# Generating Correct Mathematical Documents

William F. Hammond

September, 2001

# **Mathematical Articles**

### **Mathematical Articles**

• Contain ordinary text and mathematical content.

#### **Mathematical Articles**

- Contain ordinary text and mathematical content.
- Need realization on paper and on computer networks.

# A Segment of an Article

The gamma function may be defined in a suitable right-half plane by an integral, which essentially amounts to Fourier transform relative to the multiplicative group of postive real numbers of the reciprocal exponential function, and then extended to a meromorphic function in the complex plane with entire reciprocal. The following formula represents a variant of its Weierstrass product expansion:

# A Segment of an Article

The gamma function may be defined in a suitable right-half plane by an integral, which essentially amounts to Fourier transform relative to the multiplicative group of postive real numbers of the reciprocal exponential function, and then extended to a meromorphic function in the complex plane with entire reciprocal. The following formula represents a variant of its Weierstrass product expansion:

$$\int_0^\infty t^x e^{-t} \frac{dt}{t} = \frac{1}{x} \prod_{k=1}^\infty \frac{\left(1 + \frac{1}{k}\right)^x}{\left(1 + \frac{x}{k}\right)} .$$

The product manifests simple poles at zero and each of the negative integers.

From a Single Source:

#### From a Single Source:

1. A typeset version of high quality for paper preprints.

#### From a Single Source:

- 1. A typeset version of high quality for paper preprints.
- 2. An online version for a web preprint.

#### From a Single Source:

- 1. A typeset version of high quality for paper preprints.
- 2. An online version for a web preprint.
- 3. A typeset version of high quality meeting the requirements of a journal chosen after the time of writing.

1. catalogable.

1. catalogable.

2. universally accessible.

1. catalogable.

2. universally accessible.

3. re-scalable and re-sizable.

- 1. catalogable.
- 2. universally accessible.
- 3. re-scalable and re-sizable.
- 4. searchable for mathematical content.

- 1. catalogable.
- 2. universally accessible.
- 3. re-scalable and re-sizable.
- 4. searchable for mathematical content.
- 5. "clippable" for mathematical content.

Donald Knuth's T<sub>E</sub>X

Donald Knuth's TEX

#### Donald Knuth's TEX

LAT<sub>E</sub>X Plain T<sub>E</sub>X

#### Donald Knuth's TEX

```
AT_{EX}
Plain T_{EX}
Other T_{EX} Variants:
```

AMSTEX, ConTeXt, Omega, Texinfo

Early Ideas:

#### Early Ideas:

1. Intuitive Authoring Systems: the WYSIWYG Idea

#### Early Ideas:

- 1. Intuitive Authoring Systems: the WYSIWYG Idea
- 2. Write HTML and Translate From There.

#### Early Ideas:

- 1. Intuitive Authoring Systems: the WYSIWYG Idea
- 2. Write HTML and Translate From There.
- 3. Write LATEX and Translate From There.

Standard with Common "Word Processors"

Standard with Common "Word Processors"

• Available for LATEX.

Standard with Common "Word Processors"

Available for LATEX.

Available for SGML Systems.

• Standard with Common "Word Processors"

Available for LATEX.

Available for SGML Systems.

• Easy for Easy Tasks.

- Standard with Common "Word Processors"
- Available for LATEX.
- Available for SGML Systems.
- Easy for Easy Tasks.
- OK for Very Simple Documents.

• Hard for Hard Tasks.

• Hard for Hard Tasks.

• Slow for Those Who Write Frequently.

- Hard for Hard Tasks.
- Slow for Those Who Write Frequently.
- Insufficiently Rich for Mathematics.

#### WYSIWYG: The Bad News

- Hard for Hard Tasks.
- Slow for Those Who Write Frequently.
- Insufficiently Rich for Mathematics.
- Insufficiently Abstract.

#### WYSIWYG: The Bad News

- Hard for Hard Tasks.
- Slow for Those Who Write Frequently.
- Insufficiently Rich for Mathematics.
- Insufficiently Abstract.
- Inapplicable to Multiple Formattings.

