

All characters
we need

All characters we need

Bogusław Jackowski, Piotr Pianowski
Piotr Strzelczyk

What is the problem?

What is the problem?

We want to typeset math.

What is the problem?

We want to typeset math.
Math font consists (for now)
of characters.

What is the problem?

We want to typeset math.
Math font consists (for now)
of characters.

Which characters we need
(and how many)?

The state of the art

D. E. Knuth needed less than
500 characters

The state of the art

D. E. Knuth needed less than
500 characters

Unicode has more than
100 thousand characters

The state of the art

D. E. Knuth needed less than
500 characters

Unicode has more than
100 thousand characters

DejaVu Sans has about **6 600**

Cambria Math has about **6 500**

STIX Math has about **4 200**

The state of the art

D. E. Knuth needed less than
500 characters

Unicode has more than
100 thousand characters

DejaVu Sans has about **6 600**

Cambria Math has about **6 500**

STIX Math has about **4 200**

e-foundry math fonts: **4 300**

(including script variants
and size-chained glyphs)

Do we need more characters?

Pro

—

Contra

Do we need more characters?

Pro

- better coverage of Unicode
- symbols for more branches of science
- more consistent charset
- and more size-chained glyphs

Contra

- heavier font file
- more complex to maintain
- more never used glyphs
(seen only twice: during designing and testing)

What kind of math font do we need?

Now there are 13 math fonts
maybe enough?

What kind of math font do we need?

Now there are 13 math fonts
but there is still space for:

What kind of math font do we need?

Now there are 13 math fonts
but there is still space for:

bold math font

sans-serif math font

math font for text editor

What kind of math font do we need?

Now there are 13 math fonts
but there is still space for:

bold math font

sans-serif math font

math font for text editor

(XITS math bold doesn't
meet real-life criteria)

**Does every font need
a new character set?**
all our math fonts have
the same set
(roughly speaking)

**Does every font need
a new character set?**

but new fonts would need
new sets because of:

Does every font need a new character set?

but new fonts would need
new sets because of:

bold alphabets in bold font
sans-serif alphabets in sans font
growing and extensible
characters in text editors font

**No one can scan thousands
characters one by one**

**No one can scan thousands
characters one by one,
so characters have to
be grouped**

Unicode has blocks
we can use them

Unicode has blocks

0000-007F	C0 Controls and Basic Latin
0080-00FF	C1 Controls and Latin-1 Supplement
0100-017F	Latin Extended-A
0180-024F	Latin Extended-B
0250-02AF	IPA Extensions
02B0-02FF	Spacing Modifier Letters
0300-036F	Combining Diacritical Marks
0370-03FF	Greek and Coptic
0400-04FF	Cyrillic
0500-052F	Cyrillic Supplement
0530-058F	Armenian
0590-05FF	Hebrew
0600-06FF	Arabic
0700-074F	Syriac
0750-077F	Arabic Supplement
...	

Unicode has blocks

0000-007F	C0 Controls and Basic Latin
0080-00FF	C1 Controls and Latin-1 Supplement
0100-017F	Latin Extended-A
0180-024F	Latin Extended-B
0250-02AF	IPA Extensions
02B0-02FF	Spacing Modifier Letters
0300-036F	Combining Diacritical Marks
0370-03FF	Greek and Coptic
0400-04FF	Cyrillic
0500-052F	Cyrillic Supplement
0530-058F	Armenian
0590-05FF	Hebrew
0600-06FF	Arabic
0700-074F	Syriac
0750-077F	Arabic Supplement

...

too many groups

Unicode has blocks
we can group them

Unicode has blocks

- ! Basic Latin + Sup. (0000-00FF)
- + Latin Extended-A (0100-017F)
- + Latin Extended-B (0180-024F)
- +– IPA Extensions (0250-02AF)
- + Spacing Modifier Letters (02B0-02FF)
- +– Combining Diacritical Marks (0300-036F)
- +– Greek and Coptic (0370-03FF)
- ? Cyrillic (0400-04FF)
- Cyrillic Sup., Armenian, Hebrew, Arabic, ...
Sinhala, Thai, Lao, Tibetan (0500-0FFF)
- Myanmar, Georgian, Hangul Jamo, Ethiopic, ...
Lepcha, Ol Chiki, Vedic Ext. (1000-1CFF)
- ? Phonetic Extensions + Sup. (1D00--1DBF)
- Combining Diacritical Marks Sup. (1DC0-1DFF)
- Latin Ext. Additional (1E00-1EFF)
- Greek Ext. (1F00-1FFF)
- ...

