ISO/IEC JTC1/SC2/WG2 N3198R
2007-01-26
L2/07-011R

Universal Multiple Octet Coded Character Set
International Organization for Standardization
Organisation internationale de normalisation
Международная организация по стандартизации

Doc Type: **Working Group Document**
Title: **29 Additional Mathematical and Symbol Characters**
Source: **Asmus Freytag** (Unicode), **Barbara Beeton** (American Mathematical Society), **Patrick Ion**, **Murray Sargent** and **David Carlisle** (MathML Working Group of the W3C), **Roozbeh Pournader** (Sharif FarsiWeb)
Status: **Expert contribution**
Action: **For consideration by JTC1/SC2/WG2 and UTC**
Related:

Summary

As the mathematical community completes its migration to Unicode, the need for additional mathematical symbols has been discovered, to complete the extension of the mathematical repertoire carried out over the last few revisions of the Standard. This document contains a request for 29 new symbols required for the publication of mathematical and technical material, to support mathematically oriented publications, or to address the needs for mathematical markup languages, such as MathML.

The STIX project (http://www.ams.org/STIX) of the STIPub consortium of technical and scientific publishers, which includes the AMS, is engaged in creating a font encompassing the characters needed in the publication of technical and mathematical journal articles and similar works. A number of the requested characters come from this project of reviewing the mathematical and technical literature, which implies that actual use can be attested and that the reviewers feel that the user community as represented by academic publishers has an interest in being able to encode the character in question.

Other requests for additional characters come from the ongoing work of the MathML Working Group of the W3C as well as from public review comments.

This document contains several numbered sections containing specific proposals for additional characters with background information, followed by three appendices containing related material for information, followed by the proposal summary form.

1 Non-combining Diacritic

A mathematical variable, or even an entire expression, can be decorated with what is called a math accent. By convention such a decoration is indicated with a Unicode character, even in markup languages, such as MathML. MathML 2.0 tags the accent with <mo> (operator) tag. As a result of this, the markup syntax character “>” is occasionally followed by a non-spacing character. Where possible, MathML avoids this by recommending the use of spacing clones of diacritics where they are available. Where no spacing clone exists, the layout of the source text may be affected by the combining accent, but the interpretation of the source and the display of the actual content of the document remain unambiguous—with one exception. The exception is the diagonal stroke overlay (U+0338) often used for negation. The problem here is that the combination of > followed by U+0338 is canonically equivalent to U+226F ≯ NOT GREATER-THAN. This means that text containing > followed by U+0338 cannot be distinguished from NOT GREATER-THAN. Therefore, the MathML working group has expressed their concern:

> It's of considerable importance that we have a uniform treatment of math accenting using non-combining characters in MathML. The interaction with Unicode Normalization, as recognized by the XML recommendations, and thus in XML parsers, is what drives the request.

In support of this request, we propose the addition of a spacing clone for 0338:

27CB / MATHEMATICAL SPACING LONG SOLIDUS OVERLAY

The recommended properties are the same as for other spacing clones of combining accents, except that this character will not have a decomposition. A general category of Sm (Symbol, math) is not recommended.

2 Invisible Plus Operator

In analogy to the other invisible operators coded in the range 2061 to 2063, it is proposed to encode an INVISIBLE PLUS OPERATOR character to be able to unambiguously represent expressions like $3\frac{1}{2}$, which occur frequently in school or engineering texts. For example, in presentation MathML notation, the markup source for this is:

```
<mrow>
   <mn>3</mn>
   <mo>&InvisiblePlus;</mo>
   <mfrac><mn>1</mn><mn>2</mn></mfrac>
</mrow>
```

Not having an operator at all would imply multiplication as in the example

$$3\ \frac{abc}{d}$$

where the 3 represents a factor multiplying the following fraction. As is the case for the existing invisible operators, this character would primarily be required for unambiguous representation of the mathematical intent, for example for machine parsing. In a publication, the operator would not be visible to the human reader, who would, as usual, rely on larger context to determine the intended meaning of the juxtaposition.

The MathML Working Group requests the addition of an invisible math operator:

2064 [+] INVISIBLE PLUS

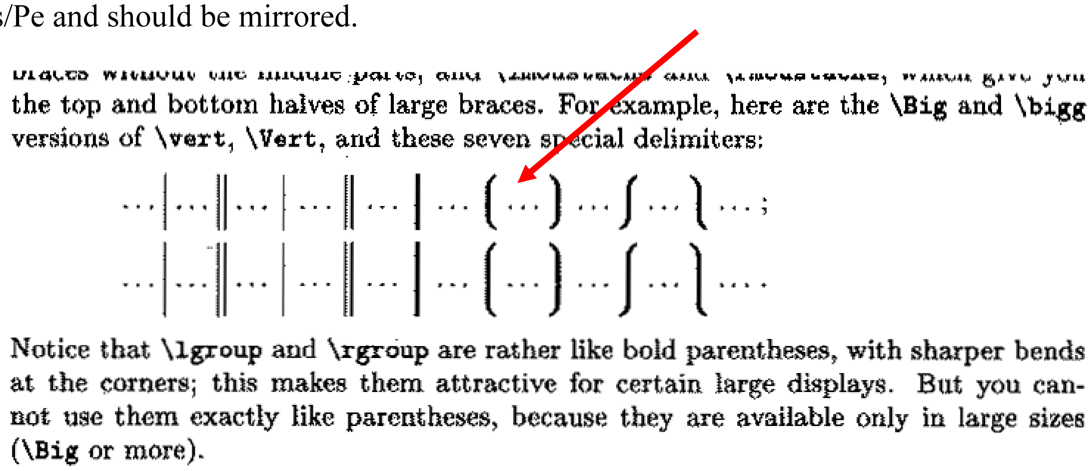
with a glyph consisting of a dashed square around a plus sign in analogy to other such operators, and General_Category=Cf and Default_Ignorable=true.

