TEX-based Production at the AMS*

Ralph Youngen†

American Mathematical Society,
P.O. Box 6248,
Providence, RI 02940, USA

June 1992

1 Background

The American Mathematical Society (AMS) is a both major publisher of mathematical research, and a professional organization whose members are mathematicians engaging in research at academic institutions and other research centers in the U.S. and around the world. A primary function of the Society is to provide channels of communication whereby these mathematicians can communicate the results of their research to each other, and to the broader scientific community. Foremost among these channels of communication is an extensive publications program which is based on the TEX typesetting system.

The AMS headquarters are in Providence, Rhode Island, and a sister office affiliated with the University of Michican is located in Ann Arbor, Michigan. The two sites employ a total of more than 200 people.

Computing hardware at the AMS consists of a VAX/VMS 6320 cluster, a 2-processor Solbourne 5E/902-64 which functions as a database server, a 4-processor Sun 690 MP which is the central server for the Ann Arbor office, a 2-processor Solbourne 5/500 which functions as a TEX server for the VMS cluster, approximately 12 other Unix and VMS workstations which serve as development machines, and approximately 20 PC-compatible and 20 Macintosh computers. High resolution output devices include two Autologic APS-μ5 phototypesetters, and an Agfa/Compugraphic 9600 PostScript imagesetter. Proofing devices include two high-speed QMS 300dpi laser printers, and approximately 20 other 300dpi laser printers. A 300dpi flatbed scanner is used for input of graphic material for the creation of line art and halftones.

The Society currently publishes 20 journals, ranging from tri-weekly's to annuals, as well as nearly 100 books per year, of which 95% are in 20 regular book series. Of the journals, six contain results of primary research, eight are translations from Russian, Japanese, and other languages, and the largest, *Mathematical Reviews (MR)*, contains reviews and extensive indexes of current mathematical literature that together occupied more than 10,000 pages in 1991. AMS books and journals combine for about 70,000 pages per year, with 62,000 typeset pages and 8,000 author-prepared pages. In addition to its own publications, the AMS also does some portion or all of the composition, editorial, printing, binding, and distribution for nine other journals published by eight scientific societies. Furthermore, the AMS acts as a service bureau for typesetting jobs composed using the TEX typesetting language. Aside from a conventional desktop publishing segment for the production of a few newsletters, brochures, and posters, and any publications produced from author-supplied copy, all other publications are typeset using TEX, with the final camera copy generated on a phototypesetter at the AMS headquarters.

2 History and Importance of TEX at AMS

The AMS has used electronic publishing methods to produce its publications since 1971. Prior to the early 1970s all mathematical typesetting was done either by using typewriters or hot lead-type machines (Monotype for example). The AMS experimented with various electronic publishing systems through the early to mid 1970s, including paper-tape systems, until settling for a time with a system developed by Science Typographers Inc. (STI). The STI system was able to handle jobs containing mathematical material, but was not useful for multi-column administrative publications such as the AMS membership list or catalog.

In 1978 the AMS learned of a new typesetting system called TEX when Prof. Donald E. Knuth of Stanford University presented the Gibbs lecture at the AMS annual meeting in January 1978. Every year, a well-known mathematician is invited to present the Gibbs lecture on a topic of his or her choosing. Knuth chose to speak about mathematics and computer science in the service of technical typesetting. In the audience were several influential members and officers of the AMS, and they undertook the investigation into TEX and encouraged its adoption as a production tool for the Society's publishing program. Knuth furthered this

*Presented at the 9^e NTG meeting, June 4, 1992, Amsterdam.

†Technical Support Manager, American Mathematical Society.

adoption process by assigning the trademark rights to TEX to the AMS.

The promise of the TEX typesetting system was to provide a new typesetting system intended for the creation of beautiful books—and especially for books that contain a lot of mathematics. Over the past 14 years TEX has truly evolved into a powerful typesetting program, generally regarded as the best system in th world for handling complicated scientific text and displayed mathematical formulas, as well as producing copy near in quality and appearance to that produced by the finest traditional compositors.

