code*(*c*) is the true substitution info plus *min_halfword* }

⟨ Put each of TeX's primitives into the hash table 252 ⟩ +≡
primitive("output", *assign_toks*, *output_routine_loc*); *primitive*("everypar", *assign_toks*, *every_par_loc*);
primitive("everymath", *assign_toks*, *every_math_loc*);
primitive("everydisplay", *assign_toks*, *every_display_loc*);
primitive("everyhbox", *assign_toks*, *every_hbox_loc*); *primitive*("everyvbox", *assign_toks*, *every_vbox_loc*);
primitive("everyjob", *assign_toks*, *every_job_loc*); *primitive*("everycr", *assign_toks*, *every_cr_loc*);
primitive("errhelp", *assign_toks*, *err_help_loc*);

262*. Region 5 of *eqtb* contains the integer parameters and registers defined here, as well as the *del_code* table. The latter table differs from the *cat_code* .. *math_code* tables that precede it, since delimiter codes are fullword integers while the other kinds of codes occupy at most a halfword. This is what makes region 5 different from region 4. We will store the *eq_level* information in an auxiliary array of quarterwords that will be defined later.

define *pretolerance_code* = 0 { badness tolerance before hyphenation }
define *tolerance_code* = 1 { badness tolerance after hyphenation }
define *line_penalty_code* = 2 { added to the badness of every line }
define *hyphen_penalty_code* = 3 { penalty for break after discretionary hyphen }
define *ex_hyphen_penalty_code* = 4 { penalty for break after explicit hyphen }
define *club_penalty_code* = 5 { penalty for creating a club line }
define *widow_penalty_code* = 6 { penalty for creating a widow line }
define *display_widow_penalty_code* = 7 { ditto, just before a display }
define *broken_penalty_code* = 8 { penalty for breaking a page at a broken line }
define *bin_op_penalty_code* = 9 { penalty for breaking after a binary operation }
define *rel_penalty_code* = 10 { penalty for breaking after a relation }
define *pre_display_penalty_code* = 11 { penalty for breaking just before a displayed formula }
define *post_display_penalty_code* = 12 { penalty for breaking just after a displayed formula }
define *inter_line_penalty_code* = 13 { additional penalty between lines }
define *double_hyphen_demerits_code* = 14 { demerits for double hyphen break }
define *final_hyphen_demerits_code* = 15 { demerits for final hyphen break }
define *adj_demerits_code* = 16 { demerits for adjacent incompatible lines }
define *mag_code* = 17 { magnification ratio }
define *delimiter_factor_code* = 18 { ratio for variable-size delimiters }
define *looseness_code* = 19 { change in number of lines for a paragraph }
define *time_code* = 20 { current time of day }
define *day_code* = 21 { current day of the month }
define *month_code* = 22 { current month of the year }
define *year_code* = 23 { current year of our Lord }
define *show_box_breadth_code* = 24 { nodes per level in *show_box* }
define *show_box_depth_code* = 25 { maximum level in *show_box* }
define *hbadness_code* = 26 { hboxes exceeding this badness will be shown by *hpack* }
define *vbadness_code* = 27 { vboxes exceeding this badness will be shown by *vpack* }
define *pausing_code* = 28 { pause after each line is read from a file }
define *tracing_online_code* = 29 { show diagnostic output on terminal }
define *tracing_macros_code* = 30 { show macros as they are being expanded }
define *tracing_stats_code* = 31 { show memory usage if TeX knows it }
define *tracing_paragraphs_code* = 32 { show line-break calculations }
define *tracing_pages_code* = 33 { show page-break calculations }
define *tracing_output_code* = 34 { show boxes when they are shipped out }
define *tracing_lost_chars_code* = 35 { show characters that aren't in the font }
define *tracing_commands_code* = 36 { show command codes at *big_switch* }
define *tracing_restores_code* = 37 { show equivalents when they are restored }
define *uc_hyph_code* = 38 { hyphenate words beginning with a capital letter }
define *output_penalty_code* = 39 { penalty found at current page break }
define *max_dead_cycles_code* = 40 { bound on consecutive dead cycles of output }
define *hang_after_code* = 41 { hanging indentation changes after this many lines }
define *floating_penalty_code* = 42 { penalty for insertions held over after a split }
define *global_defs_code* = 43 { override `\global` specifications }
define *cur_fam_code* = 44 { current family }
define *escape_char_code* = 45 { escape character for token output }
define *default_hyphen_char_code* = 46 { value of `\hyphenchar` when a font is loaded }

define *default_skew_char_code* = 47 { value of `\skewchar` when a font is loaded }
define *end_line_char_code* = 48 { character placed at the right end of the buffer }
define *new_line_char_code* = 49 { character that prints as *print_ln* }
define *language_code* = 50 { current hyphenation table }
define *left_hyphen_min_code* = 51 { minimum left hyphenation fragment size }
define *right_hyphen_min_code* = 52 { minimum right hyphenation fragment size }
define *holding_inserts_code* = 53 { do not remove insertion nodes from `\box255` }
define *error_context_lines_code* = 54 { maximum intermediate line pairs shown }
define *tex_int_pars* = 55 { total number of TeX's integer parameters }

define *web2c_int_base* = *tex_int_pars* { base for web2c's integer parameters }
define *char_sub_def_min_code* = *web2c_int_base* { smallest value in the charsubdef list }
define *char_sub_def_max_code* = *web2c_int_base* + 1 { largest value in the charsubdef list }
define *tracing_char_sub_def_code* = *web2c_int_base* + 2 { traces changes to a charsubdef def }
define *tracing_stack_levels_code* = *web2c_int_base* + 3
{ tracing *input_stack* level if *tracingmacros* positive }
define *web2c_int_pars* = *web2c_int_base* + 4 { total number of web2c's integer parameters }

define *etex_int_base* = *web2c_int_pars* { base for ε-TeX's integer parameters }
define *tracing_assigns_code* = *etex_int_base* { show assignments }
define *tracing_groups_code* = *etex_int_base* + 1 { show save/restore groups }
define *tracing_ifs_code* = *etex_int_base* + 2 { show conditionals }
define *tracing_scan_tokens_code* = *etex_int_base* + 3 { show pseudo file open and close }
define *tracing_nesting_code* = *etex_int_base* + 4 { show incomplete groups and ifs within files }
define *pre_display_direction_code* = *etex_int_base* + 5 { text direction preceding a display }
define *last_line_fit_code* = *etex_int_base* + 6 { adjustment for last line of paragraph }
define *saving_vdiscards_code* = *etex_int_base* + 7 { save items discarded from vlists }
define *saving_hyph_codes_code* = *etex_int_base* + 8 { save hyphenation codes for languages }
define *suppress_fontnotfound_error_code* = *etex_int_base* + 9 { suppress errors for missing fonts }
define *XeTeX_linebreak_locale_code* = *etex_int_base* + 10
{ string number of locale to use for linebreak locations }
define *XeTeX_linebreak_penalty_code* = *etex_int_base* + 11
{ penalty to use at locale-dependent linebreak locations }
define *XeTeX_protrude_chars_code* = *etex_int_base* + 12
{ protrude chars at left/right edge of paragraphs }
define *eTeX_state_code* = *etex_int_base* + 13 { ε-TeX state variables }
define *etex_int_pars* = *eTeX_state_code* + *eTeX_states* { total number of ε-TeX's integer parameters }

define *synctex_code* = *etex_int_pars*
define *int_pars* = *synctex_code* + 1 { total number of integer parameters }
define *count_base* = *int_base* + *int_pars* { *number_regs* user `\count` registers }
define *del_code_base* = *count_base* + *number_regs* { *number_usvs* delimiter code mappings }
define *dimen_base* = *del_code_base* + *number_usvs* { beginning of region 6 }

define *del_code*(#) ≡ *eqtb*[*del_code_base* + #].*int*
define *count*(#) ≡ *eqtb*[*count_base* + #].*int*
define *int_par*(#) ≡ *eqtb*[*int_base* + #].*int* { an integer parameter }
define *pretolerance* ≡ *int_par*(*pretolerance_code*)
define *tolerance* ≡ *int_par*(*tolerance_code*)
define *line_penalty* ≡ *int_par*(*line_penalty_code*)
define *hyphen_penalty* ≡ *int_par*(*hyphen_penalty_code*)
define *ex_hyphen_penalty* ≡ *int_par*(*ex_hyphen_penalty_code*)
define *club_penalty* ≡ *int_par*(*club_penalty_code*)
define *widow_penalty* ≡ *int_par*(*widow_penalty_code*)
define *display_widow_penalty* ≡ *int_par*(*display_widow_penalty_code*)

define *broken_penalty* ≡ *int_par*(*broken_penalty_code*)
define *bin_op_penalty* ≡ *int_par*(*bin_op_penalty_code*)
define *rel_penalty* ≡ *int_par*(*rel_penalty_code*)
define *pre_display_penalty* ≡ *int_par*(*pre_display_penalty_code*)
define *post_display_penalty* ≡ *int_par*(*post_display_penalty_code*)
define *inter_line_penalty* ≡ *int_par*(*inter_line_penalty_code*)
define *double_hyphen_demerits* ≡ *int_par*(*double_hyphen_demerits_code*)
define *final_hyphen_demerits* ≡ *int_par*(*final_hyphen_demerits_code*)
define *adj_demerits* ≡ *int_par*(*adj_demerits_code*)
define *mag* ≡ *int_par*(*mag_code*)
define *delimiter_factor* ≡ *int_par*(*delimiter_factor_code*)
define *looseness* ≡ *int_par*(*looseness_code*)
define *time* ≡ *int_par*(*time_code*)
define *day* ≡ *int_par*(*day_code*)
define *month* ≡ *int_par*(*month_code*)
define *year* ≡ *int_par*(*year_code*)
define *show_box_breadth* ≡ *int_par*(*show_box_breadth_code*)
define *show_box_depth* ≡ *int_par*(*show_box_depth_code*)
define *hbadness* ≡ *int_par*(*hbadness_code*)
define *vbadness* ≡ *int_par*(*vbadness_code*)
define *pausing* ≡ *int_par*(*pausing_code*)
define *tracing_online* ≡ *int_par*(*tracing_online_code*)
define *tracing_macros* ≡ *int_par*(*tracing_macros_code*)
define *tracing_stats* ≡ *int_par*(*tracing_stats_code*)
define *tracing_paragraphs* ≡ *int_par*(*tracing_paragraphs_code*)
define *tracing_pages* ≡ *int_par*(*tracing_pages_code*)
define *tracing_output* ≡ *int_par*(*tracing_output_code*)
define *tracing_lost_chars* ≡ *int_par*(*tracing_lost_chars_code*)
define *tracing_commands* ≡ *int_par*(*tracing_commands_code*)
define *tracing_restores* ≡ *int_par*(*tracing_restores_code*)
define *uc_hyph* ≡ *int_par*(*uc_hyph_code*)
define *output_penalty* ≡ *int_par*(*output_penalty_code*)
define *max_dead_cycles* ≡ *int_par*(*max_dead_cycles_code*)
define *hang_after* ≡ *int_par*(*hang_after_code*)
define *floating_penalty* ≡ *int_par*(*floating_penalty_code*)
define *global_defs* ≡ *int_par*(*global_defs_code*)
define *cur_fam* ≡ *int_par*(*cur_fam_code*)
define *escape_char* ≡ *int_par*(*escape_char_code*)
define *default_hyphen_char* ≡ *int_par*(*default_hyphen_char_code*)
define *default_skew_char* ≡ *int_par*(*default_skew_char_code*)
define *end_line_char* ≡ *int_par*(*end_line_char_code*)
define *new_line_char* ≡ *int_par*(*new_line_char_code*)
define *language* ≡ *int_par*(*language_code*)
define *left_hyphen_min* ≡ *int_par*(*left_hyphen_min_code*)
define *right_hyphen_min* ≡ *int_par*(*right_hyphen_min_code*)
define *holding_inserts* ≡ *int_par*(*holding_inserts_code*)
define *error_context_lines* ≡ *int_par*(*error_context_lines_code*)
define *synctex* ≡ *int_par*(*synctex_code*)
define *char_sub_def_min* ≡ *int_par*(*char_sub_def_min_code*)
define *char_sub_def_max* ≡ *int_par*(*char_sub_def_max_code*)
define *tracing_char_sub_def* ≡ *int_par*(*tracing_char_sub_def_code*)
define *tracing_stack_levels* ≡ *int_par*(*tracing_stack_levels_code*)

define *tracing_assigns* ≡ *int_par*(*tracing_assigns_code*)
define *tracing_groups* ≡ *int_par*(*tracing_groups_code*)
define *tracing_ifs* ≡ *int_par*(*tracing_ifs_code*)
define *tracing_scan_tokens* ≡ *int_par*(*tracing_scan_tokens_code*)
define *tracing_nesting* ≡ *int_par*(*tracing_nesting_code*)
define *pre_display_direction* ≡ *int_par*(*pre_display_direction_code*)
define *last_line_fit* ≡ *int_par*(*last_line_fit_code*)
define *saving_vdiscards* ≡ *int_par*(*saving_vdiscards_code*)
define *saving_hyph_codes* ≡ *int_par*(*saving_hyph_codes_code*)
define *suppress_fontnotfound_error* ≡ *int_par*(*suppress_fontnotfound_error_code*)
define *XeTeX_linebreak_locale* ≡ *int_par*(*XeTeX_linebreak_locale_code*)
define *XeTeX_linebreak_penalty* ≡ *int_par*(*XeTeX_linebreak_penalty_code*)
define *XeTeX_protrude_chars* ≡ *int_par*(*XeTeX_protrude_chars_code*)

⟨ Assign the values *depth_threshold* ← *show_box_depth* and *breadth_max* ← *show_box_breadth* 262* ⟩ ≡
depth_threshold ← *show_box_depth*; *breadth_max* ← *show_box_breadth*

This code is used in section 224.

263* We can print the symbolic name of an integer parameter as follows.

```
procedure print_param(n : integer);
  begin case n of
  pretolerance_code: print_esc("pretolerance");
  tolerance_code: print_esc("tolerance");
  line_penalty_code: print_esc("linepenalty");
  hyphen_penalty_code: print_esc("hyphenpenalty");
  ex_hyphen_penalty_code: print_esc("exhyphenpenalty");
  club_penalty_code: print_esc("clubpenalty");
  widow_penalty_code: print_esc("widowpenalty");
  display_widow_penalty_code: print_esc("displaywidowpenalty");
  broken_penalty_code: print_esc("brokenpenalty");
  bin_op_penalty_code: print_esc("binoppenalty");
  rel_penalty_code: print_esc("relpenalty");
  pre_display_penalty_code: print_esc("predisplaypenalty");
  post_display_penalty_code: print_esc("postdisplaypenalty");
  inter_line_penalty_code: print_esc("interlinepenalty");
  double_hyphen_demerits_code: print_esc("doublehyphendemerits");
  final_hyphen_demerits_code: print_esc("finalhyphendemerits");
  adj_demerits_code: print_esc("adjdemerits");
  mag_code: print_esc("mag");
  delimiter_factor_code: print_esc("delimiterfactor");
  looseness_code: print_esc("looseness");
  time_code: print_esc("time");
  day_code: print_esc("day");
  month_code: print_esc("month");
  year_code: print_esc("year");
  show_box_breadth_code: print_esc("showboxbreadth");
  show_box_depth_code: print_esc("showboxdepth");
  hbadness_code: print_esc("hbadness");
  vbadness_code: print_esc("vbadness");
  pausing_code: print_esc("pausing");
  tracing_online_code: print_esc("tracingonline");
  tracing_macros_code: print_esc("tracingmacros");
  tracing_stats_code: print_esc("tracingstats");
  tracing_paragraphs_code: print_esc("tracingparagraphs");
  tracing_pages_code: print_esc("tracingpages");
  tracing_output_code: print_esc("tracingoutput");
  tracing_lost_chars_code: print_esc("tracinglostchars");
  tracing_commands_code: print_esc("tracingcommands");
  tracing_restores_code: print_esc("tracingrestores");
  uc_hyph_code: print_esc("uchyph");
  output_penalty_code: print_esc("outputpenalty");
  max_dead_cycles_code: print_esc("maxdeadcycles");
  hang_after_code: print_esc("hangafter");
  floating_penalty_code: print_esc("floatingpenalty");
  global_defs_code: print_esc("globaldefs");
  cur_fam_code: print_esc("fam");
  escape_char_code: print_esc("escapechar");
  default_hyphen_char_code: print_esc("defaulthyphenchar");
  default_skew_char_code: print_esc("defaultskewchar");
  end_line_char_code: print_esc("endlinechar");
```

new_line_char_code: *print_esc*("newlinechar");
language_code: *print_esc*("language");
left_hyphen_min_code: *print_esc*("lefthyphenmin");
right_hyphen_min_code: *print_esc*("righthyphenmin");
holding_inserts_code: *print_esc*("holdinginserts");
error_context_lines_code: *print_esc*("errorcontextlines");
char_sub_def_min_code: *print_esc*("charsubdefmin");
char_sub_def_max_code: *print_esc*("charsubdefmax");
tracing_char_sub_def_code: *print_esc*("tracingcharsubdef");
tracing_stack_levels_code: *print_esc*("tracingstacklevels");
XeTeX_linebreak_penalty_code: *print_esc*("XeTeXlinebreakpenalty");
XeTeX_protrude_chars_code: *print_esc*("XeTeXprotrudechars");
⟨ synctex case for *print_param* 1707* ⟩
⟨ Cases for *print_param* 1469 ⟩
othercases *print*("[unknown␣integer␣parameter!]")
endcases;
end;

264* The integer parameter names must be entered into the hash table.

⟨ Put each of TeX's primitives into the hash table 252 ⟩ +≡
primitive("pretolerance", *assign_int*, *int_base* + *pretolerance_code*);
primitive("tolerance", *assign_int*, *int_base* + *tolerance_code*);
primitive("linepenalty", *assign_int*, *int_base* + *line_penalty_code*);
primitive("hyphenpenalty", *assign_int*, *int_base* + *hyphen_penalty_code*);
primitive("exhyphenpenalty", *assign_int*, *int_base* + *ex_hyphen_penalty_code*);
primitive("clubpenalty", *assign_int*, *int_base* + *club_penalty_code*);
primitive("widowpenalty", *assign_int*, *int_base* + *widow_penalty_code*);
primitive("displaywidowpenalty", *assign_int*, *int_base* + *display_widow_penalty_code*);
primitive("brokenpenalty", *assign_int*, *int_base* + *broken_penalty_code*);
primitive("binoppenalty", *assign_int*, *int_base* + *bin_op_penalty_code*);
primitive("relpenalty", *assign_int*, *int_base* + *rel_penalty_code*);
primitive("predisplaypenalty", *assign_int*, *int_base* + *pre_display_penalty_code*);
primitive("postdisplaypenalty", *assign_int*, *int_base* + *post_display_penalty_code*);
primitive("interlinepenalty", *assign_int*, *int_base* + *inter_line_penalty_code*);
primitive("doublehyphendemerits", *assign_int*, *int_base* + *double_hyphen_demerits_code*);
primitive("finalhyphendemerits", *assign_int*, *int_base* + *final_hyphen_demerits_code*);
primitive("adjdemerits", *assign_int*, *int_base* + *adj_demerits_code*);
primitive("mag", *assign_int*, *int_base* + *mag_code*);
primitive("delimiterfactor", *assign_int*, *int_base* + *delimiter_factor_code*);
primitive("looseness", *assign_int*, *int_base* + *looseness_code*);
primitive("time", *assign_int*, *int_base* + *time_code*);
primitive("day", *assign_int*, *int_base* + *day_code*);
primitive("month", *assign_int*, *int_base* + *month_code*);
primitive("year", *assign_int*, *int_base* + *year_code*);
primitive("showboxbreadth", *assign_int*, *int_base* + *show_box_breadth_code*);
primitive("showboxdepth", *assign_int*, *int_base* + *show_box_depth_code*);
primitive("hbadness", *assign_int*, *int_base* + *hbadness_code*);
primitive("vbadness", *assign_int*, *int_base* + *vbadness_code*);
primitive("pausing", *assign_int*, *int_base* + *pausing_code*);
primitive("tracingonline", *assign_int*, *int_base* + *tracing_online_code*);
primitive("tracingmacros", *assign_int*, *int_base* + *tracing_macros_code*);
primitive("tracingstats", *assign_int*, *int_base* + *tracing_stats_code*);
primitive("tracingparagraphs", *assign_int*, *int_base* + *tracing_paragraphs_code*);
primitive("tracingpages", *assign_int*, *int_base* + *tracing_pages_code*);
primitive("tracingoutput", *assign_int*, *int_base* + *tracing_output_code*);
primitive("tracinglostchars", *assign_int*, *int_base* + *tracing_lost_chars_code*);
primitive("tracingcommands", *assign_int*, *int_base* + *tracing_commands_code*);
primitive("tracingrestores", *assign_int*, *int_base* + *tracing_restores_code*);
primitive("uchyph", *assign_int*, *int_base* + *uc_hyph_code*);
primitive("outputpenalty", *assign_int*, *int_base* + *output_penalty_code*);
primitive("maxdeadcycles", *assign_int*, *int_base* + *max_dead_cycles_code*);
primitive("hangafter", *assign_int*, *int_base* + *hang_after_code*);
primitive("floatingpenalty", *assign_int*, *int_base* + *floating_penalty_code*);
primitive("globaldefs", *assign_int*, *int_base* + *global_defs_code*);
primitive("fam", *assign_int*, *int_base* + *cur_fam_code*);
primitive("escapechar", *assign_int*, *int_base* + *escape_char_code*);
primitive("defaulthyphenchar", *assign_int*, *int_base* + *default_hyphen_char_code*);
primitive("defaultskewchar", *assign_int*, *int_base* + *default_skew_char_code*);
primitive("endlinechar", *assign_int*, *int_base* + *end_line_char_code*);
primitive("newlinechar", *assign_int*, *int_base* + *new_line_char_code*);

primitive("language", *assign_int*, *int_base* + *language_code*);
primitive("lefthyphenmin", *assign_int*, *int_base* + *left_hyphen_min_code*);
primitive("righthyphenmin", *assign_int*, *int_base* + *right_hyphen_min_code*);
primitive("holdinginserts", *assign_int*, *int_base* + *holding_inserts_code*);
primitive("errorcontextlines", *assign_int*, *int_base* + *error_context_lines_code*);
if *mltex_p* **then**
  begin *mltex_enabled_p* ← *true*; { enable character substitution }
  if *false* **then** { remove the if-clause to enable \charsubdefmin }
    primitive("charsubdefmin", *assign_int*, *int_base* + *char_sub_def_min_code*);
  primitive("charsubdefmax", *assign_int*, *int_base* + *char_sub_def_max_code*);
  primitive("tracingcharsubdef", *assign_int*, *int_base* + *tracing_char_sub_def_code*);
  end;
primitive("tracingstacklevels", *assign_int*, *int_base* + *tracing_stack_levels_code*);
primitive("XeTeXlinebreakpenalty", *assign_int*, *int_base* + *XeTeX_linebreak_penalty_code*);
primitive("XeTeXprotrudechars", *assign_int*, *int_base* + *XeTeX_protrude_chars_code*);

266*. The integer parameters should really be initialized by a macro package; the following initialization does the minimum to keep TeX from complete failure.

⟨ Initialize table entries (done by INITEX only) 189 ⟩ +≡
for *k* ← *int_base* **to** *del_code_base* − 1 **do** *eqtb*[*k*].*int* ← 0;
char_sub_def_min ← 256; *char_sub_def_max* ← −1; { allow \charsubdef for char 0 }
  { *tracing_char_sub_def* ← 0 is already done }
mag ← 1000; *tolerance* ← 10000; *hang_after* ← 1; *max_dead_cycles* ← 25; *escape_char* ← "\";
end_line_char ← *carriage_return*;
for *k* ← 0 **to** *number_usvs* − 1 **do** *del_code*(*k*) ← −1;
del_code(".") ← 0; { this null delimiter is used in error recovery }

267*. The following procedure, which is called just before TeX initializes its input and output, establishes the initial values of the date and time. It calls a *date_and_time* C macro (a.k.a. *dateandtime*), which calls the C function *get_date_and_time*, passing it the addresses of *sys_time*, etc., so they can be set by the routine. *get_date_and_time* also sets up interrupt catching if that is conditionally compiled in the C code.

We have to initialize the *sys_* variables because that is what gets output on the first line of the log file. (New in 2021.)

procedure *fix_date_and_time*;
begin *date_and_time*(*sys_time*, *sys_day*, *sys_month*, *sys_year*); *time* ← *sys_time*;
  { minutes since midnight }
day ← *sys_day*; { day of the month }
month ← *sys_month*; { month of the year }
year ← *sys_year*; { Anno Domini }
end;

278* Here is a procedure that displays the contents of $eqtb[n]$ symbolically.

```
⟨ Declare the procedure called print_cmd_chr 328 ⟩
  stat procedure show_eqtb(n : pointer);
  begin if n < active_base then print_char("?")   { this can't happen }
  else if (n < glue_base) ∨ ((n > eqtb_size) ∧ (n ≤ eqtb_top)) then ⟨ Show equivalent n, in region 1 or 2 249 ⟩
    else if n < local_base then ⟨ Show equivalent n, in region 3 255 ⟩
      else if n < int_base then ⟨ Show equivalent n, in region 4 259 ⟩
        else if n < dimen_base then ⟨ Show equivalent n, in region 5 268 ⟩
          else if n ≤ eqtb_size then ⟨ Show equivalent n, in region 6 277 ⟩
            else print_char("?");   { this can't happen either }
  end;
  tats
```

279* The last two regions of *eqtb* have fullword values instead of the three fields *eq_level*, *eq_type*, and *equiv*. An *eq_type* is unnecessary, but TeX needs to store the *eq_level* information in another array called *xeq_level*.

```
⟨ Global variables 13 ⟩ +≡
zeqtb: ↑memory_word;
xeq_level: array [int_base .. eqtb_size] of quarterword;
```

282* The hash table. Control sequences are stored and retrieved by means of a fairly standard hash table algorithm called the method of "coalescing lists" (cf. Algorithm 6.4C in *The Art of Computer Programming*). Once a control sequence enters the table, it is never removed, because there are complicated situations involving `\gdef` where the removal of a control sequence at the end of a group would be a mistake preventable only by the introduction of a complicated reference-count mechanism.

The actual sequence of letters forming a control sequence identifier is stored in the *str_pool* array together with all the other strings. An auxiliary array *hash* consists of items with two halfword fields per word. The first of these, called *next*(*p*), points to the next identifier belonging to the same coalesced list as the identifier corresponding to *p*; and the other, called *text*(*p*), points to the *str_start* entry for *p*'s identifier. If position *p* of the hash table is empty, we have *text*(*p*) = 0; if position *p* is either empty or the end of a coalesced hash list, we have *next*(*p*) = 0. An auxiliary pointer variable called *hash_used* is maintained in such a way that all locations $p \geq hash\_used$ are nonempty. The global variable *cs_count* tells how many multiletter control sequences have been defined, if statistics are being kept.

A global boolean variable called *no_new_control_sequence* is set to *true* during the time that new hash table entries are forbidden.

define *next*(#) ≡ *hash*[#].*lh* { link for coalesced lists }
define *text*(#) ≡ *hash*[#].*rh* { string number for control sequence name }
define *hash_is_full* ≡ (*hash_used* = *hash_base*) { test if all positions are occupied }
define *font_id_text*(#) ≡ *text*(*font_id_base* + #) { a frozen font identifier's name }

⟨ Global variables 13 ⟩ +≡
hash: ↑*two_halves*; { the hash table }
yhash: ↑*two_halves*; { auxiliary pointer for freeing hash }
hash_used: *pointer*; { allocation pointer for *hash* }
hash_extra: *pointer*; { *hash_extra* = *hash* above *eqtb_size* }
hash_top: *pointer*; { maximum of the hash array }
eqtb_top: *pointer*; { maximum of the *eqtb* }
hash_high: *pointer*; { pointer to next high hash location }
no_new_control_sequence: *boolean*; { are new identifiers legal? }
cs_count: *integer*; { total number of known identifiers }

284* ⟨ Set initial values of key variables 23* ⟩ +≡
no_new_control_sequence ← *true*; { new identifiers are usually forbidden }
prim_next(0) ← 0; *prim_text*(0) ← 0;
for *k* ← 1 **to** *prim_size* **do** *prim*[*k*] ← *prim*[0];

285* ⟨ Initialize table entries (done by `INITEX` only) 189 ⟩ +≡
prim_used ← *prim_size*; { nothing is used }
hash_used ← *frozen_control_sequence*; { nothing is used }
hash_high ← 0; *cs_count* ← 0; *eq_type*(*frozen_dont_expand*) ← *dont_expand*;
text(*frozen_dont_expand*) ← `"notexpanded:"`; *eq_type*(*frozen_primitive*) ← *ignore_spaces*;
equiv(*frozen_primitive*) ← 1; *eq_level*(*frozen_primitive*) ← *level_one*;
text(*frozen_primitive*) ← `"primitive"`;

287* ⟨ Insert a new control sequence after p, then make p point to it 287* ⟩ ≡

```
begin if text(p) > 0 then
  begin if hash_high < hash_extra then
    begin incr(hash_high); next(p) ← hash_high + eqtb_size; p ← hash_high + eqtb_size;
    end
  else begin repeat if hash_is_full then overflow("hash␣size", hash_size + hash_extra);
      decr(hash_used);
    until text(hash_used) = 0;  { search for an empty location in hash }
    next(p) ← hash_used; p ← hash_used;
    end;
  end;
str_room(ll); d ← cur_length;
while pool_ptr > str_start_macro(str_ptr) do
  begin decr(pool_ptr); str_pool[pool_ptr + l] ← str_pool[pool_ptr];
  end;  { move current string up to make room for another }
for k ← j to j + l − 1 do
  begin if buffer[k] < "10000 then append_char(buffer[k])
  else begin append_char("D800 + (buffer[k] − "10000) div "400);
    append_char("DC00 + (buffer[k] − "10000) mod "400);
    end
  end;
text(p) ← make_string; pool_ptr ← pool_ptr + d;
stat incr(cs_count); tats
end
```

This code is used in section 286.

292* Single-character control sequences do not need to be looked up in a hash table, since we can use the character code itself as a direct address. The procedure *print_cs* prints the name of a control sequence, given a pointer to its address in *eqtb*. A space is printed after the name unless it is a single nonletter or an active character. This procedure might be invoked with invalid data, so it is "extra robust." The individual characters must be printed one at a time using *print*, since they may be unprintable.

⟨ Basic printing procedures 57 ⟩ +≡

```
procedure print_cs(p : integer);  { prints a purported control sequence }
  begin if p < hash_base then  { single character }
    if p ≥ single_base then
      if p = null_cs then
        begin print_esc("csname"); print_esc("endcsname"); print_char("␣");
        end
      else begin print_esc(p − single_base);
        if cat_code(p − single_base) = letter then print_char("␣");
        end
    else if p < active_base then print_esc("IMPOSSIBLE.")
      else print_char(p − active_base)
  else if ((p ≥ undefined_control_sequence) ∧ (p ≤ eqtb_size)) ∨ (p > eqtb_top) then
      print_esc("IMPOSSIBLE.")
    else if (text(p) ≥ str_ptr) then print_esc("NONEXISTENT.")
      else begin if (p ≥ prim_eqtb_base) ∧ (p < frozen_null_font) then
          print_esc(prim_text(p − prim_eqtb_base) − 1)
        else print_esc(text(p));
        print_char("␣");
        end;
  end;
```

301* ⟨ Global variables 13 ⟩ +≡
```
save_stack: ↑memory_word;
save_ptr: 0 .. save_size;  { first unused entry on save_stack }
max_save_stack: 0 .. save_size;  { maximum usage of save stack }
cur_level: quarterword;  { current nesting level for groups }
cur_group: group_code;  { current group type }
cur_boundary: 0 .. save_size;  { where the current level begins }
```

313* A global definition, which sets the level to *level_one*, will not be undone by *unsave*. If at least one global definition of *eqtb*[*p*] has been carried out within the group that just ended, the last such definition will therefore survive.

⟨ Store *save_stack*[*save_ptr*] in *eqtb*[*p*], unless *eqtb*[*p*] holds a global value 313* ⟩ ≡
```
if (p < int_base) ∨ (p > eqtb_size) then
  if eq_level(p) = level_one then
    begin eq_destroy(save_stack[save_ptr]);  { destroy the saved value }
    stat if tracing_restores > 0 then restore_trace(p, "retaining");
    tats
    end
  else begin eq_destroy(eqtb[p]);  { destroy the current value }
    eqtb[p] ← save_stack[save_ptr];  { restore the saved value }
    stat if tracing_restores > 0 then restore_trace(p, "restoring");
    tats
    end
else if xeq_level[p] ≠ level_one then
    begin eqtb[p] ← save_stack[save_ptr]; xeq_level[p] ← l;
    stat if tracing_restores > 0 then restore_trace(p, "restoring");
    tats
    end
  else begin stat if tracing_restores > 0 then restore_trace(p, "retaining");
    tats
    end
```
This code is used in section 312.

320* ⟨ Check the "constant" values for consistency 14 ⟩ +≡
if *cs_token_flag* + *eqtb_size* + *hash_extra* > *max_halfword* **then** *bad* ← 21;
if (*hash_offset* < 0) ∨ (*hash_offset* > *hash_base*) **then** *bad* ← 42;

330* Input stacks and states. This implementation of TeX uses two different conventions for representing sequential stacks.

1) If there is frequent access to the top entry, and if the stack is essentially never empty, then the top entry is kept in a global variable (even better would be a machine register), and the other entries appear in the array *stack*[0 .. (*ptr* − 1)]. For example, the semantic stack described above is handled this way, and so is the input stack that we are about to study.

2) If there is infrequent top access, the entire stack contents are in the array *stack*[0 .. (*ptr* − 1)]. For example, the *save_stack* is treated this way, as we have seen.

The state of TeX's input mechanism appears in the input stack, whose entries are records with six fields, called *state*, *index*, *start*, *loc*, *limit*, and *name*. This stack is maintained with convention (1), so it is declared in the following way:

```
⟨ Types in the outer block 18 ⟩ +≡
  in_state_record = record state_field, index_field: quarterword;
    start_field, loc_field, limit_field, name_field: halfword;
    synctex_tag_field: integer;  { stack the tag of the current file }
    end;
```

331* ⟨ Global variables 13 ⟩ +≡

```
input_stack: ↑in_state_record;
input_ptr: 0 .. stack_size;  { first unused location of input_stack }
max_in_stack: 0 .. stack_size;  { largest value of input_ptr when pushing }
cur_input: in_state_record;  { the "top" input state, according to convention (1) }
```

332* We've already defined the special variable *loc* ≡ *cur_input.loc_field* in our discussion of basic input-output routines. The other components of *cur_input* are defined in the same way:

```
define state ≡ cur_input.state_field  { current scanner state }
define index ≡ cur_input.index_field  { reference for buffer information }
define start ≡ cur_input.start_field  { starting position in buffer }
define limit ≡ cur_input.limit_field  { end of current line in buffer }
define name ≡ cur_input.name_field  { name of the current file }
define synctex_tag ≡ cur_input.synctex_tag_field  { SyncTeX tag of the current file }
```

334* Additional information about the current line is available via the *index* variable, which counts how many lines of characters are present in the buffer below the current level. We have $index = 0$ when reading from the terminal and prompting the user for each line; then if the user types, e.g., '`\input paper`', we will have $index = 1$ while reading the file `paper.tex`. However, it does not follow that *index* is the same as the input stack pointer, since many of the levels on the input stack may come from token lists. For example, the instruction '`\input paper`' might occur in a token list.

The global variable *in_open* is equal to the *index* value of the highest non-token-list level. Thus, the number of partially read lines in the buffer is $in\_open + 1$, and we have $in\_open = index$ when we are not reading a token list.

If we are not currently reading from the terminal, or from an input stream, we are reading from the file variable *input_file*[*index*]. We use the notation *terminal_input* as a convenient abbreviation for $name = 0$, and *cur_file* as an abbreviation for *input_file*[*index*].

The global variable *line* contains the line number in the topmost open file, for use in error messages. If we are not reading from the terminal, *line_stack*[*index*] holds the line number for the enclosing level, so that *line* can be restored when the current file has been read. Line numbers should never be negative, since the negative of the current line number is used to identify the user's output routine in the *mode_line* field of the semantic nest entries.

If more information about the input state is needed, it can be included in small arrays like those shown here. For example, the current page or segment number in the input file might be put into a variable *page*, maintained for enclosing levels in '*page_stack*: **array** [1 .. *max_in_open*] **of** *integer*' by analogy with *line_stack*.

define *terminal_input* ≡ (*name* = 0) { are we reading from the terminal? }
define *cur_file* ≡ *input_file*[*index*] { the current *alpha_file* variable }

⟨ Global variables 13 ⟩ +≡
in_open: 0 .. *max_in_open*; { the number of lines in the buffer, less one }
open_parens: 0 .. *max_in_open*; { the number of open text files }
input_file: ↑*unicode_file*;
line: *integer*; { current line number in the current source file }
line_stack: ↑*integer*;
source_filename_stack: ↑*str_number*;
full_source_filename_stack: ↑*str_number*;

336*. Here is a procedure that uses *scanner_status* to print a warning message when a subfile has ended, and at certain other crucial times:

```
⟨ Declare the procedure called runaway 336* ⟩ ≡
procedure runaway;
  var p: pointer; { head of runaway list }
  begin if scanner_status > skipping then
    begin case scanner_status of
    defining: begin print_nl("Runaway␣definition"); p ← def_ref;
      end;
    matching: begin print_nl("Runaway␣argument"); p ← temp_head;
      end;
    aligning: begin print_nl("Runaway␣preamble"); p ← hold_head;
      end;
    absorbing: begin print_nl("Runaway␣text"); p ← def_ref;
      end;
    end; { there are no other cases }
    print_char("?"); print_ln; show_token_list(link(p), null, error_line − 10);
    end;
  end;
```

This code is used in section 141.