Unicode has blocks

- ! Basic Latin + Sup. (0000-00FF)
- + Latin Extended-A (0100-017F)
- + Latin Extended-B (0180-024F)
- +– IPA Extensions (0250-02AF)
- + Spacing Modifier Letters (02B0-02FF)
- +– Combining Diacritical Marks (0300-036F)
- +– Greek and Coptic (0370-03FF)
- ? Cyrillic (0400-04FF)
- Cyrillic Sup., Armenian, Hebrew, Arabic, ...
Sinhala, Thai, Lao, Tibetan (0500-0FFF)
- Myanmar, Georgian, Hangul Jamo, Ethiopic, ...
Lepcha, Ol Chiki, Vedic Ext. (1000-1CFF)
- ? Phonetic Extensions + Sup. (1D00--1DBF)
- Combining Diacritical Marks Sup. (1DC0-1DFF)
- Latin Ext. Additional (1E00-1EFF)
- Greek Ext. (1F00-1FFF)
- ...

still not usable

Unicode has blocks

Unicode has blocks

some of them are too large

Mathematical Alphanumeric Symbols (1D40–1D7FF)

Unicode has blocks

some of them are too large

Mathematical Alphanumeric Symbols (1D40–1D7FF)

has many alphabets:

bold, italic, bold italic, script, bold script,
fraktur, bold fraktur, double-struck,
sans-serif (four alphabets: regular,
bold, italic, bold italic), typewriter;
some of them with greek and/or figures

our fonts are split into files

our fonts are split into files

`dvu-lat.mp` - Latin
`dvu-gre.mp` - Greek
`dvu-heb.mp` - Hebrew
`dvu-ss.mp` - sans-serif
`dvu-ds.mp` - double-struck
`dvu-cal.mp` - script
`dvu-fra.mp` - fraktur
`dvu-tt.mp` - typewriter
`dvu--arr.mp` - arrows
`dvu--etc.mp` - accents
`dvu--bra.mp` - braces
`dvu--ops.mp` - mathematical operators
`dvu--rel.mp` - mathematical relations
`dvu--til.mp` - tildes
`dvu--pri.mp` - primes
`dvu--xgl.mp` - geometric shapes
`dvu--ygl.mp` - integrals, sums and radicals
`dvu--spc.mp` - spaces

our fonts are split into files

`dvu-lat.mp` - Latin
`dvu-gre.mp` - Greek
`dvu-heb.mp` - Hebrew
`dvu-ss.mp` - sans-serif
`dvu-ds.mp` - double-struck
`dvu-cal.mp` - script
`dvu-fra.mp` - fraktur
`dvu-tt.mp` - typewriter
`dvu--arr.mp` - arrows
`dvu--etc.mp` - accents
`dvu--bra.mp` - braces
`dvu--ops.mp` - mathematical operators
`dvu--rel.mp` - mathematical relations
`dvu--til.mp` - tildes
`dvu--pri.mp` - primes
`dvu--xgl.mp` - geometric shapes
`dvu--ygl.mp` - integrals, sums and radicals
`dvu--spc.mp` - spaces

also not usable

our fonts are split into files
we can group and split them

our fonts are split into files

*-lat.mp - Latin (4-members family)
= other basic chars + accented letters
+ roman + italic + bold + bold italic

*-gre.mp - Greek
= roman + italic + bold + bold italic

*-heb.mp - Hebrew (only few letters)

*-ss.mp - Sans-serif (4-members family)
= roman + italic + bold + bold italic

*-ds.mp - Double-struck
:

*-etc.mp, *-til.mp, *-pri.mp - accents, tildes, primes
+ wide forms + extensible

*--bra.mp - braces
+ high forms + extensible

*--ops.mp, *--rel.mp - mathematical operators and relations
+ big operators

*--xgl.mp - geometric shapes

*--ygl.mp - integrals, sums and radicals
+ big forms + extensible

*--spc.mp - spaces

our fonts are split into files

*-lat.mp - Latin (4-members family)
= other basic chars + accented letters
+ roman + italic + bold + bold italic

*-gre.mp - Greek
= roman + italic + bold + bold italic

*-heb.mp - Hebrew (only few letters)

*-ss.mp - Sans-serif (4-members family)
= roman + italic + bold + bold italic

*-ds.mp - Double-struck
:

