Abstract

This document contains notes which are intended for those who are interested in \TeX programming. It is valuable for beginners as a first start with a lot of examples, and it is also valuable for experienced \TeXnicians who are interested in details about \TeX programming. However, it is neither a complete reference, nor a complete manual of \TeX.

Contents

1 Introduction 2

2 Programming in \TeX 2
  2.1 Variables in Registers 2
      2.1.1 Allocating Registers 3
      2.1.2 Using More than 256 Registers 4
  2.2 Arithmetics in \TeX 4
  2.3 Expansion Control 5
      2.3.1 Macros 6
      2.3.2 Token Registers 10
      2.3.3 Summary of macro definition commands 11
      2.3.4 Debugging Tools – Understanding and Tracing What \TeX Does 12
  2.4 The Scope of a Variable 12
      2.4.1 Global Variables 13
      2.4.2 Transporting Changes to an Outer Group 14
  2.5 Branching 15
      2.5.1 Boolean Variables 16
      2.5.2 Special Cases for Conditionals 16
  2.6 Loops 17
      2.6.1 Counting loops 17
      2.6.2 Loops over list of items 19
  2.7 More On \TeX 19

3 Survey of Key–Value Handling using pgfKeys 19
  3.1 The Low–Level (Direct) API of pgfKeys 20
  3.2 The Standard API of pgfKeys 22
      3.2.1 The Current Key Path 24
      3.2.2 Key Filtering 24

4 Special Tricks 24
  4.1 Handling # in Arguments 24

Index 25
1 Introduction

This document is intended to provide a direct start with \TeX programming (not necessarily \TeX typesetting). The addressed audience consists of people interested in package or library writing.

At the time of this writing, this document is far from complete. Nevertheless, it might be a good starting point for interested readers. Consult the literature given below for more details.

2 Programming in \TeX

2.1 Variables in Registers

\TeX provides several different variables and associated registers which can be manipulated freely.

\count\langle num\rangle

There are 256 Integer registers which provide 32 Bit Integer arithmetics. The registers can be used for example with \count0=42 or \count7=\macro where \macro expands to a number.

The value of a register can be typeset using \the\langle register\rangle.

\count0=42

The value is now ‘42’.

The value is now ‘-123456’.

\count0=\macro

The value is now ‘\the\count0’.

The ‘=’ sign is optional and can be omitted. One thing is common among the registers: an assignment of the form \count0=⟨···⟩ expands everything which follows until the expansion doesn’t need more numbers – even more than one following macro.

\def\firstmacro{123}
\def\secondmacro{456}
\def\thirdmacro{789}
\count0=\firstmacro\secondmacro\thirdmacro

The value is now ‘123456789’.

The precise rules can be found in [2], but it should be kept in mind that care needs to be taken here. More than once, my code failed to produce the expected result because \TeX kept expanding macros and the registers got unexpected results. Here is the correct method:

1. The value is now ‘42’.
2. The following code will absorb the ‘3’ of ‘3.’:
   \count0=3
   The value is now ‘12343’.
4. Use \textbackslash relax after an assignment to end scanning:
5. The value is now ‘1234’.

1. \count0=42 \% a white space after the number aborts the reading process. It is discarded.
   The value is now ‘\the\count0’.
2. The following code will absorb the ‘3’ of ‘3.’:
   \def\macro{1234}
   \count0=\macro \% a white space after a macro will be absorbed by \TeX, so this is wrong.
3. The value is now ‘\the\count0’.
4. Use \textbackslash relax after an assignment to end scanning:
   \count0=\macro\relax
   The value is now ‘\the\count0’.
5. The value is now ‘\the\count0’.

The command \textbackslash relax tells \TeX to “relax”: it stops scanning for tokens, but \textbackslash relax doesn’t expand to anything.
\dimen(num)
There are also 255 registers for fixed point numbers which are used pretty much in the same way as the \count registers – but \dimen register assignments require a unit like ‘cm’ or ‘pt’.
String access with ‘\the’ works in exactly the same way as for \count registers.

\dimen0=1pt
The value is now \the\dimen0.
\dimen0=0.0001pt
The value is now \the\dimen0.
\def\macro{1234.5678}
\dimen0=\macro pt
The value is now \the\dimen0.

The same rules with expansion of macros after assignments apply here as well.
The \dimen registers perform their arithmetics internally with 32 bit scaled integers, so called ‘scaled point’ with unit ‘sp’. It holds 1pt=65536sp=2^{16}sp. One of the 32 bits is used as sign. The total number range in pt is $[-(2^{30} - 1)/2^{16}, (2^{30} - 1)/2^{16}] = [-16383.9998, 16383.9998]$.

\toks(number)
There are also 255 token registers which can be thought of as special string variables. Of course, every macro assignment \def\macro{⟨content⟩} is also some kind of string variable, but token registers are special: their contents won’t be expanded when used with \the\toks⟨number⟩. This can be used for fine grained expansion control, see Section 2.3 below.

\toks0={abc}% 
\toks1={DEF}%
The value is now \the\toks0 \the\toks1.

Note the white space after \the\toks0: its purpose is to stop the number parsing when T\hbox{\TeX} scans for 0. The white space is discarded.

Token registers can also contain the special token # which would typically have a special meaning inside of macros:

\toks0={#1}% 
\message{Meaning is \the\toks0}%
This outputs “\texttt{Meaning is ##1}” in your log file.

Token registers are mainly useful when it comes to fine grained expansion control and are discussed in more depth in Section 2.3.

2.1.1 Allocating Registers
There is a very limited number of registers. Consequently, one has to think carefully how to allocate them. Typical use–cases for registers are temporary variables (like some intermediate result) and long–living resources which are to be accumulated while the document or some part of it is to be generated.

It is clearly a bad idea to carelessly overwrite a register. T\hbox{\TeX} comes with a single way to “allocate” registers:

\newdimen⟨macroname⟩
\newcount⟨macroname⟩
\newtoks⟨macroname⟩
These macros allocate a new register which is then accessible as \langle macroname \rangle.

The value is now 42.0pt.

1Please note that this does not cover the complete range of a 32 bit integer, I do not know why.
The resulting ($\texttt{\macroname}$) can now be used in the same way as if you used the register directly. In fact, it is often simpler because you do not need to worry about the register’s number.

The allocation relies on some global integer variable which is increased for each allocation. This ensures that variables stored in such allocated variables do not accidentally overwrite the contents of some other variable.

Note that deallocation is impossible.

While it is perfectly reasonable to allocate long-living variables, one should avoid the allocation of a new variable just because one needs a “new” temporary variable.

It makes sense to allocate a couple of named variables like \texttt{\tempa}, \texttt{\tempb}, or something like that and reuse these values for every temporary evaluation. Clearly, care needs to be taken to avoid unintended overwrites.

It is also possible to use token registers as explained above. However, the usage should be protected by means of groups:

\begin{verbatim}
\toks3 inside of group: Value inside of group
\toks3 outside of group: Value outside of group
\toks3={Value outside of group}
\begingroup
\toks3={Value inside of group}
\endgroup
\toks3 inside of group: \the\toks3
\toks3 outside of group: \the\toks3
\end{verbatim}

Groups constitute T\TeX{}’s concept of “scope” and are explained somewhere else in this document.

### 2.1.2 Using More than 256 Registers

T\TeX{} on its own is restricted to 256 registers. However, you can manually activate “extended T\TeX{} mode” by using

\begin{verbatim}
\usepackage{etex}
\end{verbatim}

early in your preamble. This is actually a very good idea: it allows access to 65536 registers. Today’s documents which involve lots of packages actually need \texttt{etex}.

Note that even \texttt{etex} does not justify wild and uncontrolled allocated of registers just to store temporary variables.

If you want almost unlimited temporary variables, you should store the temporaries in macros. This, of course, involves conversion from numbers to string, but it is the only save way which avoids the limited number of registers.