Beginning around 1983, the AMS began using TEX for a small portion of its technical book and journal publishing. The AMS found TEX to be very well suited for a high volume of technical typesetting. Thus, the percentage of typesetting done in TEX increased over the next few years, until, by mid 1987, TEX was used for 100% of typeset book and journal production.

There are several factors which have contributed to TEX's huge success. Certainly one of these is the fact that unlike most proprietary composition systems, TEX can be used on almost any known computer hardware from personal microcomputers to Cray supercomputers. The fact that Knuth placed the source code for TEX into the public domain meant that any skilled systems programmer could easily tailor the TEX program to run on any computer. Also unlike many proprietary systems, the cost of most TEX implementations is reasonably low to minimal.

Another significant contribution to TEX's success is that it provided mathematicians with a tool for expressing mathematical notation in a simple text file on a computer. This allowed the mathematician to use the computer both as a means of typesetting mathematics, and for communicating mathematics to other colleagues' computers via electronic mail or other electronic transfers. It is not surprising that TEX is now pervasive throughout academia.

TEX's coding mechanism, however, sometimes appears overly complex and gives the impression to some that it would be difficult to learn. This is true if one wishes to become a devoted TEXnician and learn all of the intricacies of TEX. But TEX's powerful macro capability, which allows a single command to be an abbreviation for a complex set of commands, enables TEX programmers to develop macro packages that greatly simplify the *use* of TEX. The AMS developed such a macro package, namely $\mathcal{AMS}$-TEX, in the early 1980s. $\mathcal{AMS}$-TEX greatly simplifies the author's use of TEX to the point that simple documents can be typeset by example. In the late 1980s the AMS developed the $\mathcal{AMS}$-LATEX macro package as a result of users requesting better mathematics typesetting for the popular macro package LATEX.

The AMS became interested in TEX both for the promise of very high quality output, and for the hope that authors would one day be able to submit their files electronically to the AMS for publication. The fact that TEX is in the public domain and therefore inexpensive to obtain, as well as the fact that TEX runs on almost any computer you can name, means that TEX is widely available to a large number of people—especially mathematicians at research centers who are potential authors of technical articles or books for publication. A mathematician at a university with a PC can be writing his or her document in TEX and obtain output from a dot matrix printer or laser printer on site. When the document is finished, the TEX source file can be sent electronically to the AMS, or any other publisher. Thus, the time it traditionally takes to typemark and keyboard the document from paper manuscript is saved, and the author gains in the assurance of knowing that when proof copy is received back from the AMS, the author is reading the same file that he or she has already carefully proofread. For example, the author doesn't have to painstakingly check all of the mathematical equations for correctness because this has already been done before submission to the AMS.

3 Electronic Submissions

The goal of receiving TEX-encoded electronic manuscripts from authors has now been realized. The AMS began its electronic submission program in 1988. The number of electronic manuscripts prepared in $\mathcal{AMS}$-TEX and $\mathcal{AMS}$-LATEX has steadily increased over the past few years. In 1991 the AMS received nearly 500 electronic submissions, accounting for 16% of total submissions for that year, and this figure has risen to 22% to date for 1992 submissions. Over 75% of these electronic submissions were received via electronic mail. The rest were received on either PC or Macintosh disks sent via postal mail, with approximately twice as many PC disks as Macintosh disks being received.

The objective of the electronic submission program is to receive documents that will merge smoothly into the production stream at the AMS. For the AMS program, the key factors that have lead to realizing this goal are specific guidelines describing the preparation of an electronic manuscript, adequate author support via telephone and e-mail, and a limitation on the types of documents that will be accepted (only $\mathcal{AMS}$-TEX and $\mathcal{AMS}$-LATEX documents are accepted electronically).

It became evident that early communication with authors, preferably before writing starts, was very important. Two documents entitled *Guidelines for Preparing Electronic Manuscripts* were written in both $\mathcal{AMS}$-TEX and $\mathcal{AMS}$-LATEX versions. These documents provide the author with both general and specific instructions for the creation of the electronic manuscript, such as:

- a short description of how an electronic manuscript is processed;

- a checklist of information that must accompany a manuscript order to avoid delays;
- instructions for using the logical tags, which ones are required, and examples;
- references to more detailed manuals for the basic TEX package ($\mathcal{AMS}$-TEX or LATEX);
- instructions for submitting the file on diskette, by e-mail or by FTP;
- where to go for help.