#### WYSIWYG: The Bad News

- Hard for Hard Tasks.
- Slow for Those Who Write Frequently.
- Insufficiently Rich for Mathematics.
- Insufficiently Abstract.
- Inapplicable to Multiple Formattings.
- Some say WYSIAYG: "what you see is all you get".

# WYSIWYG vs. Structured Markup

# WYSIWYG vs. Structured Markup

• Format-specific hanging indentation commands.

## WYSIWYG vs. Structured Markup

- Format-specific hanging indentation commands.
- Use of abstract list structures.

Reliable

Reliable

• But:

Reliable

• But:

1. No Math in HTML

Reliable

• But:

1. No Math in HTML

2. HTML Generally Less Rich Than LATEX

Reliable

• But:

1. No Math in HTML

2. HTML Generally Less Rich Than LATEX

3. Special characters are translation headaches. Examples: # \$ % & ~ \_ ^ \ { } < >

• Almost Impossible

Almost Impossible

Good Structure a Help

Almost Impossible

Good Structure a Help

May Require Human Intervention

Almost Impossible

Good Structure a Help

May Require Human Intervention

Need to Proof Read Twice

- Almost Impossible
- Good Structure a Help
- May Require Human Intervention
- Need to Proof Read Twice
- htlatex in the TEXLive 6 distribution is remarkably good.

eXtensible Markup Language

#### eXtensible Markup Language

• Data Under a **Template for Translation**.

#### eXtensible Markup Language

- Data Under a **Template for Translation**.
- Enforced Separation of Content and Presentation.

#### eXtensible Markup Language

- Data Under a **Template for Translation**.
- Enforced Separation of Content and Presentation.
- Universal Exchange.

#### eXtensible Markup Language

- Data Under a **Template for Translation**.
- Enforced Separation of Content and Presentation.
- Universal Exchange.
- Originated by

World Wide Web Consortium (W3C) Sun Microsystems

Many Templates

- Many Templates
- Synonym for XML Template:

Document Type

- Many Templates
- Synonym for XML Template:
   Document Type

Two worlds

- Many Templates
- Synonym for XML Template:

Document Type

- Two worlds
  - 1. Classical Documents:

Examples: HTML, Docbook, TEI, ...

- Many Templates
- Synonym for XML Template:

Document Type

- Two worlds
  - 1. Classical Documents: Examples: HTML, *Docbook*, TEI, . . .
  - 2. Electronic Data Interchange (EDI)
    Example: Graham William's TEX Catalogue found on CTAN
    help/Catalogue/catalogue.html

• Superseded my earlier ideas:

- Superseded my earlier ideas:
  - \* Strictly controlled dialect of LATEX.

- Superseded my earlier ideas:
  - \* Strictly controlled dialect of LATEX.
  - \* Adapting Texinfo (already suitable for multiple formattings).

- Superseded my earlier ideas:
  - \* Strictly controlled dialect of LATEX.
  - \* Adapting Texinfo (already suitable for multiple formattings).

Relation to the Goals:

- Superseded my earlier ideas:
  - ★ Strictly controlled dialect of LATEX.
  - \* Adapting Texinfo (already suitable for multiple formattings).

- Relation to the Goals:
  - \* No present full realization of online version goals.

#### The GELLMU Project

- Superseded my earlier ideas:
  - ★ Strictly controlled dialect of LATEX.
  - \* Adapting Texinfo (already suitable for multiple formattings).

- Relation to the Goals:
  - ⋆ No present full realization of online version goals.
  - \* Proof that full realization of all goals and more is possible.

Generalized Extensible  $\text{LAT}_{E}X$ -Like MarkUp

Generalized Extensible LATEX-Like MarkUp

• A markup interface for writing (SGML or) XML.

Generalized Extensible LATEX-Like MarkUp

- A markup interface for writing (SGML or) XML.
- LATEX-like notation more succinct than that of XML.

