RESEARCH ARTICLE | MAY 01 1994

# How AIP Handles Electronic Submissions of Articles FREE

Christopher B. Hamlin; Peggy Judd



*Comput. Phys.* 8, 255–258 (1994)
https://doi.org/10.1063/1.4823296

# HOW AIP HANDLES ELECTRONIC SUBMISSIONS OF ARTICLES

Christopher B. Hamlin and Peggy Judd

With the wide availability of computers and word-processing software, it can now be assumed that material submitted for publication has been produced from a computer file of some sort. Whenever possible it makes sense to reuse the author's file in the production process rather than starting over and rekeying the entire file. In the future, the author's file—an electronic document—may be the basis for submitting, refereeing, editing, distributing, and archiving the article.

The American Institute of Physics (AIP) has worked to accommodate various types of electronic files as input for both books and journals. TeX and variations on TeX such as REVTeX and LaTeX (see below) have proven to be the most popular formats. A total of 152 REVTeX and LaTeX files were submitted to AIP-produced journals in 1993, and the number of such submissions continues to rise.

AIP has developed routine production procedures to handle REVTeX and LaTeX files. Considering TeX's excellent typesetting capabilities, one might be tempted to use TeX output directly for journal pages. We decided not to do so because, as discussed below, we need the capabilities of a full-featured composition system to produce AIP's printed journals, and we wish to standardize on the output of Xyvision, the composition system that we have in place, as the basis for electronic archiving and potential online delivery of journal information.

Our approach to compuscripts therefore involves translating REVTeX and LaTeX files so they can drive the Xyvision system. The translation program is not trivial, but some of the common features of TeX and Xyvision (especially high-level markup) have helped to keep it manageable in scope. Following brief discussions of TeX, Xyvision, and the composition process, we will describe some of the technical challenges we faced in creating the translation program. The importance of this subject can be appreciated through its role, as outlined above, in the ongoing development of AIP's electronic-publishing capabilities.

## REVTeX

The TeX typesetting system was created by mathematician and computer scientist Donald Knuth. TeX is a text-processing program rather than an interactive word processor, but it has become very popular for several reasons. Anyone is allowed to use the ideas and code of TeX freely, and TeX—along with its supporting programs—is now available on almost all computer systems in use today. Though commercial versions exist, there are also many excellent free distributions available. TeX is easily programmable, and the macro files written for TeX can be distributed and used on various systems with a minimum of compatibility problems. [The article in CIP 8:1 (1994), p. 20, describes how to transfer TeX files on the Internet.] In addition, TeX excels at typesetting mathematics. For those people with connections to networks like Internet or Bitnet, a great deal of support is available from other users.

LaTeX is a set of macros that run on top of TeX. The macros provide a number of higher-level commands that make the software easier to use. REVTeX [originally developed as an electronic publishing product by the American Physical Society (APS)] is a set of macros that extend the normal LaTeX environment for articles. The first version of REVTeX was released by the APS early in 1988, and version 2.0 was released in April 1991. APS REVTeX-prepared author submissions quickly grew to comprise in excess of 20% of the finished *Physical Review* journal pages.

APS's favorable experience led AIP and its Member Societies to form a Joint Society Working Group to develop a standardized author tool. This project got underway in 1991. In 1992, AIP, in collaboration with APS and the Optical Society of America, released REVTeX 3.0, a fairly universal author tool for authors who publish physics articles in scholarly journals. Benjamin Bederson, APS Editor-in-Chief recently stated, "The use of REVTeX continues to burgeon, now accounting for about 25% of all manuscripts submitted to APS journals, and REVTeX has become a virtual standard for author-prepared physics manuscripts for AIP as well as APS journals."

The REVTeX 3.0 software and associated "toolkit" information include society-specific style files and instructions on how and when to use them. An author guide is available electronically as part of the software distribution (see "Getting REVTeX 3.0," p. 257). Authors must have LaTeX installed and running in order to use REVTeX.

REVTeX 3.0 has many advantages for physics authors. Especially useful are the many special macros, keystroke savers, and starting templates stylized specifically for physics information. Authors can also benefit from LaTeX and REVTeX features such as automatic numbering and cross

*Christopher B. Hamlin is the technical author of REVTeX and manager of Publishing Applications Support at AIP. E-mail: chamlin@aip.org*

*Peggy Judd is the director of Information Technology at AIP. E-mail: judd@aip.org*

referencing of section headings, equations, and references. By submitting a compuscript, the author reduces the proofreading task at "page proof" stage, and fewer author corrections are necessary prior to publication.