The Guidelines do not try to show an author how to use TEX itself (or $\mathcal{AMS}$-TEX or LATEX), but only what is specific to AMS publications and built into the author's macro packages. Although electronic submissions were being received as early as 1988, no formal guidelines existed until 1990. Since the publication of these Guidelines the quality of electronic submissions has improved greatly, as measured by the number of them that require special handling and the amount of attention required from editorial and composition staff.

In addition to these guidelines, generic documentstyles for both $\mathcal{AMS}$-TEX and $\mathcal{AMS}$-LATEX are also available to the author. Both these generic documentstyles and the ones use in production at the AMS are based on generic (SGML-like) markup. In these documentstyles, the elements of a document are tagged for logical structure, not for typographic appearance. In this way an author need not know exactly what publication his or her article will appear in while writing the paper, because logical tagging provides the flexibility of moving a paper from one publication to another by simply changing the documentstyle being used.

The documentstyles provided to authors define the appearance of a "generic journal" format, whose appearance is based loosely on the *Journal of the American Mathematical Society.* Again it is important to emphasize that a paper typeset with this generic documentstyle will resemble an article from the AMS journals, but it most certainly will not be identical in appearance. The purpose of the generic documentstyle is to provide the author with a logical framework into which the document should be encoded. Both $\mathcal{AMS}$-TEX and $\mathcal{AMS}$-LATEX contain definitions for mathematical elements such as displayed and aligned equations, matrices, case statements, etc. Macros for theorems, proofs, section headings, lists, and so forth are provided by the generic documentstyle, and the author is instructed by the Guidelines that these macros **must** be used. A heading in the form `\centerline{\bf Introduction}` is unacceptable because it defines the heading's typographic appearance but not its logical structure. The proper form would be `\heading{Introduction}`.

In fact any typographic-specific TEX commands such as specific font calls, vertical or horizontal space, and page and line breaks are specifically prohibited. The reason for this is twofold: AMS publications are typeset using Times Roman fonts and the differences in character widths between Times and Computer Modern will mean different line and page breaks than what the author sees, and the editorial staff may request changes to the text itself that would affect line and page breaks.

Occasionally an author who is also a TEX expert is troubled by the restrictive posture the AMS has taken toward electronic submissions. These authors generally wish to base their submissions on a favorite macro package they have personally developed or acquired elsewhere. While a few simple definitions to act as substitutes for repeated phrases or other constructs are allowed (even encouraged), submissions which are dependent on any additional packages other than those supplied by the AMS frequently will not merge into the AMS production system without attention from a skilled TEX programmer. This does not result in the anticipated savings of time and money for such submissions.

4 AMS Production Cycle

4.1 Typeset Books and Journals

The Society's TEX-based production cycle is really quite traditional. Manuscripts are refereed and accepted for publication by an editorial board for the book or journal. Members of these editorial boards are mathematicians from academic or research sites, and are not employees of the AMS. Whether or not an author used TEX to prepare their submission has no bearing in the acceptance process.

When an accepted manuscript is forwarded by the editorial board to the Providence office, it is logged into a tracking system which will monitor the paper's progress through various correction runs, until the paper is finally typeset as part of a specific publication. Upon receipt of the paper manuscript the existence of an electronic version is determined. If an electronic version is unavailable, the paper manuscript is delivered to the Editorial Department at the AMS, where the paper is typemarked and copyedited.

Typemarking refers to the process of labeling the logical components of the paper; all headings and constructs such as theorems, definitions, proofs, etc., are clearly marked by the editorial staff. At this point the paper is forwarded to the Composition Department where it is keyed in $\mathcal{AMS}$-TEX. The composition staff consists of about 15 keyboarders, between a third and a half of whom work from home and use terminals connected to the AMS computers via modems and phone lines. The use of terminals and modems rather than PCs was deemed preferable for several reasons, such as centralization of data and ease of communication via electronic mail. Keyboarders working from home have access to the technical staff with electronic mail, and documentation and other information can easily be distributed to the at-home staff electronically.