338*. The *param_stack* is an auxiliary array used to hold pointers to the token lists for parameters at the current level and subsidiary levels of input. This stack is maintained with convention (2), and it grows at a different rate from the others.

```
⟨ Global variables 13 ⟩ +≡
param_stack: ↑pointer; { token list pointers for parameters }
param_ptr: 0 .. param_size; { first unused entry in param_stack }
max_param_stack: integer; { largest value of param_ptr, will be ≤ param_size + 9 }
```

358*. The *begin_file_reading* procedure starts a new level of input for lines of characters to be read from a file, or as an insertion from the terminal. It does not take care of opening the file, nor does it set *loc* or *limit* or *line*.

```
procedure begin_file_reading;
  begin if in_open = max_in_open then overflow("text␣input␣levels", max_in_open);
  if first = buf_size then overflow("buffer␣size", buf_size);
  incr(in_open); push_input; index ← in_open; source_filename_stack[index] ← 0;
  full_source_filename_stack[index] ← 0; eof_seen[index] ← false; grp_stack[index] ← cur_boundary;
  if_stack[index] ← cond_ptr; line_stack[index] ← line; start ← first; state ← mid_line; name ← 0;
      { terminal_input is now true }
  ⟨ Prepare terminal input SyncTeX information 1717* ⟩;
  end;
```

361*. To get TeX's whole input mechanism going, we perform the following actions.

```
⟨ Initialize the input routines 361* ⟩ ≡
  begin input_ptr ← 0; max_in_stack ← 0; source_filename_stack[0] ← 0;
  full_source_filename_stack[0] ← 0; in_open ← 0; open_parens ← 0; max_buf_stack ← 0; grp_stack[0] ← 0;
  if_stack[0] ← null; param_ptr ← 0; max_param_stack ← 0; first ← buf_size;
  repeat buffer[first] ← 0; decr(first);
  until first = 0;
  scanner_status ← normal; warning_index ← null; first ← 1; state ← new_line; start ← 1; index ← 0;
  line ← 0; name ← 0; force_eof ← false; align_state ← 1000000;
  if ¬init_terminal then goto final_end;
  limit ← last; first ← last + 1;  { init_terminal has set loc and last }
  end
```

This code is used in section 1391*.

368* ⟨ Tell the user what has run away and try to recover 368* ⟩ ≡

```
begin runaway; { print a definition, argument, or preamble }
if cur_cs = 0 then print_err("File␣ended")
else begin cur_cs ← 0; print_err("Forbidden␣control␣sequence␣found");
  end;
⟨ Print either 'definition' or 'use' or 'preamble' or 'text', and insert tokens that should lead to
    recovery 369* ⟩;
print("␣of␣"); sprint_cs(warning_index);
help4("I␣suspect␣you␣have␣forgotten␣a␣`}´,␣causing␣me")
("to␣read␣past␣where␣you␣wanted␣me␣to␣stop.")
("I´ll␣try␣to␣recover;␣but␣if␣the␣error␣is␣serious,")
("you´d␣better␣type␣`E´␣or␣`X´␣now␣and␣fix␣your␣file.");
error;
end
```

This code is used in section 366.

369* The recovery procedure can't be fully understood without knowing more about the TeX routines that should be aborted, but we can sketch the ideas here: For a runaway definition or a runaway balanced text we will insert a right brace; for a runaway preamble, we will insert a special `\cr` token and a right brace; and for a runaway argument, we will set *long_state* to *outer_call* and insert `\par`.

⟨ Print either 'definition' or 'use' or 'preamble' or 'text', and insert tokens that should lead to recovery 369* ⟩ ≡

```
p ← get_avail;
case scanner_status of
defining: begin print("␣while␣scanning␣definition"); info(p) ← right_brace_token + "}";
  end;
matching: begin print("␣while␣scanning␣use"); info(p) ← par_token; long_state ← outer_call;
  end;
aligning: begin print("␣while␣scanning␣preamble"); info(p) ← right_brace_token + "}"; q ← p;
  p ← get_avail; link(p) ← q; info(p) ← cs_token_flag + frozen_cr; align_state ← −1000000;
  end;
absorbing: begin print("␣while␣scanning␣text"); info(p) ← right_brace_token + "}";
  end;
end; { there are no other cases }
ins_list(p)
```

This code is used in section 368*.

396* Expanding the next token. Only a dozen or so command codes > *max_command* can possibly be returned by *get_next*; in increasing order, they are *undefined_cs*, *expand_after*, *no_expand*, *input*, *if_test*, *fi_or_else*, *cs_name*, *convert*, *the*, *top_bot_mark*, *call*, *long_call*, *outer_call*, *long_outer_call*, and *end_template*.

The *expand* subroutine is used when *cur_cmd* > *max_command*. It removes a "call" or a conditional or one of the other special operations just listed. It follows that *expand* might invoke itself recursively. In all cases, *expand* destroys the current token, but it sets things up so that the next *get_next* will deliver the appropriate next token. The value of *cur_tok* need not be known when *expand* is called.

Since several of the basic scanning routines communicate via global variables, their values are saved as local variables of *expand* so that recursive calls don't invalidate them.

```
⟨ Declare the procedure called macro_call 423 ⟩
⟨ Declare the procedure called insert_relax 413 ⟩
⟨ Declare ε-TeX procedures for expanding 1563 ⟩
procedure pass_text; forward;
procedure start_input; forward;
procedure conditional; forward;
procedure get_x_token; forward;
procedure conv_toks; forward;
procedure ins_the_toks; forward;
procedure expand;
  label reswitch;
  var t: halfword; { token that is being "expanded after" }
    b: boolean; { keep track of nested csnames }
    p, q, r: pointer; { for list manipulation }
    j: 0 .. buf_size; { index into buffer }
    cv_backup: integer; { to save the global quantity cur_val }
    cvl_backup, radix_backup, co_backup: small_number; { to save cur_val_level, etc. }
    backup_backup: pointer; { to save link(backup_head) }
    save_scanner_status: small_number; { temporary storage of scanner_status }
  begin incr(expand_depth_count);
  if expand_depth_count ≥ expand_depth then overflow("expansion␣depth", expand_depth);
  cv_backup ← cur_val; cvl_backup ← cur_val_level; radix_backup ← radix; co_backup ← cur_order;
  backup_backup ← link(backup_head);
reswitch: if cur_cmd < call then ⟨ Expand a nonmacro 399 ⟩
  else if cur_cmd < end_template then macro_call
    else ⟨ Insert a token containing frozen_endv 409 ⟩;
  cur_val ← cv_backup; cur_val_level ← cvl_backup; radix ← radix_backup; cur_order ← co_backup;
  link(backup_head) ← backup_backup; decr(expand_depth_count);
  end;
```

434*. If the parameter consists of a single group enclosed in braces, we must strip off the enclosing braces. That's why *rbrace_ptr* was introduced.

```
⟨ Tidy up the parameter just scanned, and tuck it away 434* ⟩ ≡
  begin if (m = 1) ∧ (info(p) < right_brace_limit) then
    begin link(rbrace_ptr) ← null; free_avail(p); p ← link(temp_head); pstack[n] ← link(p); free_avail(p);
    end
  else pstack[n] ← link(temp_head);
  incr(n);
  if tracing_macros > 0 then
    if (tracing_stack_levels = 0) ∨ (input_ptr < tracing_stack_levels) then
      begin begin_diagnostic; print_nl(match_chr); print_int(n); print("<-");
      show_token_list(pstack[n − 1], null, 1000); end_diagnostic(false);
      end;
  end
```

This code is used in section 426.

435*. ⟨ Show the text of the macro being expanded 435* ⟩ ≡

```
  begin begin_diagnostic;
  if tracing_stack_levels > 0 then
    if input_ptr < tracing_stack_levels then
      begin v ← input_ptr; print_ln; print_char("~");
      while v > 0 do
        begin print_char("."); decr(v);
        end;
      print_cs(warning_index); token_show(ref_count);
      end
    else begin print_char("~"); print_char("~"); print_cs(warning_index);
      end
  else begin print_ln; print_cs(warning_index); token_show(ref_count);
    end;
  end_diagnostic(false);
  end
```

This code is used in section 423.

536* ⟨ Either process `\ifcase` or set b to the value of a boolean condition 536* ⟩ ≡

```
case this_if of
if_char_code, if_cat_code: ⟨ Test if two characters match 541 ⟩;
if_int_code, if_dim_code: ⟨ Test relation between integers or dimensions 538 ⟩;
if_odd_code: ⟨ Test if an integer is odd 539 ⟩;
if_vmode_code: b ← (abs(mode) = vmode);
if_hmode_code: b ← (abs(mode) = hmode);
if_mmode_code: b ← (abs(mode) = mmode);
if_inner_code: b ← (mode < 0);
if_void_code, if_hbox_code, if_vbox_code: ⟨ Test box register status 540 ⟩;
ifx_code: ⟨ Test if two tokens match 542 ⟩;
if_eof_code: begin scan_four_bit_int_or_18;
  if cur_val = 18 then b ← ¬shellenabledp
  else b ← (read_open[cur_val] = closed);
  end;
if_true_code: b ← true;
if_false_code: b ← false;
  ⟨ Cases for conditional 1577 ⟩
if_case_code: ⟨ Select the appropriate case and return or goto common_ending 544 ⟩;
if_primitive_code: begin save_scanner_status ← scanner_status; scanner_status ← normal; get_next;
  scanner_status ← save_scanner_status;
  if cur_cs < hash_base then m ← prim_lookup(cur_cs − single_base)
  else m ← prim_lookup(text(cur_cs));
  b ← ((cur_cmd ≠ undefined_cs) ∧ (m ≠ undefined_primitive) ∧ (cur_cmd = prim_eq_type(m)) ∧ (cur_chr =
      prim_equiv(m)));
  end;
end { there are no other cases }
```

This code is used in section 533.

548* The file names we shall deal with have the following structure: If the name contains '`/`' or '`:`' (for Amiga only), the file area consists of all characters up to and including the final such character; otherwise the file area is null. If the remaining file name contains '`.`', the file extension consists of all such characters from the last '`.`' to the end, otherwise the file extension is null.

We can scan such file names easily by using two global variables that keep track of the occurrences of area and extension delimiters:

```
⟨ Global variables 13 ⟩ +≡
area_delimiter: pool_pointer; { the most recent '/', if any }
ext_delimiter: pool_pointer; { the most recent '.', if any }
file_name_quote_char: UTF16_code;
```

549* Input files that can't be found in the user's area may appear in a standard system area called *TEX_area*. Font metric files whose areas are not given explicitly are assumed to appear in a standard system area called *TEX_font_area*. These system area names will, of course, vary from place to place.

In C, the default paths are specified separately.

550* Here now is the first of the system-dependent routines for file name scanning.

```
procedure begin_name;
  begin area_delimiter ← 0; ext_delimiter ← 0; quoted_filename ← false; file_name_quote_char ← 0;
  end;
```

551* And here's the second. The string pool might change as the file name is being scanned, since a new `\csname` might be entered; therefore we keep *area_delimiter* and *ext_delimiter* relative to the beginning of the current string, instead of assigning an absolute address like *pool_ptr* to them.

```
function more_name(c : ASCII_code): boolean;
  begin if stop_at_space ∧ (c = " ") ∧ (file_name_quote_char = 0) then more_name ← false
  else if stop_at_space ∧ (file_name_quote_char ≠ 0) ∧ (c = file_name_quote_char) then
      begin file_name_quote_char ← 0; more_name ← true;
      end
    else if stop_at_space ∧ (file_name_quote_char = 0) ∧ ((c = """") ∨ (c = "´")) then
        begin file_name_quote_char ← c; quoted_filename ← true; more_name ← true;
        end
      else begin str_room(1); append_char(c); { contribute c to the current string }
        if IS_DIR_SEP(c) then
          begin area_delimiter ← cur_length; ext_delimiter ← 0;
          end
        else if c = "." then ext_delimiter ← cur_length;
        more_name ← true;
        end;
  end;
```

552* The third. If a string is already in the string pool, the function *slow_make_string* does not create a new string but returns this string number, thus saving string space. Because of this new property of the returned string number it is not possible to apply *flush_string* to these strings.

```
procedure end_name;
  var temp_str: str_number; { result of file name cache lookups }
    j: pool_pointer; { running index }
  begin if str_ptr + 3 > max_strings then overflow("number␣of␣strings", max_strings − init_str_ptr);
  if area_delimiter = 0 then cur_area ← ""
  else begin cur_area ← str_ptr; str_start_macro(str_ptr + 1) ← str_start_macro(str_ptr) + area_delimiter;
    incr(str_ptr); temp_str ← search_string(cur_area);
    if temp_str > 0 then
      begin cur_area ← temp_str; decr(str_ptr); { no flush_string, pool_ptr will be wrong! }
      for j ← str_start_macro(str_ptr + 1) to pool_ptr − 1 do
        begin str_pool[j − area_delimiter] ← str_pool[j];
        end;
      pool_ptr ← pool_ptr − area_delimiter; { update pool_ptr }
      end;
    end;
  if ext_delimiter = 0 then
    begin cur_ext ← ""; cur_name ← slow_make_string;
    end
  else begin cur_name ← str_ptr;
    str_start_macro(str_ptr + 1) ← str_start_macro(str_ptr) + ext_delimiter − area_delimiter − 1;
    incr(str_ptr); cur_ext ← make_string; decr(str_ptr); { undo extension string to look at name part }
    temp_str ← search_string(cur_name);
    if temp_str > 0 then
      begin cur_name ← temp_str; decr(str_ptr); { no flush_string, pool_ptr will be wrong! }
      for j ← str_start_macro(str_ptr + 1) to pool_ptr − 1 do
        begin str_pool[j − ext_delimiter + area_delimiter + 1] ← str_pool[j];
        end;
      pool_ptr ← pool_ptr − ext_delimiter + area_delimiter + 1; { update pool_ptr }
      end;
    cur_ext ← slow_make_string; { remake extension string }
    end;
  end;
```

553* Conversely, here is a routine that takes three strings and prints a file name that might have produced them. (The routine is system dependent, because some operating systems put the file area last instead of first.)

```
define check_quoted(#) ≡ { check if string # needs quoting }
    if # ≠ 0 then
      begin j ← str_start_macro(#);
      while ((¬must_quote) ∨ (quote_char = 0)) ∧ (j < str_start_macro(# + 1)) do
        begin if str_pool[j] = " " then must_quote ← true
        else if (str_pool[j] = """") ∨ (str_pool[j] = "'") then
            begin must_quote ← true; quote_char ← """" + "'" − str_pool[j];
            end;
        incr(j);
        end;
      end
define print_quoted(#) ≡ { print string #, omitting quotes }
    if # ≠ 0 then
      for j ← str_start_macro(#) to str_start_macro(# + 1) − 1 do
        begin if str_pool[j] = quote_char then
          begin print(quote_char); quote_char ← """" + "'" − quote_char; print(quote_char);
          end;
        print(str_pool[j]);
        end
```

⟨ Basic printing procedures 57 ⟩ +≡

```
procedure print_file_name(n, a, e : integer);
  var must_quote: boolean; { whether to quote the filename }
    quote_char: integer; { current quote char (single or double) }
    j: pool_pointer; { index into str_pool }
  begin must_quote ← false; quote_char ← 0; check_quoted(a); check_quoted(n); check_quoted(e);
  if must_quote then
    begin if quote_char = 0 then quote_char ← """";
    print_char(quote_char);
    end;
  print_quoted(a); print_quoted(n); print_quoted(e);
  if quote_char ≠ 0 then print_char(quote_char);
  end;
```

554* Another system-dependent routine is needed to convert three internal TEX strings into the *name_of_file* value that is used to open files. The present code allows both lowercase and uppercase letters in the file name.

```
define append_to_name(#) ≡
    begin c ← #; incr(k);
    if k ≤ file_name_size then
      begin if (c < 128) then name_of_file[k] ← c
      else if (c < "800) then
          begin name_of_file[k] ← "C0 + c div "40; incr(k); name_of_file[k] ← "80 + c mod "40;
          end
        else begin name_of_file[k] ← "E0 + c div "1000; incr(k);
          name_of_file[k] ← "80 + (c mod "1000) div "40; incr(k);
          name_of_file[k] ← "80 + (c mod "1000) mod "40;
          end
      end
    end
procedure pack_file_name(n, a, e : str_number);
  var k: integer; { number of positions filled in name_of_file }
    c: ASCII_code; { character being packed }
    j: pool_pointer; { index into str_pool }
  begin k ← 0;
  if name_of_file then libc_free(name_of_file);
  name_of_file ← xmalloc_array(UTF8_code, (length(a) + length(n) + length(e)) * 3 + 1);
  for j ← str_start_macro(a) to str_start_macro(a + 1) − 1 do append_to_name(so(str_pool[j]));
  for j ← str_start_macro(n) to str_start_macro(n + 1) − 1 do append_to_name(so(str_pool[j]));
  for j ← str_start_macro(e) to str_start_macro(e + 1) − 1 do append_to_name(so(str_pool[j]));
  if k ≤ file_name_size then name_length ← k else name_length ← file_name_size;
  name_of_file[name_length + 1] ← 0;
  end;
```

555* A messier routine is also needed, since format file names must be scanned before TEX's string mechanism has been initialized. We shall use the global variable *TEX_format_default* to supply the text for default system areas and extensions related to format files.

Under UNIX we don't give the area part, instead depending on the path searching that will happen during file opening. Also, the length will be set in the main program.

```
define format_area_length = 0 { length of its area part }
define format_ext_length = 4 { length of its '.fmt' part }
define format_extension = ".fmt" { the extension, as a WEB constant }
⟨ Global variables 13 ⟩ +≡
format_default_length: integer;
TEX_format_default: cstring;
```

556* We set the name of the default format file and the length of that name in C, instead of Pascal, since we want them to depend on the name of the program.

558* Here is the messy routine that was just mentioned. It sets *name_of_file* from the first n characters of *TEX_format_default*, followed by *buffer*[a .. b], followed by the last *format_ext_length* characters of *TEX_format_default*.

We dare not give error messages here, since TeX calls this routine before the *error* routine is ready to roll. Instead, we simply drop excess characters, since the error will be detected in another way when a strange file name isn't found.

```
procedure pack_buffered_name(n : small_number; a, b : integer);
  var k: integer;  { number of positions filled in name_of_file }
    c: ASCII_code;  { character being packed }
    j: integer;  { index into buffer or TEX_format_default }
  begin if n + b − a + 1 + format_ext_length > file_name_size then
    b ← a + file_name_size − n − 1 − format_ext_length;
  k ← 0;
  if name_of_file then libc_free(name_of_file);
  name_of_file ← xmalloc_array(UTF8_code, n + (b − a + 1) + format_ext_length + 1);
  for j ← 1 to n do append_to_name(TEX_format_default[j]);
  for j ← a to b do append_to_name(buffer[j]);
  for j ← format_default_length − format_ext_length + 1 to format_default_length do
    append_to_name(TEX_format_default[j]);
  if k ≤ file_name_size then name_length ← k else name_length ← file_name_size;
  name_of_file[name_length + 1] ← 0;
  end;
```

559* Here is the only place we use *pack_buffered_name*. This part of the program becomes active when a "virgin" TeX is trying to get going, just after the preliminary initialization, or when the user is substituting another format file by typing '`&`' after the initial '`**`' prompt. The buffer contains the first line of input in *buffer*[*loc* .. (*last* − 1)], where *loc* < *last* and *buffer*[*loc*] ≠ "␣".

```
⟨ Declare the function called open_fmt_file 559* ⟩ ≡
function open_fmt_file: boolean;
  label found, exit;
  var j: 0 .. buf_size;  { the first space after the format file name }
  begin j ← loc;
  if buffer[loc] = "&" then
    begin incr(loc); j ← loc; buffer[last] ← "␣";
    while buffer[j] ≠ "␣" do incr(j);
    pack_buffered_name(0, loc, j − 1);  { Kpathsea does everything }
    if w_open_in(fmt_file) then goto found;
    wake_up_terminal; wterm('Sorry,␣I␣can''t␣find␣the␣format␣'');
    fputs(stringcast(name_of_file + 1), stdout); wterm('''';␣will␣try␣'');
    fputs(TEX_format_default + 1, stdout); wterm_ln('''.'); update_terminal;
    end;  { now pull out all the stops: try for the system plain file }
  pack_buffered_name(format_default_length − format_ext_length, 1, 0);
  if ¬w_open_in(fmt_file) then
    begin wake_up_terminal; wterm('I␣can''t␣find␣the␣format␣file␣'');
    fputs(TEX_format_default + 1, stdout); wterm_ln('''!'); open_fmt_file ← false; return;
    end;
found: loc ← j; open_fmt_file ← true;
exit: end;
```

This code is used in section 1357*.

560* Operating systems often make it possible to determine the exact name (and possible version number) of a file that has been opened. The following routine, which simply makes a TEX string from the value of *name_of_file*, should ideally be changed to deduce the full name of file f, which is the file most recently opened, if it is possible to do this in a Pascal program.

This routine might be called after string memory has overflowed, hence we dare not use '*str_room*'.

```
function make_name_string: str_number;
  var k: 0 .. file_name_size;  { index into name_of_file }
    save_area_delimiter, save_ext_delimiter: pool_pointer;
    save_name_in_progress, save_stop_at_space: boolean;
  begin if (pool_ptr + name_length > pool_size) ∨ (str_ptr = max_strings) ∨ (cur_length > 0) then
    make_name_string ← "?"
  else begin make_utf16_name;
    for k ← 0 to name_length16 − 1 do append_char(name_of_file16[k]);
    make_name_string ← make_string;  { At this point we also set cur_name, cur_ext, and cur_area to
        match the contents of name_of_file. }
    save_area_delimiter ← area_delimiter; save_ext_delimiter ← ext_delimiter;
    save_name_in_progress ← name_in_progress; save_stop_at_space ← stop_at_space;
    name_in_progress ← true; begin_name; stop_at_space ← false; k ← 0;
    while (k < name_length16) ∧ (more_name(name_of_file16[k])) do incr(k);
    stop_at_space ← save_stop_at_space; end_name; name_in_progress ← save_name_in_progress;
    area_delimiter ← save_area_delimiter; ext_delimiter ← save_ext_delimiter;
    end;
  end;
function u_make_name_string(var f : unicode_file): str_number;
  begin u_make_name_string ← make_name_string;
  end;
function a_make_name_string(var f : alpha_file): str_number;
  begin a_make_name_string ← make_name_string;
  end;
function b_make_name_string(var f : byte_file): str_number;
  begin b_make_name_string ← make_name_string;
  end;
function w_make_name_string(var f : word_file): str_number;
  begin w_make_name_string ← make_name_string;
  end;
```

561* Now let's consider the "driver" routines by which TeX deals with file names in a system-independent manner. First comes a procedure that looks for a file name in the input by calling *get_x_token* for the information.

```
procedure scan_file_name;
  label done;
  var save_warning_index: pointer;
  begin save_warning_index ← warning_index; warning_index ← cur_cs;
      { store cur_cs here to remember until later }
  ⟨ Get the next non-blank non-relax non-call token 438 ⟩;
      { here the program expands tokens and removes spaces and \relaxes from the input. The \relax
      removal follows LuaTeX"s implementation, and other cases of balanced text scanning. }
  back_input; { return the last token to be read by either code path }
  if cur_cmd = left_brace then scan_file_name_braced
  else begin name_in_progress ← true; begin_name; ⟨ Get the next non-blank non-call token 440 ⟩;
    loop begin if (cur_cmd > other_char) ∨ (cur_chr > biggest_char) then { not a character }
        begin back_input; goto done;
        end;
      if ¬more_name(cur_chr) then goto done;
      get_x_token;
      end;
    end;
done: end_name; name_in_progress ← false; warning_index ← save_warning_index;
      { restore warning_index }
  end;
```

565* If some trouble arises when TEX tries to open a file, the following routine calls upon the user to supply another file name. Parameter s is used in the error message to identify the type of file; parameter e is the default extension if none is given. Upon exit from the routine, variables *cur_name*, *cur_area*, *cur_ext*, and *name_of_file* are ready for another attempt at file opening.

```
procedure prompt_file_name(s, e : str_number);
  label done;
  var k: 0 .. buf_size;  { index into buffer }
    saved_cur_name: str_number;  { to catch empty terminal input }
    saved_cur_ext: str_number;  { to catch empty terminal input }
    saved_cur_area: str_number;  { to catch empty terminal input }
  begin if interaction = scroll_mode then wake_up_terminal;
  if s = "input␣file␣name" then print_err("I␣can´t␣find␣file␣`")
  else print_err("I␣can´t␣write␣on␣file␣`");
  print_file_name(cur_name, cur_area, cur_ext); print("´.");
  if (e = ".tex") ∨ (e = "") then show_context;
  print_ln; print_c_string(prompt_file_name_help_msg);
  if (e ≠ "") then
    begin print("';␣default␣file␣extension␣is␣`"); print(e); print("´");
    end;
  print(")"); print_ln; print_nl("Please␣type␣another␣"); print(s);
  if interaction < scroll_mode then fatal_error("***␣(job␣aborted,␣file␣error␣in␣nonstop␣mode)");
  saved_cur_name ← cur_name; saved_cur_ext ← cur_ext; saved_cur_area ← cur_area; clear_terminal;
  prompt_input(":␣"); ⟨ Scan file name in the buffer 566 ⟩;
  if (length(cur_name) = 0) ∧ (cur_ext = "") ∧ (cur_area = "") then
    begin cur_name ← saved_cur_name; cur_ext ← saved_cur_ext; cur_area ← saved_cur_area;
    end
  else if cur_ext = "" then cur_ext ← e;
  pack_cur_name;
  end;
```

567* Here's an example of how these conventions are used. Whenever it is time to ship out a box of stuff, we shall use the macro *ensure_dvi_open*.

```
  define log_name ≡ texmf_log_name
  define ensure_dvi_open ≡
          if output_file_name = 0 then
            begin if job_name = 0 then open_log_file;
            pack_job_name(output_file_extension);
            while ¬dvi_open_out(dvi_file) do
              prompt_file_name("file␣name␣for␣output", output_file_extension);
            output_file_name ← b_make_name_string(dvi_file);
            end
⟨ Global variables 13 ⟩ +≡
output_file_extension: str_number;
no_pdf_output: boolean;
dvi_file: byte_file;  { the device-independent output goes here }
output_file_name: str_number;  { full name of the output file }
log_name: str_number;  { full name of the log file }
```

569*. The *open_log_file* routine is used to open the transcript file and to help it catch up to what has previously been printed on the terminal.

```
procedure open_log_file;
  var old_setting: 0 .. max_selector; { previous selector setting }
    k: 0 .. buf_size; { index into months and buffer }
    l: 0 .. buf_size; { end of first input line }
    months: const_cstring;
  begin old_setting ← selector;
  if job_name = 0 then job_name ← get_job_name("texput");
  pack_job_name(".fls"); recorder_change_filename(stringcast(name_of_file + 1)); pack_job_name(".log");
  while ¬a_open_out(log_file) do ⟨ Try to get a different log file name 570 ⟩;
  log_name ← a_make_name_string(log_file); selector ← log_only; log_opened ← true;
  ⟨ Print the banner line, including the date and time 571* ⟩;
  if mltex_enabled_p then
    begin wlog_cr; wlog('MLTeX v2.2 enabled');
    end;
  input_stack[input_ptr] ← cur_input; { make sure bottom level is in memory }
  print_nl("**"); l ← input_stack[0].limit_field; { last position of first line }
  if buffer[l] = end_line_char then decr(l);
  for k ← 1 to l do print(buffer[k]);
  print_ln; { now the transcript file contains the first line of input }
  selector ← old_setting + 2; { log_only or term_and_log }
  end;
```

571*. ⟨ Print the banner line, including the date and time 571* ⟩ ≡

```
begin if src_specials_p ∨ file_line_error_style_p ∨ parse_first_line_p then wlog(banner_k)
else wlog(banner);
wlog(version_string); slow_print(format_ident); print("␣␣"); print_int(sys_day); print_char("␣");
months ← ´␣JANFEBMARAPRMAYJUNJULAUGSEPOCTNOVDEC´;
for k ← 3 * sys_month − 2 to 3 * sys_month do wlog(months[k]);
print_char("␣"); print_int(sys_year); print_char("␣"); print_two(sys_time div 60); print_char(":");
print_two(sys_time mod 60);
if eTeX_ex then
  begin ; wlog_cr; wlog(´entering␣extended␣mode´);
  end;
if shellenabledp then
  begin wlog_cr; wlog(´␣´);
  if restrictedshell then
    begin wlog(´restricted␣´);
    end;
  wlog(´\write18␣enabled.´)
  end;
if src_specials_p then
  begin wlog_cr; wlog(´␣Source␣specials␣enabled.´)
  end;
if file_line_error_style_p then
  begin wlog_cr; wlog(´␣file:line:error␣style␣messages␣enabled.´)
  end;
if parse_first_line_p then
  begin wlog_cr; wlog(´␣%&-line␣parsing␣enabled.´);
  end;
if translate_filename then
  begin wlog_cr; wlog(´␣(WARNING:␣translate-file␣"´); fputs(translate_filename, log_file);
  wlog(´"␣ignored)´);
  end;
end
```

This code is used in section 569*.

572* Let's turn now to the procedure that is used to initiate file reading when an '`\input`' command is being processed. Beware: For historic reasons, this code foolishly conserves a tiny bit of string pool space; but that can confuse the interactive '`E`' option.

```
procedure start_input; { TeX will \input something }
  label done;
  var temp_str: str_number; v: pointer; k: 0 .. file_name_size; { index into name_of_file16 }
  begin scan_file_name; { set cur_name to desired file name }
  pack_cur_name;
  loop begin begin_file_reading; { set up cur_file and new level of input }
    tex_input_type ← 1; { Tell open_input we are \input. }
      { Kpathsea tries all the various ways to get the file. }
    if kpse_in_name_ok(stringcast(name_of_file + 1)) ∧ u_open_in(cur_file, kpse_tex_format,
        XeTeX_default_input_mode, XeTeX_default_input_encoding) then
        { At this point name_of_file contains the actual name found, as a UTF8 string. We convert to
        UTF16, then extract the cur_area, cur_name, and cur_ext from it. }
      begin make_utf16_name; name_in_progress ← true; begin_name; stop_at_space ← false; k ← 0;
      while (k < name_length16) ∧ (more_name(name_of_file16[k])) do incr(k);
      stop_at_space ← true; end_name; name_in_progress ← false; goto done;
      end;
    end_file_reading; { remove the level that didn't work }
    prompt_file_name("input␣file␣name", "");
    end;
done: name ← a_make_name_string(cur_file); source_filename_stack[in_open] ← name;
  full_source_filename_stack[in_open] ← make_full_name_string;
  if name = str_ptr − 1 then { we can try to conserve string pool space now }
    begin temp_str ← search_string(name);
    if temp_str > 0 then
      begin name ← temp_str; flush_string;
      end;
    end;
  if job_name = 0 then
    begin job_name ← get_job_name(cur_name); open_log_file;
    end; { open_log_file doesn't show_context, so limit and loc needn't be set to meaningful values yet }
  if term_offset + length(full_source_filename_stack[in_open]) > max_print_line − 2 then print_ln
  else if (term_offset > 0) ∨ (file_offset > 0) then print_char("␣");
  print_char("("); incr(open_parens); slow_print(full_source_filename_stack[in_open]); update_terminal;
  if tracing_stack_levels > 0 then
    begin begin_diagnostic; print_ln; print_char("~"); v ← input_ptr − 1;
    if v < tracing_stack_levels then
      while v > 0 do
        begin print_char("."); decr(v);
        end
    else print_char("~");
    slow_print("INPUT␣"); slow_print(cur_name); slow_print(cur_ext); print_ln; end_diagnostic(false);
    end;
  state ← new_line; ⟨ Prepare new file SyncTeX information 1716* ⟩;
  ⟨ Read the first line of the new file 573 ⟩;
  end;
```

583* So that is what `TFM` files hold. Since TEX has to absorb such information about lots of fonts, it stores most of the data in a large array called *font_info*. Each item of *font_info* is a *memory_word*; the *fix_word* data gets converted into *scaled* entries, while everything else goes into words of type *four_quarters*.

When the user defines `\font\f`, say, TEX assigns an internal number to the user's font `\f`. Adding this number to *font_id_base* gives the *eqtb* location of a "frozen" control sequence that will always select the font.

⟨ Types in the outer block 18 ⟩ +≡

```
internal_font_number = integer;   { font in a char_node }
font_index = integer;   { index into font_info }
nine_bits = min_quarterword .. non_char;
```

584.* Here now is the (rather formidable) array of font arrays.

define *otgr_font_flag* = "FFFE
define *aat_font_flag* = "FFFF
define *is_aat_font*(#) ≡ (*font_area*[#] = *aat_font_flag*)
define *is_ot_font*(#) ≡ ((*font_area*[#] = *otgr_font_flag*) ∧ (*usingOpenType*(*font_layout_engine*[#])))
define *is_gr_font*(#) ≡ ((*font_area*[#] = *otgr_font_flag*) ∧ (*usingGraphite*(*font_layout_engine*[#])))
define *is_otgr_font*(#) ≡ (*font_area*[#] = *otgr_font_flag*)
define *is_native_font*(#) ≡ (*is_aat_font*(#) ∨ *is_otgr_font*(#)) { native fonts have *font_area* = 65534 or 65535, which would be a string containing an invalid Unicode character }
define
is_new_mathfont(#) ≡ ((*font_area*[#] = *otgr_font_flag*) ∧
(*isOpenTypeMathFont*(*font_layout_engine*[#])))
define *non_char* ≡ *qi*(*too_big_char*) { a *halfword* code that can't match a real character }
define *non_address* = 0 { a spurious *bchar_label* }

⟨ Global variables 13 ⟩ +≡
font_info: ↑*fmemory_word*; { the big collection of font data }
fmem_ptr: *font_index*; { first unused word of *font_info* }
font_ptr: *internal_font_number*; { largest internal font number in use }
font_check: ↑*four_quarters*; { check sum }
font_size: ↑*scaled*; { "at" size }
font_dsize: ↑*scaled*; { "design" size }
font_params: ↑*font_index*; { how many font parameters are present }
font_name: ↑*str_number*; { name of the font }
font_area: ↑*str_number*; { area of the font }
font_bc: ↑*UTF16_code*; { beginning (smallest) character code }
font_ec: ↑*UTF16_code*; { ending (largest) character code }
font_glue: ↑*pointer*; { glue specification for interword space, *null* if not allocated }
font_used: ↑*boolean*; { has a character from this font actually appeared in the output? }
hyphen_char: ↑*integer*; { current \hyphenchar values }
skew_char: ↑*integer*; { current \skewchar values }
bchar_label: ↑*font_index*;
{ start of *lig_kern* program for left boundary character, *non_address* if there is none }
font_bchar: ↑*nine_bits*; { boundary character, *non_char* if there is none }
font_false_bchar: ↑*nine_bits*; { *font_bchar* if it doesn't exist in the font, otherwise *non_char* }
font_layout_engine: ↑*void_pointer*; { either an CFDictionaryRef or a XeTeXLayoutEngine }
font_mapping: ↑*void_pointer*; { *TECkit_Converter* or 0 }
font_flags: ↑*char*; { flags: 0x01: *font_colored* 0x02: *font_vertical* }
font_letter_space: ↑*scaled*; { letterspacing to be applied to the font }
loaded_font_mapping: *void_pointer*; { used by *load_native_font* to return mapping, if any }
loaded_font_flags: *char*; { used by *load_native_font* to return flags }
loaded_font_letter_space: *scaled*;
loaded_font_design_size: *scaled*;
mapped_text: ↑*UTF16_code*; { scratch buffer used while applying font mappings }
xdv_buffer: ↑*char*; { scratch buffer used in generating XDV output }

585*. Besides the arrays just enumerated, we have directory arrays that make it easy to get at the individual entries in *font_info*. For example, the *char_info* data for character c in font f will be in $font\_info[char\_base[f]+c].qqqq$; and if w is the *width_index* part of this word (the *b0* field), the width of the character is $font\_info[width\_base[f]+w].sc$. (These formulas assume that *min_quarterword* has already been added to c and to w, since TeX stores its quarterwords that way.)

⟨ Global variables 13 ⟩ +≡
char_base: ↑*integer*; { base addresses for *char_info* }
width_base: ↑*integer*; { base addresses for widths }
height_base: ↑*integer*; { base addresses for heights }
depth_base: ↑*integer*; { base addresses for depths }
italic_base: ↑*integer*; { base addresses for italic corrections }
lig_kern_base: ↑*integer*; { base addresses for ligature/kerning programs }
kern_base: ↑*integer*; { base addresses for kerns }
exten_base: ↑*integer*; { base addresses for extensible recipes }
param_base: ↑*integer*; { base addresses for font parameters }

586*. ⟨ Set initial values of key variables 23* ⟩ +≡

587*. TeX always knows at least one font, namely the null font. It has no characters, and its seven parameters are all equal to zero.

