

Author Information: How to Avoid Common Conversion Problems \LaTeX -XML

Stefan M. Moser

3 November 2011

Abstract

This document is directed to authors of the *IEEE Transactions on Information Theory* who are preparing the final version of their paper to be sent to IEEE for publication. At IEEE, the submitted source file will be converted to XML and processed to make the paper comply with IEEE style and formatting. This document tries to provide instructions that hopefully will lead to a smoother conversion process without many conversion errors.

Contents

1	Source File	2
1.1	Completeness	2
1.2	Style	2
1.3	Author Names	2
1.4	Abstract and Keywords	3
1.5	Bibliography	3
1.6	Author Biographies	4
2	\LaTeX: What to Do and What to Avoid	4
2.1	\LaTeX Packages	4
2.2	Programming Style	5
2.3	Font Issues	6
2.4	Color	7
2.5	picture-Environment	7
3	Graphics	7
3.1	Format	7
3.2	Naming	7
3.3	Preparation of the Additional Graphic Files	8
4	Common Issues in Galley-Proofs	9
5	Acknowledgments	10

1 Source File

While it is possible to submit the final manuscript as a Microsoft Word document, we discourage this and strongly recommend to provide a \LaTeX source file. The following information refers mainly to \LaTeX source files.

1.1 Completeness

The source file needs to contain *all* information (beside the source file, only additional graphic files are to be submitted; for more about these graphic files, see Section 3). This means that the author needs to upload *one* source file that contains

- the complete paper including title, author names, abstract, keywords, all sections and possibly appendices (see Sections 1.3 and 1.4);
- the bibliography with all references (see Section 1.5);
- all personal \LaTeX definitions (macros) (see also Section 2.1); and
- short biographies of all authors (see Section 1.6).

1.2 Style

To generate a correct layout and style, the \LaTeX source should use the `IEEEtran` class:¹

```
\documentclass[twoside]{IEEEtran}
```

This will yield the correct format and style of IEEE publications, in particular, it generates a **double-column** format. Note that the \LaTeX source should not be submitted in single-column format, because the conversion to double-column format will cause many problems. For example all equations need to be newly typeset with different and new line-breaks to adapt to the shorter column width.

In addition to `twoside`, the option `a4paper` can be specified if the source will be printed on A4 (the layout is not changed, only the margins are adapted).

More information about the `IEEEtran` class and its options and commands can be found at http://mirrors.ctan.org/macros/latex/contrib/IEEEtran/IEEEtran_HOWTO.pdf

1.3 Author Names

If an author is member of IEEE, his or her name should be followed by the corresponding IEEE membership grade. Affiliations and references to financial support can be given as a `\thanks{}` comment. See the following example:

¹The class definition `IEEEtran.cls` is already by default available in a modern \LaTeX distribution. If not, it can be found online at <http://www.ctan.org/tex-archive/macros/latex/contrib/IEEEtran/>

```
\author{Lin~Gu-Rong,~\IEEEmembership{Student Member,~IEEE} and
Stefan~M.~Moser,~\IEEEmembership{Senior Member,~IEEE}\thanks{%
S.~M.~Moser is with the Department of Electrical Engineering
at National Chiao Tung University (NCTU) in Hsinchu, Taiwan.
This work was supported in part by ...}}
```

1.4 Abstract and Keywords

The abstract and the keywords should be provided using the following environments:

```
\begin{abstract}
This paper ...
\end{abstract}

\begin{IEEEkeywords}
Keyword 1, keyword 2, ...
\end{IEEEkeywords}
```

IEEE provides a list of official keywords², but other keywords specified by the authors are also acceptable. Note that the keywords need to be ordered alphabetically.

1.5 Bibliography

The bibliography needs to be specified using the `thebibliography`-environment:

```
\begin{thebibliography}{1}
\bibitem{lapidothmoser03}
A.~Lapidoth and S.~M. Moser, ‘‘Capacity bounds via duality with
applications to multiple-antenna systems on flat fading channels,’’
\emph{IEEE Trans.~Inf.~Theory}, vol.~49, no.~10, pp. 2426--2467,
Oct.~2003.

\bibitem{...}
\end{thebibliography}
```

where the argument `{1}` of `\begin{thebibliography}{1}` specifies the number of references.³

It is strongly recommended, however, not to typeset the bibliography manually, but to rely on the tool BiBTeX that will automatically create the correct IEEE-style formatting. The procedure is as follows. First, the following directives have to be included in the L^AT_EX source:

```
\bibliographystyle{IEEEtran}
\bibliography{referencefile}
```

²To obtain it, just send a blank email to keywords@ieee.org.