The at-home keyboarders code the papers in $\mathcal{AMS}$-TEX and run the paper through a syntax check when

finished. The paper must successfully pass the syntax checker before it can be submitted for proof copy. These keyboarders do not have graphic terminals nor printers at their disposal, so they are never able to actually view the output of their $\mathcal{AMS}$-TEX coding. They are encouraged to key the paper as quickly as possible, and are paid by the keystroke.

Every night proof copy is generated for all papers keyed that day, and these are delivered back to the editorial staff along with the original manuscript. The proof copy is read against the orginal manuscript, and corrections are marked on the proof copy. The paper is then given back to the same keyboarder who originally keyed the paper for corrections. Payment at this stage is a small, fixed amount per page. When these corrections have been made a set of proof copy is sent to the author for approval.

In the case of an electronic manuscript, after it is logged into the tracking database the file is given to a keyboarder who changes the documentstyle being used from a generic style to the appropriate publication style. A utility is run from within the keyboarder's editor which searches the file for certain unwanted typesetting codes such as those dealing with page and line breaks, vertical and horizontal space, magnification, `\hsize` and `\vsize`, and control characters. This utility produces a report at the top of the data file listing what it located, and the keyboarder searches the file for these items and alters them as needed. The keyboarder must also check any local definitions made by the author to make sure they don't conflict with AMS definitions.

At this point proof copy is generated and sent to the editorial staff, where it is copyedited only to the extent necessary to mark for change anything that does not conform to AMS style. After these corrections are made a proof copy is sent to the author for approval.

An electronic submission can be rejected if the author has not used $\mathcal{AMS}$-TEX or $\mathcal{AMS}$-LATEX, or if they did use one of these packages but did not follow the suggestions in the Guidelines. The overriding consideration in making this decision is whether it would be quicker to alter the author's file to make it compatible with the production system, or to re-key the author's paper. A short paper (5–10 pages) submitted in plain TEX without a great amount of mathematics will invariably be quicker to re-key.

At the point in which a paper is sent to the author, the production cycle becomes identical for papers originating either as electronic submissions or as paper manuscripts. After the author has approved the proof copy and perhaps requested some alterations, corrections are made and proof is again returned to the editorial staff. At this stage the editorial staff will combine individual papers to comprise an issue of a book or journal, and from this point onward these papers will be processed as a unit.

The issue is now paged, and specific formatting instructions such as line and page breaks are inserted. No line and page breaks are allowed in any correction runs prior to this stage, since the author may request alterations which would affect paging. These specific formatting instructions are held to a minimum, however, and the camera copy is sometimes altered by hand in preference to additional TEX runs.

4.2 Author-Prepared Books

The AMS also chooses to publish certain books directly from camera copy supplied by the author as a cost saving device, and to speed publication. The popularity of TEX among academic and other research centers has allowed for the production of these low-cost books with little sacrifice in quality.

Aside from the generic documentstyles described above, the Society also distributes $\mathcal{AMS}$-TEX and $\mathcal{AMS}$-LATEX documentstyles for a few specific book series which publish author-prepared copy. Authors are encouraged to use these documentstyles in the preparation of their work, and submission of both the TEX source file and the DVI file is requested. The DVI file is simply spooled to a typesetter, and the TEX source file is archived for possible future use. If the author has followed the suggestions in the Guidelines and used the package correctly, the final result of books published in this fashion will have the appearance of a comparable book produced via the traditional route, with the exception that the author-prepared work will be published in Computer Modern fonts while all other AMS publications use Times Roman.

5 Future Directions

5.1 Production System Enhancements

Several major enhancements to our current production are either under way or being planned for the next few years. The TEX macros which comprise the current system have been altered many times by many different people over the past decade. A study last year revealed that somewhere around 50% of the code in the current system was either obsolete or never executed. A rewrite of the system is currently underway.

