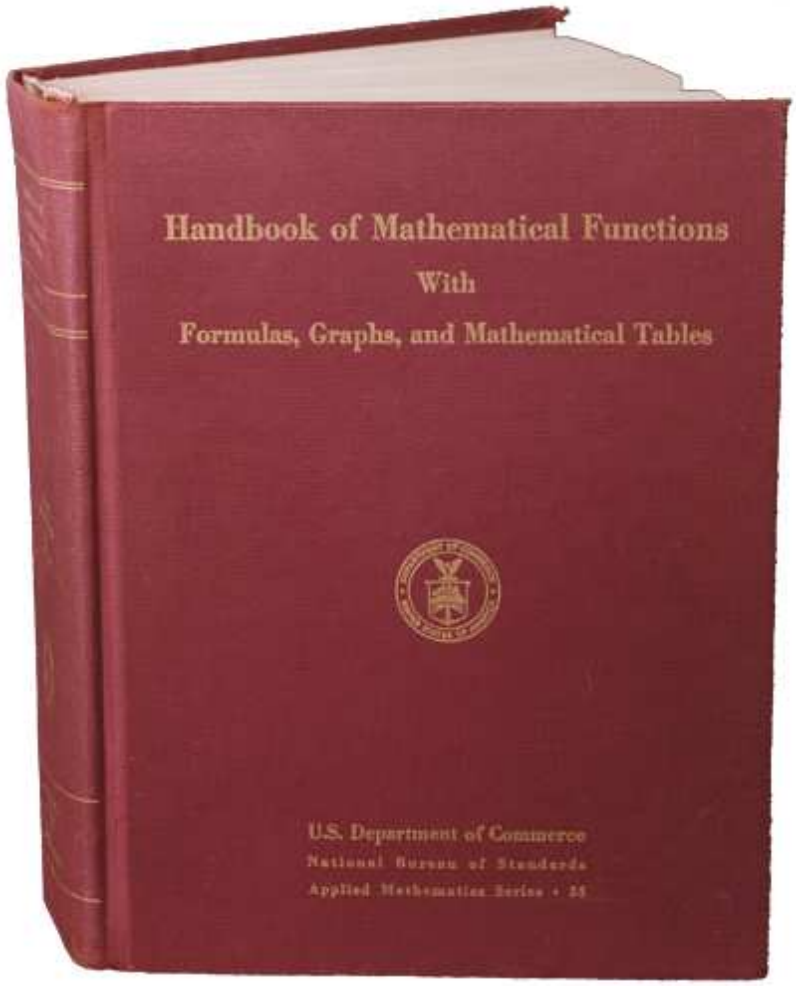
Digital Library of Mathematical Functions: LaTeX, MathML and . . . OpenMath?

Bruce R. Miller
NIST

Digital Library of Mathematical Functions

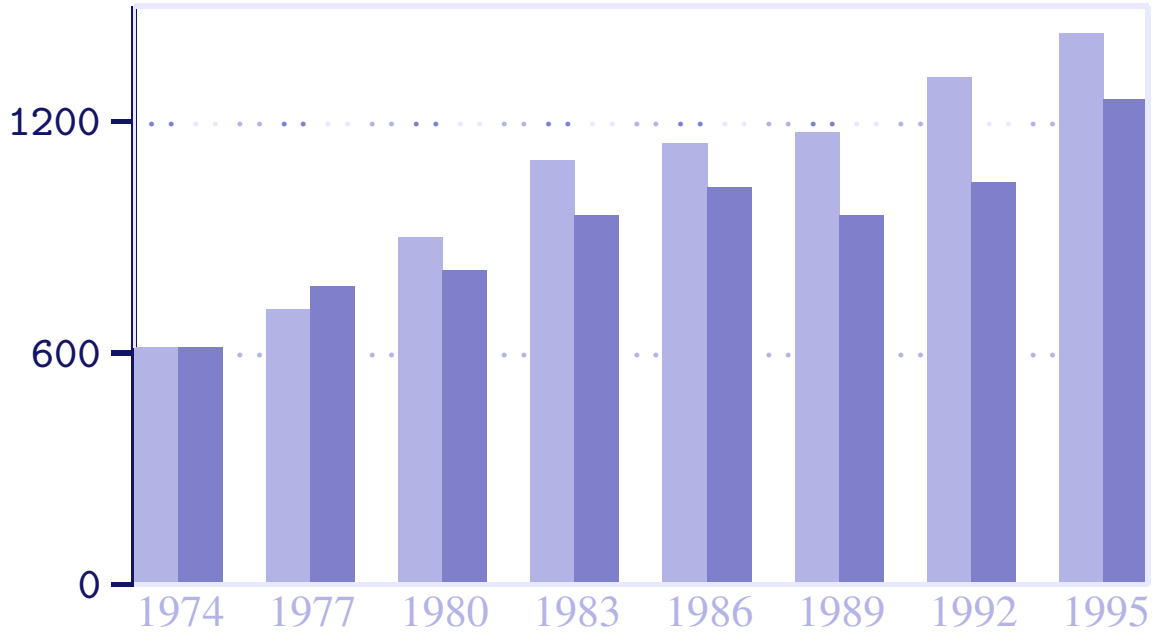
Needing no introduction...