³This number is only needed for spacing reasons, i.e., if the number of references is less than 10, the number should be single-digit (e.g., 1); if the number of references is 10 or more, the number should be double-digit (e.g., 10); or if the number of references is 100 or more, the number should be triple-digit. The exact value is irrelevant.

Here, `IEEEtran`⁴ specifies the style in which the references are formatted, and it is assumed that the BiBTeX database is provided by a file

```
referencefile.bib
```

Second, the source has to be processed both by \LaTeX and BiBTeX.⁵ Finally, the BiBTeX directives must be commented out

```
%\bibliographystyle{IEEEtran}  
%\bibliography{referencefile}
```

and the contents of the generated `.bbl`-file (consisting of the `thebibliography`-environment) must be copy-pasted into the source file.

1.6 Author Biographies

The biographies of every author (in the same order as given below the title) should be added at the very end of the source file in the following form:

```
\begin{IEEEbiographynophoto}{Stefan M. Moser}  
  (S'01--M'05--SM'10) received the diploma...  
\end{IEEEbiographynophoto}
```

Note the specification of the time when the author has reached the different IEEE membership grades (student member “S”, member “M”, senior member “SM”, fellow “F”, life fellow “LF”). If an author is not an IEEE member, this time specification can be omitted.

2 \LaTeX : What to Do and What to Avoid

As mentioned, the \LaTeX source file submitted by the authors will be converted to XML. In that process, all \LaTeX directives need to be translated into corresponding XML constructs and the behavior of \LaTeX packages needs to be mimicked. In general this works quite well, however, some special \LaTeX commands and \LaTeX packages can cause problems.

2.1 \LaTeX Packages

While any personal \LaTeX style file must be copy-pasted into the source file, many standard \LaTeX packages need not be included in the source, but can simply be referred to by using the `\usepackage` directive.

Unfortunately, there are a couple of packages that will cause problems during the conversion process of the \LaTeX source into XML. In particular, the following packages *should be avoided*:

⁴This style is defined in a file called `IEEEtran.bst` and is by default available in a modern \LaTeX distribution. If not, the file can be found online at <http://www.ctan.org/tex-archive/macros/latex/contrib/IEEEtran/bibtex>

⁵Usually, the correct compilation order is `latex, bibtex, latex, latex`.

- `hyperref.sty` (since IEEE does not provide hyperlinks, this package is not needed anyway)
- `arydshln.sty`
- the options `vlined` or `ruled` for the package `algorithm2e.sty`
- `acronym.sty`
- `amscd.sty`

There are some more packages related to fonts that should be avoided, see Section 2.3 for more details.

2.2 Programming Style

It is a general platitude for any kind of programming that the better a source code is human readable, the fewer errors it contains. A clean source code will also help the conversion process from L^AT_EX to XML. We therefore list here some basic rules for the programming style:

- Do not use overly long lines, but introduce line-breaks. This will prevent automatic line-breaks at wrong positions during the conversion process.
- Do not redefine fundamental parts of L^AT_EX. For example, it is bad style and can cause serious troubles if one redefines

```
\renewcommand{\l}{\left(}
\renewcommand{\r}{\right)}
```

or similar.

- Do only use letters, numerals and punctuation signs as label names. In particular, do not use brackets within labels. E.g., `\label{eq:low2}` or `\label{item:gg_3}` are OK, but `\label{thm:I{a}}` is not recommended (even though technically it is allowed in L^AT_EX).
- If possible, try to avoid nesting definitions of expressions that can be used both in text- and in math-mode. For example,

```
\newcommand{\G}[1]{\ensuremath{\mathcal{G}(\#1)}}
```

can be used both as `\G{1}` and `$$\G{1}$$` (both resulting in “ $\mathcal{G}(1)$ ”). However, again, the XML conversion process can be confused by such commands. Define instead

```
\newcommand{\G}[1]{\mathcal{G}(\#1)}
```

and always refer to it as `$$\G{1}$$`.

2.3 Font Issues

A special nuisance in the conversion from L^AT_EX to XML are unwanted changes in the used fonts. Particularly, if two different mathematical symbols are based on the same letter, but in different fonts, then a change of font of one of these symbols might cause a clash that fundamentally alters the meaning of mathematical expressions.

As an example consider the case where E (`\mathsf{E}`) denotes expectation, while E (`\eulerrm{E}`) stands for energy. XML does not provide Euler Roman, so E will be changed. If this change turns out to be `\mathsf`, we end up with a second E of completely different meaning and with very weird looking expressions like E[E].

