The LaTeX3 Project

Frank Mittelbach* and Chris Rowley†

26 February 1993
Version 1.00

Abstract

This is a brief sketch of the LaTeX3 Project: background, history, principles, aims and functionality. The new version of LaTeX is, like the current version, a freely available system for automated processing of structured documents, formatting them to the highest typographic standards by use of the TeX typesetting software.
Although its uses include a very large range of published documents, the importance of its unsurpassed ability to format mathematical formulas will not be forgotten in producing the new version.
It is being produced by an international group of volunteers under the technical direction of Frank Mittelbach.

1 Why a new version?

With TeX, Knuth designed a formatting system [7] that is able to produce a large range of documents typeset to extremely high quality standards. For various reasons, including its quality, portability, stability and availability, TeX spread very rapidly and can nowadays be best described as a world-wide de facto standard for high quality typesetting. Although it is most famous for its ability to typeset mathematics, it is being used for many other types of document, particularly those with multi-lingual requirements.

The TeX system is fully programmable. This allows the development of high-level user interfaces whose input is processed by TeX's interpreter to produce low-level typesetting instructions; these are input to TeX's typesetting engine which outputs the format of each page in a device-independent page-description language.

Many people have made use of this powerful feature of TeX and developed their own front-ends; the most commonly used such packages are AMS-TeX and LAMS-TeX by Michael Spivak [30, 29] and LaTeX by Leslie Lamport [8]. The development of AMS-TeX was sponsored by the American Mathematical Society, a publishing house which now processes and typesets the majority of its output using TeX.

The principal aim of AMS-TeX was to simplify the user interface to the sophisticated built-in math-typesetting capabilities of TeX. It therefore does not provide support for certain more general document processing requirements (such as symbolic cross-references, automatic numbering, etc.); it is therefore most appropriate for short articles which contain lots of formulas.

The LaTeX system, on the other hand, supports the needs of long documents such as textbooks and manuals. It was designed to separate content and form as much as possible by providing the user with a generic (i.e. logical rather than visual) markup interface; this is combined with style files in which the formatting is specified. Nevertheless, LaTeX also provides a complete set of direct formatting instructions; these allow the user to access the full power of TeX's typesetting expertise when preparing the final version of the document.

Recent years have shown that the concepts and approach of LaTeX have become widely accepted. Indeed, LaTeX has become the standard method of communicating and publishing documents in many academic disciplines. This has led to many publishers accepting LaTeX source for articles and books. The American Mathematical Society, for example, now provides a LaTeX style option [1] which makes the math-typesetting features of AMS-TeX available to users of LaTeX. But the use of LaTeX in the publishing industry goes further: Elsevier Science Publishers, for example, are developing a system [24] which links SGML (for electronic storage of documents in databases) and LaTeX (for formatting these documents).

The use of LaTeX, together with SGML, is now also spreading into industrial environments, where the technical qualities of TeX together with the concepts of

*Eichenweg 29, W-6500 Mainz 1, Federal Republic of Germany, `mittelbach@mzdmza.zdv.uni-mainz.de`
†Open University, Walton Hall, Milton Keynes MK7 6AA, United Kingdom, `c.a.rowley@open.ac.uk`

LaTeX are considered a powerful combination of great potential importance to such areas as corporate documentation [33] and database publishing.

With the spreading use of SGML-compliant systems, such as the Grif editor, LaTeX again is a common choice as the formatter for high quality typeset output [9]. A typical SGML Document Type Definition (DTD) uses concepts similar to those of LaTeX. Therefore, as in the Euromath system, the formatting is often implemented by simply mapping document elements to LaTeX constructs rather than directly to 'raw TeX' [31]. This enables the sophisticated analytical and processing techniques built in to much of the LaTeX software to be exploited; and it avoids the need to program in TeX.

Such developments ensured that the uses of LaTeX have become widespread, both geographically and typographically. However, they have also pushed the system way beyond its original intended purpose. Moreover, as most people who have been involved in the production of style files will agree, they have demonstrated that the current version, whilst making the author's job easier, provides little assistance to developers of new applications. Thus there can be no doubts about the need for the continued development of LaTeX into a new, improved, front-end to TeX—one that will serve the typesetting needs of the nineties and beyond—so that is what we intend the LaTeX3 project to provide.

2 History

The original goals for the development of a new version of LaTeX are described in a paper [21] presented in 1989 at the 10th annual meeting of the TeX Users Group at Stanford—it was also at this time that Leslie Lamport expressed his full support for such a project. However, we have since discovered that those original goals did not touch many of the deficiencies of the current system.