### 2.2 Arithmetics in T\TeX{}

\begin{verbatim}
\advance(register) by(quantity)
\end{verbatim}

The value is now 52.

\begin{verbatim}
\count0=42
\advance\count0 by 10
\end{verbatim}

The value is now 110pt.
\texttt{\textbackslash multiply}(\textit{register}) \textit{by}\langle\textit{integer}\rangle

The value is now -420.

\begin{verbatim}
\count0=42
\multiply\count0 by -10
The value is now \the\count0.
\end{verbatim}

The value is now 10.0pt.

\begin{verbatim}
\dimen0=0.5pt
\multiply\dimen0 by 20
The value is now \the\dimen0.
\end{verbatim}

\texttt{\textbackslash divide}(\textit{register}) \textit{by}\langle\textit{integer}\rangle

This allows integer division by \langle\textit{integer}\rangle with truncation.

The value is now 2.

\begin{verbatim}
\count0=5
\divide\count0 by 2
The value is now \the\count0.
\end{verbatim}

Scaling of \texttt{\textbackslash dimen} registers:

\begin{verbatim}
\dimen0=10pt
\divide\dimen0 by 20
The value is now \the\dimen0.
\end{verbatim}

It is impossible to divide by some non–integer number.

\texttt{\textbackslash dimen}(\textit{number})=\langle\textit{fixed point number without unit}\rangle\textbackslash dimen(\textit{number})

This allows fixed point multiplication in \texttt{\textbackslash dimen} registers.

\begin{verbatim}
\dimen1=50pt
\dimen0=0.6\dimen1
The value is now \the\dimen0.
\end{verbatim}

This is actually all that \TeX allows. One needs powerful macro packages like PGF with its \texttt{\pgfmathparse{\langle\textit{expression}\rangle}} to do some “real” arithmetics.

Note that the limited number range of these registers also applies to the result of any numerical operation.

\subsection*{2.3 Expansion Control}

Expansion is what \TeX does all the time. Thus, expansion control is a key concept for understanding how to program in \TeX.

The first thing to know is: \TeX deals the input as a long, long sequence of “tokens”. A token is the smallest unit which is understood by \TeX. Each character becomes a token the first time it is seen by \TeX. Every macro becomes a (single!) token the first time it is seen by \TeX.

The second thing to know is what characters are before \TeX has seen them. Although this knowledge is rarely needed in every day’s life, it is nevertheless important. The characters which are in the input document are nothing but characters at first. Even the characters known to have a special meaning like ‘\%’, ‘\ ’ or the braces ‘{\ }’ are not special – until they have been converted to a token. This happens when \TeX encounters them the first time during its linear processing of the character stream. A token stays a token - and it will remain the same token forever. If you manage to tell \TeX that ‘\ ’ is a normal character and \TeX sees just one backslash, this backslash will be a normal character token – even if the meaning of all following backslashes is again special.

Now, we are given a very long list of tokens \langle\textit{token1}\rangle\langle\textit{token2}\rangle\langle\textit{token3}\rangle\langle\textit{token4}\rangle\langle\textit{token5}\rangle\cdots. \TeX processes these tokens one-by-one in linear sequence. If \langle\textit{token1}\rangle is a character token like ‘a’, it is typeset. This is not what I want to write about here now; my main point is how to program in \TeX. So, the interesting thing in these notes is when \langle\textit{token1}\rangle is a macro.

\footnote{Of course, typesetting is an art in itself and there is a lot to read about it. Just not here in these notes.}
### 2.3.1 Macros

We have already seen some applications of macros above. Actually, most users who are willing to read notes about \TeX programming will have seen macros and may have written some on their own – for example using \newcommand (\newcommand is a “more high-level” version of \def used only in \LaTeX).

A macro has a name and is treated as an elementary token in \TeX (even if the name is very long). A macro has replacement text. As soon as \TeX encounters a macro, it replaces its occurrence with the replacement text. Furthermore, a macro can consume one or more of the following tokens as arguments.

**Executing it:** ‘This here is actually the replacement text.’.

\begin{verbatim}
\def\macro{This here is actually the replacement text.}
Executing it: \macro.
\end{verbatim}

**Invoking it:** replacement with first argument=hello!.

\begin{verbatim}
\def\macro#1{replacement with first argument=#1}
Invoking it: \macro{hello!}.
\end{verbatim}

This here is not really a surprise. What might come as a surprise is that the accepted arguments can be pretty much anything.

**Invoking it:** replacement with arguments: ‘a’ and ‘sign’.

\begin{verbatim}
\def\macro#1-#2.{replacement with arguments: '#1' and '#2'.}
Invoking it: \macro a-sign.
\end{verbatim}

The last example \macro runs through the token list which follows the occurrence of \macro. This token list is “a-sign.”. Macro expansion is greedy, that means the first matching pattern is used. Now, our \macro expected something, then a minus sign ‘-’, then another (possibly long) argument, then a period ‘.’. The argument between \macro and the minus sign is available as #1 and the tokens between the minus sign and the period as #2.

I found arguments ‘42’, ‘43’ and ‘44’.

\begin{verbatim}
\def\macro(#1,#2,#3){I found arguments '#1', '#2' and '#3'.}
\macro(42,43,44)
\end{verbatim}

As we have seen, macros can be used to manipulate the input tokens by expansion: they take some input arguments (maybe none) away and insert other tokens into the input token list. These tokens will be the next to process. We will soon learn more about that.

There is a command which helps to understand what \TeX does here:

\meaning⟨macro⟩

This command expands to the contents of ⟨macro⟩ as it is seen by \TeX.

\begin{verbatim}
\def\macro{Replacement \textmacro text \count0=42 \the\count0.}
\meaning{Debug message: '\meaning\macro'}
\end{verbatim}

As result, the log file and terminal output will contain

```
Debug message: 'macro:->Replacement \textmacro text \count 0=42 \the \count 0.'
```

The last example already shows something about \def: the replacement text can still contain other macros.

\begin{verbatim}
\def\⟨macroname⟩{⟨argument pattern⟩}{⟨replacement text⟩}
A new macro named ⟨macroname⟩ will be defined (or re-defined). The {⟨replacement text⟩} is the macro body, whenever the macro is executed, it expands to {⟨replacement text⟩}. The {⟨replacement text⟩} is a token list which can contain other macros. On the time of the definition, \TeX does not process (expand) the {⟨replacement text⟩}.

The {⟨replacement text⟩} will only be expanded if the macro is executed. This does also apply to any macros which are inside of {⟨replacement text⟩}.

Now, I execute it: Macro two contains This is macro one..
Now, I execute the second macro again: Macro two contains Redefined macroone..
\end{verbatim}
Macros can be defined almost everywhere in a \TeX{} document. They can also be invoked almost everywhere.

The \textit{(argument pattern)} is a token list which can contain simple strings or macro parameters ‘#\textit{(number)}’ or other macro tokens. The \textit{(number)} of the first parameter is always 1, the second must have 2 and so on up to at most 9. Valid argument patterns are ‘#1#2#3’, ‘(#1,#2,#3)’ or ‘---\relax’. If \TeX{} executes a macro, it searches for \textit{(argument pattern)} in the input token list until the first match is found. If no match can be found, it aborts with a (more or less helpful) error message.

The last example contains three macro definitions. Then, \TeX{} encounters \texttt{\macroone}. The input token list is now ‘\macroone abcdefg’.

The space(s) following \texttt{\macroone} are ignored by \TeX{}, they delimit the \texttt{\macroname{\macroone}}. Now, \TeX{} attempts to find matches for \textit{(argument pattern)}. It expects ‘abc’ – and it finds ‘abc’. These three tokens are removed from the input token list, and \TeX{} inserts the replacement text of \texttt{\macroone} which is \texttt{\macrotwo}. At that time, the input token list is ‘\macrotwo defg’.

Now, the same game continues with \texttt{\macrotwo}: \TeX{} searches for the expected \texttt{\{argument pattern\}} which is ‘def’, erases these tokens from the input token list and inserts the replacement text of \texttt{\macrotwo} instead. This yields ‘\macrothree g’.

Finally, \texttt{\macrothree} expects one parameter token (or a token list enclosed in parenthesis). The next token is ‘g’, which is consumed from the input token list and the replacement text is inserted – and ‘#1’ is replaced by ‘g’. Then, the token list is ‘Got ‘g’’.

This text is finally typeset (because it doesn’t expand further).

What we have seen now is how \TeX{} macros can be used to modify the token list. It should be noted explicitly that macro expansion does is in no way limited to those tokens provided inside of \texttt{\{replacement text\}} – if the last argument in \texttt{\{replacement text\}} is a macro which requires arguments, these arguments will be taken from the following tokens. Using nested macros, one can even process a complete part of the token list, in a manner of loops (but we don’t know yet how to influence macro expansion conditionally, that comes later).