It is therefore crucial to make sure that only symbols and fonts are used that are also available on the XML side.

In short, the following fonts are available in XML:

<code>\textrm:</code>	A and a	(default font in text-mode)
<code>\textsf:</code>	A and a	
<code>\texttt:</code>	A and a	
<code>\textsc:</code>	A AND A	
<code>\textbf:</code>	A and a	
<code>\textit:</code>	<i>A and a</i>	
<code>\textit{\textbf}:</code>	<i>A and a</i>	
<code>\mathit:</code>	A + a	(default font in math-mode)
<code>\mathrm:</code>	A + a	
<code>\mathsf:</code>	A + a	
<code>\mathhtt:</code>	A + a	
<code>\mathbf:</code>	A + a	
<code>\boldsymbol:</code>	A + a	
<code>\mathbb:</code>	ℤ	(no lowercase)
<code>\mathfrak:</code>	℔ + a	
<code>\mathcal:</code>	ℳ	(no lowercase)
<code>\mathscr:</code>	ℒ	(no lowercase)

These commands require the following packages:⁶

```
\usepackage{amsmath,amsfonts,amssymb,mathrsfs}
```

i.e., these packages are OK.

Any other font is not available in XML and will be changed into one of the fonts given above. In particular, the following fonts and packages *cannot* be used:

- Euler Roman and Euler bold Roman are not available.

⁶The package `amsfonts.sty` is actually implicitly loaded by `amsmath.sty`. The commands `\mathbb` and `\mathfrak` require `amssymb.sty`, and `\mathscr` requires `mathrsfs.sty`.

- Euler Script will be changed⁷ into `\mathcal`.
- The font defined by `stmaryrd.sty` cannot be used.
- The symbols defined by `pifont.sty` (command `\ding{..}`) cannot be used.
- No ASCII symbols above 127 can be used (use `amssymb.sty` for special characters instead).

2.4 Color

Color cannot be used apart inside the graphic files.

2.5 picture-Environment

Graphical drawings based on the `picture`-environment or similar cannot be used within equations or within the text, but must be done as a figure (i.e., within the `figure`-environment). It is therefore also necessary to create a separate graphic file for it, see Section 3.3 for more details.

3 Graphics

Graphic files are the only files that need to be submitted separately from the single source file. As a “graphic” we understand every floating figure (`figure`- or `figure*`-environment) or any part of the manuscript that is not generated by \LaTeX , but that is included from an external file.⁸ For each such graphic in the manuscript, a separate file has to be prepared. Note that the caption of a figure does not belong to the graphic (because the caption is generated by \LaTeX) and therefore must not be included in the corresponding graphic file.

For submission, all these graphic files should then be combined into a single `zip` or `tar` archive file.

3.1 Format

The preferred graphic format is **PDF** or **TIF**. Also acceptable⁹ is **EPS**. Any other format is not accepted, in particular, `GIF` or `JPG` cannot be used.

3.2 Naming

The different graphic files should be numbered sequentially according to the order of their appearance in the manuscript. We recommend to use the following naming convention:

⁷This change might be OK as long as `\mathcal` has not been used for other symbols, but in this case it is better to typeset these characters in `\mathcal` directly and to avoid Euler Script.

⁸Note that an externally included graphic should usually be typeset inside of a floating figure to make sure that \LaTeX can shift it around slightly to accommodate it nicely. Also note that there are floating figures that do not include an external graphic, e.g., a figure might contain a `picture`-environment (see Section 2.5).

⁹If the source is a Microsoft Office document, those native formats `DOC`, `DOCX`, `PPT`, `PPTX`, `XLS`, and `XLSX` are also accepted. But again, we strongly recommend the use of \LaTeX .

- take the first 5 letters of the first author’s last name, and
- append the graphic’s corresponding sequential number.

For example, in author “Anderson’s” paper, the first three figures would be named `ander1.pdf`, `ander2.tif`, and `ander3.pdf`.

3.3 Preparation of the Additional Graphic Files

The additional graphic file must look exactly the way the graphic is supposed to appear in the published manuscript. This in particular means that any text-replacement by `\psfrag` must have taken place already.