*-etc.mp, *-til.mp, *-pri.mp - accents, tildes, primes
+ wide forms + extensible

*--bra.mp - braces
+ high forms + extensible

*--ops.mp, *--rel.mp - mathematical operators and relations
+ big operators

*--xgl.mp - geometric shapes

*--ygl.mp - integrals, sums and radicals
+ big forms + extensible

*--spc.mp - spaces

better, but still not good enough

New approach

New approach

5 main groups

New approach
5 main groups
BASE TEXT
MATH ALPHABETS
SYMBOLS
CHAINED SYMBOLS
OTHER CHARACTERS

New approach

BASE TEXT

New approach

BASE TEXT

- basic punctuation and symbols
(without math symbols: plus, equal etc.)
- latin letters and figures
- accented letters and basic text accents

New approach **MATH ALPHABETS**

New approach

MATH ALPHABETS

- latin letters and figures
- latin bold letters and figures
- latin italic letters
- latin bold italic letters
- greek roman letters
- greek bold letters
- greek italic letters
- greek bold italic letters
- sans-serif latin letters and figures
- sans-serif latin bold letters and figures
- sans-serif greek bold letters
- sans-serif latin italic letters
- sans-serif latin bold italic letters
- sans-serif greek bold italic letters

all greek alphabets with `partialdiff` and `nabla`
all latin alphabets with `dotlessi` and `dotlessj`

New approach

MATH ALPHABETS

- double-struck latin letters and figures
- script latin letters
- script latin bold letters
- fraktur latin letters
- fraktur latin bold letters
- typewriter latin letters and figures
- hebrew symbols (aleph, beth, gimel, daleth)

all latin alphabets with dotlessi and dotlessj

New approach

SYMBOLS

New approach

SYMBOLS

- mathematical operators and relations
- rare operators
- arrows
- rare arrows
- math accents (including arrow accents, tilde accents and prime symbols)
- braces
- horizontal braces

New approach

CHAINED SYMBOLS

New approach

CHAINED SYMBOLS

- big mathematical operators
- extensible mathematical operators (integrals, sums, radicals)
- arrows long and extensible
- math accents wide and extensible
- braces high and extensible
- horizontal braces extensible

New approach

OTHER CHARACTERS

New approach

OTHER CHARACTERS

- geometric shapes
- box drawing and block elements
- braille patterns
- combining enclosing characters
- alphabetic symbols (ie. angstrom, centigrade; ohm inverted; double-struck Pi and summation)
- extra symbols (card sutis, musical notes, euler, weierstrass symbols etc.)
- spaces (invisible characters – joiners, separators etc.)
- superscripts and subscripts
- number forms (fractional figures)

New approach

OTHER CHARACTERS

- geometric shapes
- box drawing and block elements
- braille patterns
- combining enclosing characters
- alphabetic symbols (ie. angstrom, centigrade; ohm inverted; double-struck Pi and summation)
- extra symbols (card sutis, musical notes, euler, weierstrass symbols etc.)
- spaces (invisible characters – joiners, separators etc.)
- superscripts and subscripts
- number forms (fractional figures)

this clasification suffices
to define usable charsets

math font
sans-serif math
heavy math
monospaced font
text font extended

BASE TEXT

- +++++ basic punctuation and symbols
(without math symbols: plus, equal etc.)
- +++++ latin letters and figures
- +++++ accented letters and basic text accents

math font
sans-serif math
heavy math
monospaced font
text font extended

BASE TEXT

- +++++ basic punctuation and symbols
(without math symbols: plus, equal etc.)
- +++++ latin letters and figures
- +++++ accented letters and basic text accents

no doubts

math font
sans-serif math
heavy math
monospaced font
text font extended

MATH ALPHABETS

- +++++ latin letters and figures
- ++**1**+ - latin bold letters and figures
- ++++- latin italic letters
- ++**1**+ - latin bold italic letters
- +++++ greek roman letters
- ++**1**+ - greek bold letters
- +++++? greek italic letters
- ++**1**+ - greek bold italic letters
- +**2**++- sans-serif latin letters and figures
- +**21**+ - sans-serif latin bold letters and figures
- +**21**+ - sans-serif greek bold letters
- +**2**++- sans-serif latin italic letters
- +**21**+ - sans-serif latin bold italic letters
- +**21**+ - sans-serif greek bold italic letters

- 1** bold in a bold font should be actually an extra bold (it has to be distinguishable in formulas)
- 2** is sans-serif needed in sans-serif font
(if so, it should be distinguishable in formulas:
light version, rounded ends, square font, ...)