Let’s try to solve the following task. Suppose you have a macro named \texttt{\point} with \texttt{\{argument pattern\}} ‘(#1,#2)’, i.e.

\begin{verbatim}
def\point(#1,#2){we do something with #1 and #2.}
\end{verbatim}

Suppose furthermore that you want to invoke \texttt{\point} with the contents which is stored in another macro. After all, macros are some kind of string variables – it makes sense to accumulate or generate string variables which will then be used as input for other macros. Let’s assume we have \texttt{\temp} and \texttt{\temp} contains ‘(42,1234)’. A first choice to invoke \texttt{\point} would be to use \texttt{\point\temp}. But: \texttt{\point} searches for an argument pattern which starts with ‘(’, not with \texttt{\temp}. The invocation fails.

\begin{verbatim}
\expandafter\(token\)(next token)
\end{verbatim}

The \texttt{\expandafter} command is an – at first sight confusing – method to alter the input token list. But: it solves our problem with \texttt{\point\temp}!

we do something with 42 and 1234
Why did that work!? The command `\expandafter` scans for the token after `\expandafter` in the input token list. This is `\point` in our case. Then, it scans for the next token which is `\temp` in our case (remember: macros are considered to be elementary tokens, just like characters ‘a’ or so). The two scanned arguments are removed from the input token list. Then, `\expandafter` expands the ⟨next token⟩ one time. In our case, ⟨next token⟩ is `\temp`. The first level of expansion of `\temp` is ‘(42,1234)’. Then, `\expansion` inserts the (unexpanded) ⟨token⟩ followed by the (expanded) contents of ⟨next token⟩ back into the input token list. In single steps:

1. `\expandafter\point\temp`
2. Expand `\expandafter`: next two tokens are ‘\point\temp’.
3. Use `\point` as ⟨token⟩ and `\temp` as ⟨next token⟩.
4. Expand `\temp` once, which leads to the tokens ‘(42,1234)’.
5. re-insert ⟨token⟩ and the expansion of ⟨next token⟩ back into the input token list. The list is then ‘\point(42,1234)’.
6. Expand `\point` as next token.

A further example: suppose we want to invoke `\theimportantmacro{⟨argument⟩}`. However, ⟨argument⟩ is contained in another macro! Furthermore, `\theimportantmacro` is defined to take exactly one parameter and our desired argument may have more than one token (which means we need to surround it with braces). This can be solved by the listing below.

I got the pre-assembled argument ‘xyz’ here.

```
def\theimportantmacro#1{I got the pre-assembled argument ‘#1’ here.}
def\temp{xyz}
def\expandafter\theimportantmacro\expandafter{\temp}
```

Now, what happens here? Let’s apply the rules step by step again:

1. After the initial definitions, the token list is `\expandafter\theimportantmacro\expandafter{\temp}`.
2. \TeX expands `\expandafter\theimportantmacro\expandafter{\temp}` as ⟨next token⟩.
3. According to the rules, \TeX expands ⟨next token⟩ once. But: ⟨next token⟩ is again a macro, namely `\expandafter`! Does that make a difference? No:
   (a) The token list after the second `\expandafter` is ‘\{\temp\}’ (3 tokens).
   (b) The ⟨token⟩ is thus ‘C’ and ⟨next token⟩ is ‘\temp’.
   (c) The expansion of ⟨next token⟩ is ‘xyz’.
   (d) The second `\expandafter` re-inserts its ⟨token⟩ and expanded ⟨next token⟩, which is ‘xyz’.
      Note that the closing brace ‘}’ has not been touched at all, \TeX hasn’t even seen it so far.
      We come back from the recursion. Remember: ⟨token⟩ is `\theimportantmacro` and the top-level expansion of ⟨next token⟩ is – as we have seen above – ‘xyz’.
4. \TeX re-inserts ⟨token⟩ and the expansion of ⟨next token⟩ to the input token list, which leads to `\theimportantmacro{xyz}`.
   The closing brace ‘}’ has not been touched, it simply resides in the input token list.
5. \TeX expands `\theimportantmacro`.

The ⟨next token⟩ is expanded exactly once. We have already seen that if ⟨next token⟩ is a macro which does substitutions on its own, these substitutions will be performed recursively. But what means ‘once’ exactly? We will need to use `\meaning` to check that (or the `\tracingmacros` tools) because we need to see what \TeX does.

So far, nothing has been typeset. But now: 4[This is macro one –2–].
\def\macroone{This is macro one \macrotwo} \\
def\macrotwo{--2--} \\
def\macrothree\#1{\def\macrofour{4[#1]}} \\
\expandafter\macrothree\expandafter{\macroone} \\
So far, nothing has been typeset. But now: \macrofour. \\
\message{We have macrofour = \meaning\macrofour} \\
The logfile (and terminal) will now contain \\
'We have macrofour = macro:->4[This is macro one \macrotwo ]'. \\
What happened? We can proceed as in the last example. After the two \expandafter expansions, \TeX finds the input token list \\
'\macrothree{This is macro one \macrotwo}' \\
which, after execution, defines \macrofour to be 'This is macro one \macrotwo'. The top-level expansion of \macroone has not expanded the nested call to \macrotwo. \\
So, \expandafter is a normal macro which can be expanded – and it is even possible to expand an \expandafter by another \expandafter. \\
What we have seen so far is \\
1. the \def command which stores unexpanded arguments in a macro variable and \\
2. the \expandafter which allows control over top-level expansion of macros (it expands one time). \\
\TeX provides two more features for expansion control: the \edef macro and token registers. \\
\edef{\macroname}{\argument pattern}{\replacement text} \\
The \edef command is the same as \def insofar as it defines a new macro. However, it expands \replacement text until only unexpandable tokens remain (\edef = expanded definition). \\
\def\a{3} \\
\def\b{2\a} \\
\def\c{1\b} \\
\def\d{value=\c} \\
\message{Macro 'd' is defined to be '\meaning\d'} \\
\edef\d{value=\c} \\
\message{Macro 'd' is e-defined to be '\meaning\d'} \\
\expandafter\def\expandafter\d\expandafter{\c} \\
\message{Macro 'd' is defined to be '\meaning\d' using expandafter} \\
This listing results in the log-file output \\
Macro 'd' is defined to be 'macro:->value=\c' \\
Macro 'd' is e-defined to be 'macro:->value=123' \\
Macro 'd' is defined to be 'macro:->1\b' using expandafter \\
So, \def does not expand at all, \edef expands until it can’t expand any further and the \expandafter construction expands \c one time and defines \d to be the result of this expansion. \\
Although possible, it might not occur too often to specify \argument pattern for an \edef because the expansion is immediate in contrast to \def. But it works in the same way: the positional arguments \#1, \#2, ..., \#9 will be replaced with their arguments. \\
The expansion of \replacement text happens in the same way as the expansion the main token list of \TeX. \\
Now, what exactly does “expands until only unexpandable tokens remain” mean? Our example indicates that the three tokens 1, 2 and 3 are not expandable while the macros \c, \b and \a could be expanded. \\
There is one large class of \TeX commands which can’t be expanded: any assignment operation. The example \\
\edef\d{\count0=42} \\
\message{Macro 'd' is defined to be '\meaning\d'} \\
\def\a{1234} \\
\edef\d{\advance\count0 by\a} \\
\message{Macro 'd' is defined to be '\meaning\d'} \\
yields the log-messages
Macro ‘d’ is defined to be ‘macro:->\count 0=42’ and
Macro ‘d’ is defined to be ‘macro:->\advance \count 0 by 1234’.

So, assignment and arithmetics operations are not expandable, they remain as executable tokens in the newly defined macro. This does also hold for \let and other assignment operations.

Interestingly, conditional expressions using \if \cdots \fi are expandable, but we will come to that later.

There is also a method to convert a macro temporarily into an unexpandable token: the \noexpand macro.

