ELECTRONIC PUBLISHING OF THE PAC95 PROCEEDINGS

L. T. Gennari, R. H. Siemann, Stanford Linear Accelerator Center,
Stanford University, Stanford, CA 94309 USA*

The Particle Accelerator Conference (PAC) series has been a successful one with a steady rise in participation and in the size of the conference proceedings. The PAC conference proceedings are primary references for accelerator engineers and physicists, and their growth has created some problems. The main one is the efficient access to and use of the information in them. In light of this, and with the uncontrollable urge to do something new and different, we proposed to the PAC95 Organizing Committee that we publish this proceedings electronically. They accepted. This paper summarizes our adventure.

I. PUBLISHING WITH ABODE ACROBAT™

Having made, in principle, the decision to publish electronically, we sought advice from a committee of accelerator scientists interested in electronic publishing. The PAC95 Electronic Publication Committee was: Joe Bisognano (CEBAF), Robert Hamm, Jr. (AccSys Technology, Inc.), Gerry Jackson (Fermilab), Bob Jameson (LANL), Kwang-Je Kim (LBNL), Tom Knight (SLAC), and Jim Simpson (ANL). We also benefited from advice from two experienced individuals outside the accelerator community, Robert Kelly of the American Physical Society and Michael Keller from Stanford University.

One of the conditions the Organizing Committee had for approving electronic publishing was that there also be a paper version for participants desiring paper rather than an electronic format. Due to the copyright issues involved in making a paper-published volume available through the World Wide Web, we chose to publish on CD-ROM.

Adobe Acrobat seemed ideal. The Portable Document Format (PDF) promised documents that could be read on any platform no matter what hardware and software combination was used for composition. PostScript™ files produced by any word processor with embedded figures from various graphics programs could be distilled by the Acrobat Distiller into PDF files that could be viewed with the Acrobat Reader, which was available on a number of platforms. Some of the features we would have hoped for, such as full text searching, were not available in January, 1994 when we made the decision to use Acrobat, but proprietary conversations with Adobe indicated they would be by the time we were ready to publish. We were not disappointed in this regard.

The PAC conference has always attracted scientists from around the world. In 1995 the Particle Accelerator Conference was combined with the International Conference on High Energy Accelerators, making the conference truly international. We had to anticipate documents prepared on a wide range of computers with a wide variety of software and by authors with a wide range of computer skills. Conferences are an activity that people attend and contribute to voluntarily,

so it was important to place as few restrictions as possible on the methods of document preparation. As a principle and as a practical matter we decided there would be *no restrictions* placed on computers or document preparation software. This made Acrobat especially attractive. Ideally, authors could prepare their papers using whatever methods they normally used, and as a last step all they had to do was produce a PostScript file that we would distill. Although this ideal was not achieved, for the reasons discussed below, we still think it was the correct decision to place as few restrictions as possible on the way authors work and that this should be a guiding principle for future conferences.

II. INITIAL EXPERIENCE

We benefited from the beginning from two important collaborations. Steve Myers, Christine Petit-Jean-Genaz, and John Poole from CERN were interested in electronic publishing of the proceedings of the 1996 European Particle Accelerator Conference. We had numerous discussions with them, and evolved a common approach to electronic publishing. They used some of our ideas as the starting point for electronic publication of the proceedings of the 1995 LEP Performance Workshop. These proceedings are available on the WWW[i] and are the first extensive use of Acrobat for publishing in the accelerator science world.

The LEP Performance Workshop was a good trial project. The contributions are predominantly from the SL Division at CERN, and John and Christine were available to the authors to produce templates, TeX macros, etc. and to solve problems as they arose. They saw the first manifestations of some of the problems we encountered in PAC95.

The second collaboration was with Jim Simpson and Paul Schoessow (Argonne National Laboratory). They were considering electronic publication of the proceedings of a small conference, the 1994 Advanced Accelerator Conference (AAC).[ii] We agreed to work together, first developing submission techniques and then evaluating the results. This collaboration was critical. We changed the PAC95 submission procedures drastically based on their experience, and we would have failed without their help.