⟨ Initialize table entries (done by `INITEX` only) 189 ⟩ +≡

589*. Of course we want to define macros that suppress the detail of how font information is actually packed, so that we don't have to write things like

$$font\_info[width\_base[f]+font\_info[char\_base[f]+c].qqqq.b0].sc$$

too often. The WEB definitions here make $char\_info(f)(c)$ the *four_quarters* word of font information corresponding to character c of font f. If q is such a word, $char\_width(f)(q)$ will be the character's width; hence the long formula above is at least abbreviated to

$$char\_width(f)(char\_info(f)(c)).$$

Usually, of course, we will fetch q first and look at several of its fields at the same time.

The italic correction of a character will be denoted by $char\_italic(f)(q)$, so it is analogous to $char\_width$. But we will get at the height and depth in a slightly different way, since we usually want to compute both height and depth if we want either one. The value of $height\_depth(q)$ will be the 8-bit quantity

$$b = height\_index \times 16 + depth\_index,$$

and if b is such a byte we will write $char\_height(f)(b)$ and $char\_depth(f)(b)$ for the height and depth of the character c for which $q = char\_info(f)(c)$. Got that?

The tag field will be called $char\_tag(q)$; the remainder byte will be called $rem\_byte(q)$, using a macro that we have already defined above.

Access to a character's *width*, *height*, *depth*, and *tag* fields is part of TeX's inner loop, so we want these macros to produce code that is as fast as possible under the circumstances.

MLTeX will assume that a character c exists iff either exists in the current font or a character substitution definition for this character was defined using `\charsubdef`. To avoid the distinction between these two cases, MLTeX introduces the notion "effective character" of an input character c. If c exists in the current font, the effective character of c is the character c itself. If it doesn't exist but a character substitution is defined, the effective character of c is the base character defined in the character substitution. If there is an effective character for a non-existing character c, the "virtual character" c will get appended to the horizontal lists.

The effective character is used within *char_info* to access appropriate character descriptions in the font. For example, when calculating the width of a box, MLTeX will use the metrics of the effective characters. For the case of a substitution, MLTeX uses the metrics of the base character, ignoring the metrics of the accent character.

If character substitutions are changed, it will be possible that a character c neither exists in a font nor there is a valid character substitution for c. To handle these cases *effective_char* should be called with its first argument set to *true* to ensure that it will still return an existing character in the font. If neither c nor the substituted base character in the current character substitution exists, *effective_char* will output a warning and return the character $font\_bc[f]$ (which is incorrect, but can not be changed within the current framework).

Sometimes character substitutions are unwanted, therefore the original definition of *char_info* can be used using the macro *orig_char_info*. Operations in which character substitutions should be avoided are, for example, loading a new font and checking the font metric information in this font, and character accesses in math mode.

define $char\_list\_exists(\#) \equiv (char\_sub\_code(\#) > hi(0))$
define $char\_list\_accent(\#) \equiv (ho(char\_sub\_code(\#)) \textbf{ div } 256)$
define $char\_list\_char(\#) \equiv (ho(char\_sub\_code(\#)) \textbf{ mod } 256)$
define $char\_info\_end(\#) \equiv \#\boxed{)}\ ].qqqq$
define $char\_info(\#) \equiv font\_info\ [\ char\_base[\#] + effective\_char\boxed{(}true, \#, char\_info\_end$
define $orig\_char\_info\_end(\#) \equiv \#\ ].qqqq$
define $orig\_char\_info(\#) \equiv font\_info\ [\ char\_base[\#] + orig\_char\_info\_end$

define *char_width_end*(#) ≡ #.*b0*] .*sc*
define *char_width*(#) ≡ *font_info* [*width_base*[#] + *char_width_end*
define *char_exists*(#) ≡ (#.*b0* > *min_quarterword*)
define *char_italic_end*(#) ≡ (*qo*(#.*b2*)) **div** 4] .*sc*
define *char_italic*(#) ≡ *font_info* [*italic_base*[#] + *char_italic_end*
define *height_depth*(#) ≡ *qo*(#.*b1*)
define *char_height_end*(#) ≡ (#) **div** 16] .*sc*
define *char_height*(#) ≡ *font_info* [*height_base*[#] + *char_height_end*
define *char_depth_end*(#) ≡ (#) **mod** 16] .*sc*
define *char_depth*(#) ≡ *font_info* [*depth_base*[#] + *char_depth_end*
define *char_tag*(#) ≡ ((*qo*(#.*b2*)) **mod** 4)

595* TEX checks the information of a TFM file for validity as the file is being read in, so that no further checks will be needed when typesetting is going on. The somewhat tedious subroutine that does this is called *read_font_info*. It has four parameters: the user font identifier u, the file name and area strings *nom* and *aire*, and the "at" size s. If s is negative, it's the negative of a scale factor to be applied to the design size; $s = -1000$ is the normal case. Otherwise s will be substituted for the design size; in this case, s must be positive and less than 2048 pt (i.e., it must be less than 2^{27} when considered as an integer).

The subroutine opens and closes a global file variable called *tfm_file*. It returns the value of the internal font number that was just loaded. If an error is detected, an error message is issued and no font information is stored; *null_font* is returned in this case.

```
define bad_tfm = 11 { label for read_font_info }
define abort ≡ goto bad_tfm { do this when the TFM data is wrong }
⟨ Declare additional functions for MLTEX 1694* ⟩
function read_font_info(u : pointer; nom, aire : str_number; s : scaled): internal_font_number;
        { input a TFM file }
  label done, bad_tfm, not_found;
  var k: font_index; { index into font_info }
    name_too_long: boolean; { nom or aire exceeds 255 bytes? }
    file_opened: boolean; { was tfm_file successfully opened? }
    lf, lh, bc, ec, nw, nh, nd, ni, nl, nk, ne, np: halfword; { sizes of subfiles }
    f: internal_font_number; { the new font's number }
    g: internal_font_number; { the number to return }
    a, b, c, d: eight_bits; { byte variables }
    qw: four_quarters; sw: scaled; { accumulators }
    bch_label: integer; { left boundary start location, or infinity }
    bchar: 0 .. 256; { boundary character, or 256 }
    z: scaled; { the design size or the "at" size }
    alpha: integer; beta: 1 .. 16; { auxiliary quantities used in fixed-point multiplication }
  begin g ← null_font;
  file_opened ← false; pack_file_name(nom, aire, cur_ext);
  if XeTeX_tracing_fonts_state > 0 then
    begin begin_diagnostic; print_nl("Requested␣font␣"""); print_c_string(stringcast(name_of_file + 1));
    print('"');
    if s < 0 then
      begin print("␣scaled␣"); print_int(−s);
      end
    else begin print("␣at␣"); print_scaled(s); print("pt");
      end;
    end_diagnostic(false);
    end;
  if quoted_filename then
    begin { quoted name, so try for a native font }
    g ← load_native_font(u, nom, aire, s);
    if g ≠ null_font then goto done;
    end; { it was an unquoted name, or not found as an installed font, so try for a TFM file }
  ⟨ Read and check the font data if file exists; abort if the TFM file is malformed; if there's no room for this
      font, say so and goto done; otherwise incr(font_ptr) and goto done 597 ⟩;
  if g ≠ null_font then goto done;
  if ¬quoted_filename then
    begin { we failed to find a TFM file, so try for a native font }
    g ← load_native_font(u, nom, aire, s);
    if g ≠ null_font then goto done
    end;
```

```
bad_tfm: if suppress_fontnotfound_error = 0 then
    begin ⟨ Report that the font won't be loaded 596* ⟩;
    end;
done: if file_opened then b_close(tfm_file);
  if XeTeX_tracing_fonts_state > 0 then
    begin if g = null_font then
      begin begin_diagnostic; print_nl("␣->␣font␣not␣found,␣using␣""nullfont""");
      end_diagnostic(false);
      end
    else if file_opened then
        begin begin_diagnostic; print_nl("␣->␣"); print_c_string(stringcast(name_of_file + 1));
        end_diagnostic(false);
        end;
    end;
  read_font_info ← g;
  end;
```

596* There are programs called `TFtoPL` and `PLtoTF` that convert between the `TFM` format and a symbolic property-list format that can be easily edited. These programs contain extensive diagnostic information, so TEX does not have to bother giving precise details about why it rejects a particular `TFM` file.

```
define start_font_error_message ≡ print_err("Font␣"); sprint_cs(u); print_char("=");
      if file_name_quote_char ≠ 0 then print_char(file_name_quote_char);
      print_file_name(nom, aire, cur_ext);
      if file_name_quote_char ≠ 0 then print_char(file_name_quote_char);
      if s ≥ 0 then
        begin print("␣at␣"); print_scaled(s); print("pt");
        end
      else if s ≠ −1000 then
          begin print("␣scaled␣"); print_int(−s);
          end
```

⟨ Report that the font won't be loaded 596* ⟩ ≡

```
  start_font_error_message;
  if file_opened then print("␣not␣loadable:␣Bad␣metric␣(TFM)␣file")
  else if name_too_long then print("␣not␣loadable:␣Metric␣(TFM)␣file␣name␣too␣long")
    else print("␣not␣loadable:␣Metric␣(TFM)␣file␣or␣installed␣font␣not␣found");
  help5("I␣wasn´t␣able␣to␣read␣the␣size␣data␣for␣this␣font,")
  ("so␣I␣will␣ignore␣the␣font␣specification.")
  ("[Wizards␣can␣fix␣TFM␣files␣using␣TFtoPL/PLtoTF.]")
  ("You␣might␣try␣inserting␣a␣different␣font␣spec;")
  ("e.g.,␣type␣`I\font<same␣font␣id>=<substitute␣font␣name>´."); error
```

This code is used in section 595*.

598* ⟨ Open *tfm_file* for input and **begin** 598* ⟩ ≡

```
  name_too_long ← (length(nom) > 255) ∨ (length(aire) > 255);
  if name_too_long then abort; { kpse_find_file will append the ".tfm", and avoid searching the disk
        before the font alias files as well. }
  pack_file_name(nom, aire, ""); check_for_tfm_font_mapping;
  if b_open_in(tfm_file) then
    begin file_opened ← true
```

This code is used in section 597.

599* Note: A malformed TFM file might be shorter than it claims to be; thus *eof*(*tfm_file*) might be true when *read_font_info* refers to *tfm_file*↑ or when it says *get*(*tfm_file*). If such circumstances cause system error messages, you will have to defeat them somehow, for example by defining *fget* to be '**begin** *get*(*tfm_file*); **if** *eof*(*tfm_file*) **then** *abort*; **end**'.

```
define fget ≡ tfm_temp ← getc(tfm_file)
define fbyte ≡ tfm_temp
define read_sixteen(#) ≡
        begin # ← fbyte;
        if # > 127 then abort;
        fget; # ← # * ´400 + fbyte;
        end
define store_four_quarters(#) ≡
        begin fget; a ← fbyte; qw.b0 ← qi(a); fget; b ← fbyte; qw.b1 ← qi(b); fget; c ← fbyte;
        qw.b2 ← qi(c); fget; d ← fbyte; qw.b3 ← qi(d); # ← qw;
        end
```

605* We want to make sure that there is no cycle of characters linked together by *list_tag* entries, since such a cycle would get TeX into an endless loop. If such a cycle exists, the routine here detects it when processing the largest character code in the cycle.

```
define check_byte_range(#) ≡
        begin if (# < bc) ∨ (# > ec) then abort
        end
define current_character_being_worked_on ≡ k + bc − fmem_ptr
```

⟨ Check for charlist cycle 605* ⟩ ≡

```
begin check_byte_range(d);
while d < current_character_being_worked_on do
  begin qw ← orig_char_info(f)(d);  { N.B.: not qi(d), since char_base[f] hasn't been adjusted yet }
  if char_tag(qw) ≠ list_tag then goto not_found;
  d ← qo(rem_byte(qw));  { next character on the list }
  end;
if d = current_character_being_worked_on then abort;  { yes, there's a cycle }
not_found: end
```

This code is used in section 604.

608* **define** *check_existence*(#) ≡
begin *check_byte_range*(#); *qw* ← *orig_char_info*(*f*)(#); { N.B.: not *qi*(#) }
if ¬*char_exists*(*qw*) **then** *abort*;
end

⟨ Read ligature/kern program 608* ⟩ ≡
bch_label ← ´77777; *bchar* ← 256;
if *nl* > 0 **then**
begin for *k* ← *lig_kern_base*[*f*] **to** *kern_base*[*f*] + *kern_base_offset* − 1 **do**
begin *store_four_quarters*(*font_info*[*k*].*qqqq*);
if *a* > 128 **then**
begin if 256 ∗ *c* + *d* ≥ *nl* **then** *abort*;
if *a* = 255 **then**
if *k* = *lig_kern_base*[*f*] **then** *bchar* ← *b*;
end
else begin if *b* ≠ *bchar* **then** *check_existence*(*b*);
if *c* < 128 **then** *check_existence*(*d*) { check ligature }
else if 256 ∗ (*c* − 128) + *d* ≥ *nk* **then** *abort*; { check kern }
if *a* < 128 **then**
if *k* − *lig_kern_base*[*f*] + *a* + 1 ≥ *nl* **then** *abort*;
end;
end;
if *a* = 255 **then** *bch_label* ← 256 ∗ *c* + *d*;
end;
for *k* ← *kern_base*[*f*] + *kern_base_offset* **to** *exten_base*[*f*] − 1 **do** *store_scaled*(*font_info*[*k*].*sc*);

This code is used in section 597.

610* We check to see that the TFM file doesn't end prematurely; but no error message is given for files having more than *lf* words.

⟨ Read font parameters 610* ⟩ ≡
begin for *k* ← 1 **to** *np* **do**
if *k* = 1 **then** { the *slant* parameter is a pure number }
begin *fget*; *sw* ← *fbyte*;
if *sw* > 127 **then** *sw* ← *sw* − 256;
fget; *sw* ← *sw* ∗ ´400 + *fbyte*; *fget*; *sw* ← *sw* ∗ ´400 + *fbyte*; *fget*;
font_info[*param_base*[*f*]].*sc* ← (*sw* ∗ ´20) + (*fbyte* **div** ´20);
end
else *store_scaled*(*font_info*[*param_base*[*f*] + *k* − 1].*sc*);
if *feof*(*tfm_file*) **then** *abort*;
for *k* ← *np* + 1 **to** 7 **do** *font_info*[*param_base*[*f*] + *k* − 1].*sc* ← 0;
end

This code is used in section 597.

611*. Now to wrap it up, we have checked all the necessary things about the TFM file, and all we need to do is put the finishing touches on the data for the new font.

```
define adjust(#) ≡ #[f] ← qo(#[f])  { correct for the excess min_quarterword that was added }
⟨ Make final adjustments and goto done 611* ⟩ ≡
  if np ≥ 7 then font_params[f] ← np else font_params[f] ← 7;
  hyphen_char[f] ← default_hyphen_char; skew_char[f] ← default_skew_char;
  if bch_label < nl then bchar_label[f] ← bch_label + lig_kern_base[f]
  else bchar_label[f] ← non_address;
  font_bchar[f] ← qi(bchar); font_false_bchar[f] ← qi(bchar);
  if bchar ≤ ec then
    if bchar ≥ bc then
      begin qw ← orig_char_info(f)(bchar);  { N.B.: not qi(bchar) }
      if char_exists(qw) then font_false_bchar[f] ← non_char;
      end;
  font_name[f] ← nom; font_area[f] ← aire; font_bc[f] ← bc; font_ec[f] ← ec; font_glue[f] ← null;
  adjust(char_base); adjust(width_base); adjust(lig_kern_base); adjust(kern_base); adjust(exten_base);
  decr(param_base[f]); fmem_ptr ← fmem_ptr + lf; font_ptr ← f; g ← f;
  font_mapping[f] ← load_tfm_font_mapping; goto done
```

This code is used in section 597.

616* When TEX wants to typeset a character that doesn't exist, the character node is not created; thus the output routine can assume that characters exist when it sees them. The following procedure prints a warning message unless the user has suppressed it.

```
⟨ Declare subroutines for new_character 616* ⟩ ≡
procedure print_ucs_code(n : UnicodeScalar);  { cf. print_hex }
  var k: 0 .. 22;  { index to current digit; we assume that 0 ≤ n < 16^22 }
  begin k ← 0; print("U+");  { prefix with U+ instead of " }
  repeat dig[k] ← n mod 16; n ← n div 16; incr(k);
  until n = 0;  { pad to at least 4 hex digits }
  while k < 4 do
    begin dig[k] ← 0; incr(k);
    end;
  print_the_digs(k);
  end;
procedure char_warning(f : internal_font_number; c : integer);
  var old_setting: integer;  { saved value of tracing_online }
  begin if tracing_lost_chars > 0 then
    begin old_setting ← tracing_online;
    if eTeX_ex ∧ (tracing_lost_chars > 1) then tracing_online ← 1;
    if tracing_lost_chars > 2 then print_err("Missing␣character:␣There␣is␣no␣")
    else begin begin_diagnostic; print_nl("Missing␣character:␣There␣is␣no␣")
      end;
    if c < "10000 then print_ASCII(c)
    else print_char(c);  { non-Plane 0 Unicodes can't be sent through print_ASCII }
    print("␣(");
    if is_native_font(f) then print_ucs_code(c)
    else print_hex(c);
    print(")"); print("␣in␣font␣"); slow_print(font_name[f]);
    if tracing_lost_chars < 3 then print_char("!");
    tracing_online ← old_setting;
    if tracing_lost_chars > 2 then
      begin help0; error;
      end
    else end_diagnostic(false);
    end;  { of tracing_lost_chars > 0 }
  end;  { of procedure }
```

See also section 744.

This code is used in section 1694*.

617* The subroutines for *new_character* have been moved.

618.* Here is a function that returns a pointer to a character node for a given character in a given font. If that character doesn't exist, *null* is returned instead.

This allows a character node to be used if there is an equivalent in the *char_sub_code* list.

```
function new_character(f : internal_font_number; c : ASCII_code): pointer;
  label exit;
  var p: pointer;  { newly allocated node }
    ec: quarterword;  { effective character of c }
  begin if is_native_font(f) then
    begin new_character ← new_native_character(f, c); return;
    end;
  ec ← effective_char(false, f, qi(c));
  if font_bc[f] ≤ qo(ec) then
    if font_ec[f] ≥ qo(ec) then
      if char_exists(orig_char_info(f)(ec)) then  { N.B.: not char_info }
        begin p ← get_avail; font(p) ← f; character(p) ← qi(c); new_character ← p; return;
        end;
  char_warning(f, c); new_character ← null;
exit: end;
```

628* Shipping pages out. After considering TeX's eyes and stomach, we come now to the bowels.

The *ship_out* procedure is given a pointer to a box; its mission is to describe that box in DVI form, outputting a "page" to *dvi_file*. The DVI coordinates $(h, v) = (0, 0)$ should correspond to the upper left corner of the box being shipped.

Since boxes can be inside of boxes inside of boxes, the main work of *ship_out* is done by two mutually recursive routines, *hlist_out* and *vlist_out*, which traverse the hlists and vlists inside of horizontal and vertical boxes.

As individual pages are being processed, we need to accumulate information about the entire set of pages, since such statistics must be reported in the postamble. The global variables *total_pages*, *max_v*, *max_h*, *max_push*, and *last_bop* are used to record this information.

The variable *doing_leaders* is *true* while leaders are being output. The variable *dead_cycles* contains the number of times an output routine has been initiated since the last *ship_out*.

A few additional global variables are also defined here for use in *vlist_out* and *hlist_out*. They could have been local variables, but that would waste stack space when boxes are deeply nested, since the values of these variables are not needed during recursive calls.

⟨ Global variables 13 ⟩ +≡
total_pages: *integer*; { the number of pages that have been shipped out }
max_v: *scaled*; { maximum height-plus-depth of pages shipped so far }
max_h: *scaled*; { maximum width of pages shipped so far }
max_push: *integer*; { deepest nesting of *push* commands encountered so far }
last_bop: *integer*; { location of previous *bop* in the DVI output }
dead_cycles: *integer*; { recent outputs that didn't ship anything out }
doing_leaders: *boolean*; { are we inside a leader box? }

{ character and font in current *char_node* }
c: *quarterword*;
f: *internal_font_number*;
rule_ht, *rule_dp*, *rule_wd*: *scaled*; { size of current rule being output }
g: *pointer*; { current glue specification }
lq, *lr*: *integer*; { quantities used in calculations for leaders }

631* Some systems may find it more efficient to make *dvi_buf* a **packed** array, since output of four bytes at once may be facilitated.

⟨ Global variables 13 ⟩ +≡
dvi_buf: ↑*eight_bits*; { buffer for DVI output }
half_buf: *integer*; { half of *dvi_buf_size* }
dvi_limit: *integer*; { end of the current half buffer }
dvi_ptr: *integer*; { the next available buffer address }
dvi_offset: *integer*; { *dvi_buf_size* times the number of times the output buffer has been fully emptied }
dvi_gone: *integer*; { the number of bytes already output to *dvi_file* }

633* The actual output of *dvi_buf*[*a* .. *b*] to *dvi_file* is performed by calling *write_dvi*(*a*, *b*). For best results, this procedure should be optimized to run as fast as possible on each particular system, since it is part of TeX's inner loop. It is safe to assume that *a* and *b* + 1 will both be multiples of 4 when *write_dvi*(*a*, *b*) is called; therefore it is possible on many machines to use efficient methods to pack four bytes per word and to output an array of words with one system call.

In C, we use a macro to call *fwrite* or *write* directly, writing all the bytes in one shot. Much better even than writing four bytes at a time.

634* To put a byte in the buffer without paying the cost of invoking a procedure each time, we use the macro *dvi_out*.

The length of *dvi_file* should not exceed ″7FFFFFFF; we set *cur_s* ← −2 to prevent further DVI output causing infinite recursion.

```
define dvi_out(#) ≡ begin dvi_buf[dvi_ptr] ← #; incr(dvi_ptr);
    if dvi_ptr = dvi_limit then dvi_swap;
    end
procedure dvi_swap; { outputs half of the buffer }
  begin if dvi_ptr > (″7FFFFFFF − dvi_offset) then
    begin cur_s ← −2; fatal_error("dvi␣length␣exceeds␣""7FFFFFFF");
    end;
  if dvi_limit = dvi_buf_size then
    begin write_dvi(0, half_buf − 1); dvi_limit ← half_buf; dvi_offset ← dvi_offset + dvi_buf_size;
    dvi_ptr ← 0;
    end
  else begin write_dvi(half_buf, dvi_buf_size − 1); dvi_limit ← dvi_buf_size;
    end;
  dvi_gone ← dvi_gone + half_buf;
  end;
```

635* Here is how we clean out the buffer when TeX is all through; *dvi_ptr* will be a multiple of 4.

```
⟨ Empty the last bytes out of dvi_buf 635* ⟩ ≡
  if dvi_limit = half_buf then write_dvi(half_buf, dvi_buf_size − 1);
  if dvi_ptr > (″7FFFFFFF − dvi_offset) then
    begin cur_s ← −2; fatal_error("dvi␣length␣exceeds␣""7FFFFFFF");
    end;
  if dvi_ptr > 0 then write_dvi(0, dvi_ptr − 1)
```

This code is used in section 680*.

638* Here's a procedure that outputs a font definition. Since TeX82 uses at most 256 different fonts per job, *fnt_def1* is always used as the command code.

```
procedure dvi_native_font_def(f : internal_font_number);
  var font_def_length, i: integer;
  begin dvi_out(define_native_font); dvi_four(f − font_base − 1); font_def_length ← make_font_def(f);
  for i ← 0 to font_def_length − 1 do dvi_out(xdv_buffer[i]);
  end;
procedure dvi_font_def(f : internal_font_number);
  var k: pool_pointer; { index into str_pool }
    l: integer; { length of name without mapping option }
    begin if is_native_font(f) then dvi_native_font_def(f)
    else begin if f ≤ 256 + font_base then
        begin dvi_out(fnt_def1); dvi_out(f − font_base − 1);
        end
      else begin dvi_out(fnt_def1 + 1); dvi_out((f − font_base − 1) div ′400);
        dvi_out((f − font_base − 1) mod ′400);
        end;
      dvi_out(qo(font_check[f].b0)); dvi_out(qo(font_check[f].b1)); dvi_out(qo(font_check[f].b2));
      dvi_out(qo(font_check[f].b3));
      dvi_four(font_size[f]); dvi_four(font_dsize[f]);
      dvi_out(length(font_area[f])); ⟨ Output the font name whose internal number is f 639 ⟩;
      end;
```

653* ⟨ Initialize variables as *ship_out* begins 653* ⟩ ≡

```
dvi_h ← 0; dvi_v ← 0; cur_h ← h_offset; dvi_f ← null_font;
⟨ Calculate page dimensions and margins 1428 ⟩;
ensure_dvi_open;
if total_pages = 0 then
  begin dvi_out(pre); dvi_out(id_byte); { output the preamble }
  dvi_four(25400000); dvi_four(473628672); { conversion ratio for sp }
  prepare_mag; dvi_four(mag); { magnification factor is frozen }
  if output_comment then
    begin l ← strlen(output_comment); dvi_out(l);
    for s ← 0 to l − 1 do dvi_out(output_comment[s]);
    end
  else begin { the default code is unchanged }
    old_setting ← selector; selector ← new_string; print("␣XeTeX␣output␣"); print_int(year);
    print_char("."); print_two(month); print_char("."); print_two(day); print_char(":");
    print_two(time div 60); print_two(time mod 60); selector ← old_setting; dvi_out(cur_length);
    for s ← str_start_macro(str_ptr) to pool_ptr − 1 do dvi_out(so(str_pool[s]));
    pool_ptr ← str_start_macro(str_ptr); { flush the current string }
    end;
  end
```

This code is used in section 678*.

655* The recursive procedures *hlist_out* and *vlist_out* each have local variables *save_h* and *save_v* to hold the values of *dvi_h* and *dvi_v* just before entering a new level of recursion. In effect, the values of *save_h* and *save_v* on TeX's run-time stack correspond to the values of h and v that a DVI-reading program will push onto its coordinate stack.

```
define move_past = 13 { go to this label when advancing past glue or a rule }
define fin_rule = 14 { go to this label to finish processing a rule }
define next_p = 15 { go to this label when finished with node p }
define check_next = 1236
define end_node_run = 1237
⟨ Declare procedures needed in hlist_out, vlist_out 1431 ⟩
procedure hlist_out; { output an hlist_node box }
  label reswitch, move_past, fin_rule, next_p, continue, found, check_next, end_node_run;
  var base_line: scaled; { the baseline coordinate for this box }
    left_edge: scaled; { the left coordinate for this box }
    save_h, save_v: scaled; { what dvi_h and dvi_v should pop to }
    this_box: pointer; { pointer to containing box }
    g_order: glue_ord; { applicable order of infinity for glue }
    g_sign: normal .. shrinking; { selects type of glue }
    p: pointer; { current position in the hlist }
    save_loc: integer; { DVI byte location upon entry }
    leader_box: pointer; { the leader box being replicated }
    leader_wd: scaled; { width of leader box being replicated }
    lx: scaled; { extra space between leader boxes }
    outer_doing_leaders: boolean; { were we doing leaders? }
    edge: scaled; { right edge of sub-box or leader space }
    prev_p: pointer; { one step behind p }
    len: integer; { length of scratch string for native word output }
    q, r: pointer; k, j: integer; glue_temp: real; { glue value before rounding }
    cur_glue: real; { glue seen so far }
    cur_g: scaled; { rounded equivalent of cur_glue times the glue ratio }
  begin cur_g ← 0; cur_glue ← float_constant(0); this_box ← temp_ptr; g_order ← glue_order(this_box);
  g_sign ← glue_sign(this_box);
  if XeTeX_interword_space_shaping_state > 1 then
    begin ⟨ Merge sequences of words using native fonts and inter-word spaces into single nodes 656 ⟩;
    end;
  p ← list_ptr(this_box); incr(cur_s);
  if cur_s > 0 then dvi_out(push);
  if cur_s > max_push then max_push ← cur_s;
  save_loc ← dvi_offset + dvi_ptr; base_line ← cur_v; prev_p ← this_box + list_offset;
  ⟨ Initialize hlist_out for mixed direction typesetting 1524 ⟩;
  left_edge ← cur_h; ⟨ Start hlist SyncTeX information record 1725* ⟩;
  while p ≠ null do ⟨ Output node p for hlist_out and move to the next node, maintaining the condition
      cur_v = base_line 658* ⟩;
  ⟨ Finish hlist SyncTeX information record 1726* ⟩;
  ⟨ Finish hlist_out for mixed direction typesetting 1525 ⟩;
  prune_movements(save_loc);
  if cur_s > 0 then dvi_pop(save_loc);
  decr(cur_s);
  end;
```

658* We ought to give special care to the efficiency of one part of *hlist_out*, since it belongs to TeX's inner loop. When a *char_node* is encountered, we save a little time by processing several nodes in succession until reaching a non-*char_node*. The program uses the fact that *set_char_0* = 0.

In MLTeX this part looks for the existence of a substitution definition for a character *c*, if *c* does not exist in the font, and create appropriate `DVI` commands. Former versions of MLTeX have spliced appropriate character, kern, and box nodes into the horizontal list. Because the user can change character substitutions or `\charsubdefmax` on the fly, we have to test a again for valid substitutions. (Additional it is necessary to be careful—if leaders are used the current hlist is normally traversed more than once!)

```
⟨ Output node p for hlist_out and move to the next node, maintaining the condition cur_v = base_line 658* ⟩ ≡
reswitch: if is_char_node(p) then
    begin synch_h; synch_v;
    repeat f ← font(p); c ← character(p);
      if (p ≠ lig_trick) ∧ (font_mapping[f] ≠ nil) then c ← apply_tfm_font_mapping(font_mapping[f], c);
      if f ≠ dvi_f then ⟨ Change font dvi_f to f 659* ⟩;
      if font_ec[f] ≥ qo(c) then
        if font_bc[f] ≤ qo(c) then
          if char_exists(orig_char_info(f)(c)) then { N.B.: not char_info }
            begin if c ≥ qi(128) then dvi_out(set1);
            dvi_out(qo(c));
            cur_h ← cur_h + char_width(f)(orig_char_info(f)(c)); goto continue;
            end;
      if mltex_enabled_p then ⟨ Output a substitution, goto continue if not possible 1695* ⟩;
    continue: prev_p ← link(prev_p); { N.B.: not prev_p ← p, p might be lig_trick }
      p ← link(p);
    until ¬is_char_node(p);
    ⟨ Record current point SyncTeX information 1728* ⟩;
    dvi_h ← cur_h;
    end
  else ⟨ Output the non-char_node p for hlist_out and move to the next node 660* ⟩
```

This code is used in section 655*.

659* ⟨ Change font *dvi_f* to *f* 659* ⟩ ≡

```
begin if ¬font_used[f] then
    begin dvi_font_def(f); font_used[f] ← true;
    end;
  if f ≤ 64 + font_base then dvi_out(f − font_base − 1 + fnt_num_0)
  else if f ≤ 256 + font_base then
      begin dvi_out(fnt1); dvi_out(f − font_base − 1);
      end
    else begin dvi_out(fnt1 + 1); dvi_out((f − font_base − 1) div ´400);
      dvi_out((f − font_base − 1) mod ´400);
      end;
  dvi_f ← f;
  end
```

This code is used in sections 658*, 1426, and 1430.

660* ⟨ Output the non-*char_node* p for *hlist_out* and move to the next node 660* ⟩ ≡

```
begin case type(p) of
hlist_node, vlist_node: ⟨ Output a box in an hlist 661* ⟩;
rule_node: begin rule_ht ← height(p); rule_dp ← depth(p); rule_wd ← width(p); goto fin_rule;
  end;
whatsit_node: ⟨ Output the whatsit node p in an hlist 1430 ⟩;
glue_node: ⟨ Move right or output leaders 663 ⟩;
margin_kern_node: begin cur_h ← cur_h + width(p);
  end;
kern_node: begin ⟨ Record kern_node SyncTeX information 1730* ⟩;
  cur_h ← cur_h + width(p);
  end;
math_node: begin ⟨ Record math_node SyncTeX information 1731* ⟩;
  ⟨ Handle a math node in hlist_out 1526 ⟩;
  end;
ligature_node: ⟨ Make node p look like a char_node and goto reswitch 692 ⟩;
  ⟨ Cases of hlist_out that arise in mixed direction text only 1530 ⟩
othercases do_nothing
endcases;
goto next_p;
fin_rule: ⟨ Output a rule in an hlist 662 ⟩;
move_past: begin cur_h ← cur_h + rule_wd;
  ⟨ Record horizontal rule_node or glue_node SyncTeX information 1729* ⟩;
  end;
next_p: prev_p ← p; p ← link(p);
  end
```

This code is used in section 658*.

661* ⟨ Output a box in an hlist 661* ⟩ ≡

```
if list_ptr(p) = null then
  begin ⟨ Record void list SyncTeX information 1727* ⟩;
  cur_h ← cur_h + width(p);
  end
else begin save_h ← dvi_h; save_v ← dvi_v; cur_v ← base_line + shift_amount(p);
      { shift the box down }
  temp_ptr ← p; edge ← cur_h + width(p);
  if cur_dir = right_to_left then cur_h ← edge;
  if type(p) = vlist_node then vlist_out else hlist_out;
  dvi_h ← save_h; dvi_v ← save_v; cur_h ← edge; cur_v ← base_line;
  end
```

This code is used in section 660*.

667*. The *vlist_out* routine is similar to *hlist_out*, but a bit simpler.

```
procedure vlist_out; { output a vlist_node box }
  label move_past, fin_rule, next_p;
  var left_edge: scaled; { the left coordinate for this box }
    top_edge: scaled; { the top coordinate for this box }
    save_h, save_v: scaled; { what dvi_h and dvi_v should pop to }
    this_box: pointer; { pointer to containing box }
    g_order: glue_ord; { applicable order of infinity for glue }
    g_sign: normal .. shrinking; { selects type of glue }
    p: pointer; { current position in the vlist }
    save_loc: integer; { DVI byte location upon entry }
    leader_box: pointer; { the leader box being replicated }
    leader_ht: scaled; { height of leader box being replicated }
    lx: scaled; { extra space between leader boxes }
    outer_doing_leaders: boolean; { were we doing leaders? }
    edge: scaled; { bottom boundary of leader space }
    glue_temp: real; { glue value before rounding }
    cur_glue: real; { glue seen so far }
    cur_g: scaled; { rounded equivalent of cur_glue times the glue ratio }
    upwards: boolean; { whether we're stacking upwards }
  begin cur_g ← 0; cur_glue ← float_constant(0); this_box ← temp_ptr; g_order ← glue_order(this_box);
  g_sign ← glue_sign(this_box); p ← list_ptr(this_box);
  upwards ← (subtype(this_box) = min_quarterword + 1); incr(cur_s);
  if cur_s > 0 then dvi_out(push);
  if cur_s > max_push then max_push ← cur_s;
  save_loc ← dvi_offset + dvi_ptr; left_edge ← cur_h; ⟨ Start vlist SyncTeX information record 1723* ⟩;
  if upwards then cur_v ← cur_v + depth(this_box)
  else cur_v ← cur_v − height(this_box);
  top_edge ← cur_v;
  while p ≠ null do ⟨ Output node p for vlist_out and move to the next node, maintaining the condition
      cur_h = left_edge 668 ⟩;
  ⟨ Finish vlist SyncTeX information record 1724* ⟩;
  prune_movements(save_loc);
  if cur_s > 0 then dvi_pop(save_loc);
  decr(cur_s);
  end;
```

670*. The *synch_v* here allows the DVI output to use one-byte commands for adjusting v in most cases, since the baselineskip distance will usually be constant.

```
⟨ Output a box in a vlist 670* ⟩ ≡
  if list_ptr(p) = null then
    begin if upwards then cur_v ← cur_v − depth(p)
    else cur_v ← cur_v + height(p);
    ⟨ Record void list SyncTeX information 1727* ⟩;
    if upwards then cur_v ← cur_v − height(p)
    else cur_v ← cur_v + depth(p);
    end
  else begin if upwards then cur_v ← cur_v − depth(p)
    else cur_v ← cur_v + height(p);
    synch_v; save_h ← dvi_h; save_v ← dvi_v;
    if cur_dir = right_to_left then cur_h ← left_edge − shift_amount(p)
    else cur_h ← left_edge + shift_amount(p); { shift the box right }
    temp_ptr ← p;
    if type(p) = vlist_node then vlist_out else hlist_out;
    dvi_h ← save_h; dvi_v ← save_v;
    if upwards then cur_v ← save_v − height(p)
    else cur_v ← save_v + depth(p);
    cur_h ← left_edge;
    end
```

This code is used in section 669.