\noexpand(expandable token)

The \noexpand command is only useful inside of the {⟨replacement text⟩} of an \edef command. As soon as \edef encounters the \noexpand, the \noexpand will be removed and the ⟨expandable token⟩ will be converted into an unexpandable token. Thus, the code

\begin{verbatim}
\edef\d{Invoke \noexpand\a another macro}
\message{Macro ‘d’ is defined to be ‘\meaning\d’}
\end{verbatim}

yields the terminal output

Macro ‘d’ is defined to be ‘macro:->Invoke \a another macro’

because \noexpand\a yields the token ‘\a’ (unexpanded).³

2.3.2 Token Registers

Now, we turn to token registers. As we have already seen in Section 2.1, a token register stores a token list. A macro does also store a token list in its {⟨replacement text⟩}, so where is the difference? There are two differences:

1. Token registers are faster.
2. The contents of token registers will never be expanded.

I can’t give numbers for the first point – I have just read it in [2]. But the second point allows expansion control. While \edef allows “infinite” expansion, token registers allow only top-level expansion, just like \expandafter. But they can be used in a more flexible (and often more efficient) way than \expandafter.

The following examples demonstrates the second point.

\begin{verbatim}
\toks0={A \token list \a \b \count0=42 will never be expanded}
\edef\d{\the\toks0 }% the space token is important!
\message{Macro ‘d’ is defined to be ‘\meaning\d’}
\end{verbatim}

Executing this code fragment yields the log output

Macro ‘d’ is defined to be ‘macro:->A \token list \a \b \count 0=42 will never be expanded’.

So, the contents of \toks0 has been copied unexpanded into \d, although we have just \edef. Note that the space token after \the\toks0 is indeed important! \TeX uses it to delimit the integer 0. Without the space token, it would have continued scanning, even beyond the boundaries of the replacement text of \edef (see Section 2.1 for details about this scanning).

The example is very simple, and we could have done the same with \expandafter as before. But let’s try something more difficult: we want to assemble a new macro which consists of different pieces. Each piece is stored in a macro, and for whatever reason, we only want top-level expansion of the single pieces. And: the pieces won’t be adjacent to each other. We can assemble the target macro using the following example listing.

\begin{verbatim}
\def\piecea{\a{xyz}}
\def\pieceb{\count0=42}
\def\piecec{string \b}
\toks0=\expandafter{\piecea}
\toks1=\expandafter{\pieceb}
\toks2=\expandafter{\piecec}
\edef\d{I have \the\toks0 and \the\toks1 and \the\toks2}
\message{Macro ‘d’ is defined to be ‘\meaning\d’}
\end{verbatim}

³The \noexpand key is actually used to implement the \protect \LTEx command \protect: \LTEx’s concept of moveable arguments is implemented with \edef.
The first three lines define our pieces. Each of the macros \pieca, \piecb and \piecc contains tokens which should not be expanded during the definition of \d. The three following lines assign the top-level expansion of our pieces into token registers. Since \toks0=\pieca would have stored ‘\pieca’ into the token register, we need to use \expandafter here\footnote{We could have eliminated the \pieca* macros by writing everything into token registers directly. But I think this example is more realistic.}. Then, we use \the\toks\number to insert the contents of a token list somewhere – in our case, into the expanded replacement text of our macro \d. Thus, the complete example yields the log–output

Macro ‘d’ is defined to be ‘macro:->I have \a {xyz}and \count 0=42 and string \b ’. It is possible to get exactly the same result using (a lot of) \expandafters. Don’t try it.

2.3.3 Summary of macro definition commands

Besides \def and \edef, there are some more commands which allow to define macros (although the main functionality is covered by \def and \edef). Here are the remaining definition commands.

\def\langle macroname\rangle\langle argument pattern\rangle\{\langle replacement text\rangle\}

Defines a new macro named \macroname without expanding {\langle replacement text\rangle}, see above.

\edef\langle macroname\rangle\langle argument pattern\rangle\{\langle replacement text\rangle\}

Defines a new macro named \macroname, expanding {\langle replacement text\rangle} completely (see above).

\let\langle newmacro\rangle=\langle token\rangle

Defines or redefines \newmacro to be an equivalent to \langle token\rangle. For example, \let\a=\b will create a new copy of macro \b. The copy is named \a, and it will have exactly the same {\langle replacement text\rangle} and \langle argument pattern\rangle as \b.

It is also possible that \langle token\rangle is something different than a macro, for example a named register or a single character.

\gdef\langle macroname\rangle\langle argument pattern\rangle\{\langle replacement text\rangle\}

A shortcut for \global\def. It defines \macroname globally, independent of the current scope.

You should avoid macros which exist in both, the global namespace and a local scope, with different meanings. Section 2.4 explains more about scoping.

\xdef\langle macroname\rangle\langle argument pattern\rangle\{\langle replacement text\rangle\}

A shortcut for \global\edef. It defines \macroname globally, independent of the current scope.

You should avoid macros which exist in both, the global namespace and a local scope, with different meanings. Section 2.4 explains more about scoping.

\csname\langle expandable tokens\rangle\endcsname

This command is not a macro definition, it is a definition of a macro’s name. The “cs” means “control sequence”. The \csname, \endcsname pair defines a control sequence name (a macro name) using \langle expandable tokens\rangle. The control sequence character ‘\’ will be prepended automatically by \csname.\footnote{In fact, the contents of \escapechar will be used here. If its value is -1, no character will be prepended. The same holds for any occurrence where a backslash would be inserted by \TeX commands.}

This here is normal usage: ‘Content’.
This here uses csname: ‘Content’.
\def\macro{Content}
This here is normal usage: ‘\macro’.
This here uses csname: ‘\csname macro\endcsname’.

The example demonstrates that \csname\langle expandable tokens\rangle\endcsname is actually the same as if you had written \langle expandable tokens\rangle directly – but the \csname construction allows much more tokens inside of macro names:

I use a strange macro. Here is it: ‘Content’.
\expandafter\def\csname a1macro with.strange.chars\endcsname\{Content\}
I use a strange macro. Here is it: ‘\csname a1macro with.strange.chars\endcsname’.
The example uses \expandafter to expand \csname one time. The top-level expansion of \csname is a single token, namely the control sequence name. Then, \def is used to define a macro with the prepared macro name.

When \csname is expanded, it parses all tokens up to the next \endcsname. Those tokens will be expanded until only unexpandable tokens remain (as in \edef). The resulting string will be used to define a macro name (with the control sequence character ‘\’ prepended). The fact that (expandable tokens) is expanded allows to use “indirect” macro names:

\begin{verbatim}
I have just defined “macroonetwothree
with replacement text ‘Content’.
\def\macro{onetwothree}
\expandafter\def\csname macro\macro\endcsname{Content}
I have just defined \expandafter\string\csname macro\macro\endcsname
with replacement text ‘\csname macro\macro\endcsname’.
\end{verbatim}

I suppose the example is self-explaining, up to the \string command which is described below.

Due do this flexibility, \csname is used to implement all (?) of the available key–value packages in TeX.

\begin{verbatim}
\string(\macro)
This command does not define a macro. Instead, it returns a macro’s name as a sequence of separate tokens, including the control sequence token ‘\’.

I have just defined ‘“macro’ using ‘“def’.
\def\macro{Content}
I have just defined ‘\string\macro’ using ‘\string\def’.
\end{verbatim}

You can also use \string on other tokens – for example characters. That doesn’t hurt, the character will be returned as-is.

2.3.4 Debugging Tools – Understanding and Tracing What TeX Does

\begin{verbatim}
\message{⟨tokens⟩}
\meaning⟨\macro⟩
\tracingmacros=2
\tracingcommands=2
\tracingrestores=1
\end{verbatim}

2.4 The Scope of a Variable

Each programming language knows the concept of a scope: they limit the effect of variables or routines. However, TeX’s scoping mechanisms have not been designed for programming – TeX is a typesetting language. Many programming languages like C, C++, java or a lot of scripting languages define the scope of a variable using the place where the variable has been defined. For example, the C fragment

\begin{verbatim}
int i = 42;
{
  ++i;
  int i = 5;
}
\end{verbatim}

changes the value of the outer i to 43. The inner i is 5, but it will be deleted as soon as the closing brace is encountered. It may even be possible to access both, the value of the inner i variable and the value of the outer i variable, at the same time.

In TeX, braces are also used for scopes. But: while TeX will also destroy any variables (macros) defined inside of a scope at the end of that scope, it will also undo any change which has been applied inside of that scope.
The value of \i is now 42.