#### Generalized Extensible LATEX-Like MarkUp

- A markup interface for writing (SGML or) XML.
- LATEX-like notation more succinct than that of XML.
- Extensible using GELLMU's \newcommand with arguments. (SGML has no analogue of macros with arguments.)

# Difference Between LATEX Source and GELLMU Source

# Difference Between LATEX Source and GELLMU Source

Article prepared under a template for processing by latex, the Program.

# Difference Between LATEX Source and GELLMU Source

- Article prepared under a template for processing by latex, the Program.
- Article prepared under a template for processing by many programs.

It is a small step from LATEX source to GELLMU source.

Text + Commands

Text + Commands

• A command is a function that operates on text.

- Text + Commands
- A command is a function that operates on text.
- A command may take a non-negative number of arguments.

Example of a command taking no argument:

\latex

**LATEX** 

Example of a command taking no argument:

\latex LAT<sub>E</sub>X

Example of a command taking one argument:

this is \emph{emphasized} text
this is emphasized text

Example of a command taking no argument:

\latex LAT<sub>E</sub>X

Example of a command taking one argument:

this is \emph{emphasized} text
this is emphasized text

Example of a command (for math) taking two arguments:

Example of a command taking no argument:

\latex LATEX

Example of a command taking one argument:

this is \emph{emphasized} text
this is emphasized text

Example of a command (for math) taking two arguments:

$$\[ \frac{z + b}{cz + d} \]$$

$$\frac{az+b}{cz+d}$$

1. Basic

1. Basic

2. Advanced

- 1. Basic
- 2. Advanced
  - (a) Regular

- 1. Basic
- 2. Advanced
  - (a) Regular
  - (b) Other (less fully developed)

1. LATEX-Like Source.

1. LATEX-Like Source.

2. Syntactic Translation to SGML.

- 1. LATEX-Like Source.
- 2. Syntactic Translation to SGML.
- 3. Translation of SGML to Enriched XML.

- 1. LATEX-Like Source.
- 2. Syntactic Translation to SGML.
- 3. Translation of SGML to Enriched XML.
- 4. Various Formattings of Enriched XML.

• Each stage presents opportunities for customizing.

- Each stage presents opportunities for customizing.
- Each output format is the result of successive transforms.

- Each stage presents opportunities for customizing.
- Each output format is the result of successive transforms.
- Additional intermediate transforms can be provided.

- Each stage presents opportunities for customizing.
- Each output format is the result of successive transforms.
- Additional intermediate transforms can be provided.
- These slides were prepared using a special formatting of regular GELLMU's article.

- Each stage presents opportunities for customizing.
- Each output format is the result of successive transforms.
- Additional intermediate transforms can be provided.
- These slides were prepared using a special formatting of regular GELLMU's article.
- The source markup\* for these slides is as readable as ordinary LATEX source.

<sup>\*</sup>URI: correct.glm

## The Syntatic Translator

## The Syntatic Translator

source markup  $\longrightarrow$  XML or SGML

### The Syntatic Translator

• Command names (element names) may contain numbers.

- Command names (element names) may contain numbers.
- Example: \frac23 is a command name.

- Command names (element names) may contain numbers.
- Example: \frac23 is a command name.
- Arguments must be delimited with braces or brackets.

- Command names (element names) may contain numbers.
- Example: \frac23 is a command name.
- Arguments must be delimited with braces or brackets.
- No white space between command name and first argument delimiter.

- Command names (element names) may contain numbers.
- Example: \frac23 is a command name.
- Arguments must be delimited with braces or brackets.
- No white space between command name and first argument delimiter.
- No white space between delimiters of successive arguments.

- Command names (element names) may contain numbers.
- Example: \frac23 is a command name.
- Arguments must be delimited with braces or brackets.
- No white space between command name and first argument delimiter.
- No white space between delimiters of successive arguments.
- Bracketed arguments may not be optional.