3 Math Delimiters

In typesetting tall mathematical expressions, a special set of delimiters that look like flattened parentheses are occasionally used. The chosen encoding model for large math delimiters is to use the ordinary character code to select the type of delimiter and have the layout software do the required stretching to scale the delimiter to the enclosed content. The delimiters considered here are special, in that they have no ordinary-sized counterparts. In TeX, these delimiters can be selected with the \lgroup or \rgroup macros, but for Unicode-based math layout, whether via MathML or other format, they cannot be used, as there are currently no corresponding characters. Accordingly we propose

27EE ⟮ MATHEMATICAL LEFT FLATTENED PARENTHESIS
= lgroup
27FF ⟯ MATHEMATICAL RIGHT FLATTENED PARENTHESIS
= rgroup

with representative glyphs that look like scaled-down versions of the shapes in Figure 1. Like other paired punctuation, these should be assigned General_Category values of Ps/Pe and should be mirrored.


the top and bottom halves of large braces. For example, here are the \Big and \bigg versions of \vert, \Vert, and these seven special delimiters:

Notice that \lgroup and \rgroup are rather like bold parentheses, with sharper bends at the corners; this makes them attractive for certain large displays. But you cannot use them exactly like parentheses, because they are available only in large sizes (\Big or more).

Figure 1. \lgroup and \rgroup from the TeXbook

3 Arrows

The Unicode Standard contains the following quadruple stemmed arrows:
U+27F0 ⟰ UPWARDS QUADRUPLE ARROW
U+27F1 ⟱ DOWNWARDS QUADRUPLE ARROW.

In reviewing the font being created for the STIX Project of the mathematical and scientific publishers, it was noted that the left/right orientations of these had been inadvertently omitted from earlier requests. The triple and double arrows are already encoded in all four orientations. Accordingly the following characters are proposed, with glyphs based on appropriately rotated shapes of the glyph for U+27F0.

2B45 ⭅ LEFTWARDS QUADRUPLE ARROW
2B46 ⭆ RIGHTWARDS QUADRUPLE ARROW

Like all other arrows in Unicode, these should not be mirrored.

4 Long Division

In texts for elementary arithmetic, there's a common notation for long division for which the correct shape looks something like this

2) $\overline{616}$

where the horizontal extent matches the operand. The occasional alternate representation

2 | $\overline{616}$

is regarded as an inferior fallback representation.

While this symbol is not used in scientific publishing, its use in educational material is very widespread and key implementers have expressed interest in supporting it. Figure 1a shows several typical ways of using this symbol, both in displayed, as well as in inline form (in the "bubble").

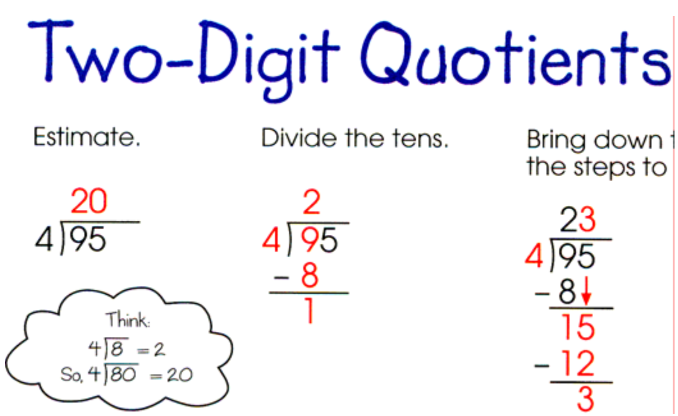


Figure 1a. From a 4th Grade Math Workbook

The proper way to support this notation would be analogous to the support of the square root which also needs to expand based on the expression of which the root is taken. In other words, the character code would be used in a context, such as MathML or other mathematical notation that supports the necessary scoping. In the linear notation developed by one of us (MS), for example, the operand would be enclosed in parens which are suppressed in display.

Accordingly we propose

27CC ⟌ LONG DIVISION

with properties to match U+221A √ SQUARE ROOT. Note that in ordinary text the character would simply act as a standalone symbol and not require special handling by non-mathematical layout engines.