\def\i{42}
{ \def\i{43} \def\b{2} }
\def\i{43}

The listing above defines \i, enters a local scope (a TeX “group”) and changes \i. However, due to TeX’s scoping rules, the old program state will be restored completely after returning from the local group! Neither the change to \i nor the definition of \b will survive. The same holds for register changes or other assignments.

TeX groups can be created in one of three ways: using curly braces\(^6\), using \begingroup or using \bgroup. Curly braces are seldom used to delimit TeX groups because the other commands are more flexible. If one uses curly braces, they need to match up – it is forbidden to have unmatched curly braces.

\begingroup
Starts a new TeX group (a local scope). The scope will be active until it will be closed by \endgroup. The \endgroup command can occur later in the main token list.

\endgroup
Ends a TeX group which has been opened with \begingroup.

\bgroup
A special variant of \begingroup which can also be used to delimit arguments to \hbox or \vbox (i.e. it avoids the necessity to provide matched curly braces in this context).

The \bgroup macro is also useful to test whether the next following character is an opening brace (see \futurelet).

If one just needs to open a TeX group, one should prefer \begingroup.

\egroup
Closes a preceding \bgroup.

TeX does not know how to write into macros of an outer scope – except for the topmost (global) scope. This restriction is quite heavy if one needs to write complex structures: local variables should be declared inside of local groups, but changes to the structure should be written to the outer group. There is no direct possibility to do such a thing (except global variables).

2.4.1 Global Variables

TeX knows only “global” variables and “local” variables. A local variable will be deleted at the end of the group in which it has been declared. All values assigned locally will also be restored to their old value at the end of the group.

A global variable, on the other hand, maintains the same value throughout every scope. Usually, the topmost scope is the same as the one used for global variables: if you define anything in your TeX document, you add commands on global scope. It is also possible to explicitly make assignments or definitions in the global scope.

\global{definition or assignment}

The definition which follows \global immediately will be done globally.

\{ \global\def\a{123} \global\advance\count0 by3 \global\toks0={34} \}

\(^6\)Or other tokens with the correct category code, compare [2].
\texttt{\globaldefs=-1|0|1} (initially 0)

I cite from \cite{2}: "If the \texttt{\globaldefs} parameter is positive at the time of an assignment, a prefix of \texttt{\global} is automatically implied; but if \texttt{\globaldefs} is negative at the time of the assignment, a prefix of \texttt{\global} is ignored. If \texttt{\globaldefs} is zero (which it usually is), the appearance of nonappearance of \texttt{\global} determines whether or not a global assignment is made."

2.4.2 Transporting Changes to an Outer Group

There are a couple of methods to “transport” changes to an outer scope. Some are copy operations, some require to redo the changes again after the end of the scope. All of them can be realized using expansion control.

Let’s start with macro definitions which should be carried over the end of the group. I see the following methods:

- Copy the macro into a global, temporary variable (or even token register) and get that value after the scope.

\begin{verbatim}
def\initialvalue{0}
{
   % do something:
   \def\initialvalue{42}
   \global\let\myglobaltemporary=\initialvalue
}
\let\initialvalue=\myglobaltemporary
\end{verbatim}

The idea is that \texttt{\myglobaltemporary} is only used temporary; its value is always undefined and can be overwritten at any time. This allows to use a local variable \texttt{\initialvalue}.

Please note that you should not use variables both globally and locally. This confuses \TeX{} and results in a slow-down at runtime.

- “Smuggle” the result outside of the current group. I know this idea from the implementation of \cite{4} written by Mark Wibrow and Till Tantau. The idea is to use several \texttt{\expandafter}s and a \texttt{\def} to redefine the macro directly after the end of the group:

\begin{verbatim}
def\smuggle#1\endgroup{%
\expandafter\endgroup\expandafter\def\expandafter#1\expandafter{#1}%
}
\begingroup
\def\variable{12}
\edef\variable{\variable34}
\edef\variable{\variable56}
\smuggle\variable
\endgroup
\end{verbatim}

The technique relies on groups started with \texttt{\begingroup} and ended with \texttt{\endgroup} because unmatched braces are not possible with \texttt{\def}. The effect is that after all those \texttt{\expandafter}s, \TeX{} encounters the token list
\texttt{\endgroup\def\variable{123456}}
at the end of the group.

- Use the aftergroup stack. \TeX{} has a special token stack of limited size which can be used to re-insert tokens after the end of a group. However, this does only work efficiently if the number of tokens which need to be transported is small and constant (say, at most three). It works by prefixing every token with \texttt{\aftergroup}, compare \cite{2} for details.

Sometimes one needs to copy other variables outside of a scope. The trick with a temporary global variable works always, of course. But it is also possible to define a macro which contains commands to apply any required changes and transport that macro out of the scope.
2.5 Branching

Here we discuss some of the available branching constructions of TeX, with emphasis on conditions involving numbers and tokens.

\texttt{ifnum}(\texttt{count/inte}\texttt{ger number})=(\texttt{count/inte}\texttt{ger number})(\texttt{true-block})\\texttt{else}(\texttt{false-block})\\texttt{fi}

\texttt{ifnum} compare integer numbers or integer registers (\texttt{count} registers) and contains two branches, one is executed in the true case, the other in the case of false:

This is shown if above results to false.

\texttt{\textbackslash ifnum}1=2 % this space is important.
This is shown if above were true.
\texttt{\textbackslash else}
This is shown if above results to false.
\texttt{\textbackslash fi}

Note that the \texttt{\else} with its (\texttt{false-block}) is optional.

\texttt{ifdim}(\texttt{dimen/fixed point number})=(\texttt{dimen/fixed point number})(\texttt{true-block})\\texttt{else}(\texttt{false-block})\\texttt{fi}

Similar to \texttt{ifnum}, \texttt{ifdim} compares two fixed point numbers or \texttt{dimen} registers. The numbers must have a unit.

This is shown if above results to false.

\texttt{\textbackslash ifdim}1pt=2pt % this space is important.
This is shown if above were true.
\texttt{\textbackslash else}
This is shown if above results to false.
\texttt{\textbackslash fi}

\texttt{ifx}(\texttt{token1})(\texttt{token2})(\texttt{true-block})\\texttt{else}(\texttt{false-block})\\texttt{fi}

\texttt{ifx} is a bit more complex: It compares two tokens up to their first-level expansion.

This is shown if the two tokens have equal expansion.

\texttt{\textbackslash def}\texttt{empty}(\texttt{\textbackslash empty})
\texttt{\textbackslash ifs}empty\texttt{empty} %
This is shown if the two tokens have equal expansion.
\texttt{\textbackslash else}
This is shown if the two tokens expand to something different.
\texttt{\textbackslash fi}

Here, we have defined a token \texttt{\textbackslash empty} to be a replacement for \texttt{\textbackslash empty} and subsequently have compared whether these two tokens are equal in first-level expansion. Note that the definition is actually nonsense. If \TeX{} ever were to go through the whole expansion – i.e. we would put \texttt{\textbackslash empty} somewhere else – it would do so indefinitely. However, with \texttt{\textbackslash ifx} only first-level expansion is done and compared. Hence, the statement evaluates to true.

Have a look at the following example:

This is shown if the two tokens expand to something different.

\texttt{\textbackslash def}\texttt{empty}(\texttt{\textbackslash relax})
\texttt{\textbackslash ifs}empty\texttt{relax} %
This is shown if the two tokens have equal expansion.
\texttt{\textbackslash else}
This is shown if the two tokens expand to something different.
\texttt{\textbackslash fi}

On first glance, this should evaluate to true: \texttt{\textbackslash empty} is defined as a replacement for \texttt{\textbackslash relax}. But it does not. Why? \texttt{\textbackslash empty} is expanded to \texttt{\textbackslash relax}, however \texttt{\textbackslash relax} expanded has a different meaning, namely stop scanning and not \texttt{\textbackslash relax} anymore. Hence, they are different and the statement is false! If the expansion in \texttt{\textbackslash ifx} were to be taken till maximum, both would be equal but not in the case of a comparison on first-level expansion only.
The `\if` comparison is closely related to the `\ifx` conditional, with one major exception: it expands tokens until it finds the next two unexpandable tokens. If these two tokens are the same, it expands to the \langle true-block \rangle, otherwise to the \langle false-block \rangle.