Digital Library of Mathematical Functions

Old, but still relevant

Citations of AMS55 relative to All Scientific.



AMS55 is apparently used more than ever.

Time for a Rewrite

- New functions;
- New properties of old functions;
- New applications.
- . . . and many opportunities.
  - The Internet;
  - Computer Algebra, Theorem Proving systems;
  - The Semantic Web.

DLMF Project

- Started looking at feasibility in 1997.
- NSF funding for authorship in 1999.
- 4 editors, $\approx$ 12 associate editors, $\approx$ 40 authors.
- Goals:
  - New mathematical content updating AMS55,
  - in form of Digital Library,
  - and in print form,
  - by 2005.

Digital Library of Mathematical Functions

Choices: LaTeX, XML, MathML, OpenMath

- LaTeX is obviously good choice for document source.
- . . . and obviously bad.
- Target: XML, MathML, and (eventually) OpenMath.

I don't need to tell *you* why. . .

Overview of talk

- LaTeXML tool.
- Metadata: markup, annotations and connections,
- Data model of the Library
- Math: Parsing, synthesizing meaning.

LaTeXML: Goals

- LaTeX ⇒ XML Transformer
  - General purpose.
  - LaTeX-like DTD (or other?)
  - Math to MathML, OpenMath
- Closely mimic TeX behaviour (& Quirks).
- Lossless.
- Extensible, Adaptable.
- Encourage higher-level markup, declarations.
- . . . and finish DLMF project!

Digital Library of Mathematical Functions

LaTeXML: DLMF Approach

To make more feasible adopt

- Modestly Content-oriented LaTeX.
- Discourage Presentation Markup
  but don't forbid.
- Encourage Content Markup,
  but keep typeable.
- Use document-specific information (internal/external)
  to resolve ambiguities.

Metadata: Making Connections

- Traditional LaTeX: `\ref`, `\cite`, `\index`.
- Leverage our mathematics markup.
- Additional markup:
  - Annotations `\note`.
  - Special metadata: Original handbook reference.
  - Additional declarations.

Digital Library of Mathematical Functions

Metadata: Using Connections

- Postprocessing XML documents.
- Disassemble XML into ‘database’.
- Note all connections.

Not really that hard.

DLMF Data Model

- Simple model (maybe too simple)
  - ID $\Rightarrow$ Object(XML)
    (Chapter, Section, Table, Equation, . . .)
  - linkages embedded within each object
    (insertion, reference, . . .)
- Can (re)construct as necessary
  - Sectional units,
  - Search 'hit-lists'
- Developing an 'Indexing' API by which search, refnum lookup, . . .
  $\Rightarrow$ ID's

Digital Library of Mathematical Functions

LaTeXML Math Processing

TeX source $\xrightarrow{\text{LaTeXML}}$ XML

- Let LaTeXML deal with TeX quirks.
- Acts as structure-preserving Lexer.
  - Possibly augmented (math) Tokens:
    - Name,
    - Unicode, Font, ...
    - PartOfSpeech (ID, Function, Operator, ...)
    - Type (eventually).
  - preserve any given structure (eg. `\frac`, ...)

Math: The Easy Stuff

```
a = b+c
```

LaTeXML produces the tokens

```
<XMTok>a</XMTok>
<XMTok>=</XMTok>
<XMTok>b</XMTok>
<XMTok>+</XMTok>
<XMTok>c</XMTok>
```

LaTeXML Math Processing *continued*

XML $\xrightarrow{\text{LaTeXMLpost}}$ XML'

- Grammar-based parser.
- Undeclared tokens get PartOfSpeech from
  - Document-specific dictionary
    (possibly sectionally scoped)
  - Default dictionary
- Resulting Expression tree
  - inspired by OpenMath.
  - $\approx$ Content MathML;
    (although we haven't done this yet).
  - Easily converted to Presentation MathML.