5 Large Squares

According to the discussion of the use of abstract geometrical shapes in UTR #25, *Unicode Support for Mathematics* (http://www.unicode.org/reports/tr25/), the "normal" size of the characters for squares in the Standard is medium large. See Table 2.4 in that report. However, the STIX project has received a specific request from the American Institute for Physics (AIP) for squares that are larger than the standard squares mapped to U+25A0 BLACK SQUARE and U+25A1 WHITE SQUARE (see figure 2, but note that the source says "should be larger" than shown).

□	25A1	□		(square, open)
&squarelg;		□	should be larger	(larger open square)
▪	25A0	■		(square, solid)
&squfbtml;	bfr /	◪		(half closed square (bottom left))
&squfbtmr;	25EA	◪		(half closed square (bottom right))
&squflg;		■		(large-closed-square)
&squftopl;	25E9	◩		(half closed square (top left))
&squftopr;	bfp	◪		(half closed square (top right))
⋆	22C6	⋆		(small star, filled)

Figure 2. AIP character request to STIX

The entity names in Figure 2 imply a systematic distinction between large and regular squares (as corroborated by the comment for &squarelg;), even though the glyphs in the request document cited are inadequate.

The character mentioned for that purpose in table 2.4 of UTR #25, U+2588 █ FULL BLOCK does not appear appropriate for several reasons: By design this character is a full display cell in the font, and for nearly all fonts, that's a rectangle and not a square. There is also no corresponding white form, as there is for all other squares. The final reason is that squares for mathematical fonts should be centered on the math centerline, while the various blocks are aligned on the font's maximum ascent and descent. We therefore propose the addition of the following characters

2B1B ⬛ BLACK LARGE SQUARE
2B1C ⬜ WHITE LARGE SQUARE

with a size in between U+25A0 ■ BLACK SQUARE and U+2588 █ FULL BLOCK. (See also the review paper by P.R. Chastney cited in a later section of this document).

6 Very Small Squares

Unlike circles, for which there are many distinct sizes encoded, the Standard currently contains no squares smaller than the U+25AA BLACK SMALL SQUARE. As figure 3 shows, the Wolfram collection (implemented in the widely used Mathematica application) specifies the use of a *very small* filled square in contrast to the *small* square.

Therefore, we propose the addition of

2B1D ▪ BLACK VERY SMALL SQUARE
2B1E ▫ WHITE VERY SMALL SQUARE

comparable in size to U+2218 BULLET OPERATOR, and including an empty form for consistency. (See also the review paper cited in the next section of this document.)

form	full name	alias	form	full name	alias
▪	\[FilledVerySmallSquare]	:fvssq:	∘	\[EmptySmallCircle]	:esci:
▫	\[EmptySmallSquare]	:essq:	•	\[FilledSmallCircle]	:fsci:
▪	\[FilledSmallSquare]	:fssq:	○	\[EmptyCircle]	:eci:
□	\[EmptySquare]	:esq:	◍	\[GrayCircle]	:gci:
▩	\[GraySquare]	:gsq:	●	\[FilledCircle]	:fci:
■	\[FilledSquare]	:fsq:	△	\[EmptyUpTriangle]	
⬚	\[DottedSquare]		▲	\[FilledUpTriangle]	
▯	\[EmptyRectangle]		▽	\[EmptyDownTriangle]	
▮	\[FilledRectangle]		▼	\[FilledDownTriangle]	
◇	\[EmptyDiamond]		★	\[FivePointedStar]	:*5:
◆	\[FilledDiamond]		✶	\[SixPointedStar]	:*6:

Figure 3. From: the Mathematica web site at URL
http://documents.wolfram.com/v3/MainBook/3.10.3.html

7 Unit Pentagon on the Plane

It has been brought to our attention that if a pentagon shape is intended for use as an avatar of the standard unit pentagon on the complex number plane, its apex would point along the real axis, and therefore the shape would be right-pointing. Our original position was that the pentagon is intended for generic use and does not need to be encoded in all four orientations. However, we find the use of it in the context mentioned here persuasive enough that we recommend that UTC consider making an exception for the right pointing pentagon. Therefore we propose:

2B52 ⭓ BLACK RIGHT-POINTING PENTAGON
2B53 ⭔ WHITE RIGHT-POINTING PENTAGON

8 Other Geometrical Shapes

In January of 2006, the Unicode Technical Committee received public review feedback by P.R. Chastney (document L2/06-34). In it he carefully reviews the set of geometric shapes and the recommendations for their sizes and alignment presented in UTR #25. He comments: “in the interests of consistency and completeness, [we recommend that] the following characters be added to the set of abstract shapes:”

BLACK LARGE CIRCLE
BLACK PENTAGON
BLACK SMALL DIAMOND
WHITE SMALL DIAMOND
BLACK SMALL LOZENGE
WHITE SMALL LOZENGE
WHITE SMALL ARABIC STAR

The original list in L2/06-304 included the two large squares. They are discussed in Section 5 of this proposal, because there is a separate concurring request for them.

Two of us (AF and MS) were asked by the Scripts Committee of the Unicode Technical Committee (UTC) to review this request and arrive at a proposed resolution. After conferring with the co-authors and other experts, it was determined that the set of requested additions strikes a careful balance between consistency and completeness on the one hand, while avoiding the addition of shapes and sizes that are likely not going to be widely used. The extension of the set of diamonds and lozenges is seen as particularly desirable and had been a long-standing request.