For the journal publisher, the primary advantage of using compuscripts in the production process is that pages are composed from the same file submitted by the author, decreasing both the copyediting and the proofreading cycles in house. A reduction in composition costs can also be achieved, but only if authors closely follow the procedures outlined in the instructions, and only if the number of REVTeX submissions rises to a significant level, so that the staff becomes proficient in handling them. We estimate that breakeven in this transition process will occur when the percentage of submissions in REVTeX rises to approximately 35% of the total.

## Composition System

AIP's present composition system is Xyvision. Xyvision is a full-featured system that integrates text, graphics, math, and tables into finished articles. Page layout is fully automatic. Complete issues of journals can be created from articles, including headers, footers, and automatic page numbering throughout the issue. We are now completing an upgrade of the Xyvision system that targets PostScript as the output device of choice. We are also in the process of integrating scanning in order to digitize all graphics in AIP journals, and so one product of the composition cycle will be a file of complete PostScript pages for each article.

Another useful feature of Xyvision is its capability for user-defined tags. A simple tag can set in effect complex typesetting commands—for example, `{h1}` to start a main section head. If a system-support person has set the tag `{h1}` to the correct typographic specifications, then, when a keyboarder enters `{h1}Introduction%`, Xyvision takes care of the typesetting appropriate for a section head. In fact, the `{h1}` tag is used across all articles regardless of fonts, spacing, and so on, and it is defined to execute the correct typesetting for the particular journal for which the article is scheduled.

Such high-level (or generalized) markup can be extremely helpful. It eliminates the need for keyboarders to memorize and input vast amounts of typographic information. Keyboarders only need to input the appropriate tags. So, generalized markup on Xyvision simplifies keyboarding, promotes consistency in typesetting, and adds information to the file that can be used in various ways.

## Translating TeX

Although there are advantages to using the output from TeX for the finished article—mainly the degree of author control over the final product and TeX's excellent math typesetting—there are disadvantages from AIP's perspective:

- TeX's page-layout mechanism is notoriously difficult to control for complex layout.
- Composition workers would need to be trained on a completely different typesetting system.
- A large amount of work would be needed to maintain a uniform appearance for material being produced by either Xyvision or TeX.
- Using TeX obviates Xyvision's advantages in assembling and managing preparation of issues as a whole, rather than just unrelated individual documents.
- AIP generates information from its composition system for secondary products, including GPAA, SPIN, and indexes. Integrating information from two separate production systems makes things much more difficult.

These considerations led us to develop procedures for integrating TeX-related input into the process of composing on the Xyvision system. The LaTeX/REVTeX-to-Xyvision "translator" that we have developed actually comprises a number of processes that run one after the other.

Some of the translation is simple one-to-one mapping. This type of conversion is handled by Xyvision's Xychange facility, which enables powerful search-and-replacement rules to operate on files. Single characters, for example, have simple, unambiguous names in both Xyvision and LaTeX: ξ is produced by `|gy` in Xyvision and `\xi` in LaTeX. This conversion is handled by a Xychange rule.

On the other hand, LaTeX has a number of complicated facilities, especially in automatic numbering and cross referencing. Such constructions cannot be handled by simple search and replace—what is needed is a program that is able to parse the complicated structures in the input file and take appropriate action. This part of the translation process is programmed in lex and yacc, the well-known compiler-construction tools most commonly associated with UNIX development systems. [A recent physics application of lex and yacc

is described in CIP 8:1(1994), p. 98.]

An example of "parsing and acting" arises in connection with LaTeX's `\label` and `\ref` commands. The `\label` command accepts one brace-delimited argument. This argument is then associated with the correct cross-referencing output for the context within which it appears, and the output may be reproduced by passing the `\ref` command the same argument.

Following the LaTeX command for a section heading, any `\label` command will create a cross reference to the number assigned to that section heading. So, for example, assuming we are at the beginning of a paper,

```
\section{Introduction}
\label{arbitrarylabel}
We discuss in Section
\ref{arbitrarylabel}...
```

could produce the output

```
I. Introduction
    We discuss in Section I . . .
```

assuming that sections have been set up as numbered with roman numbers.

Actually, the `\label` command reacts to a kind of scoping. A main section is one block of scope. A subsection within that main section defines its own, more-restrictive block wherein `\label` will cross reference the subsection number.

An equation within the subsection is yet more restrictive. Using `\label` within a numbered equation will create a cross reference to the equation's automatically assigned number, not the section containing the equation.

Immediately following the end of the equation, the equation scope disappears and is replaced by the less-restrictive scope surrounding it. For example, at the start of an article, the LaTeX code

```
\section{Introduction}
\label{myseclabel}
Here in Section \ref{myseclabel} we
show that
\begin{equation}
E=mc+2 \label{myeqlabel}
\end{equation}
\label{myotherseclabel}
is Equation (\ref{myeqlabel})
in Section \ref{myotherseclabel}.
```

produces

```
I. Introduction
   Here in Section I we show that

   E=mc^2                    (1)

   is Equation (1) in Section I.
```

Note that the automatic numbering and cross referencing are all done with just `\label` and `\ref`, which also work for subsections, tables, and figures.

