UnicodeMathML

Murray Sargent III

Unicode Technical Workshop, Mountain View, CA, 2025

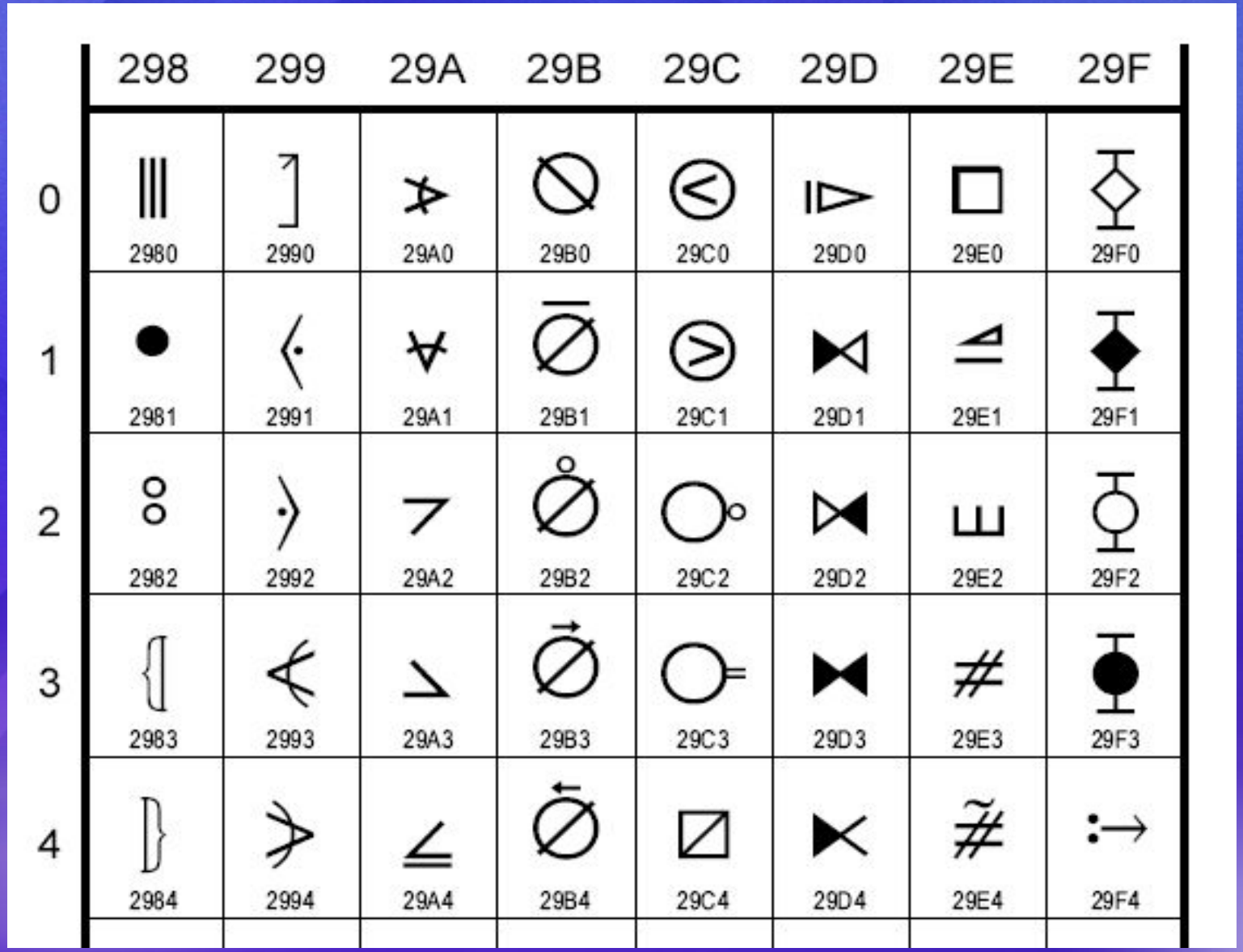
Overview

- What is UnicodeMath?
- Widely used in Microsoft Office apps
- Now available as open source on web
- UnicodeMathML: UnicodeMath, LaTeX, braille, dictation, Markdown → MathML 4
- Also vice versa
- Demo

Unicode

- 426 math chars exist in ASCII, Greek, combining marks, arrows, U+2200 block
- 1163 math alphanumeric characters are in Unicode Plane 1 or Letterlike Symbols
- 701 math symbols in U+2300—U+2BFF
- 23 math variants
- 2310 symbols with Math property
- Many math characters were requested by STIX