Authors submitted PostScript files prepared with only a few restrictions to the AAC. The hope was that these files would be distilled into PDF files with little trouble. This was not the case due to three major problems:

- Most PostScript files prepared from TeX source used Type 3 fonts. The distilled files looked great when printed but were almost illegible on the screen.
- There were problems incorporating graphics files into documents.
- Many authors were inexperienced at producing electronic documents.

These are the problems we would encounter also, but we were able to modify the submission procedure to reduce their

impact based on the AAC experience. We requested the following be submitted:

1. A diskette containing the complete printable paper, i.e., containing all figures, graphics, diagrams etc., as a single PostScript file. This file was to fit on a single DS, HD floppy disk (1.4 MB) either full size or using standard compression methods, i.e., Stuffit, etc. Authors who had Adobe Acrobat 2.0 software could submit a PDF file instead of a PostScript file.
2. One or more diskettes containing the native format (WORD, TeX, etc.) source code for the document. This would be used to reconstruct the paper if we were not successful distilling the submitted PostScript file.
3. One or more diskettes containing the native format (Illustrator, EXCEL, etc.) source code for the figures and charts. These would also be used to reconstruct the paper. Almost everyone submitted the native format source code for the document as requested in *2* above, but few people submitted the native format source code for the figures and charts. Our instructions were not clear enough. Fortunately, we developed techniques for recovering the graphics from the submitted PostScript files and had access to a scanner.
4. A paper copy of the paper. This was needed for showing us what the authors thought the paper looked like, for providing text, figures, and equations to use in proofing the PDF file, for providing us with figures that could be scanned if necessary, and, in a few cases, for use in an optical character recognition (OCR) program.

An FTP site was available for authors who wished to submit by FTP rather than diskettes. We asked for paper copies of these contributions at the conference.

III. PAC95 EXPERIENCE

There are 1099 papers totaling 3429 pages in the proceedings. Most were submitted following guidelines 1, 2, and 4 above. There were only hard copies for a few contributions from Third World authors who did not have the hardware and software to meet the submission requirements. The programs used to compose these papers are summarized in Table 1.

Table 1. Document Source Codes

Ami Pro	4
Claris Works	2
FrameMaker	51
Island Write	2
MacWrite Pro	4
Microsoft Word, v. 2	30
Microsoft Word, v. 5 & 6	505
Nisus Write	1
PageMaker	5
Page Stream	1
TeX	382
WordPerfect	54
Unknown	58

About one-third of the PostScript files could be distilled to produce a satisfactory PDF file. Distilling times, using Acrobat Distiller version 2.0, ranged from one to three minutes on Power Macintosh 7100s. Each PDF file was proofed on the screen and on paper with particular attention to figures and equations. Distilling took a small fraction of the total time needed to handle a paper. Proofing the PDF file and entering and checking manuscript information in a database were much more time consuming.

Some PostScript files could be recovered by editing them with Adobe LaserTalk™ on the Macintosh or emacs on the Unix system, and the OCR program Text Bridge was used in a few cases. Finally, one or two papers were typed in from scratch because the extensive use of equations made OCR impractical. We had to generate new PostScript files from the supplied source files for the remaining papers, roughly two-thirds of the total.

The first major problem that caused us to reprocess a large number of papers was the extensive use of Type 3 fonts in PostScript files produced by TeX. Type 3 fonts become bitmaps in PDF documents. At 72 dpi, Macintosh screen resolution, they are completely illegible. Figure 1 shows a document created with Type 3 fonts and a document created with Type 1 fonts. We recompiled these papers using Type 1 rather than Type 3 fonts. This problem could be minimized in the future by making authors aware of the problem and providing macros and easy access to Type 1 fonts.