The conclusion is that the authors support the recommendation in L2/06-304, with addition of the inclusion of the missing solid equivalent of 22C4 and white equivalent of 22C6. Further, in addition to the bullet sized small lozenges, a set of medium lozenges is known to be in use as notation for composition laws and should also be added. Note that the proposed set might be affected slightly by the resolution of the glyph issue for the existing character U+22C4 DIAMOND OPERATOR described in Section 11.

2B1F ⬟ BLACK PENTAGON
x 2B20 ⬠ white pentagon
2B24 ⬤ BLACK LARGE CIRCLE
x 25EF ◯ large white circle

2B25 ⬥ BLACK MEDIUM DIAMOND
2B26 ⬦ WHITE MEDIUM DIAMOND
2B27 ⬧ BLACK MEDIUM LOZENGE
2B28 ⬨ WHITE MEDIUM LOZENGE
2B29 ⬩ BLACK SMALL DIAMOND
x 22C4 ⋄ diamond operator
2B2A ⬪ BLACK SMALL LOZENGE
2B2B ⬫ WHITE SMALL LOZENGE
2B50 ⭐ WHITE MEDIUM STAR
x 22C6 ⋆ star operator
2B51 ⭑ WHITE SMALL STAR
x 066D ٭ arabic star

The proposed General_Category for these characters should be So, in keeping with most other geometric shapes.

9 Ellipse Shapes

The generic geometric shapes contained in the Unicode Standard contain circles, but not ellipses. Ellipses are used in many contexts, for example the white ellipse is frequently used in forms in both orientations. Evidence for ellipses in mathematical publication had been originally found in the symbol list from the Chicago Manual of Styles (see Figure 5). More recently, evidence has surfaced that that set of symbols is used in other countries as well (see Figure 4).

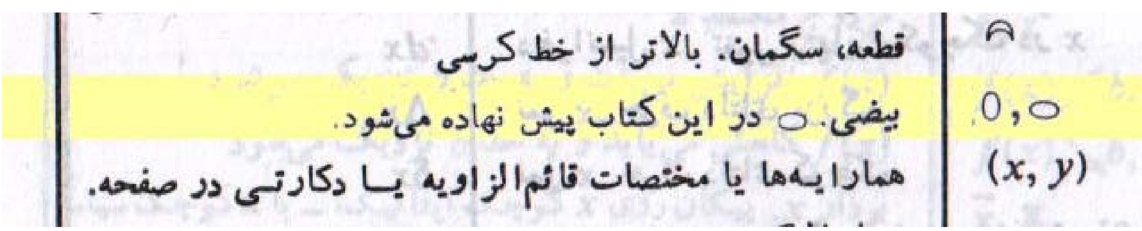


Figure 4. From "A Guide to Book Preparation", by Mir Šamseddin Adib-Soltani.

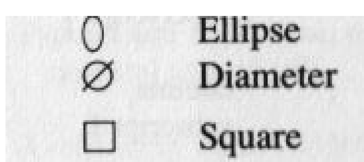


Figure 5. From the 14th edition of the Chicago Manual of Style

We propose the addition of

2B2C ⬬ BLACK HORIZONTAL ELLIPSE
2B2D ⬭ WHITE HORIZONTAL ELLIPSE
2B2E ⬮ BLACK VERTICAL ELLIPSE
2B2F ⬯ WHITE VERTICAL ELLIPSE

The proposed General_Category for these characters should be So, in keeping with most other geometric shapes.

10 Asterisk Accent

STIX has received a request from IEEE (see figure 6) for a combining asterisk accent (asterisk above) to be used as a math accent. TeX or LATEX users not using the IEEE macro package referenced in the request, can use such accents by constructing them dynamically using special macro packages (the \accentset macro is widely documented). MathML or other formats require that a corresponding combining character code is used. We therefore propose:

20F0 * COMBINING ASTERISK ABOVE

with properties analogous to other characters in that block.

27. \dwedge
28. \precstar
29. \asteraccent $\overset{*}{x}$
30. \ZED
31. \CTHRU
32. \semaphorel
33. \semaphorer

Figure 6. From IEEE request to STIX

11 Representative Glyph Shape of U+22C4 Diamond Operator

The character U+22C4 ◊ DIAMOND OPERATOR is one of a group of characters added between “Unicode 0.9” and Unicode 1.0. In Unicode 0.9, or formally, the *Unicode 1.0 Final Review Draft*, the Mathematical Operators block ended at 22C3. In the published Unicode 1.0 book, the 10646-1:1993 standard and the AFII registry, the glyph for this character is given as a fairly small lozenge and that is still the shape shown in the standard today, over fifteen years later. The use of a lozenge shape is in apparent contrast to the naming conventions for the standard. However, note that U+2662 ♢ WHITE DIAMOND SUIT is also not a rotated square.