Write:

Write:

```
the WWW \a[href="http://www.w3.org/"
]{Consortium} site
```

for generating the XML:

Write:

```
the WWW \a[href="http://www.w3.org/"
]{Consortium} site
```

for generating the XML:

```
the WWW <a href="http://www.w3.org/"
>Consortium</a> site
```

to produce:

#### Write:

```
the WWW \a[href="http://www.w3.org/"
]{Consortium} site
```

for generating the XML:

```
the WWW <a href="http://www.w3.org/"
>Consortium</a> site
```

to produce:

the WWW Consortium\* site

<sup>\*</sup>URI: http://www.w3.org/

**Definitions** 

#### **Definitions**

```
\newcommand{\emph}[1]{\em{#1}}
\newcommand{\w3ref}[2][]{%
\a[href="http://www.w3.org/#1"]{#2}}
```

#### **Definitions**

```
\newcommand{\emph}[1]{\em{#1}}
\newcommand{\w3ref}[2][]{%
\a[href="http://www.w3.org/#1"]{#2}}
```

#### **Invocations**

#### **Definitions**

```
\newcommand{\emph}[1]{\em{#1}}
\newcommand{\w3ref}[2][]{%
\a[href="http://www.w3.org/#1"]{#2}}
```

#### **Invocations**

```
Using GELLMU's \emph{newcommand} one can reduce the markup required for an anchor to \w3ref{W3C}'s \w3ref{Math/}{MathML} site.
```

#### **Definitions**

```
\newcommand{\emph}[1]{\em{#1}}
\newcommand{\w3ref}[2][]{%
\a[href="http://www.w3.org/#1"]{#2}}
```

#### **Invocations**

```
Using GELLMU's \emph{newcommand} one can reduce the markup required for an anchor to \w3ref{W3C}'s \w3ref{Math/}{MathML} site.
```

**Rendering**: Using GELLMU's *newcommand* one can reduce the markup required for an anchor to W3C's MathML site.

• Intended as a first XML document type for LATEX authors

Intended as a first XML document type for LATEX authors

• Sits in the middle between

- Intended as a first XML document type for LATEX authors
- Sits in the middle between
  - 1. What LATEX authors are accustomed to.

- Intended as a first XML document type for LATEX authors
- Sits in the middle between
  - 1. What LATEX authors are accustomed to.
  - 2. What high end XML people think is needed.

- Intended as a first XML document type for LATEX authors
- Sits in the middle between
  - 1. What LATEX authors are accustomed to.
  - 2. What high end XML people think is needed.

Room to adjust and expand.

# The Gamma Function: Its Weierstrass Product

# The Gamma Function: Its Weierstrass Product

$$\int_0^\infty t^x e^{-t} \frac{dt}{t} = \frac{1}{x} \prod_{k=1}^\infty \frac{\left(1 + \frac{1}{k}\right)^x}{\left(1 + \frac{x}{k}\right)}$$

## Markup for the Gamma Identity

## Markup for the Gamma Identity

Regular GELLMU source for the identity:

### Markup for the Gamma Identity

#### Regular GELLMU source for the identity:

# Gamma: Derived XML Markup

#### Gamma: Derived XML Markup

```
<displaymath>
<int>
 <msub>0</msub>
 <msup><infty/></msup>
 t<pow>x</pow> e<pow><minus/>t</pow>
 <frac>
  <numr>d t</numr>
 <denm>t</denm>
 </frac>
</int>
<equals/>
<frac><numr>1</numr><denm>x</denm></frac>
od>
 <msub>k<equals/>1</msub>
 <msup><infty/></msup>
 <frac>
  <numr>
   <bal>1<plus/>
    <frac>
     <numr>1</numr>
     <denm>k</denm>
    </frac>
```

## Gamma: in MathML

### Gamma: in MathML

(not by automatic translation)