New applications of LaTeX have highlighted many limitations of its interface (both for authors of documents and for designers of styles); and further research on such problems led us to the conclusion that one gains very little by just providing more and more specialized style files to solve this or that special problem. This is because many of these deficiencies and limitations have their source in LaTeX's internal concepts and design [21, 22, 23].

The most important of those original goals, and one that is still a core part of the LaTeX3 system and a central concern of the project team, is the provision of a good style design interface—one that allows easy implementation of various layouts. (Easy, of course, is relative: we mean 'as easy as possible, given the complexity of the task'.) In order to make LaTeX3 a fully flexible and extensible system, a major effort is needed in the near future in order to get this interface 'right'.

3 Aims

The principle aims guiding our work on the project's development are as follows.

- The LaTeX3 system will provide high quality typesetting for a wide variety of document types and typographic requirements.
- For authors, it will be easy to use since it will be highly automated, but controllable.
- For editors and designers, it will support the direct formatting commands which are essential to the fine-tuning of document layout.
- It will process complex structured documents and support a document syntax that allows automatic translation of documents conforming to commonly used SGML document type definitions into LaTeX documents—this syntax will therefore, for example, support the SGML concepts of 'attribute' (or 'named argument') and 'short reference', in such a way that these can be easily linked to the corresponding SGML features.
- LaTeX3 will be designed as an open system and, like the present version, it will be usable with any standard TeX system and will thus be available on a very wide range of platforms.
- Its highly modular design will provide a system that is flexible and extensible, with well-defined and fully documented interfaces.
- The code itself will also be thoroughly documented and the modular design will help to make the system easy to maintain and enhance.
- We shall also provide extensive catalogues containing many examples that are carefully designed to make the learning time for new users (including designers, editors and programmers) as short as possible.

4 Available now!

In some important areas, the extra facilities of LaTeX3 are already available to current LaTeX users.

- The New Font Selection Scheme (NFSS) is now in widespread use, providing very general and powerful tools for setting up and accessing all available fonts. A new version (NFSS2) has recently been released, which extends the flexibility of the system in the following areas of font management.
  - Scalable fonts, e.g., support for PostScript fonts.
  - Encodings, i.e., support for multilingual documents allowing change of font encoding (code page) within a document.
  - Math symbols, facilitating access to a wide range of symbols.
  - Math typesetting using any suitable family of fonts, eg Lucida or Adobe Times.
- With AMS-LaTeX [1] the capabilities of LaTeX for mathematical typesetting have reached at least the standard of AMS-TeX.

- Work by Frank Mittelbach and David Carlisle, together with valuable suggestions by several others, has made available more sophisticated tabular processing [15].
- Various forms of multiple-column formatting are also now available [14].
- Johannes Braams [2] has produced a new version of the 'Babel system' to support a wide range of languages.

5 New features

LaTeX3 will provide much new and enhanced functionality in addition to the above facilities; and its 'LaTeX3 programming language' will enable it to be further extended in a controlled way.

Here are further details of some of the many planned improvements.

- A robust author interface: providing interactive error recovery linked to an on-line help system.
- Languages: the typesetting will be customisable for use with different languages (as will the whole system, e.g. the error/help components). In particular, there will be support for mixed language documents and for multiple languages within one paragraph to be correctly hyphenated.
- Table formatting: a large number of extensions in this area will be available, including:
  - multi-page tables;
  - automated column width calculation;
  - a variety of designs for ruled tables;
- Float handling: the major problem in processing floats (e.g. tables and figures) is the precise specification of the designer's rules concerning how to position them. The new system will make it possible to implement a large range of such positioning algorithms, including the requirements of multi-column formatting, whilst also supporting explicit positioning commands.
- Many of the non-typesetting aspects of document processing will be automated, providing a flexible interface to cover a wide range of requirements and styles in the following areas:
  - citations and bibliographies;
  - cross-references;
  - indexing, etc.;
  - tables of contents, etc.;
  - multiple marks.
- The enhanced functionality in the area of mathematical typesetting will include:
  - alignments in displayed formulas;
  - alphabets, symbols, embellishments;
  - commutative (arrow) diagrams.

  These will extend many of the features which are already available in the current version of AMS-LaTeX [1].

  The system will also support the elements defined in the 'standard SGML DTD for mathematical expressions' which is currently under development [32].
- The typesetting requirements of many other areas will also be addressed—some examples:
  - technical documentation (e.g. offset layout, change bars);
  - academic publishing in the humanities (e.g. critical text editions);
  - structural formulas in chemistry;
  - integration of graphical structures, including shading and colour;

6 Interfaces

As we said earlier, the design specification interface is probably the most significant single new feature of LaTeX3. It will have two clearly separated parts:

- the creation of the generic mark-up (e.g. environments) used by the author in creating the LaTeX form of the document;
- the specification of how (in SGML language) the document elements will be formatted.