One major difficulty that the rewrite will address is the ability to easily produce a set of typographic specifications for each publication. The Editorial Department has a set of written specifications for each publication, but it is a difficult task to ensure that the TEX macros which produce the publication exactly match the written specifications. Work is being done to have the TEX macros actually read a file containing the typographic specifications for each publication.

A prototype of this file is being developed. Every logical component of a book or journal article is listed in this file, along with such attributes as the font to use, the amount of space (maximum, minimum, and optimal) before and after the element, penalties for page

breaks before and after, other attributes such as linespacing, wordspacing, case, justification, etc., are listed in the file in a format that is both human readable and can be parsed by TEX.

Another problem area being worked on is the ability to change a documentstyle for a given publication without affecting papers for that publication which are already in the production stream. Changes to publications are sometimes requested to begin with a specific issues (such as the first issue of a new year), but when papers are accepted for many journals, they are not accepted for a specific issue but instead go into a pool for that journal. Therefore, it is often the case that papers for several different issues of a journal are already in process at various points along the production stream at any given time.

The idea here is to embed a date into the data file which TEX will read. In addition, whenever changes are made to any macro files in the production system, a copy of the code is first copied to the bottom of the macro file before any changes are made. The unaltered code is marked with the date the revisions are being made to the file, and then changes to this code are made in the main section of the file. When a data file is processed, TEX reads the date in the data file and then reads the macro file(s). After the main section of the macro files are read, TEX checks the dates on the altered code at the bottom of the macro file, and stops reading when the date in the macro file is less than the date in the data file.

For example, supposed `\heading` is to be changed from boldface to italic type, and suppose there are two data files, one with a date of 1-MAY-1992 and the other with 1-JUN-1992. Suppose also that the change to `\heading` was made on 15-MAY-1992. The desired result is that the data file keyed on 1-MAY-1992 should retain boldface headings, while the 1-JUN-1992 should pick up the new italic headings. The diagram below shows how the above scheme accomplishes this.

macro file: `\def\heading#1{\noindent\bf#1}`
(additional code)

`\dateofchange{15-MAY-1992}`
`\def\heading#1{\noindent\it#1}`

data file #1: `\filedate{1-MAY-1992}`
ITALIC Headings
data file #2: `\filedate{1-JUN-1992}`
BOLD Headings

5.2 Changing Fonts: CM, Autologic Times Roman, PostScript Times Roman

The STI system used during the late 1970s and into the early 1980s typeset the AMS publications using Times Roman fonts. For a period of several years during the mid 1980s some AMS journals were published which contained articles typeset with both STI and TEX in the same issue. The enthusiastic move toward a TEX-based production system was slightly diminished by the lack of TEX-compatible Times Roman fonts at that time. Many people involved in the publications process felt that Times Roman had become an industry standard for technical publications for its readability and appearance, and they were somewhat reluctant to publish using TEX's Computer Modern fonts.

In 1987 the first Autologic APS-μ5 was purchased, and work was immediately begun to create a set of TEX-compatible Autologic Times Roman fonts. Accessing the Autologic Times Roman, Italic, and Bold fonts were fairly straightforward, but the creation of an Autologic Times Math Italic font for use with TEX was a major effort. The details of this work were presented at tenth annual TEX User's Group meeting at Stanford University in August of 1989. A published version of this presentation, entitled *Migration from Computer Modern Fonts to Times Fonts* by R. E. Youngen, W. B. Woolf, and D. C. Latterner can be found in TUGboat, volume 10, number 4.

At the time when the APS was purchased, high-resolution PostScript typesetters were available in the marketplace, but they were all much too slow for the volume of production at the AMS. However, since that time the hardware has improved dramatically, and PostScript has assumed its place as an industry standard. The Society is now preparing to move book and journal production to PostScript fonts over the next few years.

Switching production to PostScript Times fonts is contingent on a PostScript Times Math Italic font, equivalent to the Autologic version in current use. The *MathTime*$^{\mathrm{TM}}$ fonts recently released by Michael Spivak contain such a font, and plans call for this to be evaluated along with the possibility of the AMS creating its own such font.