676*. The *hlist_out* and *vlist_out* procedures are now complete, so we are ready for the *ship_out* routine that gets them started in the first place.

```
procedure ship_out(p : pointer); { output the box p }
  label done;
  var page_loc: integer; { location of the current bop }
    j, k: 0 .. 9; { indices to first ten count registers }
    s: pool_pointer; { index into str_pool }
    old_setting: 0 .. max_selector; { saved selector setting }
  begin ⟨ Start sheet SyncTeX information record 1721* ⟩;
  begin if job_name = 0 then open_log_file;
  if tracing_output > 0 then
    begin print_nl("");  print_ln;  print("Completed␣box␣being␣shipped␣out");
    end;
  if term_offset > max_print_line − 9 then print_ln
  else if (term_offset > 0) ∨ (file_offset > 0) then print_char("␣");
  print_char("[");  j ← 9;
  while (count(j) = 0) ∧ (j > 0) do decr(j);
  for k ← 0 to j do
    begin print_int(count(k));
    if k < j then print_char(".");
    end;
  update_terminal;
  if tracing_output > 0 then
    begin print_char("]");  begin_diagnostic;  show_box(p);  end_diagnostic(true);
    end;
  ⟨ Ship box p out 678* ⟩;
  if eTeX_ex then ⟨ Check for LR anomalies at the end of ship_out 1541 ⟩;
  if tracing_output ≤ 0 then print_char("]");
  dead_cycles ← 0;  update_terminal; { progress report }
  ⟨ Flush the box from memory, showing statistics if requested 677 ⟩;
  end; ⟨ Finish sheet SyncTeX information record 1722* ⟩;
  end;
```

678* ⟨ Ship box p out 678* ⟩ ≡

```
⟨ Update the values of max_h and max_v; but if the page is too large, goto done 679 ⟩;
⟨ Initialize variables as ship_out begins 653* ⟩;
page_loc ← dvi_offset + dvi_ptr; dvi_out(bop);
for k ← 0 to 9 do dvi_four(count(k));
dvi_four(last_bop); last_bop ← page_loc;  { generate a pagesize special at start of page }
old_setting ← selector; selector ← new_string; print("pdf:pagesize␣");
if (pdf_page_width > 0) ∧ (pdf_page_height > 0) then
  begin print("width"); print("␣"); print_scaled(pdf_page_width); print("pt"); print("␣");
  print("height"); print("␣"); print_scaled(pdf_page_height); print("pt");
  end
else print("default");
selector ← old_setting; dvi_out(xxx1); dvi_out(cur_length);
for s ← str_start_macro(str_ptr) to pool_ptr − 1 do dvi_out(so(str_pool[s]));
pool_ptr ← str_start_macro(str_ptr);  { erase the string }
cur_v ← height(p) + v_offset;  { does this need changing for upwards mode ???? }
temp_ptr ← p;
if type(p) = vlist_node then vlist_out else hlist_out;
dvi_out(eop); incr(total_pages); cur_s ← −1;
if ¬no_pdf_output then fflush(dvi_file);
ifdef('IPC')
  if ipc_on > 0 then
    begin if dvi_limit = half_buf then
      begin write_dvi(half_buf, dvi_buf_size − 1); flush_dvi; dvi_gone ← dvi_gone + half_buf;
      end;
    if dvi_ptr > ("7FFFFFFF − dvi_offset) then
      begin cur_s ← −2; fatal_error("dvi␣length␣exceeds␣""7FFFFFFF");
      end;
    if dvi_ptr > 0 then
      begin write_dvi(0, dvi_ptr − 1); flush_dvi; dvi_offset ← dvi_offset + dvi_ptr;
      dvi_gone ← dvi_gone + dvi_ptr;
      end;
    dvi_ptr ← 0; dvi_limit ← dvi_buf_size; ipc_page(dvi_gone);
    end;
endif('IPC');
done:
```

This code is used in section 676*.

680*. At the end of the program, we must finish things off by writing the postamble. If *total_pages* = 0, the DVI file was never opened. If *total_pages* ≥ 65536, the DVI file will lie. And if *max_push* ≥ 65536, the user deserves whatever chaos might ensue.

An integer variable k will be declared for use by this routine.

```
⟨ Finish the DVI file 680* ⟩ ≡
  while cur_s > −1 do
    begin if cur_s > 0 then dvi_out(pop)
    else begin dvi_out(eop); incr(total_pages);
      end;
    decr(cur_s);
    end;
  if total_pages = 0 then print_nl("No␣pages␣of␣output.")
  else if cur_s ≠ −2 then
      begin dvi_out(post); { beginning of the postamble }
      dvi_four(last_bop); last_bop ← dvi_offset + dvi_ptr − 5; { post location }
      dvi_four(25400000); dvi_four(473628672); { conversion ratio for sp }
      prepare_mag; dvi_four(mag); { magnification factor }
      dvi_four(max_v); dvi_four(max_h);
      dvi_out(max_push div 256); dvi_out(max_push mod 256);
      dvi_out((total_pages div 256) mod 256); dvi_out(total_pages mod 256);
      ⟨ Output the font definitions for all fonts that were used 681 ⟩;
      dvi_out(post_post); dvi_four(last_bop); dvi_out(id_byte);
      ifdef('IPC')k ← 7 − ((3 + dvi_offset + dvi_ptr) mod 4); { the number of 223's }
      endif('IPC')ifndef('IPC')k ← 4 + ((dvi_buf_size − dvi_ptr) mod 4); { the number of 223's }
      endifn('IPC')
        while k > 0 do
          begin dvi_out(223); decr(k);
          end;
      ⟨ Empty the last bytes out of dvi_buf 635* ⟩;
      k ← dvi_close(dvi_file);
      if k = 0 then
        begin print_nl("Output␣written␣on␣"); print(output_file_name); print("␣(");
        print_int(total_pages);
        if total_pages ≠ 1 then print("␣pages")
        else print("␣page");
        if no_pdf_output then
          begin print(",␣"); print_int(dvi_offset + dvi_ptr); print("␣bytes).");
          end
        else print(").");
        end
      else begin print_nl("Error␣"); print_int(k); print("␣(");
        if no_pdf_output then print_c_string(strerror(k))
        else print("driver␣return␣code");
        print(")␣generating␣output;"); print_nl("file␣"); print(output_file_name);
        print("␣may␣not␣be␣valid."); history ← output_failure;
        end;
      end
```

This code is used in section 1387*.

751* ⟨ Look at the list of characters starting with x in font g; set f and c whenever a better character is found; **goto** *found* as soon as a large enough variant is encountered 751* ⟩ ≡

```
if is_ot_font(g) then
  begin x ← map_char_to_glyph(g, x); f ← g; c ← x; w ← 0; n ← 0;
  repeat y ← get_ot_math_variant(g, x, n, addressof(u), 0);
    if u > w then
      begin c ← y; w ← u;
      if u ≥ v then goto found;
      end;
    n ← n + 1;
  until u < 0; { if we get here, then we didn't find a big enough glyph; check if the char is extensible }
  ot_assembly_ptr ← get_ot_assembly_ptr(g, x, 0);
  if ot_assembly_ptr ≠ nil then goto found;
  end
else begin y ← x;
  if (qo(y) ≥ font_bc[g]) ∧ (qo(y) ≤ font_ec[g]) then
    begin continue: q ← orig_char_info(g)(y);
    if char_exists(q) then
      begin if char_tag(q) = ext_tag then
        begin f ← g; c ← y; goto found;
        end;
      hd ← height_depth(q); u ← char_height(g)(hd) + char_depth(g)(hd);
      if u > w then
        begin f ← g; c ← y; w ← u;
        if u ≥ v then goto found;
        end;
      if char_tag(q) = list_tag then
        begin y ← rem_byte(q); goto continue;
        end;
      end;
    end
```

This code is used in section 750.

764* Here we save memory space in a common case.

```
⟨ Simplify a trivial box 764* ⟩ ≡
  q ← list_ptr(x);
  if is_char_node(q) then
    begin r ← link(q);
    if r ≠ null then
      if link(r) = null then
        if ¬is_char_node(r) then
          if type(r) = kern_node then { unneeded italic correction }
            begin free_node(r, medium_node_size); link(q) ← null;
            end;
    end
```

This code is used in section 763.

765* It is convenient to have a procedure that converts a *math_char* field to an "unpacked" form. The *fetch* routine sets *cur_f*, *cur_c*, and *cur_i* to the font code, character code, and character information bytes of a given noad field. It also takes care of issuing error messages for nonexistent characters; in such cases, *char_exists*(*cur_i*) will be *false* after *fetch* has acted, and the field will also have been reset to *empty*.

```
procedure fetch(a : pointer); { unpack the math_char field a }
  begin cur_c ← cast_to_ushort(character(a)); cur_f ← fam_fnt(fam(a) + cur_size);
  cur_c ← cur_c + (plane_and_fam_field(a) div "100) * "10000;
  if cur_f = null_font then ⟨ Complain about an undefined family and set cur_i null 766 ⟩
  else if is_native_font(cur_f) then
      begin cur_i ← null_character;
      end
    else begin if (qo(cur_c) ≥ font_bc[cur_f]) ∧ (qo(cur_c) ≤ font_ec[cur_f]) then
        cur_i ← orig_char_info(cur_f)(cur_c)
      else cur_i ← null_character;
      if ¬(char_exists(cur_i)) then
        begin char_warning(cur_f, qo(cur_c)); math_type(a) ← empty; cur_i ← null_character;
        end;
      end;
  end;
```

784* ⟨ Switch to a larger accent if available and appropriate 784* ⟩ ≡

```
  loop begin if char_tag(i) ≠ list_tag then goto done;
    y ← rem_byte(i); i ← orig_char_info(f)(y);
    if ¬char_exists(i) then goto done;
    if char_width(f)(i) > w then goto done;
    c ← y;
    end;
done:
```

This code is used in section 781.

793* If the nucleus of an *op_noad* is a single character, it is to be centered vertically with respect to the axis, after first being enlarged (via a character list in the font) if we are in display style. The normal convention for placing displayed limits is to put them above and below the operator in display style.

The italic correction is removed from the character if there is a subscript and the limits are not being displayed. The *make_op* routine returns the value that should be used as an offset between subscript and superscript.

After *make_op* has acted, *subtype*(*q*) will be *limits* if and only if the limits have been set above and below the operator. In that case, *new_hlist*(*q*) will already contain the desired final box.

```
⟨ Declare math construction procedures 777 ⟩ +≡
function make_op(q : pointer): scaled;
  label found;
  var delta: scaled; { offset between subscript and superscript }
    p, v, x, y, z: pointer; { temporary registers for box construction }
    c: quarterword; i: four_quarters; { registers for character examination }
    shift_up, shift_down: scaled; { dimensions for box calculation }
    h1, h2: scaled; { height of original text-style symbol and possible replacement }
    n, g: integer; { potential variant index and glyph code }
    ot_assembly_ptr: void_pointer; save_f: internal_font_number;
  begin if (subtype(q) = normal) ∧ (cur_style < text_style) then subtype(q) ← limits;
  delta ← 0; ot_assembly_ptr ← nil;
  if math_type(nucleus(q)) = math_char then
    begin fetch(nucleus(q));
    if ¬is_ot_font(cur_f) then
      begin if (cur_style < text_style) ∧ (char_tag(cur_i) = list_tag) then { make it larger }
        begin c ← rem_byte(cur_i); i ← orig_char_info(cur_f)(c);
        if char_exists(i) then
          begin cur_c ← c; cur_i ← i; character(nucleus(q)) ← c;
          end;
        end;
      delta ← char_italic(cur_f)(cur_i);
      end;
    x ← clean_box(nucleus(q), cur_style);
    if is_new_mathfont(cur_f) then
      begin p ← list_ptr(x);
      if is_glyph_node(p) then
        begin if cur_style < text_style then
          begin { try to replace the operator glyph with a display-size variant, ensuring it is larger
              than the text size }
          h1 ← get_ot_math_constant(cur_f, displayOperatorMinHeight);
          if h1 < (height(p) + depth(p)) * 5/4 then h1 ← (height(p) + depth(p)) * 5/4;
          c ← native_glyph(p); n ← 0;
          repeat g ← get_ot_math_variant(cur_f, c, n, addressof(h2), 0);
            if h2 > 0 then
              begin native_glyph(p) ← g; set_native_glyph_metrics(p, 1);
              end;
            incr(n);
          until (h2 < 0) ∨ (h2 ≥ h1);
          if (h2 < 0) then
            begin
              { if we get here, then we didn't find a big enough glyph; check if the char is extensible }
            ot_assembly_ptr ← get_ot_assembly_ptr(cur_f, c, 0);
            if ot_assembly_ptr ≠ nil then
```

```
            begin free_node(p, glyph_node_size);
            p ← build_opentype_assembly(cur_f, ot_assembly_ptr, h1, 0); list_ptr(x) ← p; delta ← 0;
            goto found;
            end;
          end
        else set_native_glyph_metrics(p, 1);
        end;
      delta ← get_ot_math_ital_corr(cur_f, native_glyph(p));
    found: width(x) ← width(p); height(x) ← height(p); depth(x) ← depth(p);
    end
  end;
if (math_type(subscr(q)) ≠ empty) ∧ (subtype(q) ≠ limits) then width(x) ← width(x) − delta;
    { remove italic correction }
shift_amount(x) ← half(height(x) − depth(x)) − axis_height(cur_size); { center vertically }
math_type(nucleus(q)) ← sub_box; info(nucleus(q)) ← x;
end;
save_f ← cur_f;
if subtype(q) = limits then ⟨ Construct a box with limits above and below it, skewed by delta 794 ⟩;
free_ot_assembly(ot_assembly_ptr); make_op ← delta;
end;
```

964* **define** *wrap_lig*(#) ≡
  if *ligature_present* **then**
    begin $p \leftarrow$ *new_ligature*(*hf*, *cur_l*, *link*(*cur_q*));
    if *lft_hit* **then**
      begin *subtype*(*p*) ← 2; *lft_hit* ← *false*;
      end;
    if # **then**
      if *lig_stack* = *null* **then**
        begin *incr*(*subtype*(*p*)); *rt_hit* ← *false*;
        end;
    link(*cur_q*) ← *p*; *t* ← *p*; *ligature_present* ← *false*;
    end

define *pop_lig_stack* ≡
  begin if *lig_ptr*(*lig_stack*) > *null* **then**
    begin *link*(*t*) ← *lig_ptr*(*lig_stack*); { this is a charnode for *hu*[*j* + 1] }
    t ← *link*(*t*); *incr*(*j*);
    end;
  p ← *lig_stack*; *lig_stack* ← *link*(*p*); *free_node*(*p*, *small_node_size*);
  if *lig_stack* = *null* **then** *set_cur_r* **else** *cur_r* ← *character*(*lig_stack*);
  end { if *lig_stack* isn't *null* we have *cur_rh* = *non_char* }

⟨ Append a ligature and/or kern to the translation; **goto** *continue* if the stack of inserted ligatures is nonempty 964* ⟩ ≡
wrap_lig(*rt_hit*);
if *w* ≠ 0 **then**
  begin *link*(*t*) ← *new_kern*(*w*); *t* ← *link*(*t*); *w* ← 0; *sync_tag*(*t* + *medium_node_size*) ← 0;
    { *SyncTeX*: do nothing, it is too late }
  end;
if *lig_stack* > *null* **then**
  begin *cur_q* ← *t*; *cur_l* ← *character*(*lig_stack*); *ligature_present* ← *true*; *pop_lig_stack*;
  goto *continue*;
  end

This code is used in section 960.

974* The patterns are stored in a compact table that is also efficient for retrieval, using a variant of "trie memory" [cf. *The Art of Computer Programming* **3** (1973), 481–505]. We can find each pattern $p_1 \dots p_k$ by letting z_0 be one greater than the relevant language index and then, for $1 \le i \le k$, setting $z_i \leftarrow trie\_link(z_{i-1}) + p_i$; the pattern will be identified by the number z_k. Since all the pattern information is packed together into a single *trie_link* array, it is necessary to prevent confusion between the data from inequivalent patterns, so another table is provided such that $trie\_char(z_i) = p_i$ for all i. There is also a table $trie\_op(z_k)$ to identify the numbers $n_0 \dots n_k$ associated with $p_1 \dots p_k$.

The theory that comparatively few different number sequences $n_0 \dots n_k$ actually occur, since most of the n's are generally zero, seems to fail at least for the large German hyphenation patterns. Therefore the number sequences cannot any longer be encoded in such a way that $trie\_op(z_k)$ is only one byte long. We have introduced a new constant *max_trie_op* for the maximum allowable hyphenation operation code value; *max_trie_op* might be different for TeX and `INITEX` and must not exceed *max_halfword*. An opcode will occupy a halfword if *max_trie_op* exceeds *max_quarterword* or a quarterword otherwise. If $trie\_op(z_k) \neq min\_trie\_op$, when $p_1 \dots p_k$ has matched the letters in $hc[(l-k+1) \,..\, l]$ of language t, we perform all of the required operations for this pattern by carrying out the following little program: Set $v \leftarrow trie\_op(z_k)$. Then set $v \leftarrow v + op\_start[t]$, $hyf[l - hyf\_distance[v]] \leftarrow \max(hyf[l - hyf\_distance[v]], hyf\_num[v])$, and $v \leftarrow hyf\_next[v]$; repeat, if necessary, until $v = min\_trie\_op$.

⟨ Types in the outer block 18 ⟩ +≡
  trie_pointer = 0 .. *ssup_trie_size*; { an index into *trie* }
  trie_opcode = 0 .. *ssup_trie_opcode*; { a trie opcode }

975* For more than 255 trie op codes, the three fields *trie_link*, *trie_char*, and *trie_op* will no longer fit into one memory word; thus using web2c we define *trie* as three array instead of an array of records. The variant will be implemented by reusing the opcode field later on with another macro.

define *trie_link*(#) ≡ *trie_trl*[#] { "downward" link in a trie }
define *trie_char*(#) ≡ *trie_trc*[#] { character matched at this trie location }
define *trie_op*(#) ≡ *trie_tro*[#] { program for hyphenation at this trie location }

⟨ Global variables 13 ⟩ +≡
  { We will dynamically allocate these arrays. }
trie_trl: ↑*trie_pointer*; { *trie_link* }
trie_tro: ↑*trie_pointer*; { *trie_op* }
trie_trc: ↑*quarterword*; { *trie_char* }
hyf_distance: **array** [1 .. *trie_op_size*] **of** *small_number*; { position $k-j$ of n_j }
hyf_num: **array** [1 .. *trie_op_size*] **of** *small_number*; { value of n_j }
hyf_next: **array** [1 .. *trie_op_size*] **of** *trie_opcode*; { continuation code }
op_start: **array** [0 .. *biggest_lang*] **of** 0 .. *trie_op_size*; { offset for current language }

977*. Assuming that these auxiliary tables have been set up properly, the hyphenation algorithm is quite short. In the following code we set $hc[hn+2]$ to the impossible value 256, in order to guarantee that $hc[hn+3]$ will never be fetched.

```
⟨ Find hyphen locations for the word in hc, or return 977* ⟩ ≡
  for j ← 0 to hn do hyf[j] ← 0;
  ⟨ Look for the word hc[1 .. hn] in the exception table, and goto found (with hyf containing the hyphens)
      if an entry is found 984* ⟩;
  if trie_char(cur_lang + 1) ≠ qi(cur_lang) then return; { no patterns for cur_lang }
  hc[0] ← 0; hc[hn + 1] ← 0; hc[hn + 2] ← max_hyph_char; { insert delimiters }
  for j ← 0 to hn − r_hyf + 1 do
    begin z ← trie_link(cur_lang + 1) + hc[j]; l ← j;
    while hc[l] = qo(trie_char(z)) do
      begin if trie_op(z) ≠ min_trie_op then ⟨ Store maximum values in the hyf table 978* ⟩;
      incr(l); z ← trie_link(z) + hc[l];
      end;
    end;
found: for j ← 0 to l_hyf − 1 do hyf[j] ← 0;
  for j ← 0 to r_hyf − 1 do hyf[hn − j] ← 0
```

This code is used in section 944.

978*. ⟨ Store maximum values in the *hyf* table 978* ⟩ ≡

```
  begin v ← trie_op(z);
  repeat v ← v + op_start[cur_lang]; i ← l − hyf_distance[v];
    if hyf_num[v] > hyf[i] then hyf[i] ← hyf_num[v];
    v ← hyf_next[v];
  until v = min_trie_op;
  end
```

This code is used in section 977*.

979*. The exception table that is built by TEX's `\hyphenation` primitive is organized as an ordered hash table [cf. Amble and Knuth, *The Computer Journal* **17** (1974), 135–142] using linear probing. If α and β are words, we will say that $\alpha < \beta$ if $|\alpha| < |\beta|$ or if $|\alpha| = |\beta|$ and α is lexicographically smaller than β. (The notation $|\alpha|$ stands for the length of α.) The idea of ordered hashing is to arrange the table so that a given word α can be sought by computing a hash address $h = h(\alpha)$ and then looking in table positions h, $h-1$, ..., until encountering the first word $\leq \alpha$. If this word is different from α, we can conclude that α is not in the table. This is a clever scheme which saves the need for a hash link array. However, it is difficult to increase the size of the hyphen exception arrays. To make this easier, the ordered hash has been replaced by a simple hash, using an additional array *hyph_link*. The value 0 in *hyph_link*$[k]$ means that there are no more entries corresponding to the specific hash chain. When *hyph_link*$[k] > 0$, the next entry in the hash chain is *hyph_link*$[k] - 1$. This value is used because the arrays start at 0.

The words in the table point to lists in *mem* that specify hyphen positions in their *info* fields. The list for $c_1 \ldots c_n$ contains the number k if the word $c_1 \ldots c_n$ has a discretionary hyphen between c_k and c_{k+1}.

```
⟨ Types in the outer block 18 ⟩ +≡
  hyph_pointer = 0 .. ssup_hyph_size;
      { index into hyphen exceptions hash table; enlarging this requires changing (un)dump code }
```

980* ⟨ Global variables 13 ⟩ +≡
hyph_word: ↑*str_number*; { exception words }
hyph_list: ↑*pointer*; { lists of hyphen positions }
hyph_link: ↑*hyph_pointer*; { link array for hyphen exceptions hash table }
hyph_count: *integer*; { the number of words in the exception dictionary }
hyph_next: *integer*; { next free slot in hyphen exceptions hash table }

982* ⟨ Set initial values of key variables 23* ⟩ +≡
for $z \leftarrow 0$ **to** *hyph_size* **do**
begin *hyph_word*[z] ← 0; *hyph_list*[z] ← *null*; *hyph_link*[z] ← 0;
end;
hyph_count ← 0; *hyph_next* ← *hyph_prime* + 1;
if *hyph_next* > *hyph_size* **then** *hyph_next* ← *hyph_prime*;

984* First we compute the hash code h, then we search until we either find the word or we don't. Words from different languages are kept separate by appending the language code to the string.

⟨ Look for the word *hc*[1 .. *hn*] in the exception table, and **goto** *found* (with *hyf* containing the hyphens) if an entry is found 984* ⟩ ≡
$h \leftarrow hc[1]$; *incr*(*hn*); $hc[hn] \leftarrow$ *cur_lang*;
for $j \leftarrow 2$ **to** *hn* **do** $h \leftarrow (h + h + hc[j])$ **mod** *hyph_prime*;
loop begin ⟨ If the string *hyph_word*[h] is less than *hc*[1 .. *hn*], **goto** *not_found*; but if the two strings are equal, set *hyf* to the hyphen positions and **goto** *found* 985* ⟩;
$h \leftarrow$ *hyph_link*[h];
if $h = 0$ **then goto** *not_found*;
decr(h);
end;
not_found: *decr*(*hn*)
This code is used in section 977*.

985* ⟨ If the string *hyph_word*[h] is less than *hc*[1 .. *hn*], **goto** *not_found*; but if the two strings are equal, set *hyf* to the hyphen positions and **goto** *found* 985* ⟩ ≡
{ This is now a simple hash list, not an ordered one, so the module title is no longer descriptive. }
$k \leftarrow$ *hyph_word*[h];
if $k = 0$ **then goto** *not_found*;
if *length*(k) = *hn* **then**
begin $j \leftarrow 1$; $u \leftarrow$ *str_start_macro*(k);
repeat if *so*(*str_pool*[u]) ≠ *hc*[j] **then goto** *done*;
incr(j); *incr*(u);
until $j >$ *hn*;
⟨ Insert hyphens as specified in *hyph_list*[h] 986 ⟩;
decr(*hn*); **goto** *found*;
end;
done:
This code is used in section 984*.

988* We have now completed the hyphenation routine, so the *line_break* procedure is finished at last. Since the hyphenation exception table is fresh in our minds, it's a good time to deal with the routine that adds new entries to it.

When TeX has scanned '`\hyphenation`', it calls on a procedure named *new_hyph_exceptions* to do the right thing.

```
define set_cur_lang ≡
    if language ≤ 0 then cur_lang ← 0
    else if language > biggest_lang then cur_lang ← 0
      else cur_lang ← language
procedure new_hyph_exceptions; { enters new exceptions }
  label reswitch, exit, found, not_found, not_found1;
  var n: 0 .. hyphenatable_length_limit + 1; { length of current word; not always a small_number }
    j: 0 .. hyphenatable_length_limit + 1; { an index into hc }
    h: hyph_pointer; { an index into hyph_word and hyph_list }
    k: str_number; { an index into str_start }
    p: pointer; { head of a list of hyphen positions }
    q: pointer; { used when creating a new node for list p }
    s: str_number; { strings being compared or stored }
    u, v: pool_pointer; { indices into str_pool }
  begin scan_left_brace; { a left brace must follow \hyphenation }
  set_cur_lang;
  init if trie_not_ready then
    begin hyph_index ← 0; goto not_found1;
    end;
  tini
  set_hyph_index;
not_found1: ⟨ Enter as many hyphenation exceptions as are listed, until coming to a right brace; then
      return 989 ⟩;
exit: end;
```

993* ⟨ Enter a hyphenation exception 993* ⟩ ≡

```
begin incr(n); hc[n] ← cur_lang; str_room(n); h ← 0;
for j ← 1 to n do
  begin h ← (h + h + hc[j]) mod hyph_prime; append_char(hc[j]);
  end;
s ← make_string; ⟨ Insert the pair (s, p) into the exception table 994* ⟩;
end
```

This code is used in section 989.

994* ⟨ Insert the pair (s,p) into the exception table 994* ⟩ ≡

```
if hyph_next ≤ hyph_prime then
  while (hyph_next > 0) ∧ (hyph_word[hyph_next − 1] > 0) do decr(hyph_next);
if (hyph_count = hyph_size) ∨ (hyph_next = 0) then overflow("exception␣dictionary", hyph_size);
incr(hyph_count);
while hyph_word[h] ≠ 0 do
  begin ⟨ If the string hyph_word[h] is less than or equal to s, interchange (hyph_word[h], hyph_list[h])
      with (s, p) 995* ⟩;
  if hyph_link[h] = 0 then
    begin hyph_link[h] ← hyph_next;
    if hyph_next ≥ hyph_size then hyph_next ← hyph_prime;
    if hyph_next > hyph_prime then incr(hyph_next);
    end;
  h ← hyph_link[h] − 1;
  end;
found: hyph_word[h] ← s; hyph_list[h] ← p
```

This code is used in section 993*.

995* ⟨ If the string $hyph\_word[h]$ is less than or equal to s, interchange $(hyph\_word[h], hyph\_list[h])$ with (s,p) 995* ⟩ ≡

```
  { This is now a simple hash list, not an ordered one, so the module title is no longer descriptive. }
k ← hyph_word[h];
if length(k) ≠ length(s) then goto not_found;
u ← str_start_macro(k); v ← str_start_macro(s);
repeat if str_pool[u] ≠ str_pool[v] then goto not_found;
  incr(u); incr(v);
until u = str_start_macro(k + 1);  { repeat hyphenation exception; flushing old data }
flush_string; s ← hyph_word[h];  { avoid slow_make_string! }
decr(hyph_count);  { We could also flush_list(hyph_list[h]);, but it interferes with trip.log. }
goto found;
not_found:
```

This code is used in section 994*.

997* Before we discuss trie building in detail, let's consider the simpler problem of creating the *hyf_distance*, *hyf_num*, and *hyf_next* arrays.

Suppose, for example, that TEX reads the pattern '`ab2cde1`'. This is a pattern of length 5, with $n_0 \ldots n_5 = 0\,0\,2\,0\,0\,1$ in the notation above. We want the corresponding *trie_op* code v to have $hyf\_distance[v] = 3$, $hyf\_num[v] = 2$, and $hyf\_next[v] = v'$, where the auxiliary *trie_op* code v' has $hyf\_distance[v'] = 0$, $hyf\_num[v'] = 1$, and $hyf\_next[v'] = min\_trie\_op$.

TEX computes an appropriate value v with the *new_trie_op* subroutine below, by setting

$$v' \leftarrow new\_trie\_op(0, 1, min\_trie\_op), \qquad v \leftarrow new\_trie\_op(3, 2, v').$$

This subroutine looks up its three parameters in a special hash table, assigning a new value only if these three have not appeared before for the current language.

The hash table is called *trie_op_hash*, and the number of entries it contains is *trie_op_ptr*.

```
⟨ Global variables 13 ⟩ +≡
  init trie_op_hash: array [neg_trie_op_size .. trie_op_size] of 0 .. trie_op_size;
        { trie op codes for quadruples }
trie_used: array [0 .. biggest_lang] of trie_opcode;  { largest opcode used so far for this language }
trie_op_lang: array [1 .. trie_op_size] of 0 .. biggest_lang;  { language part of a hashed quadruple }
trie_op_val: array [1 .. trie_op_size] of trie_opcode;  { opcode corresponding to a hashed quadruple }
trie_op_ptr: 0 .. trie_op_size;  { number of stored ops so far }
  tini
max_op_used: trie_opcode;  { largest opcode used for any language }
small_op: boolean;  { flag used while dumping or undumping }
```

998* It's tempting to remove the *overflow* stops in the following procedure; *new_trie_op* could return *min_trie_op* (thereby simply ignoring part of a hyphenation pattern) instead of aborting the job. However, that would lead to different hyphenation results on different installations of TeX using the same patterns. The *overflow* stops are necessary for portability of patterns.

```
⟨ Declare procedures for preprocessing hyphenation patterns 998* ⟩ ≡
function new_trie_op(d, n : small_number; v : trie_opcode): trie_opcode;
  label exit;
  var h: neg_trie_op_size .. trie_op_size;  { trial hash location }
    u: trie_opcode;  { trial op code }
    l: 0 .. trie_op_size;  { pointer to stored data }
  begin h ← abs(n+313*d+361*v+1009*cur_lang) mod (trie_op_size − neg_trie_op_size)+neg_trie_op_size;
  loop begin l ← trie_op_hash[h];
    if l = 0 then  { empty position found for a new op }
      begin if trie_op_ptr = trie_op_size then overflow("pattern␣memory␣ops", trie_op_size);
      u ← trie_used[cur_lang];
      if u = max_trie_op then
        overflow("pattern␣memory␣ops␣per␣language", max_trie_op − min_trie_op);
      incr(trie_op_ptr); incr(u); trie_used[cur_lang] ← u;
      if u > max_op_used then max_op_used ← u;
      hyf_distance[trie_op_ptr] ← d; hyf_num[trie_op_ptr] ← n; hyf_next[trie_op_ptr] ← v;
      trie_op_lang[trie_op_ptr] ← cur_lang; trie_op_hash[h] ← trie_op_ptr; trie_op_val[trie_op_ptr] ← u;
      new_trie_op ← u; return;
      end;
    if (hyf_distance[l] = d) ∧ (hyf_num[l] = n) ∧ (hyf_next[l] = v) ∧ (trie_op_lang[l] = cur_lang) then
      begin new_trie_op ← trie_op_val[l]; return;
      end;
    if h > −trie_op_size then decr(h) else h ← trie_op_size;
    end;
exit: end;
```

See also sections 1002, 1003, 1007, 1011, 1013, 1014*, and 1020*.

This code is used in section 996.

999* After *new_trie_op* has compressed the necessary opcode information, plenty of information is available to unscramble the data into the final form needed by our hyphenation algorithm.

```
⟨ Sort the hyphenation op tables into proper order 999* ⟩ ≡
  op_start[0] ← −min_trie_op;
  for j ← 1 to biggest_lang do op_start[j] ← op_start[j − 1] + qo(trie_used[j − 1]);
  for j ← 1 to trie_op_ptr do trie_op_hash[j] ← op_start[trie_op_lang[j]] + trie_op_val[j];  { destination }
  for j ← 1 to trie_op_ptr do
    while trie_op_hash[j] > j do
      begin k ← trie_op_hash[j];
      t ← hyf_distance[k]; hyf_distance[k] ← hyf_distance[j]; hyf_distance[j] ← t;
      t ← hyf_num[k]; hyf_num[k] ← hyf_num[j]; hyf_num[j] ← t;
      t ← hyf_next[k]; hyf_next[k] ← hyf_next[j]; hyf_next[j] ← t;
      trie_op_hash[j] ← trie_op_hash[k]; trie_op_hash[k] ← k;
      end
```

This code is used in section 1006.

1000* Before we forget how to initialize the data structures that have been mentioned so far, let's write down the code that gets them started.

```
⟨ Initialize table entries (done by INITEX only) 189 ⟩ +≡
  for k ← −trie_op_size to trie_op_size do trie_op_hash[k] ← 0;
  for k ← 0 to biggest_lang do trie_used[k] ← min_trie_op;
  max_op_used ← min_trie_op; trie_op_ptr ← 0;
```

1001* The linked trie that is used to preprocess hyphenation patterns appears in several global arrays. Each node represents an instruction of the form "if you see character c, then perform operation o, move to the next character, and go to node l; otherwise go to node r." The four quantities c, o, l, and r are stored in four arrays *trie_c*, *trie_o*, *trie_l*, and *trie_r*. The root of the trie is *trie_l*[0], and the number of nodes is *trie_ptr*. Null trie pointers are represented by zero. To initialize the trie, we simply set *trie_l*[0] and *trie_ptr* to zero. We also set *trie_c*[0] to some arbitrary value, since the algorithm may access it.

The algorithms maintain the condition

$$trie\_c[trie\_r[z]] > trie\_c[z] \qquad \text{whenever } z \neq 0 \text{ and } trie\_r[z] \neq 0;$$

in other words, sibling nodes are ordered by their c fields.

```
define trie_root ≡ trie_l[0] { root of the linked trie }
⟨ Global variables 13 ⟩ +≡
  init trie_c: ↑packed_ASCII_code; { characters to match }
  trie_o: ↑trie_opcode; { operations to perform }
  trie_l: ↑trie_pointer; { left subtrie links }
  trie_r: ↑trie_pointer; { right subtrie links }
  trie_ptr: trie_pointer; { the number of nodes in the trie }
  trie_hash: ↑trie_pointer; { used to identify equivalent subtries }
  tini
```

1004* The compressed trie will be packed into the *trie* array using a "top-down first-fit" procedure. This is a little tricky, so the reader should pay close attention: The *trie_hash* array is cleared to zero again and renamed *trie_ref* for this phase of the operation; later on, *trie_ref*[p] will be nonzero only if the linked trie node p is the smallest character in a family and if the characters c of that family have been allocated to locations *trie_ref*[p] + c in the *trie* array. Locations of *trie* that are in use will have *trie_link* = 0, while the unused holes in *trie* will be doubly linked with *trie_link* pointing to the next larger vacant location and *trie_back* pointing to the next smaller one. This double linking will have been carried out only as far as *trie_max*, where *trie_max* is the largest index of *trie* that will be needed. To save time at the low end of the trie, we maintain array entries *trie_min*[c] pointing to the smallest hole that is greater than c. Another array *trie_taken* tells whether or not a given location is equal to *trie_ref*[p] for some p; this array is used to ensure that distinct nodes in the compressed trie will have distinct *trie_ref* entries.

```
define trie_ref ≡ trie_hash { where linked trie families go into trie }
define trie_back(#) ≡ trie_tro[#] { use the opcode field now for backward links }
⟨ Global variables 13 ⟩ +≡
  init trie_taken: ↑boolean; { does a family start here? }
  trie_min: array [ASCII_code] of trie_pointer; { the first possible slot for each character }
  trie_max: trie_pointer; { largest location used in trie }
  trie_not_ready: boolean; { is the trie still in linked form? }
  tini
```

1005* Each time `\patterns` appears, it contributes further patterns to the future trie, which will be built only when hyphenation is attempted or when a format file is dumped. The boolean variable *trie_not_ready* will change to *false* when the trie is compressed; this will disable further patterns.

```
⟨ Initialize table entries (done by INITEX only) 189 ⟩ +≡
  trie_not_ready ← true;
```

1012* When the whole trie has been allocated into the sequential table, we must go through it once again so that *trie* contains the correct information. Null pointers in the linked trie will be represented by the value 0, which properly implements an "empty" family.

```
  define clear_trie ≡ { clear trie[r] }
      begin trie_link(r) ← 0; trie_op(r) ← min_trie_op; trie_char(r) ← min_quarterword;
        { trie_char ← qi(0) }
      end
⟨ Move the data into trie 1012* ⟩ ≡
  if trie_max = 0 then { no patterns were given }
    begin for r ← 0 to max_hyph_char do clear_trie;
    trie_max ← max_hyph_char;
    end
  else begin if hyph_root > 0 then trie_fix(hyph_root);
    if trie_root > 0 then trie_fix(trie_root); { this fixes the non-holes in trie }
    r ← 0; { now we will zero out all the holes }
    repeat s ← trie_link(r); clear_trie; r ← s;
    until r > trie_max;
    end;
  trie_char(0) ← qi("?"); { make trie_char(c) ≠ c for all c }
```

This code is used in section 1020*.