The `\if` conditional should be handled with care as it might produce undesirable effects. Use it only if you know what you do.

A useful example is if you know that a macro contains at most one character, and you want to test for a particular one:

```
\def\choice{a}
\if b\choice
This is shown for the 'b' choice.
\else
This is shown for all other choices.
\fi
```

```
i
\fi
```

A “conditional” which always invokes the \langle true-block \rangle.

```
\fi
```

A “conditional” which always invokes the \langle false-block \rangle.

### 2.5.1 Boolean Variables

```
\newif(if-name)
```

You can declare a new “boolean variable ‘if\supermanmode by means of \newif\ifsupermanmode. Afterwards, you can use the \supermanmodetrue and \supermanmodefalse switches to assign the boolean and \ifsupermanmode to check it.

The \langle if-name \rangle has to start with \if (to support scans for nested \if...\fi pairs, see below).

### 2.5.2 Special Cases for Conditionals

Whenever you work with `\if...` and friends, you should know the following features:

1. `\if...\else...\fi` is expandable (including each of the single macros `\if...`, `\else` and `\fi`), which means you can even use it inside of `\edef`:

```
\def\choice(a)
\if \choice
This is shown for the 'b' choice.
\else
This is shown for all other choices.
\fi
```

```
i
\fi
```

A “conditional” which always invokes the \langle true-block \rangle.

```
i
\fi
```

A “conditional” which always invokes the \langle false-block \rangle.

This example is tricky. What would have happened without the \expandafter?! Well, \shownexttoken would be invoked with \#1=\fi. This would lead to an error because the \fi would be missing, and it would spoil the effect since we do not want the \fi to be seen – we expected \#1=2.

The \expandafter first expands \fi (which simply removes the \fi without further effect) such that \shownexttoken will see the 2 token in our example above. This would also have worked if there was an \else branch instead of \fi.
You should generally make sure that the matching `\else` or `\fi` tokens are “directly reachable”, i.e. without token expansion.

The background here is that \TeX\ works on a token–based level: Whenever it encounters an `\if`... statement, it evaluates it and scans tokens to find the matching end part (either an `\else` or an `\fi` token). But it will not expand tokens during this scan, although it will count nested `\if`...`\fi` pairs! Thus, if you are careless, it might become confused and your conditional will go awry.

### 2.6 Loops

As you have seen, in \TeX\ we have a very specific control over token expansion. This makes it possible to construct even loops via means of recursion. In essence, a loop consist of the following parts:

- counter or, more generally, list of items
- incrementor, or more generally, a next item picker
- threshold or, more generally, an end list marker
- a check of the threshold or end marker, respectively

Leafing through the sections above, we realize that all of this is actually in place: We do know about counters, we do know about branching. Only the specifics of how to create these loops is still to be made clear. We will show both cases, the counting loop and the loop over a list of items in the following in detail.

In general, for a loop done via a recursion we need two definitions: One for the loop start and another for the loop step.

#### 2.6.1 Counting loops

For a counting loop, we need a counter `\count0`, an incrementor `\advance`, a threshold `3` and a check `\ifnum\count=10` if the threshold has been reached.

```
\long\def\loopcounter#1#2#3{%
  #3%
  \ifnum#1=#2 %
    \else%
      \advance#1 by1 %
      \loopcounter{#1}{#2}{#3}%
    \fi%
%
\long\def\countingloop#1 in #2:#3#4{%
  #1=#2 %
  \loopcounter{#1}{#2}{#3}%
%
\countingloop{\count0} in 0:{3}{%

The current value is ‘0’
The current value is ‘1’
The current value is ‘2’
The current value is ‘3’
```

There are some subtleties to the above example:

- We put a lot of `%` in the example. Why? Note that whenever \TeX\ scans for a number – e.g. as in the case of `#1=#2` – it will continue scanning token by token, that is digit by digit, till he is sure that the number has ended, even over white space, and even expanding macros in case they themselves might not represent numbers again. Hence, `%` tells \TeX\ to stop scanning. It is generally good practice to place `%` to tell \TeX\ to stop scanning for more digits. However, there are some exceptions to it as well: In case of `\advance#1 by1` one should keep a white space in between, as well as in the case of `\ifnum#1=#2`.

- We placed the threshold `3` in `\countingloop{\count0} in 0:{3}` in curly brackets. Why? \TeX\ otherwise will recognize only the token `1` if a threshold of e.g. `10` is given and stumble over the now remnant ‘extra’ argument `0`. That is because a single letter represents a token to \TeX. Hence, two
letters are two tokens and – ungrouped – become two arguments. Here, we have to group the threshold to make clear what we mean.

- One last thing becomes clear first when debugging is activated: As loops are done by recursion, i.e. by expansion followed by expansion till some threshold is reached, we will end with a lot of \if...\fi in the above case. If we place \tracingmacros=2 \tracingcommands=2 before the \countingloop call and inspect the log file this will become apparent. This is bad because \TeX{} will keep a stack frame open for each \if...\fi sequence. If we now have a loop over 10,000 items ...

- It is not good practice to use one of the system counters, here \count0, because one can never be sure that is not used for something else or changed somewhere else. E.g. when the page is full, \TeX{} will interrupt the current sequence of tokens to deal with creating a new page and finishing the old one, in this course changing \count0. Hence, we should also create our own counter.

Hence, we modify the example as follows:

```latex
\def\loopcounter#1#2#3{\ifnum#1=#2\let\next=\relax\else\advance#1 by1\def\next{\loopcounter{#1}{#2}{#3}}\fi}\next
\newcount\ourcounter
\long\def\countingloop#1 in #2:#3#4{% #1=#2 
\loopcounter(#1)(#3)(#4)%}
\long\def\countingloop(#1)#2#3{% #3\ifnum#1=#2\let\next=\relax\else\advance#1 by1\def\next{\loopcounter(#1)(#2)(#3)}\fi\next\newcount\ourcounter
\countingloop(\ourcounter) in 0:(% The current value is \texttt{\the\ourcounter}\par)}
```

Principally, nothing has changed in terms of the output. However, notice that we have introduced the macro \next which either recurses into the next level – but after the \fi statement has been given – or ends the recursion by simply containing \relax. Also, we have declared a new counter called \ourcounter that is safe from harm.

Finally, let us briefly summarize what happens in detail:

1. \countingloop... is expanded to an assignment #1=#2 and another macro \loopcounter....
2. The assignment is done: \ourcounter is set to the starting value 0.
3. The actual loop macro is expanded to the command block – printing the current value – and an if statement.
4. The current value is printed.
5. \ourcounter is compared to the threshold 3 and ...
   - False, i.e. the if statement is expanded to an \advance statement followed by defining \next to be another call of the same macro loop.
   - True, i.e. \next is set to be just \relax.
6. The statement is still false: \advance will increase \ourcounter by one, it is now 1. \next is set to the loop macro.
7. The loop macro is again expanded, go to step 3. \ourcounter is ...2...\ourcounter is 3.
8. Now the statement is true: \next is expanded to \relax and nothing happens.
### 2.6.2 Loops over list of items

Looping over a list of items is very similar, only we will need \texttt{\ifx} in place of \texttt{\ifnum} and we need some end marker instead of the threshold value. However, how do we specify the list itself? Let’s make some comma-separated list, e.g. \{a,b,c,d\} and call the end marker \texttt{\listingloopENDMARKER}.

```latex
\def\listingloopENDMARKER{\par \listingloopENDMARKER}
\long\def\listingloop#1\#2#3{%
  \looppicker{#1}{#3}#2,\listingloopENDMARKER,%
}\long\def\looppicker#1#2#3,{%
  \def\tempitem{#3}%
  \ifx\tempitem\listingloopENDMARKER%
    \let\next=\relax%
  \else%
    \def#1{#3}%
    #2%
    \def\next{\looppicker{#1}{#2}}%
  \fi%
  \next%
}\listingloop\x in\{a,b,c,,d,e\}{% 
  The current item is ‘\x’ 
}
```

Again, we make clear the subtleties contained therein:

- We have defined \texttt{\listingloopENDMARKER} to replace itself. This is possible because \texttt{\ifx} will only compare first-level expansion, see Section 2.5.
- We seem to miss a white space in \ldots\#1in\#2\ldots. However, tokens are always ending with an additional white space as \texttt{\xin} is not equal to \texttt{\x in}. Hence, none is needed here and more than one white space would probably get gobbled.
- The definition \texttt{\looppicker\#1\#2\#3,\ldots} has three arguments but the recursive call \texttt{\looppicker\#1\#2} only gives two arguments!? This is the actual magic making this type of list possible! \TeX{} is actually scanning beyond the scope of the given token to obtain the third argument. In effect, we are biting off piece by piece, list item by list item off the given list. All because we have stated an additional , – comma being the item separator – in the definition of the \texttt{\looppicker} macro. The expansion of the loop macro will always pick up one more item from the list concatenated to its end until it has reached the \texttt{\ENDMARKER}. This is added to the list’s very end on the loop’s start, and there it stops.