When the character U+22C4 was added to the standard, it was cross-mapped in Unicode 1.0 to the SGML *diam* entity; in many other mappings published since then it is also mapped to the AMSTeX \diamond macro. (The 1997 AFII registry is one example of such a mapping). The definitions of AMSTeX macros, unlike SGML mappings, are backed up by identifiable font resources that have been widely available and in use for mathematical publications and may be considered stable. Reliable information on the intended shapes for SGML entities is generally more difficult to obtain, The entity set just says <!ENTITY diam "[SDATA]" --/diamond B: open diamond-->, which isn't very informative, however the full document of ISO/IEC TR 9573-13:1991 does show it as a rotated square. From evidence by participants it was established that the mapping of *diam* to \diamond is in itself deliberate, i.e. *diam* is an implementation of \diamond.

Figure 6a shows a recent example of use of the AMSTeX \diamond, a small rotated square that has approximately the size of U+2218 RING OPERATOR (second to last line).

Proof: First we show that for each $j \in Obj\,\mathcal{J}$ we have a morphism $\pi_j : L = \Delta_L(j) \to Fj = A_j$ in the 2-category $\mathcal{C}$. We claim that π_j is a morphism, *i.e.* a functor. Let $1_\eta = \Theta : \eta \rightsquigarrow \eta$ be the identity modification of the pseudo cone $\eta : \Delta_1 \Rightarrow F$. This means $\Theta_j = i_{\eta_j} : \eta_j \Rightarrow \eta_j$ is the identity natural transformation for all $j \in Obj\,\mathcal{J}$. Then $\pi_j(1_\eta) = \pi_j(\Theta) = \Theta_j(*) = 1_{\eta_j(*)} = 1_{\pi_j(\eta)}$ and π_j preserves identities. Now let Θ, Ξ denote modifications in L such that $\Xi \diamond \Theta$ exists. Then $\pi_j(\Xi \diamond \Theta) = (\Xi \diamond \Theta)_j(*) = \Xi_j \odot \Theta_j(*) = \Xi_j(*) \circ \Theta_j(*) = \pi_j(\Xi) \circ \pi_j(\Theta)$. Thus $\pi_j : L \to A_j$ is a functor.

Figure 6a: From Thomas M. Fiore, *Pseudo Limits, Biadjoints, and Pseudo Algebras: Categorical Foundations of Conformal Field Theory*.

Figure 6b shows the same operator in context with a large set of other binary operators, some of which also appear in Figure 6a. The size is consistently a medium size and the glyph is as tall as it is wide.

Table 39: Binary Operators

$\amalg$	\amalg	$\cup$	\cup	$\oplus$	\oplus	$\times$	\times
$\ast$	\ast	$\dagger$	\dagger	$\oslash$	\oslash	$\triangleleft$	\triangleleft
$\bigcirc$	\bigcirc	$\ddagger$	\ddagger	$\otimes$	\otimes	$\triangleright$	\triangleright
$\bigtriangledown$	\bigtriangledown	$\diamond$	\diamond	$\pm$	\pm	$\unlhd$	\unlhd*
$\bigtriangleup$	\bigtriangleup	$\div$	\div	$\rhd$	\rhd*	$\unrhd$	\unrhd*
$\bullet$	\bullet	$\lhd$	\lhd*	$\setminus$	\setminus	$\uplus$	\uplus
$\cap$	\cap	$\mp$	\mp	$\sqcap$	\sqcap	$\vee$	\vee
$\cdot$	\cdot	$\odot$	\odot	$\sqcup$	\sqcup	$\wedge$	\wedge
$\circ$	\circ	$\ominus$	\ominus	$\star$	\star	$\wr$	\wr

Figure 6b: Excerpt from Scott Pakin's list of 3,300 symbols available to LaTeX users.

The character 22C4 has been implemented variously in fonts. Lucida Sans shows a large rotated square, Arial Unicode MS shows a small rotated square matching that in Figure 6.a. Microsoft's new Cambria font shows a small lozenge, as depicted in the current code charts. MS Mincho also shows a lozenge shape, but a larger one (yet smaller than U+25CA LOZENGE).

Because U+22C4 was added to the standard in contrast to U+25C7 ◇ WHITE DIAMOND the fact that some implementations map a larger glyph to U+22C4 can be ignored for the remainder of our analysis. Therefore, the question to be addressed is primarily one of the correct aspect ratio.

In principle there are two alternatives:

A) Leave the shape of U+22C4 unchanged.
B) Change the shape of U+22C4 to a more symmetric aspect ratio

The following two sections outline the consequences of either approach. The authors are unanimous in seeking a speedy resolution of this issue, a majority prefer option B.

11.A Consequences of Retaining the Current Glyph for 22C4

There are a number of consequences of retaining the current glyph for U+22C4:

1. This will result in an inconsistency in the naming of 'diamond' shapes for mathematical use. This requires adding some annotation.
2. The existing mapping of \diamond to 22C4 needs to be changed. This might cause considerable difficulty as these mappings are not centrally controlled.
3. The existing mapping of *diam* to 22C4 would need to be changed away from what is documented in Unicode 1.0 – this can be considered a change in character identity. (In principle the need for that would depend on whether equating *diam* with \diamond is correct, however, as outlined above, we consider that mapping to be beyond question.)
4. The current shape matches at least one recently released font for mathematical use, however, another font for mathematical use, in late stages of quality assurance, is about to be released with the opposite glyph choice.
5. Other than an annotation, the Unicode Standard does not change; this could be considered a benefit.