1014* Now let's go back to the easier problem, of building the linked trie. When `INITEX` has scanned the '`\patterns`' control sequence, it calls on *new_patterns* to do the right thing.

```
⟨ Declare procedures for preprocessing hyphenation patterns 998* ⟩ +≡
procedure new_patterns; { initializes the hyphenation pattern data }
  label done, done1;
  var k, l: 0 .. hyphenatable_length_limit + 1;
      { indices into hc and hyf; not always in small_number range }
    digit_sensed: boolean; { should the next digit be treated as a letter? }
    v: trie_opcode; { trie op code }
    p, q: trie_pointer; { nodes of trie traversed during insertion }
    first_child: boolean; { is p = trie_l[q]? }
    c: ASCII_code; { character being inserted }
  begin if trie_not_ready then
    begin set_cur_lang; scan_left_brace; { a left brace must follow \patterns }
    ⟨ Enter all of the patterns into a linked trie, until coming to a right brace 1015 ⟩;
    if saving_hyph_codes > 0 then ⟨ Store hyphenation codes for current language 1666 ⟩;
    end
  else begin print_err("Too late for "); print_esc("patterns");
    help1("All patterns must be given before typesetting begins."); error;
    link(garbage) ← scan_toks(false, false); flush_list(def_ref);
    end;
  end;
```

1017*. When the following code comes into play, the pattern $p_1 \ldots p_k$ appears in $hc[1 \,..\, k]$, and the corresponding sequence of numbers $n_0 \ldots n_k$ appears in $hyf[0 \,..\, k]$.

```
⟨ Insert a new pattern into the linked trie 1017* ⟩ ≡
  begin ⟨ Compute the trie op code, v, and set l ← 0 1019* ⟩;
  q ← 0; hc[0] ← cur_lang;
  while l ≤ k do
    begin c ← hc[l]; incr(l); p ← trie_l[q]; first_child ← true;
    while (p > 0) ∧ (c > so(trie_c[p])) do
      begin q ← p; p ← trie_r[q]; first_child ← false;
      end;
    if (p = 0) ∨ (c < so(trie_c[p])) then
      ⟨ Insert a new trie node between q and p, and make p point to it 1018* ⟩;
    q ← p;  { now node q represents p_1 ... p_{l-1} }
    end;
  if trie_o[q] ≠ min_trie_op then
    begin print_err("Duplicate␣pattern"); help1("(See␣Appendix␣H.)"); error;
    end;
  trie_o[q] ← v;
  end
```

This code is used in section 1015.

1018*. ⟨ Insert a new trie node between q and p, and make p point to it 1018* ⟩ ≡

```
  begin if trie_ptr = trie_size then overflow("pattern␣memory", trie_size);
  incr(trie_ptr); trie_r[trie_ptr] ← p; p ← trie_ptr; trie_l[p] ← 0;
  if first_child then trie_l[q] ← p else trie_r[q] ← p;
  trie_c[p] ← si(c); trie_o[p] ← min_trie_op;
  end
```

This code is used in sections 1017*, 1666, and 1667.

1019*. ⟨ Compute the trie op code, v, and set $l \leftarrow 0$ 1019* ⟩ ≡

```
  if hc[1] = 0 then hyf[0] ← 0;
  if hc[k] = 0 then hyf[k] ← 0;
  l ← k; v ← min_trie_op;
  loop begin if hyf[l] ≠ 0 then v ← new_trie_op(k − l, hyf[l], v);
    if l > 0 then decr(l) else goto done1;
    end;
done1:
```

This code is used in section 1017*.

1020*. Finally we put everything together: Here is how the trie gets to its final, efficient form. The following packing routine is rigged so that the root of the linked tree gets mapped into location 1 of *trie*, as required by the hyphenation algorithm. This happens because the first call of *first_fit* will "take" location 1.

```
⟨ Declare procedures for preprocessing hyphenation patterns 998* ⟩ +≡
procedure init_trie;
  var p: trie_pointer; { pointer for initialization }
    j, k, t: integer; { all-purpose registers for initialization }
    r, s: trie_pointer; { used to clean up the packed trie }
  begin incr(max_hyph_char); ⟨ Get ready to compress the trie 1006 ⟩;
  if trie_root ≠ 0 then
    begin first_fit(trie_root); trie_pack(trie_root);
    end;
  if hyph_root ≠ 0 then ⟨ Pack all stored hyph_codes 1668 ⟩;
  ⟨ Move the data into trie 1012* ⟩;
  trie_not_ready ← false;
  end;
```

1042*. Pages are built by appending nodes to the current list in TeX's vertical mode, which is at the outermost level of the semantic nest. This vlist is split into two parts; the "current page" that we have been talking so much about already, and the "contribution list" that receives new nodes as they are created. The current page contains everything that the page builder has accounted for in its data structures, as described above, while the contribution list contains other things that have been generated by other parts of TeX but have not yet been seen by the page builder. The contribution list starts at *link*(*contrib_head*), and it ends at the current node in TeX's vertical mode.

When TeX has appended new material in vertical mode, it calls the procedure *build_page*, which tries to catch up by moving nodes from the contribution list to the current page. This procedure will succeed in its goal of emptying the contribution list, unless a page break is discovered, i.e., unless the current page has grown to the point where the optimum next page break has been determined. In the latter case, the nodes after the optimum break will go back onto the contribution list, and control will effectively pass to the user's output routine.

We make *type*(*page_head*) = *glue_node*, so that an initial glue node on the current page will not be considered a valid breakpoint.

⟨ Initialize the special list heads and constant nodes 838 ⟩ +≡
type(*page_head*) ← *glue_node*; *subtype*(*page_head*) ← *normal*;
{ *SyncTeX* watch point: box(*page_head*) size ¿= *glue_node* size }

1088* We leave the *space_factor* unchanged if $sf\_code(cur\_chr) = 0$; otherwise we set it equal to $sf\_code(cur\_chr)$, except that it should never change from a value less than 1000 to a value exceeding 1000. The most common case is $sf\_code(cur\_chr) = 1000$, so we want that case to be fast.

The overall structure of the main loop is presented here. Some program labels are inside the individual sections.

```
define adjust_space_factor ≡
    main_s ← sf_code(cur_chr) mod "10000;
    if main_s = 1000 then space_factor ← 1000
    else if main_s < 1000 then
        begin if main_s > 0 then space_factor ← main_s;
        end
      else if space_factor < 1000 then space_factor ← 1000
        else space_factor ← main_s
define check_for_inter_char_toks(#) ≡ { check for a spacing token list, goto # if found, or big_switch in
      case of the initial letter of a run }
    cur_ptr ← null; space_class ← sf_code(cur_chr) div "10000;
    if XeTeX_inter_char_tokens_en ∧ space_class ≠ char_class_ignored then
      begin { class 4096 = ignored (for combining marks etc) }
      if prev_class = char_class_boundary then
        begin { boundary }
        if (state ≠ token_list) ∨ (token_type ≠ backed_up_char) then
          begin find_sa_element(inter_char_val, char_class_boundary * char_class_limit + space_class,
              false);
          if (cur_ptr ≠ null) ∧ (sa_ptr(cur_ptr) ≠ null) then
            begin if cur_cmd ≠ letter then cur_cmd ← other_char;
            cur_tok ← (cur_cmd * max_char_val) + cur_chr; back_input;
            token_type ← backed_up_char; begin_token_list(sa_ptr(cur_ptr), inter_char_text);
            goto big_switch;
            end
          end
        end
      else begin find_sa_element(inter_char_val, prev_class * char_class_limit + space_class, false);
        if (cur_ptr ≠ null) ∧ (sa_ptr(cur_ptr) ≠ null) then
          begin if cur_cmd ≠ letter then cur_cmd ← other_char;
          cur_tok ← (cur_cmd * max_char_val) + cur_chr; back_input; token_type ← backed_up_char;
          begin_token_list(sa_ptr(cur_ptr), inter_char_text); prev_class ← char_class_boundary;
          goto #;
          end;
        end;
      prev_class ← space_class;
      end
define check_for_post_char_toks(#) ≡
      if XeTeX_inter_char_tokens_en ∧ (space_class ≠ char_class_ignored) ∧ (prev_class ≠
          char_class_boundary) then
        begin prev_class ← char_class_boundary;
        find_sa_element(inter_char_val, space_class * char_class_limit + char_class_boundary, false);
          { boundary }
        if (cur_ptr ≠ null) ∧ (sa_ptr(cur_ptr) ≠ null) then
          begin if cur_cs = 0 then
            begin if cur_cmd = char_num then cur_cmd ← other_char;
            cur_tok ← (cur_cmd * max_char_val) + cur_chr;
            end
```

```
      else cur_tok ← cs_token_flag + cur_cs;
      back_input; begin_token_list(sa_ptr(cur_ptr), inter_char_text); goto #;
      end;
    end
```

⟨ Append character *cur_chr* and the following characters (if any) to the current hlist in the current font; **goto** *reswitch* when a non-character has been fetched 1088* ⟩ ≡

```
if ((head = tail) ∧ (mode > 0)) then
  begin if (insert_src_special_auto) then append_src_special;
  end;
prev_class ← char_class_boundary;  { boundary }
  { added code for native font support }
if is_native_font(cur_font) then
  begin if mode > 0 then
    if language ≠ clang then fix_language;
  main_h ← 0; main_f ← cur_font; native_len ← 0;
collect_native: adjust_space_factor; check_for_inter_char_toks(collected);
  if (cur_chr > "FFFF) then
    begin native_room(2); append_native((cur_chr − "10000) div 1024 + "D800);
    append_native((cur_chr − "10000) mod 1024 + "DC00);
    end
  else begin native_room(1); append_native(cur_chr);
    end;
  is_hyph ← (cur_chr = hyphen_char[main_f]) ∨ (XeTeX_dash_break_en ∧ ((cur_chr = "2014) ∨ (cur_chr =
      "2013)));
  if (main_h = 0) ∧ is_hyph then main_h ← native_len;
      { try to collect as many chars as possible in the same font }
  get_next;
  if (cur_cmd = letter) ∨ (cur_cmd = other_char) ∨ (cur_cmd = char_given) then goto collect_native;
  x_token;
  if (cur_cmd = letter) ∨ (cur_cmd = other_char) ∨ (cur_cmd = char_given) then goto collect_native;
  if cur_cmd = char_num then
    begin scan_usv_num; cur_chr ← cur_val; goto collect_native;
    end;
  check_for_post_char_toks(collected);
collected: if (font_mapping[main_f] ≠ 0) then
    begin main_k ← apply_mapping(font_mapping[main_f], native_text, native_len); native_len ← 0;
    native_room(main_k); main_h ← 0;
    for main_p ← 0 to main_k − 1 do
      begin append_native(mapped_text[main_p]);
      if (main_h = 0) ∧ ((mapped_text[main_p] = hyphen_char[main_f]) ∨ (XeTeX_dash_break_en ∧
          ((mapped_text[main_p] = "2014) ∨ (mapped_text[main_p] = "2013)))) then
        main_h ← native_len;
      end
    end;
  if tracing_lost_chars > 0 then
    begin temp_ptr ← 0;
    while (temp_ptr < native_len) do
      begin main_k ← native_text[temp_ptr]; incr(temp_ptr);
      if (main_k ≥ "D800) ∧ (main_k < "DC00) then
        begin main_k ← "10000 + (main_k − "D800) * 1024;
        main_k ← main_k + native_text[temp_ptr] − "DC00; incr(temp_ptr);
        end;
```

```
      if map_char_to_glyph(main_f, main_k) = 0 then char_warning(main_f, main_k);
      end
    end;
  main_k ← native_len; main_pp ← tail;
  if mode = hmode then
    begin main_ppp ← head; { find node preceding tail, skipping discretionaries }
    while (main_ppp ≠ main_pp) ∧ (link(main_ppp) ≠ main_pp) do
      begin if (¬is_char_node(main_ppp)) ∧ (type(main_ppp) = disc_node) then
        begin temp_ptr ← main_ppp;
        for main_p ← 1 to replace_count(temp_ptr) do main_ppp ← link(main_ppp);
        end;
      if main_ppp ≠ main_pp then main_ppp ← link(main_ppp);
      end;
    temp_ptr ← 0;
    repeat if main_h = 0 then main_h ← main_k;
      if is_native_word_node(main_pp) ∧ (native_font(main_pp) = main_f) ∧ (main_ppp ≠
          main_pp) ∧ (¬is_char_node(main_ppp)) ∧ (type(main_ppp) ≠ disc_node) then
        begin { make a new temp string that contains the concatenated text of tail + the current
            word/fragment }
        main_k ← main_h + native_length(main_pp); native_room(main_k);
        save_native_len ← native_len;
        for main_p ← 0 to native_length(main_pp) − 1 do
          append_native(get_native_char(main_pp, main_p));
        for main_p ← 0 to main_h − 1 do append_native(native_text[temp_ptr + main_p]);
        do_locale_linebreaks(save_native_len, main_k); native_len ← save_native_len;
            { discard the temp string }
        main_k ← native_len − main_h − temp_ptr;
            { and set main_k to remaining length of new word }
        temp_ptr ← main_h; { pointer to remaining fragment }
        main_h ← 0;
        while (main_h < main_k) ∧ (native_text[temp_ptr + main_h] ≠
            hyphen_char[main_f]) ∧ ((¬XeTeX_dash_break_en) ∨ ((native_text[temp_ptr + main_h] ≠
            "2014) ∧ (native_text[temp_ptr + main_h] ≠ "2013))) do incr(main_h);
            { look for next hyphen or end of text }
        if (main_h < main_k) then incr(main_h); { remove the preceding node from the list }
        link(main_ppp) ← link(main_pp); link(main_pp) ← null; flush_node_list(main_pp);
        main_pp ← tail;
        while (link(main_ppp) ≠ main_pp) do main_ppp ← link(main_ppp);
        end
      else begin do_locale_linebreaks(temp_ptr, main_h); { append fragment of current word }
        temp_ptr ← temp_ptr + main_h; { advance ptr to remaining fragment }
        main_k ← main_k − main_h; { decrement remaining length }
        main_h ← 0;
        while (main_h < main_k) ∧ (native_text[temp_ptr + main_h] ≠
            hyphen_char[main_f]) ∧ ((¬XeTeX_dash_break_en) ∨ ((native_text[temp_ptr + main_h] ≠
            "2014) ∧ (native_text[temp_ptr + main_h] ≠ "2013))) do incr(main_h);
            { look for next hyphen or end of text }
        if (main_h < main_k) then incr(main_h);
        end;
      if (main_k > 0) ∨ is_hyph then
        begin tail_append(new_disc); { add a break if we aren't at end of text (must be a hyphen), or
            if last char in original text was a hyphen }
```

```
      main_pp ← tail;
      end;
  until main_k = 0;
  end
else begin { must be restricted hmode, so no need for line-breaking or discretionaries }
  { but there might already be explicit disc_nodes in the list }
  main_ppp ← head; { find node preceding tail, skipping discretionaries }
  while (main_ppp ≠ main_pp) ∧ (link(main_ppp) ≠ main_pp) do
    begin if (¬is_char_node(main_ppp)) ∧ (type(main_ppp) = disc_node) then
      begin temp_ptr ← main_ppp;
      for main_p ← 1 to replace_count(temp_ptr) do main_ppp ← link(main_ppp);
      end;
    if main_ppp ≠ main_pp then main_ppp ← link(main_ppp);
    end;
  if is_native_word_node(main_pp) ∧ (native_font(main_pp) = main_f) ∧ (main_ppp ≠
      main_pp) ∧ (¬is_char_node(main_ppp)) ∧ (type(main_ppp) ≠ disc_node) then
    begin { total string length for the new merged whatsit }
    link(main_pp) ← new_native_word_node(main_f, main_k + native_length(main_pp));
    tail ← link(main_pp); { copy text from the old one into the new }
    for main_p ← 0 to native_length(main_pp) − 1 do
      set_native_char(tail, main_p, get_native_char(main_pp, main_p)); { append the new text }
    for main_p ← 0 to main_k − 1 do
      set_native_char(tail, main_p + native_length(main_pp), native_text[main_p]);
    set_native_metrics(tail, XeTeX_use_glyph_metrics); { remove the preceding node from the list }
    main_p ← head;
    if main_p ≠ main_pp then
      while link(main_p) ≠ main_pp do main_p ← link(main_p);
    link(main_p) ← link(main_pp); link(main_pp) ← null; flush_node_list(main_pp);
    end
  else begin { package the current string into a native_word whatsit }
    link(main_pp) ← new_native_word_node(main_f, main_k); tail ← link(main_pp);
    for main_p ← 0 to main_k − 1 do set_native_char(tail, main_p, native_text[main_p]);
    set_native_metrics(tail, XeTeX_use_glyph_metrics);
    end
  end;
if XeTeX_interword_space_shaping_state > 0 then
  begin { tail is a word we have just appended. If it is preceded by another word with a normal
      inter-word space between (all in the same font), then we will measure that space in context and
      replace it with an adjusted glue value if it differs from the font's normal space. }
    { First we look for the most recent native_word in the list and set main_pp to it. This is potentially
      expensive, in the case of very long paragraphs, but in practice it's negligible compared to the
      cost of shaping and measurement. }
  main_p ← head; main_pp ← null;
  while main_p ≠ tail do
    begin if is_native_word_node(main_p) then main_pp ← main_p;
    main_p ← link(main_p);
    end;
  if (main_pp ≠ null) then
    begin { check if the font matches; if so, check the intervening nodes }
    if (native_font(main_pp) = main_f) then
      begin main_p ← link(main_pp);
          { Skip nodes that should be invisible to inter-word spacing, so that e.g., '\nobreak\ '
```

```
      doesn't prevent contextual measurement. This loop is guaranteed to end safely because it'll
      eventually hit tail, which is a native_word node, if nothing else intervenes. }
    while node_is_invisible_to_interword_space(main_p) do main_p ← link(main_p);
    if ¬is_char_node(main_p) ∧ (type(main_p) = glue_node) then
      begin { We found a glue node: we might have an inter-word space to deal with. Again,
        skip nodes that should be invisible to inter-word spacing. We leave main_p pointing to
        the glue node; main_pp is the preceding word. }
      main_ppp ← link(main_p);
      while node_is_invisible_to_interword_space(main_ppp) do main_ppp ← link(main_ppp);
      if main_ppp = tail then
        begin { We found a candidate inter-word space! Collect the characters of both words,
          separated by a single space, into a native_word node and measure its overall width. }
        temp_ptr ← new_native_word_node(main_f, native_length(main_pp)+1+native_length(tail));
        main_k ← 0;
        for t ← 0 to native_length(main_pp) − 1 do
          begin set_native_char(temp_ptr, main_k, get_native_char(main_pp, t)); incr(main_k);
          end;
        set_native_char(temp_ptr, main_k, "␣"); incr(main_k);
        for t ← 0 to native_length(tail) − 1 do
          begin set_native_char(temp_ptr, main_k, get_native_char(tail, t)); incr(main_k);
          end;
        set_native_metrics(temp_ptr, XeTeX_use_glyph_metrics); { The contextual space width is
          the difference between this width and the sum of the two words measured separately. }
        t ← width(temp_ptr) − width(main_pp) − width(tail);
        free_node(temp_ptr, native_size(temp_ptr)); { If the desired width differs from the font's
          default word space, we will insert a suitable kern after the existing glue. Because kerns
          are discardable, this will behave OK during line breaking, and it's easier than actually
          modifying/replacing the glue node. }
        if t ≠ width(font_glue[main_f]) then
          begin temp_ptr ← new_kern(t − width(font_glue[main_f]));
          subtype(temp_ptr) ← space_adjustment; link(temp_ptr) ← link(main_p);
          link(main_p) ← temp_ptr;
          end
        end
      end
    end;
  if cur_ptr ≠ null then goto big_switch
  else goto reswitch;
  end; { End of added code for native fonts }
adjust_space_factor;
check_for_inter_char_toks(big_switch); main_f ← cur_font; bchar ← font_bchar[main_f];
false_bchar ← font_false_bchar[main_f];
if mode > 0 then
  if language ≠ clang then fix_language;
fast_get_avail(lig_stack); font(lig_stack) ← main_f; cur_l ← qi(cur_chr); character(lig_stack) ← cur_l;
cur_q ← tail;
if cancel_boundary then
  begin cancel_boundary ← false; main_k ← non_address;
  end
else main_k ← bchar_label[main_f];
```

if *main_k* = *non_address* **then goto** *main_loop_move* + 2; { no left boundary processing }
cur_r ← *cur_l*; *cur_l* ← *non_char*; **goto** *main_lig_loop* + 1; { begin with cursor after left boundary }

main_loop_wrapup: ⟨ Make a ligature node, if *ligature_present*; insert a null discretionary, if appropriate 1089 ⟩;
main_loop_move: ⟨ If the cursor is immediately followed by the right boundary, **goto** *reswitch*; if it's followed by an invalid character, **goto** *big_switch*; otherwise move the cursor one step to the right and **goto** *main_lig_loop* 1090* ⟩;
main_loop_lookahead: ⟨ Look ahead for another character, or leave *lig_stack* empty if there's none there 1092 ⟩;
main_lig_loop: ⟨ If there's a ligature/kern command relevant to *cur_l* and *cur_r*, adjust the text appropriately; exit to *main_loop_wrapup* 1093 ⟩;
main_loop_move_lig: ⟨ Move the cursor past a pseudo-ligature, then **goto** *main_loop_lookahead* or *main_lig_loop* 1091* ⟩

This code is used in section 1084.

1090* ⟨ If the cursor is immediately followed by the right boundary, **goto** *reswitch*; if it's followed by an invalid character, **goto** *big_switch*; otherwise move the cursor one step to the right and **goto** *main_lig_loop* 1090* ⟩ ≡

```
if lig_stack = null then goto reswitch;
cur_q ← tail; cur_l ← character(lig_stack);
main_loop_move + 1: if ¬is_char_node(lig_stack) then goto main_loop_move_lig;
main_loop_move + 2: if (qo(effective_char(false, main_f,
        qi(cur_chr))) > font_ec[main_f])∨(qo(effective_char(false, main_f, qi(cur_chr))) < font_bc[main_f])
        then
    begin char_warning(main_f, cur_chr); free_avail(lig_stack); goto big_switch;
    end;
main_i ← effective_char_info(main_f, cur_l);
if ¬char_exists(main_i) then
    begin char_warning(main_f, cur_chr); free_avail(lig_stack); goto big_switch;
    end;
link(tail) ← lig_stack; tail ← lig_stack   { main_loop_lookahead is next }
```

This code is used in section 1088*.

1091* Here we are at *main_loop_move_lig*. When we begin this code we have *cur_q* = *tail* and *cur_l* = *character*(*lig_stack*).

⟨ Move the cursor past a pseudo-ligature, then **goto** *main_loop_lookahead* or *main_lig_loop* 1091* ⟩ ≡

```
main_p ← lig_ptr(lig_stack);
if main_p > null then tail_append(main_p);   { append a single character }
temp_ptr ← lig_stack; lig_stack ← link(temp_ptr); free_node(temp_ptr, small_node_size);
    { SyncTeX watch point: proper size! }
main_i ← char_info(main_f)(cur_l); ligature_present ← true;
if lig_stack = null then
    if main_p > null then goto main_loop_lookahead
    else cur_r ← bchar
else cur_r ← character(lig_stack);
goto main_lig_loop
```

This code is used in section 1088*.

1103*. The '*you_cant*' procedure prints a line saying that the current command is illegal in the current mode; it identifies these things symbolically.

⟨ Declare action procedures for use by *main_control* 1097 ⟩ +≡

```
procedure you_cant;
  begin print_err("You can't use `"); print_cmd_chr(cur_cmd, cur_chr); print_in_mode(mode);
  end;
```

1145* ⟨ Declare action procedures for use by *main_control* 1097 ⟩ +≡

```
function norm_min(h : integer): small_number;
  begin if h ≤ 0 then norm_min ← 1 else if h ≥ 63 then norm_min ← 63 else norm_min ← h;
  end;
procedure new_graf(indented : boolean);
  begin prev_graf ← 0;
  if (mode = vmode) ∨ (head ≠ tail) then tail_append(new_param_glue(par_skip_code));
  push_nest; mode ← hmode; space_factor ← 1000; set_cur_lang; clang ← cur_lang;
  prev_graf ← (norm_min(left_hyphen_min) * ´100 + norm_min(right_hyphen_min)) * ´200000 + cur_lang;
  if indented then
    begin tail ← new_null_box; link(head) ← tail; width(tail) ← par_indent;
    if (insert_src_special_every_par) then insert_src_special;
    end;
  if every_par ≠ null then begin_token_list(every_par, every_par_text);
  if nest_ptr = 1 then build_page; { put par_skip glue on current page }
  end;
```

1193* ⟨ Go into ordinary math mode 1193* ⟩ ≡
begin *push_math*(*math_shift_group*); *eq_word_define*(*int_base* + *cur_fam_code*, −1);
if (*insert_src_special_every_math*) **then** *insert_src_special*;
if *every_math* ≠ *null* **then** *begin_token_list*(*every_math*, *every_math_text*);
end

This code is used in sections 1192 and 1196.

1221* ⟨ Cases of *main_control* that build boxes and lists 1110 ⟩ +≡
mmode + *vcenter*: **begin** *scan_spec*(*vcenter_group*, *false*); *normal_paragraph*; *push_nest*; *mode* ← −*vmode*;
prev_depth ← *ignore_depth*;
if (*insert_src_special_every_vbox*) **then** *insert_src_special*;
if *every_vbox* ≠ *null* **then** *begin_token_list*(*every_vbox*, *every_vbox_text*);
end;

1269* When a control sequence is to be defined, by `\def` or `\let` or something similar, the *get_r_token* routine will substitute a special control sequence for a token that is not redefinable.

```
⟨ Declare subprocedures for prefixed_command 1269* ⟩ ≡
procedure get_r_token;
  label restart;
  begin restart: repeat get_token;
  until cur_tok ≠ space_token;
  if (cur_cs = 0) ∨ (cur_cs > eqtb_top) ∨ ((cur_cs > frozen_control_sequence) ∧ (cur_cs ≤ eqtb_size)) then
    begin print_err("Missing␣control␣sequence␣inserted");
    help5("Please␣don´t␣say␣`\def␣cs{...}´,␣say␣`\def\cs{...}´.")
    ("I´ve␣inserted␣an␣inaccessible␣control␣sequence␣so␣that␣your")
    ("definition␣will␣be␣completed␣without␣mixing␣me␣up␣too␣badly.")
    ("You␣can␣recover␣graciously␣from␣this␣error,␣if␣you´re")
    ("careful;␣see␣exercise␣27.2␣in␣The␣TeXbook.");
    if cur_cs = 0 then back_input;
    cur_tok ← cs_token_flag + frozen_protection; ins_error; goto restart;
    end;
  end;
```

See also sections 1283, 1290, 1297, 1298, 1299, 1300, 1301, 1311*, and 1319*.

This code is used in section 1265.

1276* A `\chardef` creates a control sequence whose *cmd* is *char_given*; a `\mathchardef` creates a control sequence whose *cmd* is *math_given*; and the corresponding *chr* is the character code or math code. A `\countdef` or `\dimendef` or `\skipdef` or `\muskipdef` creates a control sequence whose *cmd* is *assign_int* or ... or *assign_mu_glue*, and the corresponding *chr* is the *eqtb* location of the internal register in question.

```
  define char_def_code = 0   { shorthand_def for \chardef }
  define math_char_def_code = 1   { shorthand_def for \mathchardef }
  define count_def_code = 2   { shorthand_def for \countdef }
  define dimen_def_code = 3   { shorthand_def for \dimendef }
  define skip_def_code = 4   { shorthand_def for \skipdef }
  define mu_skip_def_code = 5   { shorthand_def for \muskipdef }
  define toks_def_code = 6   { shorthand_def for \toksdef }
  define char_sub_def_code = 7   { shorthand_def for \charsubdef }
  define XeTeX_math_char_num_def_code = 8
  define XeTeX_math_char_def_code = 9
⟨ Put each of TeX's primitives into the hash table 252 ⟩ +≡
  primitive("chardef", shorthand_def, char_def_code);
  primitive("mathchardef", shorthand_def, math_char_def_code);
  primitive("XeTeXmathcharnumdef", shorthand_def, XeTeX_math_char_num_def_code);
  primitive("Umathcharnumdef", shorthand_def, XeTeX_math_char_num_def_code);
  primitive("XeTeXmathchardef", shorthand_def, XeTeX_math_char_def_code);
  primitive("Umathchardef", shorthand_def, XeTeX_math_char_def_code);
  primitive("countdef", shorthand_def, count_def_code);
  primitive("dimendef", shorthand_def, dimen_def_code);
  primitive("skipdef", shorthand_def, skip_def_code);
  primitive("muskipdef", shorthand_def, mu_skip_def_code);
  primitive("toksdef", shorthand_def, toks_def_code);
  if mltex_p then
    begin primitive("charsubdef", shorthand_def, char_sub_def_code);
    end;
```

1277* ⟨ Cases of *print_cmd_chr* for symbolic printing of primitives 253 ⟩ +≡

```
shorthand_def: case chr_code of
  char_def_code: print_esc("chardef");
  math_char_def_code: print_esc("mathchardef");
  XeTeX_math_char_def_code: print_esc("Umathchardef");
  XeTeX_math_char_num_def_code: print_esc("Umathcharnumdef");
  count_def_code: print_esc("countdef");
  dimen_def_code: print_esc("dimendef");
  skip_def_code: print_esc("skipdef");
  mu_skip_def_code: print_esc("muskipdef");
  char_sub_def_code: print_esc("charsubdef");
  othercases print_esc("toksdef")
  endcases;
char_given: begin print_esc("char"); print_hex(chr_code);
  end;
math_given: begin print_esc("mathchar"); print_hex(chr_code);
  end;
XeTeX_math_given: begin print_esc("Umathchar"); print_hex(math_class_field(chr_code));
  print_hex(math_fam_field(chr_code)); print_hex(math_char_field(chr_code));
  end;
```

1278* We temporarily define p to be *relax*, so that an occurrence of p while scanning the definition will simply stop the scanning instead of producing an "undefined control sequence" error or expanding the previous meaning. This allows, for instance, '`\chardef\foo=123\foo`'.

```
⟨ Assignments 1271 ⟩ +≡
shorthand_def: if cur_chr = char_sub_def_code then
    begin scan_char_num; p ← char_sub_code_base + cur_val; scan_optional_equals; scan_char_num;
    n ← cur_val; { accent character in substitution }
    scan_char_num;
    if (tracing_char_sub_def > 0) then
      begin begin_diagnostic; print_nl("New␣character␣substitution:␣");
      print_ASCII(p − char_sub_code_base); print("␣=␣"); print_ASCII(n); print_char("␣");
      print_ASCII(cur_val); end_diagnostic(false);
      end;
    n ← n * 256 + cur_val; define(p, data, hi(n));
    if (p − char_sub_code_base) < char_sub_def_min then
      word_define(int_base + char_sub_def_min_code, p − char_sub_code_base);
    if (p − char_sub_code_base) > char_sub_def_max then
      word_define(int_base + char_sub_def_max_code, p − char_sub_code_base);
    end
  else begin n ← cur_chr; get_r_token; p ← cur_cs; define(p, relax, too_big_usv); scan_optional_equals;
    case n of
    char_def_code: begin scan_usv_num; define(p, char_given, cur_val);
      end;
    math_char_def_code: begin scan_fifteen_bit_int; define(p, math_given, cur_val);
      end;
    XeTeX_math_char_num_def_code: begin scan_xetex_math_char_int;
      define(p, XeTeX_math_given, cur_val);
      end;
    XeTeX_math_char_def_code: begin scan_math_class_int; n ← set_class_field(cur_val);
      scan_math_fam_int; n ← n + set_family_field(cur_val); scan_usv_num; n ← n + cur_val;
      define(p, XeTeX_math_given, n);
      end;
    othercases begin scan_register_num;
      if cur_val > 255 then
        begin j ← n − count_def_code; { int_val .. box_val }
        if j > mu_val then j ← tok_val; { int_val .. mu_val or tok_val }
        find_sa_element(j, cur_val, true); add_sa_ref(cur_ptr);
        if j = tok_val then j ← toks_register else j ← register;
        define(p, j, cur_ptr);
        end
      else case n of
        count_def_code: define(p, assign_int, count_base + cur_val);
        dimen_def_code: define(p, assign_dimen, scaled_base + cur_val);
        skip_def_code: define(p, assign_glue, skip_base + cur_val);
        mu_skip_def_code: define(p, assign_mu_glue, mu_skip_base + cur_val);
        toks_def_code: define(p, assign_toks, toks_base + cur_val);
        end; { there are no other cases }
      end
    endcases;
    end;
```

1306* ⟨ Assignments 1271 ⟩ +≡
```
hyph_data: if cur_chr = 1 then
    begin Init new_patterns; goto done; Tini
    print_err("Patterns␣can␣be␣loaded␣only␣by␣INITEX"); help0; error;
    repeat get_token;
    until cur_cmd = right_brace;  { flush the patterns }
    return;
    end
  else begin new_hyph_exceptions; goto done;
    end;
```

1311* ⟨ Declare subprocedures for *prefixed_command* 1269* ⟩ +≡
```
procedure new_font(a : small_number);
  label common_ending;
  var u: pointer;  { user's font identifier }
    s: scaled;  { stated "at" size, or negative of scaled magnification }
    f: internal_font_number;  { runs through existing fonts }
    t: str_number;  { name for the frozen font identifier }
    old_setting: 0 .. max_selector;  { holds selector setting }
  begin if job_name = 0 then open_log_file;  { avoid confusing texput with the font name }
  get_r_token; u ← cur_cs;
  if u ≥ hash_base then t ← text(u)
  else if u ≥ single_base then
      if u = null_cs then t ← "FONT" else t ← u − single_base
    else begin old_setting ← selector; selector ← new_string; print("FONT"); print(u − active_base);
      selector ← old_setting; str_room(1); t ← make_string;
      end;
  define(u, set_font, null_font); scan_optional_equals; scan_file_name;
  ⟨ Scan the font size specification 1312 ⟩;
  ⟨ If this font has already been loaded, set f to the internal font number and goto common_ending 1314* ⟩;
  f ← read_font_info(u, cur_name, cur_area, s);
common_ending: define(u, set_font, f); eqtb[font_id_base + f] ← eqtb[u]; font_id_text(f) ← t;
  end;
```

1314* When the user gives a new identifier to a font that was previously loaded, the new name becomes the font identifier of record. Font names '`xyz`' and '`XYZ`' are considered to be different.

```
⟨ If this font has already been loaded, set f to the internal font number and goto common_ending 1314* ⟩ ≡
  for f ← font_base + 1 to font_ptr do
    begin if str_eq_str(font_name[f],
          cur_name) ∧ (((cur_area = "") ∧ is_native_font(f)) ∨ str_eq_str(font_area[f], cur_area)) then
      begin if s > 0 then
        begin if s = font_size[f] then goto common_ending;
        end
      else if font_size[f] = xn_over_d(font_dsize[f], −s, 1000) then goto common_ending;
      end; { could be a native font whose "name" ended up partly in area or extension }
    append_str(cur_area); append_str(cur_name); append_str(cur_ext);
    if str_eq_str(font_name[f], make_string) then
      begin flush_string;
      if is_native_font(f) then
        begin if s > 0 then
          begin if s = font_size[f] then goto common_ending;
          end
        else if font_size[f] = xn_over_d(font_dsize[f], −s, 1000) then goto common_ending;
        end
      end
    else flush_string;
    end
```

This code is used in section 1311*.