This will simplify the production of different layouts for the same document type. It will also enable LaTeX documents to be created and modified by structured document editors, such as Arbortext Publisher and Grif. Indeed, these two parts are direct analogues of the following components of the Grif system: the 'generic structure definition (or S) language' and the 'presentation description (or P) language' [4].

However, in order to support the requirements of high quality typesetting, the LaTeX3 equivalent of the P language will need to be even richer than the Grif language. Despite the complexity of the task, this part of the 'designer interface' will help to make the following straightforward:

- specification of a wide variety of typographic design rules;
- linking of the elements in a document type to the desired formatting;
- specifying and modifying all of the parameters that influence the layout.

Another important interface will be the LaTeX3 programming language, used for producing enhancements and extensions: it will be an entirely new language based on data structures and operations suited to the kind of programming required by document processing applications and to the expression of visual components of the layout process. Built on this language there will be high-level generic functions that allow the straightforward expression of common layout components.

7 Resources

The majority of the work, including conceptualisation, modelling, prototyping, implementation and testing, is being undertaken by a dozen individuals under the technical direction of Frank Mittelbach. This work is, in most cases including that of the technical director, done entirely in their spare time and so involves, as you can imagine, a lot of enthusiasm to keep the project alive. A large number of other individuals and organisations have contributed in one way or another to the effort, and we are confident that this number will continue to rise.

There are, nevertheless, many other tasks still to be done in support of the LATEX3 project. These can be worked on concurrently with the development of the LATEX3 kernel system. Furthermore, some of these tasks require special expertise not found among the core programming team. Initial research, analysis, and other work on these tasks by volunteers will, we are sure, greatly speed up the process of integrating a number of desirable features into LATEX3.

For this reason a 'volunteer task list' has been set up: this describes briefly the individual tasks, which require a wide variety of expertise and time involvement. It will be updated at regular intervals and a copy is available, via anonymous `ftp` from the following sites: `niord.shsu.edu`, directory `[fileserv.vol-task]`; `ftp.uni-stuttgart.de`, directory `soft/tex/vol-task`. For access via mail server, send mail to `fileserv@shsu.bitnet`, with no subject line and in the body write: `sendme vol-task`, or to `mail-server@rus.uni-stuttgart.de`, with no subject line and in the body write: `send soft/tex/vol-task/vol-task.tex`. If you are unable to retrieve a copy via electronic networks, please contact Chris Rowley.

To provide a means of communication with a large number of LATEX users an electronic mailing list has been installed at Heidelberg. To subscribe to this list send a mail message to `listserv@vm.urz.uni-heidelberg.de`, with one line as the body of the message (substituting your own names):

```
subscribe LaTeX-L
        Your_first_name Your-surname
```

One of the major, and growing, problems is how to bring people from all over the world together to discuss the open questions and find new solutions. It is important that these meetings involve people from outside the project since we very much need the views and experience of typesetters, designers, publishers, etc. to help eliminate the flaws in the system and to find new and better solutions.

We have so far held two 'open workshops', in London, UK and Boston, USA; these were hugely successful and showed that further workshops of this kind are essential if we are to provide LATEX3 with a good designer interface.

It is now clear that our ability to maintain the current progress will depend on adequate financial support being available for the following purposes:

- enhancement of computer equipment and software for the core development team;
- purchase of books on typography and other related subjects;
- essential expenses (travel, accommodation, etc.) for meetings of the project's core development team; and also for meetings with others, for example: testers of early versions; publishers; typographic designers; suppliers of related software, etc.

References and Bibliography

This bibliography contains some items which are not specifically referenced in this article: these contain further information about the LATEX3 project. It also contains entries concerning BIBTEX, which will be reimplemented and enhanced by Oren Patashnik for use with LATEX3.

[1] American Mathematical Society, Providence, Rode Island. *AMS-LATEX Version 1.1 User's Guide*, December 1990.

[2] Johannes Braams An update on the `babel` system *TUGboat*, 14, to appear 1993.

[3] Malcolm Clark. What is the LATEX3 project? *TEX and TUG NEWS*, 1(1):4, February 1992.

[4] Grif S.A, St Quentin en Yvelines. *Grif Languages: Grif 2.2*, 1992.