A further complication is that LaTeX allows forward cross references—that is, a `\ref{xxx}` command can be used in the file before the corresponding `\label{xxx}`. This is implemented in LaTeX by using a two-pass system: at the end of each run of the file, LaTeX writes all defined labels to an auxiliary file. At the start of the next run, LaTeX reads in the auxiliary file, effectively defining all the tags from all the `\label` commands in the file at the beginning of the run. This allows LaTeX to resolve any possible `\ref` command at the very start of the run. (Note: When authors add new `\label` commands, they must first run the file to create a complete list of tags in the auxiliary file, and then rerun the file to resolve them.)

Given all this complication, it is clear that a very sophisticated translation of the LaTeX file is required. Numbering and cross referencing are done in two parts of the translation process. The first part parses the file according to a yacc grammar that describes the structure of LaTeX involving automatic numbering and cross referencing. During this pass, a symbol table that links author tags with output is written to a file. In a later pass, the argument to a `\ref` command is found in the symbol table and the appropriate output is inserted in place of the `\ref` command.

We have seen that LaTeX's "higher-level" features can present difficulties in translation, but some of these features also can be very useful. For example, the generalized markup in LaTeX may often correspond closely to that in Xyvision, so that, for example, `\section{Acknowledgments}`

## Getting REVTeX 3.0

Anyone can get a copy of REVTeX free of charge. It is easily accessible by anonymous FTP over the Internet or by e-mail, or you can order a floppy-disk version from AIP. Remember: you must have LaTeX running in order to be able to use REVTeX.

Internet users can retrieve REVTeX by anonymous FTP from PINET, AIP's Physics Information Network. REVTeX is in the directory `/pub/revtex` on `pinet.aip.org`. Get all the files in this directory and read the `readme` file.

REVTeX is also available by e-mail. Just send a mail message to *fileserv@shsu.edu* (Internet). In the body of the message, write simply

```
SENDME REVTeX
```

The file server will queue your request and send the files to you within a couple of days. Note that some assembly of the files will be needed if they are fetched by e-mail.

If you would like more detailed instructions for electronic retrieval, or require a DOS diskette version of REVTeX, contact Katie Corey, 500 Sunnyside Boulevard, Woodbury, NY 11797; phone: 516/576-2335; e-mail: *katie@aip.org*.

translates directly to `{h1}Acknowledgments%`. Without the generalized markup denoting a section heading in LaTeX, it would have been more difficult to locate the heading and mark it correctly for Xyvision composition. Therefore, success in converting author files (whether from LaTeX or another form) depends to a great degree upon the author's cooperation and consistency in following the publisher's required formatting.

## Translations and books

Most AIP journals (including this one) now accept compuscripts, and efforts are underway to extend this capability to the remaining few that do not. We also are extending electronic procedures in the translations and books areas.

Within the Translations Division, AIP has been moving toward greater use of LaTeX and REVTeX. At present, one Russian journal is keyed directly into LaTeX as it is translated. The file can then be translated into Xyvision and used for production. This project is working very smoothly, and the next step, now underway, is to broaden the use of LaTeX and REVTeX among translators and involve editors. When translators, editors, and AIP can all use the same tool, the translator's original file will be sent to the editor, who can edit the file directly and forward the corrected file to Production.

The books division of AIP—AIP Press—has a number of book projects in TeX and has already developed a standard, REVTeX-based toolbox with macros, a sample file, template, and author instructions. The nature of book production is different from that of journal production. AIP can provide different services, depending on the project, including custom macros, TeX support, and Xyvision typesetting from LaTeX or REVTeX files. The author, the editor, or AIP may create the final output, and either TeX or Xyvision may be used. Authors and editors need to discuss options early on with AIP Press, so that proper arrangements can be made to accommodate their projects.

## The future

AIP is in the process of expanding its author submission program to include articles prepared in WordPerfect and Microsoft Word. Some of our journal editors feel that broadening our acceptance criteria to include these word-processing programs will increase the number of compuscripts we receive. During the past year, we solicited submissions for test purposes and found that neither of these programs handles the complex math as easily as TeX. Because we are convinced, however, that use of both WordPerfect and Word will increase within our author community, we are developing conversion routines that will allow successful acceptance of these files. As part of this project, at present both the *Journal of Vacuum Science* and the *Journal of Applied Physics* accept compuscripts prepared in Word or WordPerfect.

We plan also to develop procedures for successfully receiving and utilizing graphics files. We are currently working with one of our journals where all the graphics files come from a single, well-defined, and reliable source. Based upon such studies, we hope to recommend at least one "standard" set of submission procedures for authors wishing to submit graphics files.