1319* ⟨ Declare subprocedures for *prefixed_command* 1269* ⟩ +≡

```
procedure new_interaction;
  begin print_ln; interaction ← cur_chr;
  if interaction = batch_mode then kpse_make_tex_discard_errors ← 1
  else kpse_make_tex_discard_errors ← 0;
  ⟨ Initialize the print selector based on interaction 79 ⟩;
  if log_opened then selector ← selector + 2;
  end;
```

1329* ⟨ Declare action procedures for use by *main_control* 1097 ⟩ +≡

```
procedure open_or_close_in;
  var c: 0 .. 1;  { 1 for \openin, 0 for \closein }
    n: 0 .. 15;  { stream number }
    k: 0 .. file_name_size;  { index into name_of_file16 }
  begin c ← cur_chr; scan_four_bit_int; n ← cur_val;
  if read_open[n] ≠ closed then
    begin u_close(read_file[n]); read_open[n] ← closed;
    end;
  if c ≠ 0 then
    begin scan_optional_equals; scan_file_name; pack_cur_name; tex_input_type ← 0;
        { Tell open_input we are \openin. }
    if kpse_in_name_ok(stringcast(name_of_file + 1)) ∧ u_open_in(read_file[n], kpse_tex_format,
          XeTeX_default_input_mode, XeTeX_default_input_encoding) then
      begin make_utf16_name; name_in_progress ← true; begin_name; stop_at_space ← false; k ← 0;
      while (k < name_length16) ∧ (more_name(name_of_file16[k])) do incr(k);
      stop_at_space ← true; end_name; name_in_progress ← false; read_open[n] ← just_open;
      end;
    end;
  end;
```

1355* ⟨ Initialize table entries (done by `INITEX` only) 189 ⟩ +≡
```
if ini_version then format_ident ← "␣(INITEX)";
```

1356* ⟨ Declare action procedures for use by *main_control* 1097 ⟩ +≡
```
init procedure store_fmt_file;
label found1, found2, done1, done2;
var j, k, l: integer; { all-purpose indices }
  p, q: pointer; { all-purpose pointers }
  x: integer; { something to dump }
  format_engine: ↑char;
begin ⟨ If dumping is not allowed, abort 1358 ⟩;
⟨ Create the format_ident, open the format file, and inform the user that dumping has begun 1382 ⟩;
⟨ Dump constants for consistency check 1361* ⟩;
⟨ Dump MLTeX-specific data 1700* ⟩;
⟨ Dump the string pool 1363* ⟩;
⟨ Dump the dynamic memory 1365* ⟩;
⟨ Dump the table of equivalents 1367 ⟩;
⟨ Dump the font information 1374* ⟩;
⟨ Dump the hyphenation tables 1378* ⟩;
⟨ Dump a couple more things and the closing check word 1380 ⟩;
⟨ Close the format file 1383 ⟩;
end;
tini
```

1357* Corresponding to the procedure that dumps a format file, we have a function that reads one in. The function returns *false* if the dumped format is incompatible with the present TeX table sizes, etc.

```
define bad_fmt = 6666 { go here if the format file is unacceptable }
define too_small(#) ≡
    begin wake_up_terminal; wterm_ln('---!␣Must␣increase␣the␣', #); goto bad_fmt;
    end
```

⟨ Declare the function called *open_fmt_file* 559* ⟩
```
function load_fmt_file: boolean;
label bad_fmt, exit;
var j, k: integer; { all-purpose indices }
  p, q: pointer; { all-purpose pointers }
  x: integer; { something undumped }
  format_engine: ↑char;
begin ⟨ Undump constants for consistency check 1362* ⟩;
⟨ Undump MLTeX-specific data 1701* ⟩;
⟨ Undump the string pool 1364* ⟩;
⟨ Undump the dynamic memory 1366* ⟩;
⟨ Undump the table of equivalents 1368* ⟩;
⟨ Undump the font information 1375* ⟩;
⟨ Undump the hyphenation tables 1379* ⟩;
⟨ Undump a couple more things and the closing check word 1381* ⟩;
load_fmt_file ← true; return; { it worked! }
bad_fmt: wake_up_terminal; wterm_ln('(Fatal␣format␣file␣error;␣I''m␣stymied)');
  load_fmt_file ← false;
exit: end;
```

1359*. Format files consist of *memory_word* items, and we use the following macros to dump words of different types:

```
⟨ Global variables 13 ⟩ +≡
fmt_file: word_file; { for input or output of format information }
```

1360*. The inverse macros are slightly more complicated, since we need to check the range of the values we are reading in. We say '*undump*(*a*)(*b*)(*x*)' to read an integer value x that is supposed to be in the range $a \le x \le b$. System error messages should be suppressed when undumping.

```
define undump_end_end(#) ≡ # ← x; end
define undump_end(#) ≡ (x > #) then goto bad_fmt else undump_end_end
define undump(#) ≡
    begin undump_int(x);
    if (x < #) ∨ undump_end
define format_debug_end(#) ≡ write_ln(stderr, ´␣=␣´, #);
    end ;
define format_debug(#) ≡
    if debug_format_file then
      begin write(stderr, ´fmtdebug:´, #); format_debug_end
define undump_size_end_end(#) ≡ too_small(#) else format_debug(#)(x); undump_end_end
define undump_size_end(#) ≡
      if x > # then undump_size_end_end
define undump_size(#) ≡
    begin undump_int(x);
    if x < # then goto bad_fmt;
    undump_size_end
```

1361*. The next few sections of the program should make it clear how we use the dump/undump macros.

```
⟨ Dump constants for consistency check 1361* ⟩ ≡
  dump_int("57325458); { Web2C TeX's magic constant: "W2TX" }
    { Align engine to 4 bytes with one or more trailing NUL }
  x ← strlen(engine_name); format_engine ← xmalloc_array(char, x + 4);
  strcpy(stringcast(format_engine), engine_name);
  for k ← x to x + 3 do format_engine[k] ← 0;
  x ← x + 4 − (x mod 4); dump_int(x); dump_things(format_engine[0], x); libc_free(format_engine);
  dump_int(@$);
  dump_int(max_halfword);
  dump_int(hash_high); ⟨ Dump the ε-TeX state 1464 ⟩
  dump_int(mem_bot);
  dump_int(mem_top);
  dump_int(eqtb_size);
  dump_int(hash_prime);
  dump_int(hyph_prime)
```

This code is used in section 1356*.

1362* Sections of a WEB program that are "commented out" still contribute strings to the string pool; therefore INITEX and TEX will have the same strings. (And it is, of course, a good thing that they do.)

```
⟨ Undump constants for consistency check 1362* ⟩ ≡ Init libc_free(font_info); libc_free(str_pool);
  libc_free(str_start); libc_free(yhash); libc_free(zeqtb); libc_free(yzmem); Tini undump_int(x);
  format_debug('format␣magic␣number')(x);
  if x ≠ "57325458 then goto bad_fmt; { not a format file }
  undump_int(x); format_debug('engine␣name␣size')(x);
  if (x < 0) ∨ (x > 256) then goto bad_fmt; { corrupted format file }
  format_engine ← xmalloc_array(char, x); undump_things(format_engine[0], x);
  format_engine[x − 1] ← 0; { force string termination, just in case }
  if strcmp(engine_name, stringcast(format_engine)) then
    begin wake_up_terminal;
    wterm_ln('---!␣', stringcast(name_of_file + 1), '␣was␣written␣by␣', format_engine);
    libc_free(format_engine); goto bad_fmt;
    end;
  libc_free(format_engine); undump_int(x); format_debug('string␣pool␣checksum')(x);
  if x ≠ @$ then
    begin { check that strings are the same }
    wake_up_terminal;
    wterm_ln('---!␣', stringcast(name_of_file + 1), '␣made␣by␣different␣executable␣version');
    goto bad_fmt;
    end;
  undump_int(x);
  if x ≠ max_halfword then goto bad_fmt; { check max_halfword }
  undump_int(hash_high);
  if (hash_high < 0) ∨ (hash_high > sup_hash_extra) then goto bad_fmt;
  if hash_extra < hash_high then hash_extra ← hash_high;
  eqtb_top ← eqtb_size + hash_extra;
  if hash_extra = 0 then hash_top ← undefined_control_sequence
  else hash_top ← eqtb_top;
  yhash ← xmalloc_array(two_halves, 1 + hash_top − hash_offset); hash ← yhash − hash_offset;
  next(hash_base) ← 0; text(hash_base) ← 0;
  for x ← hash_base + 1 to hash_top do hash[x] ← hash[hash_base];
  zeqtb ← xmalloc_array(memory_word, eqtb_top + 1); eqtb ← zeqtb;
  eq_type(undefined_control_sequence) ← undefined_cs; equiv(undefined_control_sequence) ← null;
  eq_level(undefined_control_sequence) ← level_zero;
  for x ← eqtb_size + 1 to eqtb_top do eqtb[x] ← eqtb[undefined_control_sequence];
  ⟨ Undump the ε-TEX state 1465 ⟩
  undump_int(x); format_debug('mem_bot')(x);
  if x ≠ mem_bot then goto bad_fmt;
  undump_int(mem_top); format_debug('mem_top')(mem_top);
  if mem_bot + 1100 > mem_top then goto bad_fmt;
  head ← contrib_head; tail ← contrib_head; page_tail ← page_head; { page initialization }
  mem_min ← mem_bot − extra_mem_bot; mem_max ← mem_top + extra_mem_top;
  yzmem ← xmalloc_array(memory_word, mem_max − mem_min + 1); zmem ← yzmem − mem_min;
      { this pointer arithmetic fails with some compilers }
  mem ← zmem; undump_int(x);
  if x ≠ eqtb_size then goto bad_fmt;
  undump_int(x);
  if x ≠ hash_prime then goto bad_fmt;
  undump_int(x);
  if x ≠ hyph_prime then goto bad_fmt
```

This code is used in section 1357*.

1363*. **define** *dump_four_ASCII* ≡ *w.b0* ← *qi*(*so*(*str_pool*[*k*])); *w.b1* ← *qi*(*so*(*str_pool*[*k* + 1])); *w.b2* ← *qi*(*so*(*str_pool*[*k* + 2])); *w.b3* ← *qi*(*so*(*str_pool*[*k* + 3])); *dump_qqqq*(*w*)

⟨ Dump the string pool 1363* ⟩ ≡
dump_int(*pool_ptr*); *dump_int*(*str_ptr*);
dump_things(*str_start_macro*(*too_big_char*), *str_ptr* + 1 − *too_big_char*); *dump_things*(*str_pool*[0], *pool_ptr*);
print_ln; *print_int*(*str_ptr*); *print*("␣strings␣of␣total␣length␣"); *print_int*(*pool_ptr*)

This code is used in section 1356*.

1364*. **define** *undump_four_ASCII* ≡ *undump_qqqq*(*w*); *str_pool*[*k*] ← *si*(*qo*(*w.b0*)); *str_pool*[*k* + 1] ← *si*(*qo*(*w.b1*)); *str_pool*[*k* + 2] ← *si*(*qo*(*w.b2*)); *str_pool*[*k* + 3] ← *si*(*qo*(*w.b3*))

⟨ Undump the string pool 1364* ⟩ ≡
undump_size(0)(*sup_pool_size* − *pool_free*)('string␣pool␣size')(*pool_ptr*);
if *pool_size* < *pool_ptr* + *pool_free* **then** *pool_size* ← *pool_ptr* + *pool_free*;
undump_size(0)(*sup_max_strings* − *strings_free*)('sup␣strings')(*str_ptr*);
if *max_strings* < *str_ptr* + *strings_free* **then** *max_strings* ← *str_ptr* + *strings_free*;
str_start ← *xmalloc_array*(*pool_pointer*, *max_strings*);
undump_checked_things(0, *pool_ptr*, *str_start_macro*(*too_big_char*), *str_ptr* + 1 − *too_big_char*);
str_pool ← *xmalloc_array*(*packed_ASCII_code*, *pool_size*); *undump_things*(*str_pool*[0], *pool_ptr*);
init_str_ptr ← *str_ptr*; *init_pool_ptr* ← *pool_ptr*

This code is used in section 1357*.

1365*. By sorting the list of available spaces in the variable-size portion of *mem*, we are usually able to get by without having to dump very much of the dynamic memory.

We recompute *var_used* and *dyn_used*, so that INITEX dumps valid information even when it has not been gathering statistics.

⟨ Dump the dynamic memory 1365* ⟩ ≡
sort_avail; *var_used* ← 0; *dump_int*(*lo_mem_max*); *dump_int*(*rover*);
if *eTeX_ex* **then**
  for *k* ← *int_val* **to** *inter_char_val* **do** *dump_int*(*sa_root*[*k*]);
p ← *mem_bot*; *q* ← *rover*; *x* ← 0;
repeat *dump_things*(*mem*[*p*], *q* + 2 − *p*); *x* ← *x* + *q* + 2 − *p*; *var_used* ← *var_used* + *q* − *p*;
  p ← *q* + *node_size*(*q*); *q* ← *rlink*(*q*);
until *q* = *rover*;
var_used ← *var_used* + *lo_mem_max* − *p*; *dyn_used* ← *mem_end* + 1 − *hi_mem_min*;
dump_things(*mem*[*p*], *lo_mem_max* + 1 − *p*); *x* ← *x* + *lo_mem_max* + 1 − *p*; *dump_int*(*hi_mem_min*);
dump_int(*avail*); *dump_things*(*mem*[*hi_mem_min*], *mem_end* + 1 − *hi_mem_min*);
x ← *x* + *mem_end* + 1 − *hi_mem_min*; *p* ← *avail*;
while *p* ≠ *null* **do**
  begin *decr*(*dyn_used*); *p* ← *link*(*p*);
  end;
dump_int(*var_used*); *dump_int*(*dyn_used*); *print_ln*; *print_int*(*x*);
print("␣memory␣locations␣dumped;␣current␣usage␣is␣"); *print_int*(*var_used*); *print_char*("&");
print_int(*dyn_used*)

This code is used in section 1356*.

1366* ⟨ Undump the dynamic memory 1366* ⟩ ≡

```
undump(lo_mem_stat_max + 1000)(hi_mem_stat_min − 1)(lo_mem_max);
undump(lo_mem_stat_max + 1)(lo_mem_max)(rover);
if eTeX_ex then
  for k ← int_val to inter_char_val do undump(null)(lo_mem_max)(sa_root[k]);
p ← mem_bot; q ← rover;
repeat undump_things(mem[p], q + 2 − p); p ← q + node_size(q);
  if (p > lo_mem_max) ∨ ((q ≥ rlink(q)) ∧ (rlink(q) ≠ rover)) then goto bad_fmt;
  q ← rlink(q);
until q = rover;
undump_things(mem[p], lo_mem_max + 1 − p);
if mem_min < mem_bot − 2 then { make more low memory available }
  begin p ← llink(rover); q ← mem_min + 1; link(mem_min) ← null; info(mem_min) ← null;
      { we don't use the bottom word }
  rlink(p) ← q; llink(rover) ← q;
  rlink(q) ← rover; llink(q) ← p; link(q) ← empty_flag; node_size(q) ← mem_bot − q;
  end;
undump(lo_mem_max + 1)(hi_mem_stat_min)(hi_mem_min); undump(null)(mem_top)(avail);
mem_end ← mem_top; undump_things(mem[hi_mem_min], mem_end + 1 − hi_mem_min);
undump_int(var_used); undump_int(dyn_used)
```

This code is used in section 1357*.

1368* ⟨ Undump the table of equivalents 1368* ⟩ ≡

```
⟨ Undump regions 1 to 6 of eqtb 1371* ⟩;
undump(hash_base)(hash_top)(par_loc); par_token ← cs_token_flag + par_loc;
undump(hash_base)(hash_top)(write_loc);
⟨ Undump the hash table 1373* ⟩
```

This code is used in section 1357*.

1369* The table of equivalents usually contains repeated information, so we dump it in compressed form: The sequence of $n+2$ values $(n, x_1, \ldots, x_n, m)$ in the format file represents $n+m$ consecutive entries of *eqtb*, with m extra copies of x_n, namely $(x_1, \ldots, x_n, x_n, \ldots, x_n)$.

⟨ Dump regions 1 to 4 of *eqtb* 1369* ⟩ ≡

```
k ← active_base;
repeat j ← k;
  while j < int_base − 1 do
    begin if (equiv(j) = equiv(j + 1)) ∧ (eq_type(j) = eq_type(j + 1)) ∧ (eq_level(j) = eq_level(j + 1))
        then goto found1;
    incr(j);
    end;
  l ← int_base; goto done1; { j = int_base − 1 }
found1: incr(j); l ← j;
  while j < int_base − 1 do
    begin if (equiv(j) ≠ equiv(j + 1)) ∨ (eq_type(j) ≠ eq_type(j + 1)) ∨ (eq_level(j) ≠ eq_level(j + 1))
        then goto done1;
    incr(j);
    end;
done1: dump_int(l − k); dump_things(eqtb[k], l − k); k ← j + 1; dump_int(k − l);
until k = int_base
```

This code is used in section 1367.

1370* ⟨ Dump regions 5 and 6 of *eqtb* 1370* ⟩ ≡

```
repeat j ← k;
  while j < eqtb_size do
    begin if eqtb[j].int = eqtb[j + 1].int then goto found2;
    incr(j);
    end;
  l ← eqtb_size + 1; goto done2; { j = eqtb_size }
found2: incr(j); l ← j;
  while j < eqtb_size do
    begin if eqtb[j].int ≠ eqtb[j + 1].int then goto done2;
    incr(j);
    end;
done2: dump_int(l − k); dump_things(eqtb[k], l − k); k ← j + 1; dump_int(k − l);
until k > eqtb_size;
if hash_high > 0 then dump_things(eqtb[eqtb_size + 1], hash_high); { dump hash_extra part }
```

This code is used in section 1367.

1371* ⟨ Undump regions 1 to 6 of *eqtb* 1371* ⟩ ≡

```
k ← active_base;
repeat undump_int(x);
  if (x < 1) ∨ (k + x > eqtb_size + 1) then goto bad_fmt;
  undump_things(eqtb[k], x); k ← k + x; undump_int(x);
  if (x < 0) ∨ (k + x > eqtb_size + 1) then goto bad_fmt;
  for j ← k to k + x − 1 do eqtb[j] ← eqtb[k − 1];
  k ← k + x;
until k > eqtb_size;
if hash_high > 0 then undump_things(eqtb[eqtb_size + 1], hash_high); { undump hash_extra part }
```

This code is used in section 1368*.

1372* A different scheme is used to compress the hash table, since its lower region is usually sparse. When *text*(*p*) ≠ 0 for *p* ≤ *hash_used*, we output two words, *p* and *hash*[*p*]. The hash table is, of course, densely packed for *p* ≥ *hash_used*, so the remaining entries are output in a block.

⟨ Dump the hash table 1372* ⟩ ≡

```
for p ← 0 to prim_size do dump_hh(prim[p]);
dump_int(hash_used); cs_count ← frozen_control_sequence − 1 − hash_used + hash_high;
for p ← hash_base to hash_used do
  if text(p) ≠ 0 then
    begin dump_int(p); dump_hh(hash[p]); incr(cs_count);
    end;
dump_things(hash[hash_used + 1], undefined_control_sequence − 1 − hash_used);
if hash_high > 0 then dump_things(hash[eqtb_size + 1], hash_high);
dump_int(cs_count);
print_ln; print_int(cs_count); print("␣multiletter␣control␣sequences")
```

This code is used in section 1367.

1373* ⟨ Undump the hash table 1373* ⟩ ≡

```
for p ← 0 to prim_size do undump_hh(prim[p]);
undump(hash_base)(frozen_control_sequence)(hash_used); p ← hash_base − 1;
repeat undump(p + 1)(hash_used)(p); undump_hh(hash[p]);
until p = hash_used;
undump_things(hash[hash_used + 1], undefined_control_sequence − 1 − hash_used);
if debug_format_file then
  begin print_csnames(hash_base, undefined_control_sequence − 1);
  end;
if hash_high > 0 then
  begin undump_things(hash[eqtb_size + 1], hash_high);
  if debug_format_file then
    begin print_csnames(eqtb_size + 1, hash_high − (eqtb_size + 1));
    end;
  end;
undump_int(cs_count)
```

This code is used in section 1368*.

1374* ⟨ Dump the font information 1374* ⟩ ≡

```
dump_int(fmem_ptr); dump_things(font_info[0], fmem_ptr); dump_int(font_ptr);
⟨ Dump the array info for internal font number k 1376* ⟩;
print_ln; print_int(fmem_ptr − 7); print("␣words␣of␣font␣info␣for␣");
print_int(font_ptr − font_base);
if font_ptr ≠ font_base + 1 then print("␣preloaded␣fonts")
else print("␣preloaded␣font")
```

This code is used in section 1356*.

1375* ⟨ Undump the font information 1375* ⟩ ≡

```
undump_size(7)(sup_font_mem_size)('font␣mem␣size')(fmem_ptr);
if fmem_ptr > font_mem_size then font_mem_size ← fmem_ptr;
font_info ← xmalloc_array(fmemory_word, font_mem_size); undump_things(font_info[0], fmem_ptr);
undump_size(font_base)(font_base + max_font_max)('font␣max')(font_ptr);
    { This undumps all of the font info, despite the name. }
⟨ Undump the array info for internal font number k 1377* ⟩;
```

This code is used in section 1357*.

1376*. ⟨ Dump the array info for internal font number k 1376* ⟩ ≡

```
begin dump_things(font_check[null_font], font_ptr + 1 − null_font);
dump_things(font_size[null_font], font_ptr + 1 − null_font);
dump_things(font_dsize[null_font], font_ptr + 1 − null_font);
dump_things(font_params[null_font], font_ptr + 1 − null_font);
dump_things(hyphen_char[null_font], font_ptr + 1 − null_font);
dump_things(skew_char[null_font], font_ptr + 1 − null_font);
dump_things(font_name[null_font], font_ptr + 1 − null_font);
dump_things(font_area[null_font], font_ptr + 1 − null_font);
dump_things(font_bc[null_font], font_ptr + 1 − null_font);
dump_things(font_ec[null_font], font_ptr + 1 − null_font);
dump_things(char_base[null_font], font_ptr + 1 − null_font);
dump_things(width_base[null_font], font_ptr + 1 − null_font);
dump_things(height_base[null_font], font_ptr + 1 − null_font);
dump_things(depth_base[null_font], font_ptr + 1 − null_font);
dump_things(italic_base[null_font], font_ptr + 1 − null_font);
dump_things(lig_kern_base[null_font], font_ptr + 1 − null_font);
dump_things(kern_base[null_font], font_ptr + 1 − null_font);
dump_things(exten_base[null_font], font_ptr + 1 − null_font);
dump_things(param_base[null_font], font_ptr + 1 − null_font);
dump_things(font_glue[null_font], font_ptr + 1 − null_font);
dump_things(bchar_label[null_font], font_ptr + 1 − null_font);
dump_things(font_bchar[null_font], font_ptr + 1 − null_font);
dump_things(font_false_bchar[null_font], font_ptr + 1 − null_font);
for k ← null_font to font_ptr do
  begin print_nl("\font"); print_esc(font_id_text(k)); print_char("=");
  if is_native_font(k) ∨ (font_mapping[k] ≠ 0) then
    begin print_file_name(font_name[k], "", "");
    print_err("Can't␣\dump␣a␣format␣with␣native␣fonts␣or␣font-mappings");
    help3("You␣really,␣really␣don't␣want␣to␣do␣this.")
    ("It␣won't␣work,␣and␣only␣confuses␣me.")
    ("(Load␣them␣at␣runtime,␣not␣as␣part␣of␣the␣format␣file.)"); error;
    end
  else print_file_name(font_name[k], font_area[k], "");
  if font_size[k] ≠ font_dsize[k] then
    begin print("␣at␣"); print_scaled(font_size[k]); print("pt");
    end;
  end
end
```

This code is used in section 1374*.

1377* This module should now be named 'Undump all the font arrays'.

⟨ Undump the array info for internal font number k 1377* ⟩ ≡

```
begin { Allocate the font arrays }
font_mapping ← xmalloc_array(void_pointer, font_max);
font_layout_engine ← xmalloc_array(void_pointer, font_max);
font_flags ← xmalloc_array(char, font_max); font_letter_space ← xmalloc_array(scaled, font_max);
font_check ← xmalloc_array(four_quarters, font_max); font_size ← xmalloc_array(scaled, font_max);
font_dsize ← xmalloc_array(scaled, font_max); font_params ← xmalloc_array(font_index, font_max);
font_name ← xmalloc_array(str_number, font_max); font_area ← xmalloc_array(str_number, font_max);
font_bc ← xmalloc_array(UTF16_code, font_max); font_ec ← xmalloc_array(UTF16_code, font_max);
font_glue ← xmalloc_array(halfword, font_max); hyphen_char ← xmalloc_array(integer, font_max);
skew_char ← xmalloc_array(integer, font_max); bchar_label ← xmalloc_array(font_index, font_max);
font_bchar ← xmalloc_array(nine_bits, font_max); font_false_bchar ← xmalloc_array(nine_bits, font_max);
char_base ← xmalloc_array(integer, font_max); width_base ← xmalloc_array(integer, font_max);
height_base ← xmalloc_array(integer, font_max); depth_base ← xmalloc_array(integer, font_max);
italic_base ← xmalloc_array(integer, font_max); lig_kern_base ← xmalloc_array(integer, font_max);
kern_base ← xmalloc_array(integer, font_max); exten_base ← xmalloc_array(integer, font_max);
param_base ← xmalloc_array(integer, font_max);
for k ← null_font to font_ptr do font_mapping[k] ← 0;
undump_things(font_check[null_font], font_ptr + 1 − null_font);
undump_things(font_size[null_font], font_ptr + 1 − null_font);
undump_things(font_dsize[null_font], font_ptr + 1 − null_font);
undump_checked_things(min_halfword, max_halfword, font_params[null_font], font_ptr + 1 − null_font);
undump_things(hyphen_char[null_font], font_ptr + 1 − null_font);
undump_things(skew_char[null_font], font_ptr + 1 − null_font);
undump_upper_check_things(str_ptr, font_name[null_font], font_ptr + 1 − null_font);
undump_upper_check_things(str_ptr, font_area[null_font], font_ptr + 1 − null_font); { There's no point in
    checking these values against the range [0, 255], since the data type is unsigned char, and all values
    of that type are in that range by definition. }
undump_things(font_bc[null_font], font_ptr + 1 − null_font);
undump_things(font_ec[null_font], font_ptr + 1 − null_font);
undump_things(char_base[null_font], font_ptr + 1 − null_font);
undump_things(width_base[null_font], font_ptr + 1 − null_font);
undump_things(height_base[null_font], font_ptr + 1 − null_font);
undump_things(depth_base[null_font], font_ptr + 1 − null_font);
undump_things(italic_base[null_font], font_ptr + 1 − null_font);
undump_things(lig_kern_base[null_font], font_ptr + 1 − null_font);
undump_things(kern_base[null_font], font_ptr + 1 − null_font);
undump_things(exten_base[null_font], font_ptr + 1 − null_font);
undump_things(param_base[null_font], font_ptr + 1 − null_font);
undump_checked_things(min_halfword, lo_mem_max, font_glue[null_font], font_ptr + 1 − null_font);
undump_checked_things(0, fmem_ptr − 1, bchar_label[null_font], font_ptr + 1 − null_font);
undump_checked_things(min_quarterword, non_char, font_bchar[null_font], font_ptr + 1 − null_font);
undump_checked_things(min_quarterword, non_char, font_false_bchar[null_font], font_ptr + 1 − null_font);
end
```

This code is used in section 1375*.

1378* ⟨ Dump the hyphenation tables 1378* ⟩ ≡

```
dump_int(hyph_count);
if hyph_next ≤ hyph_prime then hyph_next ← hyph_size;
dump_int(hyph_next); { minimum value of hyphen_size needed }
for k ← 0 to hyph_size do
  if hyph_word[k] ≠ 0 then
    begin dump_int(k + 65536 * hyph_link[k]);
      { assumes number of hyphen exceptions does not exceed 65535 }
    dump_int(hyph_word[k]); dump_int(hyph_list[k]);
    end;
print_ln; print_int(hyph_count);
if hyph_count ≠ 1 then print("␣hyphenation␣exceptions")
else print("␣hyphenation␣exception");
if trie_not_ready then init_trie;
dump_int(trie_max); dump_int(hyph_start); dump_things(trie_trl[0], trie_max + 1);
dump_things(trie_tro[0], trie_max + 1); dump_things(trie_trc[0], trie_max + 1); dump_int(max_hyph_char);
dump_int(trie_op_ptr); dump_things(hyf_distance[1], trie_op_ptr); dump_things(hyf_num[1], trie_op_ptr);
dump_things(hyf_next[1], trie_op_ptr); print_nl("Hyphenation␣trie␣of␣length␣"); print_int(trie_max);
print("␣has␣"); print_int(trie_op_ptr);
if trie_op_ptr ≠ 1 then print("␣ops")
else print("␣op");
print("␣out␣of␣"); print_int(trie_op_size);
for k ← biggest_lang downto 0 do
  if trie_used[k] > min_quarterword then
    begin print_nl("␣␣"); print_int(qo(trie_used[k])); print("␣for␣language␣"); print_int(k);
    dump_int(k); dump_int(qo(trie_used[k]));
    end
```

This code is used in section 1356*.

1379* Only "nonempty" parts of *op_start* need to be restored.

⟨ Undump the hyphenation tables 1379* ⟩ ≡
```
undump_size(0)(hyph_size)(´hyph_size´)(hyph_count);
undump_size(hyph_prime)(hyph_size)(´hyph_size´)(hyph_next); j ← 0;
for k ← 1 to hyph_count do
  begin undump_int(j);
  if j < 0 then goto bad_fmt;
  if j > 65535 then
    begin hyph_next ← j div 65536; j ← j − hyph_next ∗ 65536;
    end
  else hyph_next ← 0;
  if (j ≥ hyph_size) ∨ (hyph_next > hyph_size) then goto bad_fmt;
  hyph_link[j] ← hyph_next; undump(0)(str_ptr)(hyph_word[j]);
  undump(min_halfword)(max_halfword)(hyph_list[j]);
  end; { j is now the largest occupied location in hyph_word }
incr(j);
if j < hyph_prime then j ← hyph_prime;
hyph_next ← j;
if hyph_next ≥ hyph_size then hyph_next ← hyph_prime
else if hyph_next ≥ hyph_prime then incr(hyph_next);
undump_size(0)(trie_size)(´trie␣size´)(j); init trie_max ← j; tini undump(0)(j)(hyph_start);
    { These first three haven't been allocated yet unless we're INITEX; we do that precisely so we don't
    allocate more space than necessary. }
if ¬trie_trl then trie_trl ← xmalloc_array(trie_pointer, j + 1);
undump_things(trie_trl[0], j + 1);
if ¬trie_tro then trie_tro ← xmalloc_array(trie_pointer, j + 1);
undump_things(trie_tro[0], j + 1);
if ¬trie_trc then trie_trc ← xmalloc_array(quarterword, j + 1);
undump_things(trie_trc[0], j + 1); undump_int(max_hyph_char);
undump_size(0)(trie_op_size)(´trie␣op␣size´)(j); init trie_op_ptr ← j; tini
    { I'm not sure we have such a strict limitation (64) on these values, so let's leave them unchecked. }
undump_things(hyf_distance[1], j); undump_things(hyf_num[1], j);
undump_upper_check_things(max_trie_op, hyf_next[1], j);
init for k ← 0 to biggest_lang do trie_used[k] ← min_quarterword;
tini
k ← biggest_lang + 1;
while j > 0 do
  begin undump(0)(k − 1)(k); undump(1)(j)(x); init trie_used[k] ← qi(x); tini
  j ← j − x; op_start[k] ← qo(j);
  end;
init trie_not_ready ← false tini
```
This code is used in section 1357*.

1381* ⟨ Undump a couple more things and the closing check word 1381* ⟩ ≡
```
undump(batch_mode)(error_stop_mode)(interaction);
if interaction_option ≠ unspecified_mode then interaction ← interaction_option;
undump(0)(str_ptr)(format_ident); undump_int(x);
if x ≠ 69069 then goto bad_fmt
```
This code is used in section 1357*.

1386* Now this is really it: TEX starts and ends here.

The initial test involving *ready_already* should be deleted if the Pascal runtime system is smart enough to detect such a "mistake."

```
define const_chk(#) ≡
        begin if # < inf@&# then # ← inf@&#
        else if # > sup@&# then # ← sup@&#
        end { setup_bound_var stuff duplicated in mf.ch. }
define setup_bound_var(#) ≡ bound_default ← #; setup_bound_var_end
define setup_bound_var_end(#) ≡ bound_name ← #; setup_bound_var_end_end
define setup_bound_var_end_end(#) ≡ setup_bound_variable(addressof(#), bound_name, bound_default)
procedure main_body;
begin { start_here }
  { Bounds that may be set from the configuration file. We want the user to be able to specify the names
    with underscores, but TANGLE removes underscores, so we're stuck giving the names twice, once as a
    string, once as the identifier. How ugly. }
setup_bound_var(0)('mem_bot')(mem_bot); setup_bound_var(250000)('main_memory')(main_memory);
    { memory_words for mem in INITEX }
setup_bound_var(0)('extra_mem_top')(extra_mem_top); { increase high mem in VIRTEX }
setup_bound_var(0)('extra_mem_bot')(extra_mem_bot); { increase low mem in VIRTEX }
setup_bound_var(200000)('pool_size')(pool_size);
setup_bound_var(75000)('string_vacancies')(string_vacancies);
setup_bound_var(5000)('pool_free')(pool_free); { min pool avail after fmt }
setup_bound_var(15000)('max_strings')(max_strings); max_strings ← max_strings + too_big_char;
    { the max_strings value doesn't include the 64K synthetic strings }
setup_bound_var(100)('strings_free')(strings_free);
setup_bound_var(100000)('font_mem_size')(font_mem_size);
setup_bound_var(500)('font_max')(font_max); setup_bound_var(20000)('trie_size')(trie_size);
    { if ssup_trie_size increases, recompile }
setup_bound_var(659)('hyph_size')(hyph_size); setup_bound_var(3000)('buf_size')(buf_size);
setup_bound_var(50)('nest_size')(nest_size); setup_bound_var(15)('max_in_open')(max_in_open);
setup_bound_var(60)('param_size')(param_size); setup_bound_var(4000)('save_size')(save_size);
setup_bound_var(300)('stack_size')(stack_size);
setup_bound_var(16384)('dvi_buf_size')(dvi_buf_size); setup_bound_var(79)('error_line')(error_line);
setup_bound_var(50)('half_error_line')(half_error_line);
setup_bound_var(79)('max_print_line')(max_print_line);
setup_bound_var(0)('hash_extra')(hash_extra);
setup_bound_var(10000)('expand_depth')(expand_depth); const_chk(mem_bot);
const_chk(main_memory); Init extra_mem_top ← 0; extra_mem_bot ← 0; Tini
if extra_mem_bot > sup_main_memory then extra_mem_bot ← sup_main_memory;
if extra_mem_top > sup_main_memory then extra_mem_top ← sup_main_memory;
    { mem_top is an index, main_memory a size }
mem_top ← mem_bot + main_memory − 1; mem_min ← mem_bot; mem_max ← mem_top;
    { Check other constants against their sup and inf. }
const_chk(trie_size); const_chk(hyph_size); const_chk(buf_size); const_chk(nest_size);
const_chk(max_in_open); const_chk(param_size); const_chk(save_size); const_chk(stack_size);
const_chk(dvi_buf_size); const_chk(pool_size); const_chk(string_vacancies); const_chk(pool_free);
const_chk(max_strings); const_chk(strings_free); const_chk(font_mem_size); const_chk(font_max);
const_chk(hash_extra);
if error_line > ssup_error_line then error_line ← ssup_error_line; { array memory allocation }
buffer ← xmalloc_array(UnicodeScalar, buf_size); nest ← xmalloc_array(list_state_record, nest_size);
save_stack ← xmalloc_array(memory_word, save_size);
input_stack ← xmalloc_array(in_state_record, stack_size);
```

```
  input_file ← xmalloc_array(unicode_file, max_in_open);
  line_stack ← xmalloc_array(integer, max_in_open); eof_seen ← xmalloc_array(boolean, max_in_open);
  grp_stack ← xmalloc_array(save_pointer, max_in_open); if_stack ← xmalloc_array(pointer, max_in_open);
  source_filename_stack ← xmalloc_array(str_number, max_in_open);
  full_source_filename_stack ← xmalloc_array(str_number, max_in_open);
  param_stack ← xmalloc_array(halfword, param_size); dvi_buf ← xmalloc_array(eight_bits, dvi_buf_size);
  hyph_word ← xmalloc_array(str_number, hyph_size);
  hyph_list ← xmalloc_array(halfword, hyph_size); hyph_link ← xmalloc_array(hyph_pointer, hyph_size);
    Init yzmem ← xmalloc_array(memory_word, mem_top − mem_bot + 1);
  zmem ← yzmem − mem_bot; { Some compilers require mem_bot = 0 }
  eqtb_top ← eqtb_size + hash_extra;
  if hash_extra = 0 then hash_top ← undefined_control_sequence
  else hash_top ← eqtb_top;
  yhash ← xmalloc_array(two_halves, 1 + hash_top − hash_offset); hash ← yhash − hash_offset;
    { Some compilers require hash_offset = 0 }
  next(hash_base) ← 0; text(hash_base) ← 0;
  for hash_used ← hash_base + 1 to hash_top do hash[hash_used] ← hash[hash_base];
  zeqtb ← xmalloc_array(memory_word, eqtb_top); eqtb ← zeqtb;
  str_start ← xmalloc_array(pool_pointer, max_strings);
  str_pool ← xmalloc_array(packed_ASCII_code, pool_size);
  font_info ← xmalloc_array(fmemory_word, font_mem_size); Tinihistory ← fatal_error_stop;
    { in case we quit during initialization }
  t_open_out; { open the terminal for output }
  if ready_already = 314159 then goto start_of_TEX;
  ⟨ Check the "constant" values for consistency 14 ⟩
  if bad > 0 then
    begin wterm_ln('Ouch---my␣internal␣constants␣have␣been␣clobbered!', '---case␣', bad : 1);
    goto final_end;
    end;
  initialize; { set global variables to their starting values }
  Init if ¬get_strings_started then goto final_end;
  init_prim; { call primitive for each primitive }
  init_str_ptr ← str_ptr; init_pool_ptr ← pool_ptr; fix_date_and_time;
  Tini
  ready_already ← 314159;
start_of_TEX: ⟨ Initialize the output routines 55 ⟩;
  ⟨ Get the first line of input and prepare to start 1391* ⟩;
  history ← spotless; { ready to go! }
  ⟨ Initialize synctex primitive 1711* ⟩main_control; { come to life }
  final_cleanup; { prepare for death }
  close_files_and_terminate;
final_end: do_final_end;
  end { main_body }
  ;
```

1387*. Here we do whatever is needed to complete TEX's job gracefully on the local operating system. The code here might come into play after a fatal error; it must therefore consist entirely of "safe" operations that cannot produce error messages. For example, it would be a mistake to call *str_room* or *make_string* at this time, because a call on *overflow* might lead to an infinite loop. (Actually there's one way to get error messages, via *prepare_mag*; but that can't cause infinite recursion.)

If *final_cleanup* is bypassed, this program doesn't bother to close the input files that may still be open.

```
⟨ Last-minute procedures 1387* ⟩ ≡
procedure close_files_and_terminate;
  var k: integer;  { all-purpose index }
  begin ⟨ Finish the extensions 1441 ⟩;
  new_line_char ← −1;
  stat if tracing_stats > 0 then ⟨ Output statistics about this job 1388* ⟩; tats
  wake_up_terminal; ⟨ Finish the DVI file 680* ⟩;
  ⟨ Close SyncTEX file and write status 1719* ⟩;
  if log_opened then
    begin wlog_cr; a_close(log_file); selector ← selector − 2;
    if selector = term_only then
      begin print_nl("Transcript␣written␣on␣"); print(log_name); print_char(".");
      end;
    end;
  print_ln;
  if (edit_name_start ≠ 0) ∧ (interaction > batch_mode) then
    call_edit(str_pool, edit_name_start, edit_name_length, edit_line);
  end;
```

See also sections 1389*, 1390, and 1392*.

This code is used in section 1384.