### 2.7 More On \TeX{}

This document is far from complete. I recommend reading about conditional expressions in [3] (German, online version) or [2] (bounded book). Hints about loops can be found in the manual of \texttt{pgfplots}, [1] and the manual of \texttt{pgf}, [4]. Moreover, \texttt{pgfplots} and \texttt{pgf} come with a whole lot of utility functions which are documented in the source .\texttt{code.tex} files.

### 3 Survey of Key–Value Handling using pgfKeys

One of the most important things for every \TeX{} package is key–value input. There is a good overview and survey over different key–value packages, among them \texttt{xkeyval} and \texttt{pgfkeys}, in [5].
In addition to the paper mentioned above and the extensive reference manual for `pdfkeys` in [4], I give a brief survey over `pgfkeys` here. The addressed audience is primarily package writers or macro programmers. This section should allow you to define your own user interfaces and styles for `PGFPLOTS` and for `PGF`. It should also improve the understanding of `pgfkeys` and how it is to be used. I also address the topic of key filtering which is mainly useful for package writers.

The package `pgfkeys` is available as stand-alone package \usepackage{pgfkeys}. However, I believe that you never need to load it explicitly as PGF will be loaded anyway and PGF always loads `pgfkeys`.

It comes with two user interfaces. I believe that it is a best-practice to use the best of both worlds; although it might be sufficient to use just one of them. Consequently, I discuss both of them and propose a best-practices afterwards.

### 3.1 The Low-Level (Direct) Api of pgfKeys

Let us start with the low-level API of PGFKeys. It consists of a couple of macros which allow to define keys, assign values, and get their values back.

\pgfkeyssetvalue{/key path/key name}{value}

This macro (re)defines a key.

It is (almost) equivalent to a macro definition of sorts

\expandafter\def\csname key@⟨/key path/key name⟩\endcsname{⟨value⟩};

i.e. it stores (value) into a new macro such that the key can be looked up in constant time in \TeX's hash map. Note that in contrast to other key–value packages like `xkeyval`, the low-level macro name which is used to store the value is not part of the PGFKeys API⁷ – use `\pgfkeysgetvalue` and its friends to access the value.

The only limit for the number of possible keys is the size of \TeX's hash map (which is very large).

You may have wondered what the slash ‘/’ means. Users which are accustomed to PGF/PGFPLOTS know that there exists some kind of “key path” which qualifies ⟨key name⟩. The ⟨key path⟩ has the purpose of providing a name space such that many many keys with the same name can exist piecefully without ever touching another – provided the correct ⟨key path⟩ has been used. It can be seen as a (unix) file path: you can have many files with the same name, provided the files reside in different directories (i.e. have different paths).

You should always provide a key path, and it is highly recommended to use a different key path than just ‘/’.

The ⟨value⟩ can be anything; it is just stored. It can even contain #.

\pgfkeysgetvalue{/key path/key name}{⟨macro⟩}

As you might have guessed, this macro allows to retrieve the value for some key and store it into ⟨macro⟩.

Now that we have read about `\pgfkeyssetvalue` and `\pgfkeysgetvalue`, we can also provide an example:

The value of key `/notes/key` is `abc`.

\pgfkeyssetvalue{/notes/key}{abc}
\pgfkeysgetvalue{/notes/key}\temp

The value of key \texttt{/notes/key} is \texttt{\temp}.

There is few magic around these two keys; it is just like a hashmap access with some special naming convention for the keys (due to the key path). Note that since “hashmap access” is what \TeX does all the time when it handles macros, we could have replaced the pair `\pgfkeyssetvalue/\pgfkeysgetvalue` by `\def` and suitable `\let` commands, perhaps combined with `\csname...\endcsname`. The advantage of PGFKeys comes into play as soon as we inspect the high-level user interface in the next section.

Note that since `\pgfkeyssetvalue` is essentially the same as a suitable `\def`, the assignment is local to the current \TeX group. In other words: the assignment will be undone by the next closing curly brace, or the next `\end{⟨environment⟩}`.

⁷Note that `key@` is unrelated to PGFKeys.
This is essentially the same as \pgfkeyssetvalue, except that the key’s value is already available inside of \(\texttt{macro}\):

\begin{verbatim}
\def\something{abc}
\pgfkeyslet{/notes/key}{\something}
\pgfkeysgetvalue{/notes/key}\temp
\end{verbatim}

The value of key \texttt{/notes/key} is ‘\texttt{\temp}’.

Just like \pgfkeyssetvalue boils down to \texttt{def}, \pgfkeyslet boils down to \texttt{let}.

This is essentially the same as \pgfkeysgetvalue\{\(\texttt{macro}\)\}; i.e. it expands to the value stored in a key.

\begin{verbatim}
\pgfkeyssetvalue{/notes/key}{abc}
\end{verbatim}

The value of key \texttt{/notes/key} is ‘\texttt{abc}’.

However, this key has one major advantage: it can be used inside of an \texttt{edef} (because it is fully expandable):

\begin{verbatim}
\edef\temp{--- \pgfkeysvalueof{/notes/key} ---}
\end{verbatim}

The value of key \texttt{/notes/key} along with dashes is \texttt{\temp}.

It boils down to a suitable \texttt{\csname} ... \texttt{\endcsname}. Consequently, it expands to \texttt{relax} if the key happens to be undefined (see \pgfkeysifdefined below).

This is a variant of \pgfkeyssetvalue. However, it has a substantial difference which appears to be unmotivated as long as we discuss the low–level API. It defines a so–called code–key.

Code–keys are executable macros. They take an argument, and they do something with it. “Assigning values” to such a key is equivalent to invoking \(\texttt{macro body}\) in a “suitable” way.

The result of this macro call is a new key named \(\langle key path/key name/\rangle\).@cmd. That key, in turn, is stored as executable macro. The macro is equivalent to the following definition (up to the name, of course):

\begin{verbatim}
\def\macro#1\pgfeov{\texttt{macro body}}
\end{verbatim}

This macro is stored (using \pgfkeyslet) under \(\langle key path/key name/\rangle\).@cmd.

We can use \pgfkeysgetvalue and/or \pgfkeysvalueof to access this special key\(^8\), even though its use becomes more apparent later in this document:

\begin{verbatim}
\pgfkeyssetvalue{/notes/code key}{Expansion with value #1---X.}\%\texttt{\pgfeov}
\end{verbatim}

We “assign a value” or “execute the code key” (which is equivalent):

Expansion with value abc—X.

\begin{verbatim}
\pgfkeyssetvalue{/notes/code key}{Expansion with value \#1---X.}%
\end{verbatim}

Note that in this case, we have to use \texttt{\pgfeov} to terminate the argument list. We could have placed our argument into curly braces, but we have to provide \texttt{\pgfeov}; just as we had to add the suffix \texttt{/.@cmd}.

\(^8\)Note that the suffix \texttt{/.@cmd} is part of the public API of pgfKeys, so it is no hackery to make use of it.
These keys provide conditionals based on existance or type of a key. Please refer to the reference manual in [4] for details.

3.2 The Standard Api of pgfKeys

Now that we have seen how things defined by PGFKeys can be accessed at a rather low level of abstraction, we will repeat the same using a higher level. This section explains the standard API of PGFKeys; this is how Keys can be defined and maintained easily, and it is also the end user interface.

PGFKeys addresses a couple of use–cases with its standard API:

1. simple key–value storage (i.e. put and get),
2. code–keys which can do some (complex) operation whenever the key is used,
3. configuration and modification of the key–value tool.