Therefore, to create each of the additional graphic files, we need a small wrapper file that contains only one of the figures. Consider the following example file called `plot_extraction.tex`:

```
\documentclass[twoside,a4paper]{IEEEtran}

\usepackage[dvips]{graphicx}
\usepackage{psfrag}
\usepackage{amsmath,amssymb,mathrsfs}
\pagestyle{empty}

\begin{document}

\begin{figure}[t]
  \centering
  \psfrag{E/sigma [dB]}[ct][ct]{\small  $\frac{E}{\sigma^2}$  [dB]}
  \psfrag{C [nats]}[cb][cb]{\small  $C$  [nats per channel use]}
  \includegraphics[width=0.45\textwidth]{plot.eps}
\end{figure}

\end{document}
```

Here, the part between `\begin{figure}[t]` and `\end{figure}` was copied from the manuscript’s L^AT_EX source file (but the caption and label command were removed). This short wrapper file needs to be compiled, converted to postscript, and then transformed into EPS format:

```
latex plot_extraction.tex
dvips plot_extraction.dvi
ps2eps -f plot_extraction.ps
```

This then yields an EPS file `plot_extraction.eps` that can now be renamed according to the naming rules of Section 3.2.

The same approach should also be used in the situation when instead of `\includegraphics`, e.g., a `picture`-environment is used inside the `figure`-environment (see Section 2.5).

This way all graphics can be handled, the resulting EPS files compressed into one single zip or tar archive file and uploaded.

A positive side-effect of this approach is that the generated graphics have the exact size¹⁰ as they appear in the published manuscript and will not have to be rescaled.

Note that if an imported PDF or EPS graphic is not changed by L^AT_EX (no `\psfrag` directive or similar), then this PDF or EPS file can be used directly as the additional graphic file and the extraction procedure described above using the wrapper file `plot_extraction.tex` can be omitted.

4 Common Issues in Galley-Proofs

Once the L^AT_EX source has been converted, the authors will receive a galley-proof of their manuscript. It is recommended that this galley-proof is read very carefully, because even if the authors did follow the hints and instructions given above, it is still possible that it contains errors and problems.

Some of the most common issues are as follows:

- Spacing issues concerning mathematical symbols that have a double-superscript or a combination of a hat/tilde or similar and of a superscript. For example,

$$\hat{X}^n$$

is wrongly printed as

$$\hat{X}^n$$

Note the n being too high up.

- Spacing issues concerning mathematical symbols that have both a super- and a subscript. For example,

$$\mathbf{H}_{a+1}^T$$

is wrongly printed as

$$\mathbf{H}_{a+1}^T$$

Note that the superscript shows up *outside* of the subscript and how the transpose- T is too big and too high up. This can be particularly ugly if the expression is inside of brackets, which in the wrongly spaced form become too small:

$$h(\mathbf{H}_{a+1}^T) \quad \text{versus} \quad h(\mathbf{H}_{a+1}^T)$$

- Spacing issues concerning mathematical expressions within the text. For example, the transpose- T in $(\mathbf{A}^T \text{diag}(\hat{\mathbf{H}})\mathbf{A})^T$ overlaps with the text of the line above.¹¹
- Spacing issues between operators and brackets. For example,

$$\log \det (\mathbf{I} + \mathbf{A}^T)$$

¹⁰Make sure that double-column graphics are generated as double-column graphics in the wrapper file, too, i.e., use the starred version `\begin{figure*}`.

¹¹Note that this example is a bit artificial because the transpose- T should be made smaller in any case.

is wrongly printed as

$$\log \det (I + A^T)$$

Note the too large gap between `det` and the opening bracket.

- Spacing issues of equations or arrays of equations that are vertically not properly centered (spaces above and below equation are not the same).
- Discrepancies between sizes of opening and closing brackets if there is a line-break in between. For example,

$$a = \log \left(1 + \frac{b}{2} \right)$$

is wrongly printed as

$$a = \log (1 + \frac{b}{2})$$

So, when proof-reading their manuscript, the authors should particularly look out for such issues and mark them.¹² Note that if the authors react quickly¹³ and provide the feedback to the galley-proof in short time, then they are allowed to ask for several rounds of galley-proofs and corrections, until they are satisfied with the produced manuscript.

5 Acknowledgments

I would like to thank Kevin Lisankie and Helmut Bölcskei for their very valuable input that was the basis of this document.

Any feedback or comments are highly appreciated! Write to

`stefan.moser@ieee.org`

Thank you!

Stefan M. Moser

¹²The comments about the galley-proof can be given in several forms. The authors can write them by hand onto the proof and then scan the commented version; they can use a tool that allows commenting and marking of a PDF file; or they can give a list of issues describing carefully the exact location and the issue.

¹³The time between the notification of the first galley-proof and the publication date is limited and relatively short.