If UTC accepts this option, the proposed characters in section 8 above would change as follows:

2B2A ⬪ WHITE SMALL DIAMOND
2B2B ⬫ BLACK SMALL LOZENGE
x 22C4 ⋄ diamond operator

11.B Consequences of Changing the Glyph for 22C4 to a Rotated Square

There are a number of consequences of changing the representative glyph for U+22C4:

1. The naming of characters for mathematical use would become more consistent, the only exception, U+2262 WHITE DIAMOND SUIT, is from a different context.
2. The mapping to \diamond would remain stable.
3. The mapping to *diam* would remain stable (this is a mapping that we assert in Unicode 1.0 from the time we added the character). In essence, this would affirm the stability of the underlying identity of the character.
4. The change conflicts with a recently released font for mathematical use, but the vendor now considers their glyph shape a bug.
5. A square shape would match another font for mathematical use about to be released and which is in late stages of quality assurance—this font will likely be influential in the mathematical community.
6. The Unicode Standard would need to change a shape that's been stable for 16+ years, which can be considered a downside.

12 Overview of Geometric Shapes, Existing and Proposed

Table 1 Proposed extensions in context and size relation with existing characters

shape	tiny	very small		small (bullet)		medium small		medium (default1)		regular (default2)		large	
triangle left				25C2	25C3					25C0	25C1		
triangle right				25B8/2023	25B9					25B6	25B7		
triangle up				25B4	25B5					25B2	25B3		
triangle down				25BE	25BF					25BC	25BD		
square		2B1D	2B1E	25AA	25AB	25FD	25FE	25FC	25FB	25A0	25A1	2B1B	2B1C
diamond				2B29	22C4	2B26	2B25			25C6	25C7		
lozenge				2B2B	2B2A	2B28	2B27			29EB	25CA		
pentagon										2B1F	2B20		
pentagon right										2B52	2B53		
hexagon horizontal										2B23	2394		
hexagon vertical										2B22	2B21		
arabic star				066D	2B51	22C6	2B50	2605	2606				
ellipse horizontal										2B2C	2B2D		
ellipse vertical										2B2E	2B2F		
circle	22C5	2219/00B7	2218	2022	25E6	2981	26AC	26AB	26AA	25CF	25CB	2B24	25EF
circled circles	2299	2609			233E								
circled circles	2A00	29BF	229A		29BE	25C9	25CE						

Table 1 shows the proposed extensions to the repertoire of geometric shapes in context with existing characters and in a suggested size relation based on input from the mathematical user and font community. Not shown are symbols based on circles or squares inscribed inside squares and the "diamond suit" symbols from the card suits. The latter fall in-between the diamond and lozenge shapes (fat lozenge) and are therefore not part of either series.

There are two columns labeled default. They represent common choices for sizes for the characters that come in a single or very restricted number of distinct sizes. L2/06-034 had argued that the natural choice for mathematical use is the size matching the "default 1" column, but our code charts currently show sizes more like the "default 2" column and that's what is shown here.

The sizes of the characters in each row as shown in Table 1 follow the geometric size progression recommended in document L2/06-034. The sizes in each column are adjusted so that they produce an effectively equivalent 'visual impact'. See Unicode Technical Report #25 "Unicode Support for Mathematics" (http://www.unicode.org/reports/tr25/) for a more extensive discussion.

The shape of some of the glyphs in our code charts and/or in common fonts differs from the shapes shown in this table: U+066D ARABIC STAR is sometimes implemented in Arabic fonts with a shape quite different than the one in our code charts, "*". U+2394 software function symbol is shown much smaller and lighter than the default geometric shapes, it is shown like this: ○ .

Note that this proposal no longer considers Block Elements to be geometrical shapes and that it considers two long-existing character pairs as duplicates. The first pair is 2023 and 25B8, with 25B8 the preferred character for mathematical use, partially because it forms part of a complete and regular set of 4 orientations in two colors. However, existing mappings from external character sets may not be consistent. The second pair is 2219 and 00B7. Here, 2219 is the mathematical character, but 00B7 will be found in most data originating from ISO 8859-1 and derived character sets.

Appendix 1 Note on the Shapes of Floors and Ceilings

It has come to the attention of the authors that some web-based publications misrepresent the shape and placement of floor and ceiling symbols (U+23xx–U+23xx). There's actually a very clear consensus in the mathematical community on the proper typographic treatment of these characters, which can be stated as follows:

> "Ideal forms of floors and ceilings are shaped like tall sans-serif L shapes, with their horizontal and vertical reflections appropriately translated about, with floors extending below the baseline and ceilings ending at about cap height. Stroke width tends to be uniform. The horizontal foot is short, but not too short."

It should be noted that because mathematical notation uses these symbols in distinction to both square brackets and ordinary (quine) corners, adherence to these specifications is critical to allow unambiguous recognition.