All of these items are possible with the same macro:

\pgfkeys{⟨comma–separated key–value pairs⟩}

This key constitutes the public API of PGFKeys. It accepts any number of key–value pairs, separated by commas.

We start with an example:

```
\texttt{The value of key /notes/key is ‘abc’. The value of key /notes/key is ‘efg’}.
```

\texttt{\% key definition:}
\texttt{\pgfkeys{ /notes/key/.initial=abc,}}
\texttt{)}
\texttt{\% key usage:}
\texttt{\pgfkeys{ /notes/key = efg ,}}
\texttt{)}

```
The value of key \texttt{/notes/key} is ‘\pgfkeysvalueof{/notes/key}’.
```

There are some items which appear to be clear, and I will briefly confirm that it really is clear: white spaces before and after the key name and before and after the value are stripped away. Furthermore, trailing commas are ignored. Note that trailing commas are a best–practice: always insert trailing commas. This simplifies the addition of further keys significantly (I can’t remember how often I added a key and wondered why it was not properly recognised until I found the missing comma). Just add the trailing comma as a habit. Another good practice is to indent code properly, i.e. to insert a tab stop for every new line. It is also a good idea to provide one key per line, although all that stuff is optional.

The first thing which is strange when inspecting the actual code is the suffix ‘/\_.initial’. This is, in fact, a consistent new system of PGFKeys: these suffixes allow to configure and modify the keys to which them apply. They are called “key handlers”. Whenever you encounter ⟨key path/key name⟩ followed by ‘/⟨handler⟩’, you can safely assume that ⟨key path/key name⟩ is about to be reconfigured or modified.

Knowledge of key handlers means control over PGFKeys. In the following, I will briefly discuss the most important handlers.

Key handler ⟨key⟩/\_.initial=⟨{value}⟩

The key handler /\_.initial defines a new ⟨key⟩ and assigns its initial ⟨value⟩. As such, it is equivalent to \texttt{\pgfkeyssetvalue{/⟨key⟩}{⟨value⟩}}.

That a key which has been defined by means of /\_.initial can be set at any time later using a simple value assignment (see the example above).

Consequently, the first definition needs the suffix, all following assignments need to assignment to set the value.
Key handler \(\langle key\rangle/.code=\{\langle body\rangle\}\)

This key handler defines a new code–key \(\langle key\rangle\) with \(\langle body\rangle\) as result.

Execute the key using the simple API: Expansion with value abc—X. ... execute the key using assignment in the standard API: Expansion with value abc—X.

```
\pgfkeys{
  /notes/code key/.code={(Expansion with value #1---X.)},
}
```

... execute the key using assignment in the standard API:
```
\pgfkeysvalueof{/notes/code key=abc}
```

We see that assignment of a code key means to executing \(\langle body\rangle\) where \#1 is set to the value assigned in the API.

A key defined by means of \/.code is equivalent to one defined by means of \pgfkeysdef.

Note that the argument \(\langle body\rangle\) can be surrounded by curly braces, but it does not need to be:

```
\pgfkeys{
  /notes/code key/.code={(Expansion with value #1---X.),
  /notes/code key/.code=Expansion with value #1---X.},
}
```

This is a common feature of pgfKeys: any kind of value assignment can use braces, but it does not need to. You only need to use curly braces if the assigned argument (in our \(\langle body\rangle\)) contains control characters of pgfKeys (i.e. = or ,).

Key handler \(\langle key\rangle/.style=\{\langle option list\rangle\}\)

This key handler defines a new code–key \(\langle key\rangle\) which sets all options in \(\langle option list\rangle\) whenever it is assigned (used).

Styles are defined in a simple way: they simply invoke \pgfkeys with \(\langle option list\rangle\) (well, almost – see below). However, they are very expressive in any kind of application.

Definition has been done. Assigning the style:OK. Value of A=42, value of B=42.

```
\pgfkeys{
  /notes/A/.initial=,
  /notes/B/.initial=,
  /notes/my style/.style={
    /notes/A=#1,
    /notes/B=#1,
  },
}
```

Definition has been done. Assigning the style:
```
\pgfkeys{/notes/my style=42}
```

OK. Value of A=\pgfkeysvalueof{/notes/A}, value of B=\pgfkeysvalueof{/notes/B}.

Our example is a very simple application of a style: it sets a bunch of other options.

Note that \(\langle option list\rangle\) can depend on \#1.

So far, this document did always provide fully qualified key paths. However \/.style explicitly supports the notion of a “current key path”: if a “current key path” is in effect, \(\langle option list\rangle\) will be set in a context which also makes use of the same current key path. Technically, this means that \/.style uses \pgfkeysalso to set \(\langle option list\rangle\), i.e. it does not use \pgfkeys as claimed above.

\pgfkeysalso{\{comma–separated key–value pairs\}}

This macro is almost the same as \pgfkeys{\{comma–separated key–value pairs\}}. In fact, if any assignments in its argument use fully–qualified paths (as we did so far in this document), both invocations are equivalent.
The difference is how they treat keys which are relative to some current key path, a concept which will be explained in the next subsection.

Here is the difference between the macros: \texttt{\pgfkeys} resets the current key path to / before processing its argument whereas \texttt{\pgfkeysalso} does not change the current key path. Consequently, \texttt{\pgfkeysalso} is only useful inside of the body of some code-key (like \texttt{/style}).

3.2.1 The Current Key Path

tbd

3.2.2 Key Filtering

tbd

4 Special Tricks

4.1 Handling # in Arguments

More than once, I encountered the following difficulty: I wanted to collect an argument which contains the hash sign, ‘#’. That’s not particularly difficult, but it can lead to a lot of strange error messages when the resulting argument shall be processed! Consider

\begin{verbatim}
def\collectargument{#1}{% def\collectedcontent{#1}% \iffalse\collectedcontent\empty \else \fi \else\fi\fi\fi It is not empty, executing it: #1. \fi}
collectargument{}% works collectargument{something}% works collectargument{% does not work! \def\something#1{which depends on #1} }
\end{verbatim}

The code in this example is relatively simple: the \texttt{\collectargument} macro expects one argument and checks if it is empty (using \texttt{\iffalse}, which is a common and reliable check for emptiness). It is is not empty, it executes it. The \texttt{\collectargument} macro works in most circumstances. More precisely: it works as long as there is no hash sign in its argument! In our example, the third call fails with “Illegal parameter number in definition of \texttt{\collectedcontent}.” which occurs during the \texttt{\def\collectedcontent{#1}} line (and \TeX has reasons for this message due to the special meaning of the parameter expansion).

The cure: redefine the \texttt{\collectargument} macro using

\begin{verbatim}
def\collectargument{#1}{% \toks0={#1}% \edef\collectedcontent{\the\toks0}% \iffalse\collectedcontent\empty \else \fi \else\fi\fi\fi It is not empty, executing it: #1. \fi}
\end{verbatim}

(you may want to allocate a temporary token register for this task). What is the difference? Well, the \texttt{\toks0={#1}} assignment introduces no special meaning for the hash sign #, and \texttt{\the\toks0} neither. Note, however, that this requires \texttt{\edef\collectedcontent} instead of \texttt{\def\collectedcontent} since the \texttt{\the} statement needs to be expanded. Everything works as expected.
Index

\advance, 4
\begingroup, 13
\bgroup, 13
.code handler, 23
\count, 2
\csname, 11
\def, 6, 11
\dimen, 3, 5
\divide, 5
\edef, 9, 11
\egroup, 13
\endgroup, 13
\expandafter, 7
\gdef, 11
\global, 13
\globaldefs, 14
.initial handler, 22

Key handlers
.code, 23
.initial, 22
.style, 23

\let, 11
\meaning, 6, 12
\message, 12
\multiply, 5
\newcount, 3
\newdimen, 3
\newif, 16
\newtoks, 3
\noexpand, 10
\pgfkeys, 22
\pgfkeysalso, 23
\pgfkeysdef, 21
\pgfkeysgetvalue, 20
\pgfkeysifassignable, 22
\pgfkeysifdefined, 22
\pgfkeyslet, 21
\pgfkeyssetvalue, 20
\pgfkeysvalueof, 21
\relax, 2
\string, 12
.style handler, 23
\toks, 3
\tracingcommands, 12
\tracingmacros, 12
\tracingrestores, 12
\xdef, 11

25
References