[5] Mary Guenther, editor. *TEX 90 Conference Proceedings*, March 1991. Published as TUGboat 12#1.

[6] Alan Hoenig. LATEX3, TUG and You. *TEX and TUG NEWS*, 2(1):2–4, July 1992.

[7] Donald E. Knuth. Computers & Typesetting. Addison-Wesley, Reading, Massachusetts, 1986.

[8] Leslie Lamport. *LATEX: A Document Preparation System*. Addison-Wesley, Reading, Massachusetts, 1986.

[9] Helmut Lenzing. The euromath project. *Euromath Bulletin*, 1(1):13–20, 1992.

[10] Frank Mittelbach. LATEX 2.10. In Lincoln K. Durst, editor, *1990 Conference Proceedings*, page 444, September 1990. Published as TUGboat 11#3.

[11] Frank Mittelbach. LATEX2.09 → LATEX3. *TEXline*, (14):15–18, February 1992.

[12] Frank Mittelbach. LATEX3. *Die TEXnische Komödie*, 4(2):15–22, August 1992.

[13] Frank Mittelbach. LATEX3 project. *TEX Gebruikers Group*, 92(1):87–90, May 1992.

[14] Frank Mittelbach. The `multicol` package. Distribution of style options that allow multi-column layout with LATEX, May 1992.

[15] Frank Mittelbach and David Carlisle. The `array` package. Distribution of style options that extend LATEX's array and tabular facilities, May 1992.

[16] Frank Mittelbach and Chris Rowley. The LATEX3 project fund. *Die TEXnische Komödie*, 3(4):13–15, December 1991.

[17] Frank Mittelbach and Chris Rowley. LaTeX209 → LaTeX3. *TUGboat*, 13(1):96–101, April 1992.

[18] Frank Mittelbach and Chris Rowley. The LaTeX3 project fund. *TeX and TUG NEWS*, 1(1):5–6, February 1992.

[19] Frank Mittelbach and Chris A. Rowley. The pursuit of quality—how can automated typesetting achieve the highest standards of craft typography? In C. Vanoirbeek and G. Coray, editors, *Electronic Publishing '92*, pages 260–273, Cambridge, April 1992. Cambridge University Press.

[20] Frank Mittelbach, Chris Rowley and Michael Downes. Volunteer work for the LaTeX3 project. LaTeX3 project paper, September 1992.

[21] Frank Mittelbach and Rainer Schöpf. With LaTeX into the nineties. In Christina Thiele, editor, *1989 Conference Proceedings*, volume 10#4 of TUGboat, pages 681–690. TeX Users Group, December 1989.

[22] Frank Mittelbach and Rainer Schöpf. LaTeX dans les années 90. *Cahiers GUTenberg*, (6):2–14, July 1990.

[23] Frank Mittelbach and Rainer Schöpf. Towards LaTeX 3.0. In Guenther [5], pages 74–79. Published as TUGboat 12#1.

[24] N.A.F.M. Poppelier. SGML and TeX in scientific publishing. In Guenther [5], pages 105–109. Published as TUGboat 12#1.

[25] David Rhead. Could LaTeX do more for chemists? *TeXline*, (12):2–4, December 1990. Suggestions for LaTeX3.

[26] David Rhead. Towards BIBTeX style-files that implement principal standards. *TeXline*, (10):2–8, May 1990.

[27] David Rhead. How might LaTeX3 deal with citations and reference lists? *TeXline*, (13):13–20, September 1991. Suggestions for LaTeX3.

[28] Chris Rowley. LaTeX209 → LaTeX3: an update. In Mimi Burbank, editor, *1992 Annual Meeting Proceedings*, pages 390–391, October 1992. Published as TUGboat 13#3.

[29] Michael D. Spivak. *LAMS-TeX The Synthesis*. The TeXplorators Corporation, Houston, 1989.

[30] M. D. Spivak, Ph.D. *The Joy of TeX, A Gourmet Guide to Typesetting with the AMS-TeX macro package*. American Mathematical Society, Providence, Rode Island, second edition, 1990.

[31] Eric van Herwijnen. *Practical SGML*. Wolters Kluwer Academic Publishers, Dordrecht, 1990.

[32] Björn von Sydow. The design of the euromath system. *Euromath Bulletin*, 1(1):39–48, 1992.

[33] Reinhard Wonneberger. Structured document processing: the LaTeX approach. *Computer Physics Communications*, (61):177–189, 1990.

[34] Reinhard Wonneberger and Frank Mittelbach. BIBTeX reconsidered. In Guenther [5], pages 111–124. Published as TUGboat 12#1.