Extensive Math Symbols



UnicodeMath

- “A good notation has a subtlety and suggestiveness which at times make it seem almost like a live teacher…and a perfect notation would be a substitute for thought.” Bertrand Russell
- Linear math format resembling math notation
- Enhanced computer-program math that uses Unicode math characters
- a/(b+c) is UnicodeMath for $\frac{a}{b+c}$

[La]TeX

- Widely used, high-quality tech document preparation language
- Simple ASCII keyboard entry
- Usage and math typography are very well documented
- Stable since 1990
- Myriad extensibility packages
- Complex scenarios are hard to edit
- Numerous dialects, user macros, complicate interchange

UnicodeMath Input Methods

- As with Asian languages, UnicodeMath needs input methods (IMEs)
- LaTeX control words with autocomplete, e.g., \alpha → α
- Operator transformations, e.g., +- → ±, - → −, /= → ≠
- Variable italicization, e.g., a → a
- Simple build up, e.g., x^2 → x^2, 1/2 → ½
- Symbol galleries
- Alt+x hot key

Input vs UnicodeMath

- **Enter ASCII**

  1/2\pi \int_0^2\pi \dd\theta/(a+b sin \theta)=1/√(a^2-b^2)

Input window shows UnicodeMath

1/2π ∫_0^2π ⅆθ/(a+b sinθ)=1/√(a²−b²)

Output window shows built-up math

$$\frac{1}{2\pi}\int_0^{2\pi}\frac{d\theta}{a+b\sin\theta}=\frac{1}{\sqrt{a^2-b^2}}$$

LaTeX/MathML Input Methods

- Symbol control words → Unicode, e.g., \theta → θ with autocomplete
- Operator transformations
- For LaTeX, include amsmath and unicode-math packages
- Symbol galleries
- LaTeX and MathML become more readable with Unicode symbols
- Alt+x hot key

Input vs Unicode LaTeX

- **Enter ASCII**

```
\frac{1}{2\pi}\int_0^{2\pi}
\frac{\dd\theta}{a+b\sin\theta}
=\frac{1}{\sqrt{a^2-b^2}}
```

Input window shows Unicode LaTeX

```
\frac{1}{2𝜋}∫_0^{2𝜋}\frac{ⅆ𝜃}{a+b\sin𝜃}
=\frac{1}{√{a^2-b^2}}
```

Output window shows built-up math

$$\frac{1}{2\pi}\int_0^{2\pi}\frac{d\theta}{a+b\sin\theta}=\frac{1}{\sqrt{a^2-b^2}}$$

What is OfficeMath?

- Convenient WYSIWYG editing using LaTeX or UnicodeMath in Microsoft Office apps
- TeX-quality math typography
- Automatic line breaking, math kerning
- More math semantics than TeX and greater interoperability (Presentation MathML)
- Input with ribbon, context menus, keyboard, ink
- International and accessible

Vagaries of Math Notation

- Choice of subscript/superscript base
- Function arguments like $\sin\frac{1}{2}(\omega-\omega_0)t$
- Integrands and *n*-aryands
- Absolute value ambiguities like $\big||a|-|b|\big|$. Actually this example is unambiguous, but $|a|b\ -\ c|d|$ has two possible meanings
- Context sensitive ellipses: … vs ⋯

Roles of Space

- Spaces are rarely needed inside math formulae, since spacing is automatic
- Spaces terminate autocorrect entries and expressions.
- A space builds up one subexpression; other operators build up as many as they can

Autocorrect Examples

- Type \delta and get δ, \Delta and get Δ
- Define \quadratic to be

  x = (-b ± √(b^2 - 4ac))/2a
- Then typing \quadratic<space> inserts

$$x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$$

Mode Locking Integral

- Define \integral to be

  1/2π ∫_0^2π dθ/(a+b sin θ)=1/√(a^2-b^2)

Type \integral<space> to insert

$$\frac{1}{2\pi}\int_0^{2\pi}\frac{d\theta}{a+b\sin\theta}=\frac{1}{\sqrt{a^2-b^2}}$$

Gives oscillation period of $\dot{\theta}=a+b\sin\theta$ for $|a|>|b|$. Describes mode locking in ring lasers, Zeeman lasers, 3-mode SW lasers, …