Appendix 2 Symbols and Shapes with Inscribed Letter 's'

The symbol list of the Chicago Manual of Style has long included a number of Symbols and Geometrical Shapes with Inscribed Letter 's'. Such symbols are intended to designate a plural. At the time that list of characters was reviewed for inclusion into the Unicode Standard, it was unclear where these characters were used and they were considered curiosities. The decision at the time was to hold off from encoding them, not least because the 's' is specific to an English plural. However, we are now seeing that sources in other languages reference that list and document the same symbols. See Figures 7 and 8. Because of the fact that the language of international science is English, it is perhaps not so surprising to see such cross-cultural reference.

Recently, additional evidence about that type of notation has come to the attention of the authors. In his two-volume *History of Mathematical Notation,* published in 1928, Florian Cajori discusses the development of notations used for "angle" and "angles", and cites different usages for the plural, such as ∠∠, with his usually careful attribution to specific authors and manuscripts. He then states that ∠s and ⊥s are "used as well", but for once does not indicate when and where. He cites an early 20th century usage of ∥*g'm* for "parallelogram," which is in the same spirit as ∥s (but, in these cases, there is no merging of the letter form with the symbol). Because of the unfortunate lack of attribution, there is only circumstantial evidence, including the listing in the Chicago Manual of Style, that such notation survived well into the 20th century.

Because of their unusual nature, and lack of definite attestation in modern usage, other than style guides, some of the authors feel that any discussion of encoding these is premature. Further evidence of publications using them in context, in whatever language context, would be required. In the meantime, we are placing the initial evidence below in the record. We do not recommend that UTC take any action on these at this time.

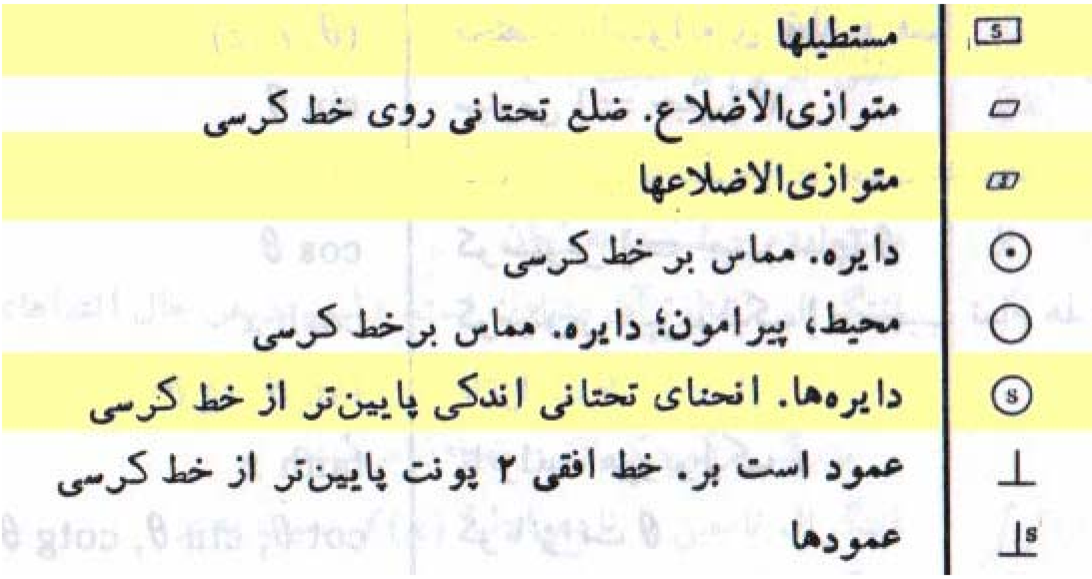


Figure 7. From "A Guide to Book Preparation", by Mir Šamseddin Adib-Soltani. (The same source also contains an s inscribed in a rectangle, a square and a triangle).

Note that in Figure 8, the “parallels” example is (except for spacing) indistinguishable from a simple sequence of “parallel” and the small letter ‘s’, whereas the other symbols are kerned and fused.

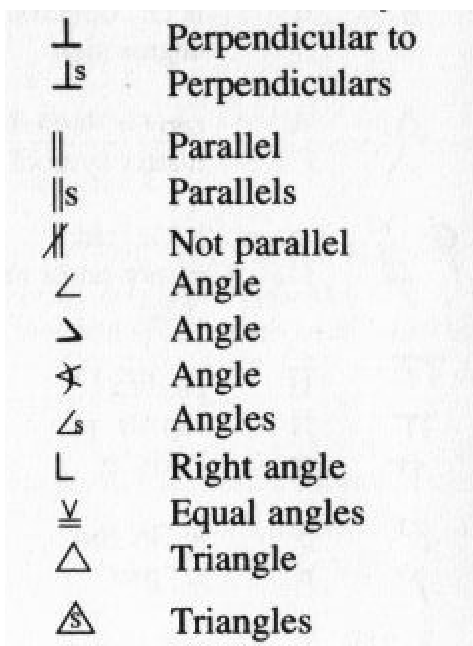


Figure 8. From the 14th edition of the Chicago Manual of Style (the same lists also contains s inscribed in a circle, a square, a parallelogram and a rectangle).