The conference proceedings will be available as a book and as a CD-ROM. Participants will choose whether they receive the proceedings as a book or/and a CD-ROM at the time they register. The advantages of the CD-ROM include its compact size, the ability to copy and paste figures and text, and searchability. We hope that many of you will find these features attractive and choose the CD-ROM.

Figure 1. Documents created with Type 3 (top) and Type 1 (bottom) fonts.

The second major problem was associated with graphics. A wide variety of plotting and illustration programs was used.[iii] These programs can have many different output formats (although Encapsulated PostScript (EPS) was most commonly used), and we encountered numerous bugs in either the way graphics files were written by the software or in the way they were incorporated into word processing and page layout documents.

Some of the lines and patterns drawn by the various graphing programs are drawn with Type 3 fonts. Since our machines did not generally have whatever the original software and its associated fonts, the distiller would substitute
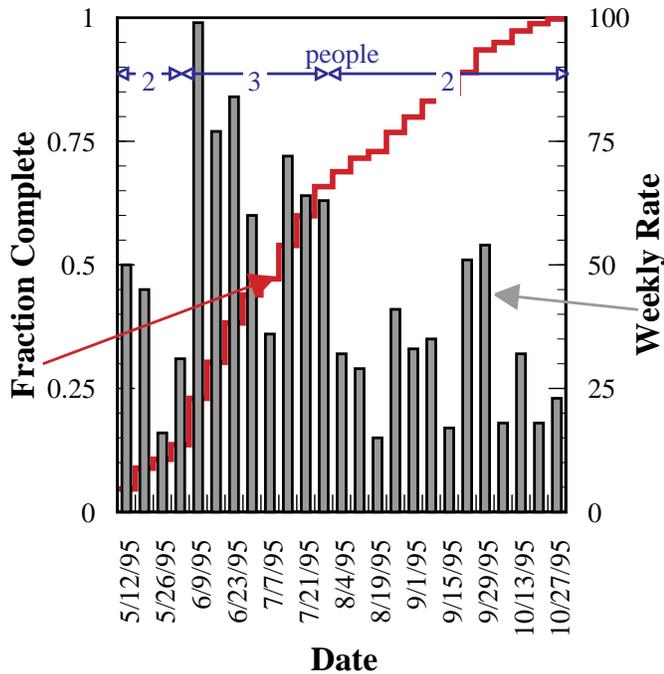
Figure 2. PAC95 proceedings processing summary. The number of people working is indicated and includes temporary workers and editorial staff.

printing times became too long. These problems would have been avoided if authors had paid closer attention to the 1.4 MB file limit. It should have told them that the graphics files they were generating were too large for this publication

The other common mistake was inadequate keying of figures. For example, a plot might look good on a computer screen with colors distinguishing different lines, but when printed in black and white, the colors are lost and the lines no longer distinct. When we encountered this we used Illustrator to provide a clearer key.

We hired two temporary workers, Jianan Lu and Jamie Walker, to process the papers. They were both experienced computer professionals looking for short term employment. They distilled documents, made a first evaluation of the quality of the PDF file, and worked with source codes to recover papers. The production history is summarized in Figure 2. Each person was able to process about 20 to 30 papers per week. The work in the beginning was mainly distilling, and in the end it was mainly recovering papers from source files. We were fortunate that most problems were not unique, and once problems were identified, the solution was generally one of a few known procedures. If this had not been the case, the production rate at the end would have fallen dramatically.

Some of the papers simply couldn't be saved with the pieces we had. We requested source files or replacement source files for about 10% of the papers.

We decided to make each paper a separate PDF file with hypertext links pointing to it from both the Table of Contents and the Index. Separate files, rather than a smaller number of larger files, make it easy to print a single paper with a few keystrokes without worrying about printing a large document by accident.

This decision did, however, create a problem—including unique page numbers in the PDF files. Unique page numbers are necessary since there will be both paper and CD-ROM proceedings, and there needs to be a way to cite a paper unambiguously. Our solution was to put the page number in the Subject field of the PDF Document Info summary. AppleScript was used to extract the title, authors, keywords, and page number for each paper from a FileMaker Pro database and insert them into the corresponding PDF file's Document Info.