The move to PostScript is driven not only by its status as an industry standard, but also for the promise of such things as auto-inclusion of PostScript graphics into TEX documents and the ability to output completed pages directly to film negatives or plate materials. Authors' graphics are currently re-rendered on a Macintosh using Adobe Illustrator, which is a PostScript-based application. These graphics are currently being output to the Society's PostScript imagesetter and stripped onto the camera copy of the TEX document by hand.

By typesetting completed pages, substantial cost savings may be realized if typesetting directly on plate materials proves practical. The current cycle is to produce camera copy from a typesetter which is taken to the print shop and photographed to produce a negative. The negative is then used to make a metal plate that can be placed on the press. Typesetting directly to plate material would save in both labor and materials costs for the intermediate steps.

5.3 Electronic Dissemination/Publication

Dissemination of information by electronic means is not a new concept to the AMS. Since 1982, *Mathematical Reviews (MR)* has been distributed electronically as well as on paper, with the data being installed in the public databases of vendors such as Dialog and ESA. Prior to 1985, MR was typeset using the STI system, and the data had to be converted into a human-readable form for database installation. Effective with the January 1985 issue, TEX became the sole input language. Since then, a set of macros has been developed to permit database users to download items from the databases to their local computer, process it with TEX, and print out the result, which is for all practical purposes identical to the published review. In 1989 MR was also released on CD-ROM which is accessible using search software on PC-compatible and Macintosh computers. Full reviews from 1980 to the present are now available on CD-ROM, along with bibliographic information from 1940–1979.

Beginning this year, the *Bulletin of the American Mathematical Society* is also being distributed electronically via the Society's e-MATH node (e-MATH.ams.com) on the Internet. The *Bulletin* is a free journal as a privilege of membership and $\mathcal{AMS}$-TEX files for this journal are installed on e-MATH for public access. Files for this journal are prepared in the same way as for any other journal, and a paper copy is still produced. Experience with this service will help determine the direction of other similar experiments.

5.4 Standard Generalized Markup Language (SGML)

The AMS has already concluded, independent of the SGML initiative, that generic markup is the only sensible way to approach electronic storage and processing of manuscripts and bibliographic data. Though strict SGML syntax is not used in the production of AMS publications, the underlying principles of SGML are in practice.

The Ann Arbor office is developing an SGML-based software tool which will perform several important document-handling functions, such as:

- accept an $\mathcal{AMS}$-LATEX file from an author, parse it into an SGML document, and store it in a document database which recognizes and manages the structural parts of the document;
- accept an SGML file and produce an $\mathcal{AMS}$-LATEX file which can be typeset through TEX;
- allow several people (cooperating authors, editors, etc.) to work together (from remote sites) on the creation of a document, maintaining *version control*, which will allow authors, editors, etc., to track the modifications each has made to the file;
- allow referees, editors, reviewers, and readers to make comments about the file, which will be stored as *annotations* attached to specific points or regions in the document, with identification of the commentator and a record of the time of annotation.

The expectation is that this tool will someday allow for on-line editing in the production of the AMS publications, replacing the cumbersome process of creating paper galleys for each correction run. Instead of corrections being marked on paper with colored pencils by an editor, then keyboarded by a typist, and finally re-proofread by an editor, the editor will be able to make corrections directly in the file and see the results instantly on the screen.

Perhaps even more importantly, this tool will define an environment for true collaborative development of electronic and paper publications by authors situated at remote sites throughout the world. *Version control* and *annotation* are the key capabilities required for a tool appropriate to collaborative research and writing in mathematics.

6 Conclusion

It is impossible to predict to what extent electronic publishing will replace traditional publishing on paper. While the advantages of electronically stored and produced documents are many and conclusive, the affinity for paper publications may be difficult to overcome. As computing devices become smaller and more portable and are linked by radio networks, the difference between sitting beside the fireplace with a book you pulled off the shelf and a book reader which electronically retrieved your favorite text becomes less obvious. TEX and SGML, or some offshoots thereof, are well positioned to play an important role in this future.