Resonance Fluorescence and Probe Absorption

Resonance fluorescence spectrum $= A_1 + \text{c. c.}$, where

$$A_1 = \frac{g^2 \mathcal{D}_1}{1 + I_2 \mathcal{L}_2} \left[\frac{I_2 \mathcal{L}_2}{2} - \frac{I_2 \mathcal{F} \gamma}{4} \frac{I_2 \mathcal{L}_2 \mathcal{D}_1 - \mathcal{D}_2^* (1 + \Gamma / i\Delta)}{1 + I_2 \mathcal{F} \frac{\gamma}{2} (\mathcal{D}_1 + \mathcal{D}_3^*)} \right]$$

I_2 = pump intensity, $\Delta = \nu_2 - \nu_1$, $\mathcal{L}_2$ = strong-field Lorentzian, $\mathcal{D}_n$ = complex Lorentzian for ν_n, $\mathcal{F}(\Delta)$ = complex Lorentzian for population pulsations

Probe absorption $= A_1 - B_1$, where B_1 looks similar to A_1

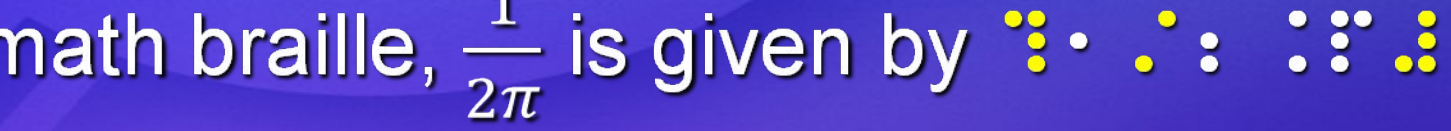
Formula Autobuildup

- User types formulas in UnicodeMath in a output window
- Expressions are built up as soon as they’re syntactically unambiguous
- Expressions can be edited in built-up form or in linear form

UnicodeMathML Speech

- Fluent math speech in English but localizable
- Two speech granularities: coarse & fine
- Coarse-grained speech speaks whole expressions and equations
- Fine-grained speech speaks one character or location at a time for editing math, e.g., “end of numerator”
- Also dictate math

Math Braille

- Nemeth math braille is the first math linear format (1952, predates Fortran)
- Math is confined to math zones, so Nemeth math braille is global
- In math braille, $\frac{1}{2\pi}$ is given by 
- Sighted users see built-up math corresponding to braille cells on refreshable braille display
- Need math to graduate from high school

MathML 4.0

- MathML 4.0 core runs on all major browsers and is browser-friendly
- Adds intent attribute to disambiguate math notation and allow precise speech
- Does $|x|$ mean absolute value of x or cardinality of x?
- Intents can emulate OMML *N*-ary expressions and math functions.

MathML for |x|

<mrow intent="absolute-value(x)">
<mo>|</mo><mi>x</mi><mo>|</mo></mrow>

or

<mrow intent="cardinality(x)">
<mo>|</mo><mi>x</mi><mo>|</mo></mrow>

Enter cardinality with \card. Else absolute value is assumed (except for determinants).

- $$\int_0^{2\pi} \frac{d\theta}{a + b\sin\theta}$$

```
< math display="block">
  <mrow intent=":nary(0,$h,$naryand)">
    <msubsup>
      <mo>∫</mo>
      <mn>0</mn>
      <mrow arg="h">
        <mn>2</mn>
        <mi>π</mi></mrow></msubsup>
    <mfrac arg="naryand">
      <mrow>
        <mi intent="ⅆ">d</mi>
        <mi>θ</mi></mrow>
      <mrow>
        <mi>a</mi>
```

```
<mo>+</mo>
<mi>b</mi>
<mrow intent=":function">
      <mi>sin</mi>
      <mo></mo>
      <mi>θ</mi>
         </mrow></mrow></mfrac></mrow>
</math>
```

Cross Platform

- Runs on the popular browsers
- User interfaces (UI) on iOS and Android are limited (no hot keys)
- Node.js npm package: unicodemathml

Credits

My UnicodeMathML started with Noah Doersing's UnicodeMathML. I learned a lot studying Noah's JavaScript and UnicodeMath grammar. Brian Kardell has given me many helpful suggestions.

Demo

Use UnicodeMathML to enter/dictate formulas like

$$\frac{1}{2\pi}\int_0^{2\pi}\frac{d\theta}{a+b\sin\theta}=\frac{1}{\sqrt{a^2-b^2}}$$

$$\mathcal{V}_{ab}(z,t)=-\frac{1}{2}\wp_{ab}\sum_{n=0}^{N}\mathcal{E}_n(t)e^{-ivt}\sin(K_n z)$$

Click the Demo button in UnicodeMathML. For more info, click the Help button.