Digital Library of Mathematical Functions

Math: The Easy Stuff *continued*

```
a = b+c
```

LaTeXMLpost parses this into

```
<XMApp><XMTok>=</XMTok>
  <XMTok>a</XMTok>
  <XMApp><XMTok>+</XMTok>
    <XMTok>b</XMTok>
    <XMTok>c</XMTok>
  </XMApp>
</XMApp>
```

Math: The Easy Stuff *continued*

```
a = b+c
```

Conversion to MathML yields

```
<math xmlns="http://www.w3.org/1998/Math/MathML">
  <mrow>
    <mi>a</mi>
    <mo>=</mo>
    <mrow>
      <mi>b</mi>
      <mo>+</mo>
      <mi>c</mi>
    </mrow>
  </mrow>
</math>
```

LaTeXML Math Processing *future*

XML' $\xrightarrow{\text{LaTeXMLpost}}$ XML"

- Extension of Dictionary to support some Type system.
- Type Analysis to further resolve 'meaning'
- $\Longrightarrow$ OpenMath.
- Any advice?

Math: Higher Level Markup

Reduce ambiguities by introducing higher-level markup:

$$\texttt{\textbackslash deriv[n]\{f\}\{x\}} \Rightarrow \frac{d^n f(x+y)}{dx^n}$$

LaTeX code:

omitted

LaTeXML declaration:

```
DefConstructor('\deriv[]{}{}',
       "<XMApp !#2(POS='BIGOP')>"
       . "<XMTok name='deriv'/>"
              . "?#2(<XMArg>#2</XMArg>)!#2(<XMTok name='Empty'
       . "<XMArg>#3</XMArg>"
              . "?#1(<XMArg>#1</XMArg>)</XMApp>");
```

Digital Library of Mathematical Functions

Math: Higher Level Markup *continued*

LaTeXML constructs the tree:

```
<XMApp><XMTok name='deriv'/>
  <XMArg><XMTok>f</XMTok>
         <XMTok>(</XMTok>
         <XMTok>x</XMTok>
         <XMTok>+</XMTok>
         <XMTok>y</XMTok>
         <XMTok>)</XMTok>
  </XMArg>
  <XMArg><XMTok>x</XMTok></XMArg>
  <XMArg><XMTok>n</XMTok></XMArg>
</XMApp>
```

Parser can treat args individually,

. . . avoiding much guesswork.

Math: Special Functions

With appropriate TEX macrology:

$$\texttt{\textbackslash HyperpFq\{p\}\{q\}} \Rightarrow {}_pF_q$$

Introduce notion of *evaluating a function at*:

$$\texttt{\textbackslash HyperpFq\{p\}\{q\}@\{a\}\{b\}\{z\}} \Rightarrow {}_pF_q\left(a;b;z\right)$$

or (alternative notation)

$$\texttt{\textbackslash HyperpFq\{p\}\{q\}@@\{a\}\{b\}\{z\}} \Rightarrow {}_pF_q\left(\begin{matrix}a\\b\end{matrix};z\right)$$

Palatable notation? Easier to type than

```
\sideset{_{p}}{_{q}}{\mathop{F}}\left({a \atop b};z\right)
```

Math: Special Functions *continued*

With the end result:

```
<XMApp>
  <XMTok name='HyperpFq'>F</XMTok>
  <XMTok>p</XMTok>
  <XMTok>q</XMTok>
  <XMTok>a</XMTok>
  <XMTok>b</XMTok>
  <XMTok>z</XMTok>
</XMApp>
```

and we know *which* 'F' is intended.

Math: Issues

- Role of text and spacing in math.
- Overloading of *symbols* (scoping?)
  - f is a function here, but a variable there.
- Palatable content math markup for LaTeX.
- For *really* meaningful math (eg. OpenMath)
  - need type analysis
  - need more info from authors
- Open ended...

Digital Library of Mathematical Functions

Trends? (Or Wishes)

- Continued development and support for MathML
- Ditto for OpenMath
- Convergience of Markup styles and DocTypes for
  - Various LaTeX⇒ XML converters
  - Richer LaTeX content markup in general (LaTeX3?)
  - Project Authors able use different tools LaTeX, CAS, Thm.Provers, Word Processors.

Digital Library of Mathematical Functions