math font
sans-serif math
heavy math
monospaced font
text font extended

MATH ALPHABETS

- ++++? double-struck latin letters and figures
- ++++? script latin letters
- ++**1**+ - script latin bold letters
- ++++? fraktur latin letters
- ++**1**+ - fraktur latin bold letters
- +++**3**- typewriter latin letters and figures
- +++++ hebrew symbols (aleph, beth, gimel, daleth)

- 1** bold in a bold font should be actually an extra bold (it has to be distinguishable in formulas)
- 3** typewriter symbols are so rare that they may be omitted in a monospaced font?

math font
sans-serif math
heavy math
monospaced font
text font extended

SYMBOLS

- +++++ mathematical operators and relations
- rare operators
- +++++ arrows
- rare arrows
- +++++ math accents (including arrow accents, tilde accents and prime symbols)
- +++++ braces
- +++++ horizontal braces

math font
sans-serif math
heavy math
monospaced font
text font extended

CHAINED SYMBOLS

- +++a** - big mathematical operators
- +++ -** - extensible mathematical operators
(integrals, sums, radicals)
- +++a** - arrows long and extensible
- +++ -** - math accents wide and extensible
- +++ -** - braces high and extensible
- +++ -** - horizontal braces extensible

a in monospaced font doesn't have enough space for big operators, but some of them have Unicode slots, so should they be included?

math font
sans-serif math
heavy math
monospaced font
text font extended

OTHER CHARACTERS

- +++++ geometric shapes
 - +++**b**- box drawing and block elements
 - - - + - braille patterns
 - +++ - - combining enclosing characters
 - +++++ alphabetic symbols (ie. angstrom, centigrade;
ohm inverted; double-struck Pi and summation)
 - +++++ extra symbols (card sutis, musical notes,
euler, weierstrass symbols etc.)
 - +++**c**+ spaces (invisible characters
- joiners, separators etc.)
 - - - **dd** superscripts and subscripts
 - - - **dd** number forms (fractional figures)
- b** box drawing is a common element of mono fonts,
but is it needed?
- c** should zero width element or invisible be added?
should they be visualised?
- d** are they needed? Unicode doesn't contain
the full superscript alphabet, but ² is useful.

Easy and hard decisions

Some decisions were obvious,
but some have to be
thought over.

Easy and hard decisions

Some decisions were obvious,
but some have to be
thought over.

And there will be, as always,
many exceptions.

The end

Control Pictures: NUL SOH STX ETX EOT ENQ ACK BEL BS HT LF VT FF CR SO SI DLE DC1 DC2 DC3 DC4 NAK SYN ETB CAN EM SUB ESC FS GS RS US SP DEL b NL 3 4 5 ?

Enclosed Alphanumerics: 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 A B C D E F G H I J K L M N O P Q R S T U V W X Y Z a b c d e f g h i j k l m n o p q r s t u v w x y z 0 1 2 3 4 5 6 7 8 9

Box Drawing, Block Elements: various geometric shapes and lines

Geometric Shapes: various geometric shapes like circles, squares, triangles

Miscellaneous Symbols, Dingbats: various symbols like stars, hearts, musical notes

Miscellaneous Mathematical Symbols-A: various mathematical symbols like less than, greater than

Supplemental Arrows-A, B: various arrow symbols

Miscellaneous Mathematical Symbols-B: various mathematical symbols like infinity, pi

Supplemental Mathematical Operators: various mathematical operators like plus, minus, multiplication

Miscellaneous Symbols and Arrows: various symbols and arrows

Latin Extended-C: various Latin characters

Supplemental Punctuation: various punctuation marks

Cyrillic Extended-B: various Cyrillic characters

Modifier Tone Letters + Latin Extended-D: various letters with tones

Alphabetic Presentation Forms: various alphabetic forms

Mathematical Alphanumeric Symbols: A-Z a-z 0-9 in various styles and colors

The colors denote characters in fonts as follows: black - present in all fonts; green - missing from some fonts; cyan - present only in e-foundry fonts; red - present in all fonts except ours; magenta - present in many fonts except ours; yellow - present only in one of the other fonts.

Kolor znaku opisuje jego występowanie w fontach, i tak: czarny - obecny we wszystkich; zielony - nieobecny w niektórych; morski - występuje tylko w naszych fontach; czerwony - występuje we wszystkich poza naszymi; purpurowy - występuje w kilku poza naszymi; złoty - obecny tylko w jednym z innych fontów.