1388*. The present section goes directly to the log file instead of using *print* commands, because there's no need for these strings to take up *str_pool* memory when a non-**stat** version of TEX is being used.

```
⟨ Output statistics about this job 1388* ⟩ ≡
  if log_opened then
    begin wlog_ln(´␣´); wlog_ln(´Here␣is␣how␣much␣of␣TeX´´s␣memory´, ´␣you␣used:´);
    wlog(´␣´, str_ptr − init_str_ptr : 1, ´␣string´);
    if str_ptr ≠ init_str_ptr + 1 then wlog(´s´);
    wlog_ln(´␣out␣of␣´, max_strings − init_str_ptr : 1);
    wlog_ln(´␣´, pool_ptr − init_pool_ptr : 1, ´␣string␣characters␣out␣of␣´, pool_size − init_pool_ptr : 1);
    wlog_ln(´␣´, lo_mem_max − mem_min + mem_end − hi_mem_min + 2 : 1,
        ´␣words␣of␣memory␣out␣of␣´, mem_end + 1 − mem_min : 1);
    wlog_ln(´␣´, cs_count : 1, ´␣multiletter␣control␣sequences␣out␣of␣´, hash_size : 1, ´+´,
        hash_extra : 1);
    wlog(´␣´, fmem_ptr : 1, ´␣words␣of␣font␣info␣for␣´, font_ptr − font_base : 1, ´␣font´);
    if font_ptr ≠ font_base + 1 then wlog(´s´);
    wlog_ln(´,␣out␣of␣´, font_mem_size : 1, ´␣for␣´, font_max − font_base : 1);
    wlog(´␣´, hyph_count : 1, ´␣hyphenation␣exception´);
    if hyph_count ≠ 1 then wlog(´s´);
    wlog_ln(´␣out␣of␣´, hyph_size : 1);
    wlog_ln(´␣´, max_in_stack : 1, ´i,´, max_nest_stack : 1, ´n,´, max_param_stack : 1, ´p,´,
        max_buf_stack + 1 : 1, ´b,´, max_save_stack + 6 : 1, ´s␣stack␣positions␣out␣of␣´,
        stack_size : 1, ´i,´, nest_size : 1, ´n,´, param_size : 1, ´p,´, buf_size : 1, ´b,´, save_size : 1, ´s´);
    end
```

This code is used in section 1387*.

1389* We get to the *final_cleanup* routine when `\end` or `\dump` has been scanned and *its_all_over*.

```
⟨ Last-minute procedures 1387* ⟩ +≡
procedure final_cleanup;
  label exit;
  var c: small_number;  { 0 for \end, 1 for \dump }
  begin c ← cur_chr;
  if c ≠ 1 then new_line_char ← −1;
  if job_name = 0 then open_log_file;
  while input_ptr > 0 do
    if state = token_list then end_token_list else end_file_reading;
  while open_parens > 0 do
    begin print("␣)"); decr(open_parens);
    end;
  if cur_level > level_one then
    begin print_nl("("); print_esc("end␣occurred␣"); print("inside␣a␣group␣at␣level␣");
    print_int(cur_level − level_one); print_char(")");
    if eTeX_ex then show_save_groups;
    end;
  while cond_ptr ≠ null do
    begin print_nl("("); print_esc("end␣occurred␣"); print("when␣"); print_cmd_chr(if_test, cur_if);
    if if_line ≠ 0 then
      begin print("␣on␣line␣"); print_int(if_line);
      end;
    print("␣was␣incomplete)"); if_line ← if_line_field(cond_ptr); cur_if ← subtype(cond_ptr);
    temp_ptr ← cond_ptr; cond_ptr ← link(cond_ptr); free_node(temp_ptr, if_node_size);
    end;
  if history ≠ spotless then
    if ((history = warning_issued) ∨ (interaction < error_stop_mode)) then
      if selector = term_and_log then
        begin selector ← term_only;
        print_nl("(see␣the␣transcript␣file␣for␣additional␣information)");
        selector ← term_and_log;
        end;
  if c = 1 then
    begin Init for c ← top_mark_code to split_bot_mark_code do
      if cur_mark[c] ≠ null then delete_token_ref(cur_mark[c]);
    if sa_mark ≠ null then
      if do_marks(destroy_marks, 0, sa_mark) then sa_mark ← null;
    for c ← last_box_code to vsplit_code do flush_node_list(disc_ptr[c]);
    if last_glue ≠ max_halfword then delete_glue_ref(last_glue);
    store_fmt_file; return; Tini
    print_nl("(\dump␣is␣performed␣only␣by␣INITEX)"); return;
    end;
exit: end;
```

1391* When we begin the following code, TEX's tables may still contain garbage; the strings might not even be present. Thus we must proceed cautiously to get bootstrapped in.

But when we finish this part of the program, TEX is ready to call on the *main_control* routine to do its work.

```
⟨ Get the first line of input and prepare to start 1391* ⟩ ≡
  begin ⟨ Initialize the input routines 361* ⟩;
  ⟨ Enable ε-TEX, if requested 1451* ⟩
  if (format_ident = 0) ∨ (buffer[loc] = "&") ∨ dump_line then
    begin if format_ident ≠ 0 then initialize;  { erase preloaded format }
    if ¬open_fmt_file then goto final_end;
    if ¬load_fmt_file then
      begin w_close(fmt_file); goto final_end;
      end;
    w_close(fmt_file); eqtb ← zeqtb;
    while (loc < limit) ∧ (buffer[loc] = "␣") do incr(loc);
    end;
  if eTeX_ex then wterm_ln('entering␣extended␣mode');
  if end_line_char_inactive then decr(limit)
  else buffer[limit] ← end_line_char;
  if mltex_enabled_p then
    begin wterm_ln('MLTeX␣v2.2␣enabled');
    end;
  fix_date_and_time;
  init if trie_not_ready then
    begin  { initex without format loaded }
    trie_trl ← xmalloc_array(trie_pointer, trie_size); trie_tro ← xmalloc_array(trie_pointer, trie_size);
    trie_trc ← xmalloc_array(quarterword, trie_size); trie_c ← xmalloc_array(packed_ASCII_code, trie_size);
    trie_o ← xmalloc_array(trie_opcode, trie_size); trie_l ← xmalloc_array(trie_pointer, trie_size);
    trie_r ← xmalloc_array(trie_pointer, trie_size); trie_hash ← xmalloc_array(trie_pointer, trie_size);
    trie_taken ← xmalloc_array(boolean, trie_size); trie_root ← 0; trie_c[0] ← si(0); trie_ptr ← 0;
    hyph_root ← 0; hyph_start ← 0;  { Allocate and initialize font arrays }
    font_mapping ← xmalloc_array(void_pointer, font_max);
    font_layout_engine ← xmalloc_array(void_pointer, font_max);
    font_flags ← xmalloc_array(char, font_max); font_letter_space ← xmalloc_array(scaled, font_max);
    font_check ← xmalloc_array(four_quarters, font_max); font_size ← xmalloc_array(scaled, font_max);
    font_dsize ← xmalloc_array(scaled, font_max); font_params ← xmalloc_array(font_index, font_max);
    font_name ← xmalloc_array(str_number, font_max);
    font_area ← xmalloc_array(str_number, font_max); font_bc ← xmalloc_array(UTF16_code, font_max);
    font_ec ← xmalloc_array(UTF16_code, font_max); font_glue ← xmalloc_array(halfword, font_max);
    hyphen_char ← xmalloc_array(integer, font_max); skew_char ← xmalloc_array(integer, font_max);
    bchar_label ← xmalloc_array(font_index, font_max); font_bchar ← xmalloc_array(nine_bits, font_max);
    font_false_bchar ← xmalloc_array(nine_bits, font_max); char_base ← xmalloc_array(integer, font_max);
    width_base ← xmalloc_array(integer, font_max); height_base ← xmalloc_array(integer, font_max);
    depth_base ← xmalloc_array(integer, font_max); italic_base ← xmalloc_array(integer, font_max);
    lig_kern_base ← xmalloc_array(integer, font_max); kern_base ← xmalloc_array(integer, font_max);
    exten_base ← xmalloc_array(integer, font_max); param_base ← xmalloc_array(integer, font_max);
    font_ptr ← null_font; fmem_ptr ← 7; font_name[null_font] ← "nullfont"; font_area[null_font] ← "";
    hyphen_char[null_font] ← "-"; skew_char[null_font] ← −1; bchar_label[null_font] ← non_address;
    font_bchar[null_font] ← non_char; font_false_bchar[null_font] ← non_char; font_bc[null_font] ← 1;
    font_ec[null_font] ← 0; font_size[null_font] ← 0; font_dsize[null_font] ← 0; char_base[null_font] ← 0;
    width_base[null_font] ← 0; height_base[null_font] ← 0; depth_base[null_font] ← 0;
    italic_base[null_font] ← 0; lig_kern_base[null_font] ← 0; kern_base[null_font] ← 0;
```

```
    exten_base[null_font] ← 0; font_glue[null_font] ← null; font_params[null_font] ← 7;
    font_mapping[null_font] ← 0; param_base[null_font] ← −1;
    for font_k ← 0 to 6 do font_info[font_k].sc ← 0;
    end;
  tini
  font_used ← xmalloc_array(boolean, font_max);
  for font_k ← font_base to font_max do font_used[font_k] ← false;
  random_seed ← (microseconds ∗ 1000) + (epochseconds mod 1000000);
  init_randoms(random_seed);
  ⟨ Compute the magic offset 813 ⟩;
  ⟨ Initialize the print selector based on interaction 79 ⟩;
  if (loc < limit) ∧ (cat_code(buffer[loc]) ≠ escape) then start_input; { \input assumed }
  end
```

This code is used in section 1386*.

1392* Debugging. Once TEX is working, you should be able to diagnose most errors with the `\show` commands and other diagnostic features. But for the initial stages of debugging, and for the revelation of really deep mysteries, you can compile TEX with a few more aids, including the Pascal runtime checks and its debugger. An additional routine called *debug_help* will also come into play when you type 'D' after an error message; *debug_help* also occurs just before a fatal error causes TEX to succumb.

The interface to *debug_help* is primitive, but it is good enough when used with a Pascal debugger that allows you to set breakpoints and to read variables and change their values. After getting the prompt '`debug #`', you type either a negative number (this exits *debug_help*), or zero (this goes to a location where you can set a breakpoint, thereby entering into dialog with the Pascal debugger), or a positive number m followed by an argument n. The meaning of m and n will be clear from the program below. (If $m = 13$, there is an additional argument, l.)

```
define breakpoint = 888 { place where a breakpoint is desirable }
⟨ Last-minute procedures 1387* ⟩ +≡
  debug procedure debug_help; { routine to display various things }
  label breakpoint, exit;
  var k, l, m, n: integer;
  begin clear_terminal;
  loop
    begin wake_up_terminal; print_nl("debug␣#␣(-1␣to␣exit):"); update_terminal; read(term_in, m);
    if m < 0 then return
    else if m = 0 then dump_core { do something to cause a core dump }
      else begin read(term_in, n);
        case m of
        ⟨ Numbered cases for debug_help 1393* ⟩
        othercases print("?")
        endcases;
        end;
    end;
exit: end;
  gubed
```

1393* ⟨ Numbered cases for *debug_help* 1393* ⟩ ≡

```
1: print_word(mem[n]);  { display mem[n] in all forms }
2: print_int(info(n));
3: print_int(link(n));
4: print_word(eqtb[n]);
5: begin print_scaled(font_info[n].sc); print_char("␣");
   print_int(font_info[n].qqqq.b0); print_char(":");
   print_int(font_info[n].qqqq.b1); print_char(":");
   print_int(font_info[n].qqqq.b2); print_char(":");
   print_int(font_info[n].qqqq.b3);
   end;
6: print_word(save_stack[n]);
7: show_box(n);  { show a box, abbreviated by show_box_depth and show_box_breadth }
8: begin breadth_max ← 10000; depth_threshold ← pool_size − pool_ptr − 10; show_node_list(n);
       { show a box in its entirety }
   end;
9: show_token_list(n, null, 1000);
10: slow_print(n);
11: check_mem(n > 0);  { check wellformedness; print new busy locations if n > 0 }
12: search_mem(n);  { look for pointers to n }
13: begin read(term_in, l); print_cmd_chr(n, l);
   end;
14: for k ← 0 to n do print(buffer[k]);
15: begin font_in_short_display ← null_font; short_display(n);
   end;
16: panicking ← ¬panicking;
```

This code is used in section 1392*.

1398* Extensions might introduce new command codes; but it's best to use *extension* with a modifier, whenever possible, so that *main_control* stays the same.

```
define immediate_code = 4 { command modifier for \immediate }
define set_language_code = 5 { command modifier for \setlanguage }
define pdftex_first_extension_code = 6
define pdf_save_pos_node ≡ pdftex_first_extension_code + 15
define reset_timer_code ≡ pdftex_first_extension_code + 25
define set_random_seed_code ≡ pdftex_first_extension_code + 27
define pic_file_code = 41 { command modifier for \XeTeXpicfile, skipping codes pdfTeX might use }
define pdf_file_code = 42 { command modifier for \XeTeXpdffile }
define glyph_code = 43 { command modifier for \XeTeXglyph }
define XeTeX_input_encoding_extension_code = 44
define XeTeX_default_encoding_extension_code = 45
define XeTeX_linebreak_locale_extension_code = 46
```

⟨ Put each of TEX's primitives into the hash table 252 ⟩ +≡

```
primitive("openout", extension, open_node);
primitive("write", extension, write_node); write_loc ← cur_val;
primitive("closeout", extension, close_node);
primitive("special", extension, special_node);
text(frozen_special) ← "special"; eqtb[frozen_special] ← eqtb[cur_val];
primitive("immediate", extension, immediate_code);
primitive("setlanguage", extension, set_language_code);
primitive("resettimer", extension, reset_timer_code);
primitive("setrandomseed", extension, set_random_seed_code);
```

1403* ⟨ Declare action procedures for use by *main_control* 1097 ⟩ +≡

```
⟨ Declare procedures needed in do_extension 1404 ⟩
procedure do_extension;
  var i, j, k: integer; { all-purpose integers }
    p: pointer; { all-purpose pointers }
  begin case cur_chr of
  open_node: ⟨ Implement \openout 1406 ⟩;
  write_node: ⟨ Implement \write 1407 ⟩;
  close_node: ⟨ Implement \closeout 1408 ⟩;
  special_node: ⟨ Implement \special 1409 ⟩;
  immediate_code: ⟨ Implement \immediate 1438 ⟩;
  set_language_code: ⟨ Implement \setlanguage 1440 ⟩;
  pdf_save_pos_node: ⟨ Implement \pdfsavepos 1450 ⟩;
  reset_timer_code: ⟨ Implement \resettimer 1414 ⟩;
  set_random_seed_code: ⟨ Implement \setrandomseed 1413 ⟩;
  pic_file_code: ⟨ Implement \XeTeXpicfile 1442 ⟩;
  pdf_file_code: ⟨ Implement \XeTeXpdffile 1443 ⟩;
  glyph_code: ⟨ Implement \XeTeXglyph 1444 ⟩;
  XeTeX_input_encoding_extension_code: ⟨ Implement \XeTeXinputencoding 1446 ⟩;
  XeTeX_default_encoding_extension_code: ⟨ Implement \XeTeXdefaultencoding 1447 ⟩;
  XeTeX_linebreak_locale_extension_code: ⟨ Implement \XeTeXlinebreaklocale 1448 ⟩;
  othercases confusion("ext1")
  endcases;
  end;
```

1405*. The next subroutine uses *cur_chr* to decide what sort of whatsit is involved, and also inserts a *write_stream* number.

```
⟨ Declare procedures needed in do_extension 1404 ⟩ +≡
procedure new_write_whatsit(w : small_number);
  begin new_whatsit(cur_chr, w);
  if w ≠ write_node_size then scan_four_bit_int
  else begin scan_int;
    if cur_val < 0 then cur_val ← 17
    else if (cur_val > 15) ∧ (cur_val ≠ 18) then cur_val ← 16;
    end;
  write_stream(tail) ← cur_val;
  end;
```

```
1433* ⟨Declare procedures needed in hlist_out, vlist_out 1431⟩ +≡
procedure write_out(p : pointer);
  var old_setting: 0 .. max_selector; { holds print selector }
    old_mode: integer; { saved mode }
    j: small_number; { write stream number }
    k: integer; q, r: pointer; { temporary variables for list manipulation }
    d: integer; { number of characters in incomplete current string }
    clobbered: boolean; { system string is ok? }
    runsystem_ret: integer; { return value from runsystem }
  begin ⟨Expand macros in the token list and make link(def_ref) point to the result 1434⟩;
  old_setting ← selector; j ← write_stream(p);
  if j = 18 then selector ← new_string
  else if write_open[j] then selector ← j
    else begin { write to the terminal if file isn't open }
      if (j = 17) ∧ (selector = term_and_log) then selector ← log_only;
      print_nl("");
      end;
  token_show(def_ref); print_ln; flush_list(def_ref);
  if j = 18 then
    begin if (tracing_online ≤ 0) then selector ← log_only { Show what we're doing in the log file. }
    else selector ← term_and_log; { Show what we're doing. }
        { If the log file isn't open yet, we can only send output to the terminal. Calling open_log_file from
          here seems to result in bad data in the log. }
    if ¬log_opened then selector ← term_only;
    print_nl("runsystem(");
    for d ← 0 to cur_length − 1 do
      begin { print gives up if passed str_ptr, so do it by hand. }
      print(so(str_pool[str_start_macro(str_ptr) + d])); { N.B.: not print_char }
      end;
    print(")...");
    if shellenabledp then
      begin str_room(1); append_char(0); { Append a null byte to the expansion. }
      clobbered ← false;
      for d ← 0 to cur_length − 1 do { Convert to external character set. }
        begin if (str_pool[str_start_macro(str_ptr) + d] = null_code) ∧ (d < cur_length − 1) then
          clobbered ← true; { minimal checking: NUL not allowed in argument string of system() }
        end;
      if clobbered then print("clobbered")
      else begin { We have the command. See if we're allowed to execute it, and report in the log. We
            don't check the actual exit status of the command, or do anything with the output. }
        if name_of_file then libc_free(name_of_file);
        name_of_file ← xmalloc(cur_length * 3 + 2); k ← 0;
        for d ← 0 to cur_length − 1 do append_to_name(str_pool[str_start_macro(str_ptr) + d]);
        name_of_file[k + 1] ← 0; runsystem_ret ← runsystem(conststringcast(name_of_file + 1));
        if runsystem_ret = −1 then print("quotation␣error␣in␣system␣command")
        else if runsystem_ret = 0 then print("disabled␣(restricted)")
          else if runsystem_ret = 1 then print("executed")
            else if runsystem_ret = 2 then print("executed␣safely␣(allowed)")
        end;
      end
    else begin print("disabled"); { shellenabledp false }
      end;
```

```
    print_char("."); print_nl(""); print_ln; pool_ptr ← str_start_macro(str_ptr); { erase the string }
    end;
  selector ← old_setting;
  end;
```

1436*. The *out_what* procedure takes care of outputting whatsit nodes for *vlist_out* and *hlist_out*.

```
⟨ Declare procedures needed in hlist_out, vlist_out 1431 ⟩ +≡
procedure pic_out(p : pointer);
  var old_setting: 0 .. max_selector; { holds print selector }
    i: integer; k: pool_pointer; { index into str_pool }
  begin synch_h; synch_v; old_setting ← selector; selector ← new_string; print("pdf:image␣");
  print("matrix␣"); print_scaled(pic_transform1(p)); print("␣"); print_scaled(pic_transform2(p));
  print("␣"); print_scaled(pic_transform3(p)); print("␣"); print_scaled(pic_transform4(p)); print("␣");
  print_scaled(pic_transform5(p)); print("␣"); print_scaled(pic_transform6(p)); print("␣");
  print("page␣"); print_int(pic_page(p)); print("␣");
  case pic_pdf_box(p) of
  pdfbox_crop: print("pagebox␣cropbox␣");
  pdfbox_media: print("pagebox␣mediabox␣");
  pdfbox_bleed: print("pagebox␣bleedbox␣");
  pdfbox_art: print("pagebox␣artbox␣");
  pdfbox_trim: print("pagebox␣trimbox␣");
  others: do_nothing;
  endcases; print("(");
  for i ← 0 to pic_path_length(p) − 1 do print_visible_char(pic_path_byte(p, i));
  print(")"); selector ← old_setting;
  if cur_length < 256 then
    begin dvi_out(xxx1); dvi_out(cur_length);
    end
  else begin dvi_out(xxx4); dvi_four(cur_length);
    end;
  for k ← str_start_macro(str_ptr) to pool_ptr − 1 do dvi_out(so(str_pool[k]));
  pool_ptr ← str_start_macro(str_ptr); { erase the string }
  end;
procedure out_what(p : pointer);
  var j: small_number; { write stream number }
    old_setting: 0 .. max_selector;
  begin case subtype(p) of
  open_node, write_node, close_node: ⟨ Do some work that has been queued up for \write 1437* ⟩;
  special_node: special_out(p);
  language_node: do_nothing;
  othercases confusion("ext4")
  endcases;
  end;
```

1437*. We don't implement `\write` inside of leaders. (The reason is that the number of times a leader box appears might be different in different implementations, due to machine-dependent rounding in the glue calculations.)

```
⟨ Do some work that has been queued up for \write 1437* ⟩ ≡
  if ¬doing_leaders then
    begin j ← write_stream(p);
    if subtype(p) = write_node then write_out(p)
    else begin if write_open[j] then
        begin a_close(write_file[j]); write_open[j] ← false;
        end;
      if subtype(p) = close_node then do_nothing { already closed }
      else if j < 16 then
          begin cur_name ← open_name(p); cur_area ← open_area(p); cur_ext ← open_ext(p);
          if cur_ext = "" then cur_ext ← ".tex";
          pack_cur_name;
          while ¬kpse_out_name_ok(stringcast(name_of_file + 1)) ∨ ¬a_open_out(write_file[j]) do
            prompt_file_name("output␣file␣name", ".tex");
          write_open[j] ← true; { If on first line of input, log file is not ready yet, so don't log. }
          if log_opened ∧ texmf_yesno(´log_openout´) then
            begin old_setting ← selector;
            if (tracing_online ≤ 0) then selector ← log_only { Show what we're doing in the log file. }
            else selector ← term_and_log; { Show what we're doing. }
            print_nl("\openout"); print_int(j); print("␣=␣`");
            print_file_name(cur_name, cur_area, cur_ext); print("´."); print_nl(""); print_ln;
            selector ← old_setting;
            end;
          end;
      end;
    end
```

This code is used in section 1436*.

1451* The extended features of ε-TEX. The program has two modes of operation: (1) In TEX compatibility mode it fully deserves the name TEX and there are neither extended features nor additional primitive commands. There are, however, a few modifications that would be legitimate in any implementation of TEX such as, e.g., preventing inadequate results of the glue to `DVI` unit conversion during *ship_out*. (2) In extended mode there are additional primitive commands and the extended features of ε-TEX are available.

The distinction between these two modes of operation initially takes place when a 'virgin' `eINITEX` starts without reading a format file. Later on the values of all ε-TEX state variables are inherited when `eVIRTEX` (or `eINITEX`) reads a format file.

The code below is designed to work for cases where '**init** ... **tini**' is a run-time switch.

```
⟨ Enable ε-TEX, if requested 1451* ⟩ ≡
  init if (etex_p ∨ (buffer[loc] = "*")) ∧ (format_ident = "␣(INITEX)") then
    begin no_new_control_sequence ← false; ⟨ Generate all ε-TEX primitives 1399 ⟩
    if buffer[loc] = "*" then incr(loc);
    eTeX_mode ← 1; { enter extended mode }
    ⟨ Initialize variables for ε-TEX extended mode 1624 ⟩
    end;
  tini
  if ¬no_new_control_sequence then { just entered extended mode ? }
    no_new_control_sequence ← true else
```

This code is used in section 1391*.

1462* define *eTeX_ex* ≡ (*eTeX_mode* = 1) { is this extended mode? }

```
⟨ Global variables 13 ⟩ +≡
eTeX_mode: 0 .. 1; { identifies compatibility and extended mode }
etex_p: boolean; { was the -etex option specified }
```

1470* In order to handle `\everyeof` we need an array *eof_seen* of boolean variables.

```
⟨ Global variables 13 ⟩ +≡
eof_seen: ↑boolean; { has eof been seen? }
```

1531* We detach the hlist, start a new one consisting of just one kern node, append the reversed list, and set the width of the kern node.

```
⟨ Reverse the complete hlist and set the subtype to reversed 1531* ⟩ ≡
  begin save_h ← cur_h; temp_ptr ← p; p ← new_kern(0); sync_tag(p + medium_node_size) ← 0;
      { SyncTEX: do nothing, it is too late }
  link(prev_p) ← p; cur_h ← 0; link(p) ← reverse(this_box, null, cur_g, cur_glue); width(p) ← −cur_h;
  cur_h ← save_h; set_box_lr(this_box)(reversed);
  end
```

This code is used in section 1524.

1532* We detach the remainder of the hlist, replace the math node by an edge node, and append the reversed hlist segment to it; the tail of the reversed segment is another edge node and the remainder of the original list is attached to it.

```
⟨ Reverse an hlist segment and goto reswitch 1532* ⟩ ≡
  begin save_h ← cur_h; temp_ptr ← link(p); rule_wd ← width(p); free_node(p, medium_node_size);
      { SyncTEX: p is a math_node }
  cur_dir ← reflected; p ← new_edge(cur_dir, rule_wd); link(prev_p) ← p;
  cur_h ← cur_h − left_edge + rule_wd; link(p) ← reverse(this_box, new_edge(reflected, 0), cur_g, cur_glue);
  edge_dist(p) ← cur_h; cur_dir ← reflected; cur_h ← save_h; goto reswitch;
  end
```

This code is used in section 1527.

1535* ⟨ Move the non-*char_node* p to the new list 1535* ⟩ ≡

```
begin q ← link(p);
case type(p) of
hlist_node, vlist_node, rule_node, kern_node: rule_wd ← width(p);
⟨ Cases of reverse that need special treatment 1536 ⟩
edge_node: confusion("LR2");
othercases goto next_p
endcases;
cur_h ← cur_h + rule_wd;
next_p: link(p) ← l;
if type(p) = kern_node then
  if (rule_wd = 0) ∨ (l = null) then
    begin free_node(p, medium_node_size); p ← l;
    end;
l ← p; p ← q;
end
```

This code is used in section 1534.

1540* Finally we have found the end of the hlist segment to be reversed; the final math node is released and the remaining list attached to the edge node terminating the reversed segment.

⟨ Finish the reversed hlist segment and **goto** *done* 1540* ⟩ ≡

```
begin free_node(p, medium_node_size);  { SyncTEX: p is a kern_node }
link(t) ← q; width(t) ← rule_wd; edge_dist(t) ← −cur_h − rule_wd; goto done;
end
```

This code is used in section 1539.

1544* When calculating the natural width, w, of the final line preceding the display, we may have to copy all or part of its hlist. We copy, however, only those parts of the original list that are relevant for the computation of *pre_display_size*.

```
⟨ Declare subprocedures for init_math 1544* ⟩ ≡
procedure just_copy(p, h, t : pointer);
  label found, not_found;
  var r: pointer; { current node being fabricated for new list }
    words: 0 .. 5; { number of words remaining to be copied }
  begin while p ≠ null do
    begin words ← 1; { this setting occurs in more branches than any other }
    if is_char_node(p) then r ← get_avail
    else case type(p) of
      hlist_node, vlist_node: begin r ← get_node(box_node_size);
        ⟨ Copy the box SyncTeX information 1733* ⟩;
        mem[r + 6] ← mem[p + 6]; mem[r + 5] ← mem[p + 5]; { copy the last two words }
        words ← 5; list_ptr(r) ← null; { this affects mem[r + 5] }
        end;
      rule_node: begin r ← get_node(rule_node_size); words ← rule_node_size;
        end;
      ligature_node: begin r ← get_avail; { only font and character are needed }
        mem[r] ← mem[lig_char(p)]; goto found;
        end;
      kern_node, math_node: begin words ← medium_node_size;
          { SyncTeX: proper size for math and kern }
        r ← get_node(words);
        end;
      glue_node: begin r ← get_node(medium_node_size); add_glue_ref(glue_ptr(p));
          { SyncTeX: proper size for glue }
        ⟨ Copy the medium sized node SyncTeX information 1735* ⟩;
        glue_ptr(r) ← glue_ptr(p); leader_ptr(r) ← null;
        end;
      whatsit_node: ⟨ Make a partial copy of the whatsit node p and make r point to it; set words to the
          number of initial words not yet copied 1417 ⟩;
      othercases goto not_found
      endcases;
    while words > 0 do
      begin decr(words); mem[r + words] ← mem[p + words];
      end;
  found: link(h) ← r; h ← r;
  not_found: p ← link(p);
    end;
  link(h) ← t;
  end;
```

See also section 1549*.

This code is used in section 1192.

1549* ⟨ Declare subprocedures for *init_math* 1544* ⟩ +≡

```
procedure just_reverse(p : pointer);
  label done;
  var l: pointer;  { the new list }
    t: pointer;  { tail of reversed segment }
    q: pointer;  { the next node }
    m, n: halfword;  { count of unmatched math nodes }
  begin m ← min_halfword; n ← min_halfword;
  if link(temp_head) = null then
    begin just_copy(link(p), temp_head, null); q ← link(temp_head);
    end
  else begin q ← link(p); link(p) ← null; flush_node_list(link(temp_head));
    end;
  t ← new_edge(cur_dir, 0); l ← t; cur_dir ← reflected;
  while q ≠ null do
    if is_char_node(q) then
      repeat p ← q; q ← link(p); link(p) ← l; l ← p;
      until ¬is_char_node(q)
    else begin p ← q; q ← link(p);
      if type(p) = math_node then ⟨ Adjust the LR stack for the just_reverse routine 1550* ⟩;
      link(p) ← l; l ← p;
      end;
  goto done; width(t) ← width(p); link(t) ← q; free_node(p, small_node_size);
done: link(temp_head) ← l;
  end;
```

1550* ⟨ Adjust the LR stack for the *just_reverse* routine 1550* ⟩ ≡

```
  if end_LR(p) then
    if info(LR_ptr) ≠ end_LR_type(p) then
      begin type(p) ← kern_node; incr(LR_problems);
          { SyncTEX node size watch point: math_node size == kern_node size }
      end
    else begin pop_LR;
      if n > min_halfword then
        begin decr(n); decr(subtype(p));  { change after into before }
        end
      else begin if m > min_halfword then decr(m) else begin width(t) ← width(p); link(t) ← q;
          free_node(p, medium_node_size);  { SyncTEX: no more "goto found", and proper node size }
          goto done;
          end;
        type(p) ← kern_node;  { SyncTEX node size watch point: math_node size == kern_node size }
        end;
      end
  else begin push_LR(p);
    if (n > min_halfword) ∨ (LR_dir(p) ≠ cur_dir) then
      begin incr(n); incr(subtype(p));  { change before into after }
      end
    else begin type(p) ← kern_node; incr(m);
          { SyncTEX node size watch point: math_node size == kern_node size }
      end;
    end
```

This code is used in section 1549*.

1566* ⟨ Initiate input from new pseudo file 1566* ⟩ ≡

```
begin_file_reading; { set up cur_file and new level of input }
line ← 0; limit ← start; loc ← limit + 1; { force line read }
if tracing_scan_tokens > 0 then
  begin if term_offset > max_print_line − 3 then print_ln
  else if (term_offset > 0) ∨ (file_offset > 0) then print_char("␣");
  name ← 19; print("(␣"); incr(open_parens); update_terminal;
  end
else begin name ← 18; ⟨ Prepare pseudo file SyncTeX information 1718* ⟩;
  end
```

This code is used in section 1564.

1584* A group entered (or a conditional started) in one file may end in a different file. Such slight anomalies, although perfectly legitimate, may cause errors that are difficult to locate. In order to be able to give a warning message when such anomalies occur, ε-TeX uses the *grp_stack* and *if_stack* arrays to record the initial *cur_boundary* and *cond_ptr* values for each input file.

```
⟨ Global variables 13 ⟩ +≡
grp_stack: ↑save_pointer; { initial cur_boundary }
if_stack: ↑pointer; { initial cond_ptr }
```

1678* System-dependent changes for Web2c. Here are extra variables for Web2c. (This numbering of the system-dependent section allows easy integration of Web2c and e-TEX, etc.)

```
⟨ Global variables 13 ⟩ +≡
edit_name_start: pool_pointer; { where the filename to switch to starts }
edit_name_length, edit_line: integer; { what line to start editing at }
ipc_on: cinttype; { level of IPC action, 0 for none [default] }
stop_at_space: boolean; { whether more_name returns false for space }
```

1679* The *edit_name_start* will be set to point into *str_pool* somewhere after its beginning if TEX is supposed to switch to an editor on exit.

```
⟨ Set initial values of key variables 23* ⟩ +≡
  edit_name_start ← 0; stop_at_space ← true;
```

1680* These are used when we regenerate the representation of the first 256 strings.

```
⟨ Global variables 13 ⟩ +≡
save_str_ptr: str_number;
save_pool_ptr: pool_pointer;
shellenabledp: cinttype;
restrictedshell: cinttype;
output_comment: ↑char;
k, l: 0 .. 255; { used by 'Make the first 256 strings', etc. }
```

1681* When debugging a macro package, it can be useful to see the exact control sequence names in the format file. For example, if ten new csnames appear, it's nice to know what they are, to help pinpoint where they came from. (This isn't a truly "basic" printing procedure, but that's a convenient module in which to put it.)

```
⟨ Basic printing procedures 57 ⟩ +≡
procedure print_csnames(hstart : integer; hfinish : integer);
  var c, h: integer;
  begin write_ln(stderr, 'fmtdebug:csnames␣from␣', hstart, '␣to␣', hfinish, ':');
  for h ← hstart to hfinish do
    begin if text(h) > 0 then
      begin { if have anything at this position }
      for c ← str_start_macro(text(h)) to str_start_macro(text(h) + 1) − 1 do
        begin put_byte(str_pool[c], stderr); { print the characters }
        end;
      write_ln(stderr, '|');
      end;
    end;
  end;
```

1682* Are we printing extra info as we read the format file?

```
⟨ Global variables 13 ⟩ +≡
debug_format_file: boolean;
```

1683.* A helper for printing file:line:error style messages. Look for a filename in *full_source_filename_stack*, and if we fail to find one fall back on the non-file:line:error style.

```
⟨ Basic printing procedures 57 ⟩ +≡
procedure print_file_line;
  var level: 0 .. max_in_open;
  begin level ← in_open;
  while (level > 0) ∧ (full_source_filename_stack[level] = 0) do decr(level);
  if level = 0 then print_nl("! ")
  else begin print_nl(""); print(full_source_filename_stack[level]); print(":");
    if level = in_open then print_int(line)
    else print_int(line_stack[level + 1]);
    print(": ");
    end;
  end;
```

1684.* To be able to determine whether `\write18` is enabled from within TeX we also implement `\eof18`. We sort of cheat by having an additional route *scan_four_bit_int_or_18* which is the same as *scan_four_bit_int* except it also accepts the value 18.

```
⟨ Declare procedures that scan restricted classes of integers 467 ⟩ +≡
procedure scan_four_bit_int_or_18;
  begin scan_int;
  if (cur_val < 0) ∨ ((cur_val > 15) ∧ (cur_val ≠ 18)) then
    begin print_err("Bad number");
    help2("Since I expected to read a number between 0 and 15,")
    ("I changed this one to zero."); int_error(cur_val); cur_val ← 0;
    end;
  end;
```

1685* The string recycling routines. TeX uses 2 upto 4 *new* strings when scanning a filename in an `\input`, `\openin`, or `\openout` operation. These strings are normally lost because the reference to them are not saved after finishing the operation. *search_string* searches through the string pool for the given string and returns either 0 or the found string number.

```
⟨ Declare additional routines for string recycling 1685* ⟩ ≡
function search_string(search : str_number): str_number;
  label found;
  var result: str_number; s: str_number; { running index }
    len: integer; { length of searched string }
  begin result ← 0; len ← length(search);
  if len = 0 then { trivial case }
    begin result ← ""; goto found;
    end
  else begin s ← search − 1; { start search with newest string below s; search > 1! }
    while s > 65535 do { first 64K strings don't really exist in the pool! }
      begin if length(s) = len then
        if str_eq_str(s, search) then
          begin result ← s; goto found;
          end;
      decr(s);
      end;
    end;
found: search_string ← result;
  end;
```

See also section 1686*.

This code is used in section 47*.

1686* The following routine is a variant of *make_string*. It searches the whole string pool for a string equal to the string currently built and returns a found string. Otherwise a new string is created and returned. Be cautious, you can not apply *flush_string* to a replaced string!

```
⟨ Declare additional routines for string recycling 1685* ⟩ +≡
function slow_make_string: str_number;
  label exit;
  var s: str_number; { result of search_string }
    t: str_number; { new string }
  begin t ← make_string; s ← search_string(t);
  if s > 0 then
    begin flush_string; slow_make_string ← s; return;
    end;
  slow_make_string ← t;
exit: end;
```

1687* More changes for Web2c. Sometimes, recursive calls to the *expand* routine may cause exhaustion of the run-time calling stack, resulting in forced execution stops by the operating system. To diminish the chance of this happening, a counter is used to keep track of the recursion depth, in conjunction with a constant called *expand_depth*.