An index for full text searching was created with Acrobat Catalog, and the search engine is included on the CD-ROM along with the Acrobat Reader.

fonts or patterns of its own. Often this resulted losing the uniqueness of the lines in the original, or disappearing patterns.

Some authors incorporated their figures by reference rather than embedding them in the word processor file. This meant that we had no electronic source for the figures. These authors also tended to be in the group that submitted native format source files of their text but not their figures.

Two procedures were used to recover graphics files. First, Adobe Illustrator™ version 5.5 has a plug-in to read PDF files. Even when the submitted PostScript file produced an unsatisfactory PDF file, some figures could be salvaged by reading the PDF file with Illustrator, selecting the figure, and deleting everything else on the page. The figure was then reduced, magnified, or edited as needed to agree with the hard copy and saved as an EPS file to be incorporated into the source. This procedure worked well in practice because the EPS files written by Illustrator are dependable. They could be included into the text files without any problems.

The other procedure was to scan the figure from the hard copy of the paper. We preferred the first procedure because it preserved the colors the authors had chosen to use. We resorted to scanning only when we did not have an adequate PDF file or when the figure in the PDF file had another problem. Most of these were associated with the inexperience of authors in preparing documents for electronic publication.

This inexperience was the third major problem, and it was associated mainly with figures. In some cases figures drawn by CAD programs with exquisite detail (and consequently large file sizes) were reduced, without downsampling, to visually tiny figures with enormous file sizes. These figures would distill properly but take forever displaying on a screen and printing. We scanned these figures when the display and

IV. SUMMARY

This CD-ROM meets our goal of providing efficient access to and use of the information in the proceedings. Features are:

- A full text search that accepts Boolean constructions and can also include titles, authors, and keywords as separate search criteria. This should make it easier to find information.
- Data tables can be electronically copied and pasted into other documents.
- Figures that are restricted to small representations on paper can be magnified on screen up to 800% to

show details, and figures can be incorporated into other documents using Adobe Illustrator.

- A compact size for the proceedings requiring only a small amount of shelf space.

There are cosmetic flaws with some of the papers. For example the sources processed with Textures have some large, odd looking characters in the upper left hand corner of each page when viewed on screen. This is due to a hack in the Textures PostScript code, which is intended to force the printer to download the document fonts. We have made Adobe and Blue Sky Research aware of this problem.

Preparing these proceedings took a considerable amount of work, but the software was relatively new and people were inexperienced. Both will improve, and preparing a conference proceedings for electronic distribution will be easier in the future. We are pleased that the 1996 European Particle Accelerator Conference is planning to use Acrobat software and many procedures similar to ours.

We are proud of this CD-ROM. We hope that you, the conference attendees, will find it opens up new and interesting ways to access and use scientific literature.

V. ACKNOWLEDGMENTS

Acrobat, Illustrator, PostScript, and Lasertalk are registered trademarks of Adobe Systems, Inc. All other software titles are registered trademarks of their respective owners.

---

[i] http://www.cern.ch/CERN/Divisions/SL/publications/ proc95/pdfproc.html

[ii] Advanced Accelerator Concepts, AIP Conf. Proc. **335**.

[iii] There was an average of 3.75 figures per paper. We do not know all the programs used because few figure source files were submitted, but we had graphics from Alchemy Mindworks, Auto CAD, Canvas, Claris CAD, Claris Draw, Cricketgraph, DeltaGraph Pro, Desk Scan II, EXCEL, Hippo Plot, Igor Pro, Illustrator, KaliedaGraph, MacDraft, MacDraw, MacView Plus, MAFIA, Mathematica, MATLAB, Ofoto, Paintbrush, PAW, Photoshop, PowerPoint, PVWave, Sigma Plot, and Top Drawer.