#### Gamma: in MathML

(not by automatic translation)

```
<math
xmlns="http://www.w3.org/1998/Math/MathML"
class="display" mode="display">
<mrow>
 <mrow>
 <msubsup>
  <mo>&Integral;</mo>
  <mrow><mn>0</mn></mrow>
  <mi>&infin;</mi>
  </msubsup>
  <mrow>
  <msup>
    <mrow><mi>t</mi></mrow>
   <mrow><mi>x</mi></mrow>
  </msup>
  < mo > </mo >
  <msup>
  <mrow><mi>e</mi></mrow>
  <mrow><mi>-t</mi>
 </msup>
 < mo > </mo >
  <mfrac>
    <mrow><mi>dt</mi></mrow>
```

```
mi>t</mi>
  </mfrac>
</mrow>
</mrow>
< mo> = </mo>
<mrow>
<mfrac>
  <mrow><mn>1</mn></mrow>
  mi>x</mi>
</mfrac>
< mo > </mo >
<msubsup>
  <mo>&Product;</mo>
  <mrow><mi>k</mi><mo>=</mo><mn>1</mn></mrow>
  <mi>&infin;</mi>
</msubsup>
<mrow>
  <mrow>
   <mfrac>
    <mrow>
     <msup>
      <mrow><mfenced>
       <mrow>
        <mn>1</mn><mo>+</mo>
        <mfrac><mn>1</mn><mi>k</mi></mfrac>
       </mrow>
      </mfenced></mrow>
      <mrow><mi>x</mi></mrow>
     </msup>
```

Viewing support for MathML in web pages is not yet widely available. The above item can be rendered by:

Viewing support for MathML in web pages is not yet widely available. The above item can be rendered by:

• W3C's Amaya: wprod.html or wprod.xml.

Viewing support for MathML in web pages is not yet widely available. The above item can be rendered by:

- W3C's Amaya: wprod.html or wprod.xml.
- Mozilla's MathML development track: wprod.xml (only).

Viewing support for MathML in web pages is not yet widely available. The above item can be rendered by:

- W3C's Amaya: wprod.html or wprod.xml.
- Mozilla's MathML development track: wprod.xml (only).
- With special plugin for MSIE: wprod.html (only).

 Ad hoc wprod.html was made from GELLMU source: wprod.glm.

 Ad hoc wprod.html was made from GELLMU source: wprod.glm.

• The short *article* form (slide 28) of GELLMU source above *could* be given automatic translation to MathML.

- Ad hoc wprod.html was made from GELLMU source: wprod.glm.
- The short *article* form (slide 28) of GELLMU source above *could* be given automatic translation to MathML.
- An automatic translation should go through content MathML and from there to presentation MathML.

- Ad hoc wprod.html was made from GELLMU source: wprod.glm.
- The short *article* form (slide 28) of GELLMU source above *could* be given automatic translation to MathML.
- An automatic translation should go through content MathML and from there to presentation MathML.
- An automatic translation would not be under the umbrella of general XML processing.

Reliable translation will require:

Reliable translation will require:

A substantial non-XML, but XML-aware, parsing of all math zones in a GELLMU source document.

Reliable translation will require:

A substantial non-XML, but XML-aware, parsing of all math zones in a GELLMU source document.

Occasional math parsing hints from authors in their markup.

Reliable translation will require:

A substantial non-XML, but XML-aware, parsing of all math zones in a GELLMU source document.

Occasional math parsing hints from authors in their markup.

Desirable, sometimes required:

Reliable translation will require:

A substantial non-XML, but XML-aware, parsing of all math zones in a GELLMU source document.

Occasional math parsing hints from authors in their markup.

Desirable, sometimes required:

1. Source markup labeling of math symbols.

Reliable translation will require:

A substantial non-XML, but XML-aware, parsing of all math zones in a GELLMU source document.

Occasional math parsing hints from authors in their markup.

Desirable, sometimes required:

- 1. Source markup labeling of math symbols.
- 2. Source markup typing of math symbols.

### MathML Generation Issues

### MathML Generation Issues

• Will authors cooperate?

#### MathML Generation Issues

Will authors cooperate?

• Will **standard** web user agents cooperate?

# **Two Final Notes**

### **Two Final Notes**

• For more information: http://www.albany.edu/~hammond/gellmu/

#### Two Final Notes

- For more information: http://www.albany.edu/~hammond/gellmu/
- GELLMU source for these slides is on the web: http://math.albany.edu:8000/~hammond/Presen/Correct/correct.glm