This does not catch all possible infinite recursion loops, just the ones that exhaust the application calling stack. The actual maximum value of *expand_depth* is outside of our control, but the initial setting of 10000 should be enough to prevent problems.

```
⟨ Global variables 13 ⟩ +≡
expand_depth_count: integer;
```

1688* ⟨ Set initial values of key variables 23* ⟩ +≡

```
expand_depth_count ← 0;
```

1689* When *scan_file_name* starts it looks for a *left_brace* (skipping `\relax`es, as other `\toks`-like primitives). If a *left_brace* is found, then the procedure scans a file name contained in a balanced token list, expanding tokens as it goes. When the scanner finds the balanced token list, it is converted into a string and fed character-by-character to *more_name* to do its job the same as in the "normal" file name scanning.

```
procedure scan_file_name_braced;
  var save_scanner_status: small_number; { scanner_status upon entry }
    save_def_ref: pointer; { def_ref upon entry, important if inside '\message }
    save_cur_cs: pointer; s: str_number; { temp string }
    p: pointer; { temp pointer }
    i: integer; { loop tally }
    save_stop_at_space: boolean; { this should be in tex.ch }
    dummy: boolean; { Initializing }
  begin save_scanner_status ← scanner_status; { scan_toks sets scanner_status to absorbing }
  save_def_ref ← def_ref; { scan_toks uses def_ref to point to the token list just read }
  save_cur_cs ← cur_cs; { we set cur_cs back a few tokens to use in runaway errors }
    { Scanning a token list }
  cur_cs ← warning_index; { for possible runaway error }
    { mimick call_func from pdfTeX }
  if scan_toks(false, true) ≠ 0 then do_nothing; { actually do the scanning }
      { s ← tokens_to_string(def_ref); }
  old_setting ← selector; selector ← new_string; show_token_list(link(def_ref), null, pool_size − pool_ptr);
  selector ← old_setting; s ← make_string; { turns the token list read in a string to input }
    { Restoring some variables }
  delete_token_ref(def_ref); { remove the token list from memory }
  def_ref ← save_def_ref; { and restore def_ref }
  cur_cs ← save_cur_cs; { restore cur_cs }
  scanner_status ← save_scanner_status; { restore scanner_status }
    { Passing the read string to the input machinery }
  save_stop_at_space ← stop_at_space; { save stop_at_space }
  stop_at_space ← false; { set stop_at_space to false to allow spaces in file names }
  begin_name;
  for i ← str_start_macro(s) to str_start_macro(s + 1) − 1 do dummy ← more_name(str_pool[i]);
      { add each read character to the current file name }
  stop_at_space ← save_stop_at_space; { restore stop_at_space }
  end;
```

1690* System-dependent changes for MLTEX. The boolean variable *mltex_p* is set by web2c according to the given command line option (or an entry in the configuration file) before any TEX function is called.

```
⟨ Global variables 13 ⟩ +≡
mltex_p: boolean;
```

1691* The boolean variable *mltex_enabled_p* is used to enable MLTEX's character substitution. It is initialized to *false*. When loading a FMT it is set to the value of the boolean *mltex_p* saved in the FMT file. Additionally it is set to the value of *mltex_p* in IniTEX.

```
⟨ Global variables 13 ⟩ +≡
mltex_enabled_p: boolean;  { enable character substitution }
native_font_type_flag: integer;
    { used by XeTeX font loading code to record which font technology was used }
xtx_ligature_present: boolean;
    { to suppress tfm font mapping of char codes from ligature nodes (already mapped) }
```

1692* ⟨ Set initial values of key variables 23* ⟩ +≡

```
  mltex_enabled_p ← false;
```

1693* The function *effective_char* computes the effective character with respect to font information. The effective character is either the base character part of a character substitution definition, if the character does not exist in the font or the character itself.

Inside *effective_char* we can not use *char_info* because the macro *char_info* uses *effective_char* calling this function a second time with the same arguments.

If neither the character c exists in font f nor a character substitution for c was defined, you can not use the function value as a character offset in *char_info* because it will access an undefined or invalid *font_info* entry! Therefore inside *char_info* and in other places, *effective_char*'s boolean parameter *err_p* is set to *true* to issue a warning and return the incorrect replacement, but always existing character *font_bc*[f].

```
⟨ Declare ε-TeX procedures for scanning 1492 ⟩ +≡
function effective_char(err_p : boolean; f : internal_font_number; c : quarterword): integer;
  label found;
  var base_c: integer;  { or eightbits: replacement base character }
    result: integer;  { or quarterword }
  begin if (¬xtx_ligature_present) ∧ (font_mapping[f] ≠ nil) then
    c ← apply_tfm_font_mapping(font_mapping[f], c);
  xtx_ligature_present ← false; result ← c;  { return c unless it does not exist in the font }
  if ¬mltex_enabled_p then goto found;
  if font_ec[f] ≥ qo(c) then
    if font_bc[f] ≤ qo(c) then
      if char_exists(orig_char_info(f)(c)) then  { N.B.: not char_info(f)(c) }
        goto found;
  if qo(c) ≥ char_sub_def_min then
    if qo(c) ≤ char_sub_def_max then
      if char_list_exists(qo(c)) then
        begin base_c ← char_list_char(qo(c)); result ← qi(base_c);  { return base_c }
        if ¬err_p then goto found;
        if font_ec[f] ≥ base_c then
          if font_bc[f] ≤ base_c then
            if char_exists(orig_char_info(f)(qi(base_c))) then goto found;
        end;
  if err_p then  { print error and return existing character? }
    begin begin_diagnostic; print_nl("Missing␣character:␣There␣is␣no␣");
    print("substitution␣for␣"); print_ASCII(qo(c)); print("␣in␣font␣"); slow_print(font_name[f]);
    print_char("!"); end_diagnostic(false); result ← qi(font_bc[f]);
        { N.B.: not non-existing character c! }
    end;
found: effective_char ← result;
  end;
```

1694*. The function *effective_char_info* is equivalent to *char_info*, except it will return *null_character* if neither the character *c* exists in font *f* nor is there a substitution definition for *c*. (For these cases *char_info* using *effective_char* will access an undefined or invalid *font_info* entry. See the documentation of *effective_char* for more information.)

```
⟨ Declare additional functions for MLTEX 1694* ⟩ ≡
function effective_char_info(f : internal_font_number; c : quarterword): four_quarters;
  label exit;
  var ci: four_quarters;  { character information bytes for c }
    base_c: integer;  { or eightbits: replacement base character }
  begin if (¬xtx_ligature_present) ∧ (font_mapping[f] ≠ nil) then
    c ← apply_tfm_font_mapping(font_mapping[f], c);
  xtx_ligature_present ← false;
  if ¬mltex_enabled_p then
    begin effective_char_info ← orig_char_info(f)(c); return;
    end;
  if font_ec[f] ≥ qo(c) then
    if font_bc[f] ≤ qo(c) then
      begin ci ← orig_char_info(f)(c);  { N.B.: not char_info(f)(c) }
      if char_exists(ci) then
        begin effective_char_info ← ci; return;
        end;
      end;
  if qo(c) ≥ char_sub_def_min then
    if qo(c) ≤ char_sub_def_max then
      if char_list_exists(qo(c)) then
        begin  { effective_char_info ← char_info(f)(qi(char_list_char(qo(c)))); }
        base_c ← char_list_char(qo(c));
        if font_ec[f] ≥ base_c then
          if font_bc[f] ≤ base_c then
            begin ci ← orig_char_info(f)(qi(base_c));  { N.B.: not char_info(f)(c) }
            if char_exists(ci) then
              begin effective_char_info ← ci; return;
              end;
            end;
        end;
  effective_char_info ← null_character;
exit: end;
```

⟨ Declare subroutines for *new_character* 616* ⟩

This code is used in section 595*.

1695*. This code is called for a virtual character c in *hlist_out* during *ship_out*. It tries to built a character substitution construct for c generating appropriate `DVI` code using the character substitution definition for this character. If a valid character substitution exists `DVI` code is created as if *make_accent* was used. In all other cases the status of the substitution for this character has been changed between the creation of the character node in the hlist and the output of the page—the created `DVI` code will be correct but the visual result will be undefined.

Former MLTeX versions have replaced the character node by a sequence of character, box, and accent kern nodes splicing them into the original horizontal list. This version does not do this to avoid a) a memory overflow at this processing stage, b) additional code to add a pointer to the previous node needed for the replacement, and c) to avoid wrong code resulting in anomalies because of the use within a `\leaders` box.

⟨ Output a substitution, **goto** *continue* if not possible 1695* ⟩ ≡
begin ⟨ Get substitution information, check it, goto *found* if all is ok, otherwise goto *continue* 1697* ⟩;
found: ⟨ Print character substitution tracing log 1698* ⟩;
⟨ Rebuild character using substitution information 1699* ⟩;
end

This code is used in section 658*.

1696*. The global variables for the code to substitute a virtual character can be declared as local. Nonetheless we declare them as global to avoid stack overflows because *hlist_out* can be called recursively.

⟨ Global variables 13 ⟩ +≡
accent_c, *base_c*, *replace_c*: *integer*;
ia_c, *ib_c*: *four_quarters*; { accent and base character information }
base_slant, *accent_slant*: *real*; { amount of slant }
base_x_height: *scaled*; { accent is designed for characters of this height }
base_width, *base_height*: *scaled*; { height and width for base character }
accent_width, *accent_height*: *scaled*; { height and width for accent }
delta: *scaled*; { amount of right shift }

1697* Get the character substitution information in *char_sub_code* for the character *c*. The current code checks that the substitution exists and is valid and all substitution characters exist in the font, so we can *not* substitute a character used in a substitution. This simplifies the code because we have not to check for cycles in all character substitution definitions.

```
⟨ Get substitution information, check it, goto found if all is ok, otherwise goto continue 1697* ⟩ ≡
  if qo(c) ≥ char_sub_def_min then
    if qo(c) ≤ char_sub_def_max then
      if char_list_exists(qo(c)) then
        begin base_c ← char_list_char(qo(c)); accent_c ← char_list_accent(qo(c));
        if (font_ec[f] ≥ base_c) then
          if (font_bc[f] ≤ base_c) then
            if (font_ec[f] ≥ accent_c) then
              if (font_bc[f] ≤ accent_c) then
                begin ia_c ← char_info(f)(qi(accent_c)); ib_c ← char_info(f)(qi(base_c));
                if char_exists(ib_c) then
                  if char_exists(ia_c) then goto found;
                end;
        begin_diagnostic; print_nl("Missing␣character:␣Incomplete␣substitution␣");
        print_ASCII(qo(c)); print("␣=␣"); print_ASCII(accent_c); print("␣"); print_ASCII(base_c);
        print("␣in␣font␣"); slow_print(font_name[f]); print_char("!"); end_diagnostic(false);
        goto continue;
        end;
  begin_diagnostic; print_nl("Missing␣character:␣There␣is␣no␣"); print("substitution␣for␣");
  print_ASCII(qo(c)); print("␣in␣font␣"); slow_print(font_name[f]); print_char("!");
  end_diagnostic(false); goto continue
```

This code is used in section 1695*.

1698* For *tracinglostchars* > 99 the substitution is shown in the log file.

```
⟨ Print character substitution tracing log 1698* ⟩ ≡
  if tracing_lost_chars > 99 then
    begin begin_diagnostic; print_nl("Using␣character␣substitution:␣"); print_ASCII(qo(c));
    print("␣=␣"); print_ASCII(accent_c); print("␣"); print_ASCII(base_c); print("␣in␣font␣");
    slow_print(font_name[f]); print_char("."); end_diagnostic(false);
    end
```

This code is used in section 1695*.

1699*. This outputs the accent and the base character given in the substitution. It uses code virtually identical to the *make_accent* procedure, but without the node creation steps.

Additionally if the accent character has to be shifted vertically it does *not* create the same code. The original routine in *make_accent* and former versions of MLTeX creates a box node resulting in *push* and *pop* operations, whereas this code simply produces vertical positioning operations. This can influence the pixel rounding algorithm in some `DVI` drivers—and therefore will probably be changed in one of the next MLTeX versions.

```
⟨ Rebuild character using substitution information 1699* ⟩ ≡
  base_x_height ← x_height(f); base_slant ← slant(f)/float_constant(65536); accent_slant ← base_slant;
      { slant of accent character font }
  base_width ← char_width(f)(ib_c); base_height ← char_height(f)(height_depth(ib_c));
  accent_width ← char_width(f)(ia_c); accent_height ← char_height(f)(height_depth(ia_c));
    { compute necessary horizontal shift (don't forget slant) }
  delta ← round((base_width − accent_width)/float_constant(2) + base_height ∗ base_slant − base_x_height ∗
      accent_slant); dvi_h ← cur_h; { update dvi_h, similar to the last statement in module 620 }
    { 1. For centering/horizontal shifting insert a kern node. }
  cur_h ← cur_h + delta; synch_h;
    { 2. Then insert the accent character possibly shifted up or down. }
  if ((base_height ≠ base_x_height) ∧ (accent_height > 0)) then
    begin { the accent must be shifted up or down }
    cur_v ← base_line + (base_x_height − base_height); synch_v;
    if accent_c ≥ 128 then dvi_out(set1);
    dvi_out(accent_c);
    cur_v ← base_line;
    end
  else begin synch_v;
    if accent_c ≥ 128 then dvi_out(set1);
    dvi_out(accent_c);
    end;
  cur_h ← cur_h + accent_width; dvi_h ← cur_h;
    { 3. For centering/horizontal shifting insert another kern node. }
  cur_h ← cur_h + (−accent_width − delta);
    { 4. Output the base character. }
  synch_h; synch_v;
  if base_c ≥ 128 then dvi_out(set1);
  dvi_out(base_c);
  cur_h ← cur_h + base_width; dvi_h ← cur_h { update of dvi_h is unnecessary, will be set in module 620 }
```

This code is used in section 1695*.

1700*. Dumping MLTeX-related material. This is just the flag in the format that tells us whether MLTeX is enabled.

```
⟨ Dump MLTeX-specific data 1700* ⟩ ≡
  dump_int("4D4C5458); { MLTeX's magic constant: "MLTX" }
  if mltex_p then dump_int(1)
  else dump_int(0);
```

This code is used in section 1356*.

1701*. Undump MLTEX-related material, which is just a flag in the format that tells us whether MLTEX is enabled.

```
⟨ Undump MLTEX-specific data 1701* ⟩ ≡
  undump_int(x); { check magic constant of MLTEX }
  if x ≠ "4D4C5458 then goto bad_fmt;
  undump_int(x); { undump mltex_p flag into mltex_enabled_p }
  if x = 1 then mltex_enabled_p ← true
  else if x ≠ 0 then goto bad_fmt;
```

This code is used in section 1357*.

1702*. **The** *Synchronize TEXnology*. This section is devoted to the *Synchronize TEXnology* - or simply *SyncTEX* - used to synchronize between input and output. This section explains how synchronization basics are implemented. Before we enter into more technical details, let us recall in a few words what is synchronization.

TEX typesetting system clearly separates the input and the output material, and synchronization will provide a new link between both that can help text editors and viewers to work together. More precisely, forwards synchronization is the ability, given a location in the input source file, to find what is the corresponding place in the output. Backwards synchronization just performs the opposite: given a location in the output, retrieve the corresponding material in the input source file.

For better code management and maintainance, we adopt a naming convention. Throughout this program, code related to the *Synchronize TEXnology* is tagged with the "*synctex*" key word. Any code extract where *SyncTEX* plays its part, either explicitly or implicitly, (should) contain the string "*synctex*". This naming convention also holds for external files. Moreover, all the code related to *SyncTEX* is gathered in this section, except the definitions.

1703*. Enabling synchronization should be performed from the command line, *synctexoption* is used for that purpose. This global integer variable is declared here but it is not used here. This is just a placeholder where the command line controller will put the *SyncTEX* related options, and the *SyncTEX* controller will read them.

1704*. ⟨ Global variables 13 ⟩ +≡
synctexoption: *integer*;

1705*. A convenient primitive is provided: `\synctex=1` in the input source file enables synchronization whereas `\synctex=0` disables it. Its memory address is *synctex_code*. It is initialized by the *SyncTEX* controller to the command-line option if given. The controller may filter some reserved bits.

1706*. ⟨ Put each of TEX's primitives into the hash table 252 ⟩ +≡
primitive(`"synctex"`, *assign_int*, *int_base* + *synctex_code*);

1707*. ⟨ synctex case for *print_param* 1707* ⟩ ≡
synctex_code: *print_esc*(`"synctex"`);
This code is used in section 263*.

1708*. In order to give the *SyncTEX* controller read and write access to the contents of the `\synctex` primitive, we declare *synctexoffset*, such that *mem*[*synctexoffset*] and `\synctex` correspond to the same memory storage. *synctexoffset* is initialized to the correct value when quite everything is initialized.

1709*. ⟨ Global variables 13 ⟩ +≡
synctexoffset: *integer*; { holds the true value of *synctex_code* }

1710*. ⟨ Initialize whatever TEX might access 8* ⟩ +≡
synctexoffset ← *int_base* + *synctex_code*;

1711*. ⟨ Initialize synctex primitive 1711* ⟩ ≡
synctex_init_command;
This code is used in section 1386*.

1712* Synchronization is achieved with the help of an auxiliary file named '*jobname*`.synctex`' (*jobname* is the contents of the `\jobname` macro), where a *SyncTEX* controller implemented in the external *synctex.c* file will store geometrical information. This *SyncTEX* controller will take care of every technical details concerning the *SyncTEX* file, we will only focus on the messages the controller will receive from the TEX program.

The most accurate synchronization information should allow to map any character of the input source file to the corresponding location in the output, if relevant. Ideally, the synchronization information of the input material consists of the file name, the line and column numbers of every character. The synchronization information in the output is simply the page number and either point coordinates, or box dimensions and position. The problem is that the mapping between these informations is only known at ship out time, which means that we must keep track of the input synchronization information until the pages ship out.

As TEX only knows about file names and line numbers, but forgets the column numbers, we only consider a restricted input synchronization information called *SyncTEX information*. It consists of a unique file name identifier, the *SyncTEX file tag*, and the line number.

Keeping track of such information, should be different whether characters or nodes are involved. Actually, only certain nodes are involved in *SyncTEX*, we call them *synchronized nodes*. Synchronized nodes store the *SyncTEX* information in their last two words: the first one contains a *SyncTEX file tag* uniquely identifying the input file, and the second one contains the current line number, as returned by the `\inputlineno` primitive. The *synctex_field_size* macro contains the necessary size to store the *SyncTEX* information in a node.

When declaring the size of a new node, it is recommanded to use the following convention: if the node is synchronized, use a definition similar to *my_synchronized_node_size*=*xxx*+*synctex_field_size*. Moreover, one should expect that the *SyncTEX* information is always stored in the last two words of a synchronized node.

1713* By default, every node with a sufficiently big size is initialized at creation time in the *get_node* routine with the current *SyncTEX* information, whether or not the node is synchronized. One purpose is to set this information very early in order to minimize code dependencies, including forthcoming extensions. Another purpose is to avoid the assumption that every node type has a dedicated getter, where initialization should take place. Actually, it appears that some nodes are created using directly the *get_node* routine and not the dedicated constructor. And finally, initializing the node at only one place is less error prone.

1714* ⟨ Initialize bigger nodes with *SyncTEX* information 1714* ⟩ ≡

```
if s ≥ medium_node_size then
  begin sync_tag(r + s) ← synctex_tag; sync_line(r + s) ← line;
  end;
```

This code is used in section 147*.

1715* Instead of storing the input file name, it is better to store just an identifier. Each time TEX opens a new file, it notifies the *SyncTEX* controller with a *synctex_start_input* message. This controller will create a new *SyncTEX* file tag and will update the current input state record accordingly. If the input comes from the terminal or a pseudo file, the *synctex_tag* is set to 0. It results in automatically disabling synchronization for material input from the terminal or pseudo files.

1716* ⟨ Prepare new file *SyncTEX* information 1716* ⟩ ≡

```
synctex_start_input; { Give control to the SyncTEX controller }
```

This code is used in section 572*.

1717* ⟨ Prepare terminal input *SyncTEX* information 1717* ⟩ ≡

```
synctex_tag ← 0;
```

This code is used in section 358*.

1718* ⟨ Prepare pseudo file *SyncTEX* information 1718* ⟩ ≡
synctex_tag ← 0;
This code is used in section 1566*.

1719* ⟨ Close *SyncTEX* file and write status 1719* ⟩ ≡
synctex_terminate(*log_opened*); { Let the *SyncTEX* controller close its files. }
This code is used in section 1387*.

1720* Synchronized nodes are boxes, math, kern and glue nodes. Other nodes should be synchronized too, in particular math noads. TEX assumes that math, kern and glue nodes have the same size, this is why both are synchronized. *In fine*, only horizontal lists are really used in *SyncTEX*, but all box nodes are considered the same with respect to synchronization, because a box node type is allowed to change at execution time.

The next sections are the various messages sent to the *SyncTEX* controller. The argument is either the box or the node currently shipped out. The vertical boxes are not recorded, but the code is available for clients.

1721* ⟨ Start sheet *SyncTEX* information record 1721* ⟩ ≡
synctex_sheet(*mag*);
This code is used in section 676*.

1722* ⟨ Finish sheet *SyncTEX* information record 1722* ⟩ ≡
synctex_teehs;
This code is used in section 676*.

1723* ⟨ Start vlist *SyncTEX* information record 1723* ⟩ ≡
synctex_vlist(*this_box*);
This code is used in section 667*.

1724* ⟨ Finish vlist *SyncTEX* information record 1724* ⟩ ≡
synctex_tsilv(*this_box*);
This code is used in section 667*.

1725* ⟨ Start hlist *SyncTEX* information record 1725* ⟩ ≡
synctex_hlist(*this_box*);
This code is used in section 655*.

1726* ⟨ Finish hlist *SyncTEX* information record 1726* ⟩ ≡
synctex_tsilh(*this_box*);
This code is used in section 655*.

1727* ⟨ Record void list *SyncTEX* information 1727* ⟩ ≡
```
if type(p) = vlist_node then
  begin synctex_void_vlist(p, this_box);
  end
else begin synctex_void_hlist(p, this_box);
  end;
```
This code is used in sections 661* and 670*.

1728* ⟨ Record current point *SyncTEX* information 1728* ⟩ ≡
synctex_current;
This code is used in section 658*.

1729* ⟨ Record horizontal *rule_node* or *glue_node* *SyncTEX* information 1729* ⟩ ≡
synctex_horizontal_rule_or_glue(*p*, *this_box*);
This code is used in section 660*.

1730* ⟨ Record *kern_node* *SyncTEX* information 1730* ⟩ ≡
synctex_kern(*p*, *this_box*);
This code is used in section 660*.

1731* ⟨ Record *math_node* *SyncTEX* information 1731* ⟩ ≡
synctex_math(*p*, *this_box*);
This code is used in section 660*.

1732* When making a copy of a synchronized node, we might also have to duplicate the *SyncTEX* information by copying the two last words. This is the case for a *box_node* and for a *glue_node*, but not for a *math_node* nor a *kern_node*. These last two nodes always keep the *SyncTEX* information they received at creation time.

1733* ⟨ Copy the box *SyncTEX* information 1733* ⟩ ≡
sync_tag(*r* + *box_node_size*) ← *sync_tag*(*p* + *box_node_size*);
sync_line(*r* + *box_node_size*) ← *sync_line*(*p* + *box_node_size*);
This code is used in sections 232* and 1544*.

1734* ⟨ Copy the rule *SyncTEX* information 1734* ⟩ ≡

{ *sync_tag*(*r* + *rule_node_size*) ← *sync_tag*(*p* + *rule_node_size*);
sync_line(*r* + *rule_node_size*) ← *sync_line*(*p* + *rule_node_size*); }
This code is used in section 232*.

1735* ⟨ Copy the medium sized node *SyncTEX* information 1735* ⟩ ≡
sync_tag(*r* + *medium_node_size*) ← *sync_tag*(*p* + *medium_node_size*);
sync_line(*r* + *medium_node_size*) ← *sync_line*(*p* + *medium_node_size*);
This code is used in sections 232* and 1544*.

1736* *Nota Bene:* The *SyncTEX* code is very close to the memory model. It is not connected to any other part of the code, except for memory management. It is possible to neutralize the *SyncTEX* code rather simply. The first step is to define a null *synctex_field_size*. The second step is to comment out the code in "Initialize bigger nodes..." and every "Copy ... *SyncTEX* information". The last step will be to comment out the *synctex_tag_field* related code in the definition of *synctex_tag* and the various "Prepare ... *SyncTEX* information". Then all the remaining code should be just harmless. The resulting program would behave exactly the same as if absolutely no *SyncTEX* related code was there, including memory management. Of course, all this assumes that *SyncTEX* is turned off from the command line.

1737* System-dependent changes.

⟨ Declare action procedures for use by *main_control* 1097 ⟩ +≡

```
procedure insert_src_special;
  var toklist, p, q: pointer;
  begin if (source_filename_stack[in_open] > 0 ∧ is_new_source(source_filename_stack[in_open], line)) then
    begin toklist ← get_avail; p ← toklist; info(p) ← cs_token_flag + frozen_special; link(p) ← get_avail;
    p ← link(p); info(p) ← left_brace_token + "{";
    q ← str_toks(make_src_special(source_filename_stack[in_open], line)); link(p) ← link(temp_head);
    p ← q; link(p) ← get_avail; p ← link(p); info(p) ← right_brace_token + "}"; ins_list(toklist);
    remember_source_info(source_filename_stack[in_open], line);
    end;
  end;
procedure append_src_special;
  var q: pointer;
  begin if (source_filename_stack[in_open] > 0 ∧ is_new_source(source_filename_stack[in_open], line)) then
    begin new_whatsit(special_node, write_node_size); write_stream(tail) ← 0; def_ref ← get_avail;
    token_ref_count(def_ref) ← null; q ← str_toks(make_src_special(source_filename_stack[in_open], line));
    link(def_ref) ← link(temp_head); write_tokens(tail) ← def_ref;
    remember_source_info(source_filename_stack[in_open], line);
    end;
  end;
```

1738* This function used to be in pdftex, but is useful in tex too.

```
function get_nullstr: str_number;
  begin get_nullstr ← "";
  end;
```

1739* Index. Here is where you can find all uses of each identifier in the program, with underlined entries pointing to where the identifier was defined. If the identifier is only one letter long, however, you get to see only the underlined entries. *All references are to section numbers instead of page numbers.*

This index also lists error messages and other aspects of the program that you might want to look up some day. For example, the entry for "system dependencies" lists all sections that should receive special attention from people who are installing TeX in a new operating environment. A list of various things that can't happen appears under "this can't happen". Approximately 40 sections are listed under "inner loop"; these account for about 60% of TeX's running time, exclusive of input and output.

The following sections were changed by the change file: 2, 4, 6, 7, 8, 11, 12, 16, 19, 20, 23, 24, 26, 27, 28, 30, 31, 32, 33, 34, 35, 37, 38, 39, 47, 49, 51, 52, 53, 54, 65, 77, 78, 80, 85, 86, 88, 97, 98, 99, 108, 113, 132, 133, 134, 135, 138, 147, 157, 160, 163, 166, 171, 176, 177, 180, 183, 190, 200, 202, 212, 228, 232, 235, 237, 239, 241, 245, 246, 248, 256, 262, 263, 264, 266, 267, 278, 279, 282, 284, 285, 287, 292, 301, 313, 320, 330, 331, 332, 334, 336, 338, 358, 361, 368, 369, 396, 434, 435, 536, 548, 549, 550, 551, 552, 553, 554, 555, 556, 558, 559, 560, 561, 565, 567, 569, 571, 572, 583, 584, 585, 586, 587, 589, 595, 596, 598, 599, 605, 608, 610, 611, 616, 617, 618, 628, 631, 633, 634, 635, 638, 653, 655, 658, 659, 660, 661, 667, 670, 676, 678, 680, 751, 764, 765, 784, 793, 964, 974, 975, 977, 978, 979, 980, 982, 984, 985, 988, 993, 994, 995, 997, 998, 999, 1000, 1001, 1004, 1005, 1012, 1014, 1017, 1018, 1019, 1020, 1042, 1088, 1090, 1091, 1103, 1145, 1193, 1221, 1269, 1276, 1277, 1278, 1306, 1311, 1314, 1319, 1329, 1355, 1356, 1357, 1359, 1360, 1361, 1362, 1363, 1364, 1365, 1366, 1368, 1369, 1370, 1371, 1372, 1373, 1374, 1375, 1376, 1377, 1378, 1379, 1381, 1386, 1387, 1388, 1389, 1391, 1392, 1393, 1398, 1403, 1405, 1433, 1436, 1437, 1451, 1462, 1470, 1531, 1532, 1535, 1540, 1544, 1549, 1550, 1566, 1584, 1678, 1679, 1680, 1681, 1682, 1683, 1684, 1685, 1686, 1687, 1688, 1689, 1690, 1691, 1692, 1693, 1694, 1695, 1696, 1697, 1698, 1699, 1700, 1701, 1702, 1703, 1704, 1705, 1706, 1707, 1708, 1709, 1710, 1711, 1712, 1713, 1714, 1715, 1716, 1717, 1718, 1719, 1720, 1721, 1722, 1723, 1724, 1725, 1726, 1727, 1728, 1729, 1730, 1731, 1732, 1733, 1734, 1735, 1736, 1737, 1738, 1739.

⟨ Display adjustment p 223 ⟩ Used in section 209.
⟨ Display box p 210 ⟩ Used in section 209.
⟨ Display choice node p 737 ⟩ Used in section 732.
⟨ Display discretionary p 221 ⟩ Used in section 209.
⟨ Display fraction noad p 739 ⟩ Used in section 732.
⟨ Display glue p 215 ⟩ Used in section 209.
⟨ Display if this box is never to be reversed 1514 ⟩ Used in section 210.
⟨ Display insertion p 214 ⟩ Used in section 209.
⟨ Display kern p 217 ⟩ Used in section 209.
⟨ Display leaders p 216 ⟩ Used in section 215.
⟨ Display ligature p 219 ⟩ Used in section 209.
⟨ Display mark p 222 ⟩ Used in section 209.
⟨ Display math node p 218 ⟩ Used in section 209.
⟨ Display node p 209 ⟩ Used in section 208.
⟨ Display normal noad p 738 ⟩ Used in section 732.
⟨ Display penalty p 220 ⟩ Used in section 209.
⟨ Display rule p 213 ⟩ Used in section 209.
⟨ Display special fields of the unset node p 211 ⟩ Used in section 210.
⟨ Display the current context 342 ⟩ Used in section 341.
⟨ Display the insertion split cost 1065 ⟩ Used in section 1064.
⟨ Display the page break cost 1060 ⟩ Used in section 1059.
⟨ Display the token (m, c) 324 ⟩ Used in section 323.
⟨ Display the value of b 537 ⟩ Used in section 533.
⟨ Display the value of *glue_set*(p) 212* ⟩ Used in section 210.
⟨ Display the whatsit node p 1416 ⟩ Used in section 209.
⟨ Display token p, and **return** if there are problems 323 ⟩ Used in section 322.
⟨ Do first-pass processing based on *type*(q); **goto** *done_with_noad* if a noad has been fully processed, **goto** *check_dimensions* if it has been translated into *new_hlist*(q), or **goto** *done_with_node* if a node has been fully processed 771 ⟩ Used in section 770.
⟨ Do ligature or kern command, returning to *main_lig_loop* or *main_loop_wrapup* or *main_loop_move* 1094 ⟩ Used in section 1093.
⟨ Do magic computation 350 ⟩ Used in section 322.
⟨ Do some work that has been queued up for `\write` 1437* ⟩ Used in section 1436*.
⟨ Drop current token and complain that it was unmatched 1120 ⟩ Used in section 1118.
⟨ Dump MLTeX-specific data 1700* ⟩ Used in section 1356*.
⟨ Dump a couple more things and the closing check word 1380 ⟩ Used in section 1356*.
⟨ Dump constants for consistency check 1361* ⟩ Used in section 1356*.
⟨ Dump regions 1 to 4 of *eqtb* 1369* ⟩ Used in section 1367.
⟨ Dump regions 5 and 6 of *eqtb* 1370* ⟩ Used in section 1367.
⟨ Dump the ε-TeX state 1464, 1569 ⟩ Used in section 1361*.
⟨ Dump the array info for internal font number k 1376* ⟩ Used in section 1374*.
⟨ Dump the dynamic memory 1365* ⟩ Used in section 1356*.
⟨ Dump the font information 1374* ⟩ Used in section 1356*.
⟨ Dump the hash table 1372* ⟩ Used in section 1367.
⟨ Dump the hyphenation tables 1378* ⟩ Used in section 1356*.
⟨ Dump the string pool 1363* ⟩ Used in section 1356*.
⟨ Dump the table of equivalents 1367 ⟩ Used in section 1356*.
⟨ Either append the insertion node p after node q, and remove it from the current page, or delete *node*(p) 1076 ⟩ Used in section 1074.
⟨ Either insert the material specified by node p into the appropriate box, or hold it for the next page; also delete node p from the current page 1074 ⟩ Used in section 1068.
⟨ Either process `\ifcase` or set b to the value of a boolean condition 536* ⟩ Used in section 533.

⟨ Implement `\XeTeXpicfile` 1442 ⟩ Used in section 1403*.
⟨ Implement `\closeout` 1408 ⟩ Used in section 1403*.
⟨ Implement `\immediate` 1438 ⟩ Used in section 1403*.
⟨ Implement `\openout` 1406 ⟩ Used in section 1403*.
⟨ Implement `\pdfsavepos` 1450 ⟩ Used in section 1403*.
⟨ Implement `\primitive` 402 ⟩ Used in section 399.
⟨ Implement `\resettimer` 1414 ⟩ Used in section 1403*.
⟨ Implement `\setlanguage` 1440 ⟩ Used in section 1403*.
⟨ Implement `\setrandomseed` 1413 ⟩ Used in section 1403*.
⟨ Implement `\special` 1409 ⟩ Used in section 1403*.
⟨ Implement `\write` 1407 ⟩ Used in section 1403*.
⟨ Incorporate a whatsit node into a vbox 1419 ⟩ Used in section 711.
⟨ Incorporate a whatsit node into an hbox 1420 ⟩ Used in section 691.
⟨ Incorporate box dimensions into the dimensions of the hbox that will contain it 693 ⟩ Used in section 691.
⟨ Incorporate box dimensions into the dimensions of the vbox that will contain it 712 ⟩ Used in section 711.
⟨ Incorporate character dimensions into the dimensions of the hbox that will contain it, then move to the next node 694 ⟩ Used in section 691.
⟨ Incorporate glue into the horizontal totals 698 ⟩ Used in section 691.
⟨ Incorporate glue into the vertical totals 713 ⟩ Used in section 711.
⟨ Increase the number of parameters in the last font 615 ⟩ Used in section 613.
⟨ Increase k until x can be multiplied by a factor of 2^{-k}, and adjust y accordingly 124 ⟩ Used in section 123.
⟨ Initialize additional fields of the first active node 1656 ⟩ Used in section 912.
⟨ Initialize bigger nodes with *SyncTeX* information 1714* ⟩ Used in section 147*.
⟨ Initialize for hyphenating a paragraph 939 ⟩ Used in section 911.
⟨ Initialize synctex primitive 1711* ⟩ Used in section 1386*.
⟨ Initialize table entries (done by `INITEX` only) 189, 248*, 254, 258, 266*, 276, 285*, 587*, 1000*, 1005*, 1270, 1355*, 1432, 1463, 1629, 1665 ⟩ Used in section 8*.
⟨ Initialize the LR stack 1520 ⟩ Used in sections 689, 1524, and 1545.
⟨ Initialize the current page, insert the `\topskip` glue ahead of p, and **goto** *continue* 1055 ⟩ Used in section 1054.
⟨ Initialize the input routines 361* ⟩ Used in section 1391*.
⟨ Initialize the output routines 55, 65*, 563, 568 ⟩ Used in section 1386*.
⟨ Initialize the print *selector* based on *interaction* 79 ⟩ Used in sections 1319* and 1391*.
⟨ Initialize the special list heads and constant nodes 838, 845, 868, 1035, 1042* ⟩ Used in section 189.
⟨ Initialize variables as *ship_out* begins 653* ⟩ Used in section 678*.
⟨ Initialize variables for ε-TeX compatibility mode 1623 ⟩ Used in sections 1463 and 1465.
⟨ Initialize variables for ε-TeX extended mode 1624 ⟩ Used in sections 1451* and 1465.
⟨ Initialize whatever TeX might access 8*, 1710* ⟩ Used in section 4*.
⟨ Initialize *hlist_out* for mixed direction typesetting 1524 ⟩ Used in section 655*.
⟨ Initiate input from new pseudo file 1566* ⟩ Used in section 1564.
⟨ Initiate or terminate input from a file 412 ⟩ Used in section 399.
⟨ Initiate the construction of an hbox or vbox, then **return** 1137 ⟩ Used in section 1133.
⟨ Input and store tokens from the next line of the file 518 ⟩ Used in section 517.
⟨ Input for `\read` from the terminal 519 ⟩ Used in section 518.
⟨ Input from external file, **goto** *restart* if no input found 373 ⟩ Used in section 371.
⟨ Input from token list, **goto** *restart* if end of list or if a parameter needs to be expanded 387 ⟩ Used in section 371.
⟨ Input the first line of *read_file*[m] 520 ⟩ Used in section 518.
⟨ Input the next line of *read_file*[m] 521 ⟩ Used in section 518.
⟨ Insert LR nodes at the beginning of the current line and adjust the LR stack based on LR nodes in this line 1517 ⟩ Used in section 928.
⟨ Insert LR nodes at the end of the current line 1519 ⟩ Used in section 928.