Appendix 3 - Typical conventions for use of mathematical alphabets

The Unicode Standard contains a number of sets of alphanumeric characters in specific type styles. The following information catalogs examples of typical uses for some of these styles without intending to be exhaustive or exclusive. The authors recommend that this information be used in updating Unicode Technical Report #25, *Unicode Support for Mathematics* when that is revised next.

- *lightface italic* -- variables
- *double-struck* -- sets
- *bold* -- vectors (more physics and applied areas, usually lowercase)
- *bold italic* -- matrices (uppercase)
- *lightface roman* -- operator names (sin, cos, etc.), some constants, units
- *lowercase greek* -- angles
- *script (caps)* -- various operators, functions and transforms
- *sans-serif* -- dimensions of SI base quantities (NIST guide, p.23; uncertain whether lightface or bold)
- *bold italic sans-serif* -- tensors (NIST guide, p.34, also NIST style sheet)

NIST Guide: NIST publication 811, Guide for the use of the international system of units. link at http://physics.nist.gov/Pubs/pdf.html

NIST Style Sheet: at http://www.physics.nist.gov/Document/typefaces.pdf

ISO/IEC JTC 1/SC 2/WG 2

PROPOSAL SUMMARY FORM TO ACCOMPANY SUBMISSIONS
FOR ADDITIONS TO THE REPERTOIRE OF ISO/IEC 10646[1]
Please fill all the sections A, B and C below.
Please read Principles and Procedures Document (P & P) from http://www.dkuug.dk/JTC1/SC2/WG2/docs/principles.html for guidelines and details before filling this form.
Please ensure you are using the latest Form from http://www.dkuug.dk/JTC1/SC2/WG2/docs/summaryform.html.
See also http://www.dkuug.dk/JTC1/SC2/WG2/docs/roadmaps.html for latest *Roadmaps*.

A. Administrative

1. **Title:** *Additional Mathematical and Symbol Characters*
2. Requester's name: *See document authors*
3. Requester type (Member body/Liaison/Individual contribution): *See document header*
4. Submission date: *See document date*
5. Requester's reference (if applicable): *L2/07-011*
6. Choose one of the following:
   This is a complete proposal: *X*
   (or) More information will be provided later:

B. Technical – General

1. Choose one of the following:
   a. This proposal is for a new script (set of characters):
      Proposed name of script:
   b. The proposal is for addition of character(s) to an existing block: *X*
      Name of the existing block: *Multiple*
2. Number of characters in proposal: *29*
3. Proposed category (select one from below - see section 2.2 of P&P document):
   A-Contemporary *X* B.1-Specialized (small collection) B.2-Specialized (large collection)
   C-Major extinct D-Attested extinct E-Minor extinct
   F-Archaic Hieroglyphic or Ideographic G-Obscure or questionable usage symbols
4. Proposed Level of Implementation (1, 2 or 3) (see Annex K in P&P document):
   Is a rationale provided for the choice?
   If Yes, reference:
5. Is a repertoire including character names provided? *Yes*
   a. If YES, are the names in accordance with the "character naming guidelines" in Annex L of P&P document? *Yes*
   b. Are the character shapes attached in a legible form suitable for review? *Yes*
6. Who will provide the appropriate computerized font (ordered preference: True Type, or PostScript format) for publishing the standard? *Contact principle author*
   If available now, identify source(s) for the font (include address, e-mail, ftp-site, etc.) and indicate the tools used:
7. References:
   a. Are references (to other character sets, dictionaries, descriptive texts etc.) provided? *Yes (STIX)*
   b. Are published examples of use (such as samples from newspapers, magazines, or other sources) of proposed characters attached? *Yes*
8. Special encoding issues:
   Does the proposal address other aspects of character data processing (if applicable) such as input, presentation, sorting, searching, indexing, transliteration etc. (if yes please enclose information)?
   Format characters analogous to already existing (see attached)

9. Additional Information:

Submitters are invited to provide any additional information about Properties of the proposed Character(s) or Script that will assist in correct understanding of and correct linguistic processing of the proposed character(s) or script. Examples of such properties are: Casing information, Numeric information, Currency information, Display behaviour information such as line breaks, widths etc., Combining behaviour, Spacing behaviour, Directional behaviour, Default Collation behaviour, relevance in Mark Up contexts, Compatibility equivalence and other Unicode normalization related information. See the Unicode standard at http://www.unicode.org for such information on other scripts. Also see http://www.unicode.org/Public/UNIDATA/UCD.html and associated Unicode Technical Reports for information needed for consideration by the Unicode Technical Committee for inclusion in the Unicode Standard.

C. Technical - Justification

1. Has this proposal for addition of character(s) been submitted before?
   If YES explain *No*
2. Has contact been made to members of the user community (for example: National Body, user groups of the script or characters, other experts, etc.)?
   If YES, with whom? *Yes, Publishers, Users, Implementers and Typographical experts*
   If YES, available relevant documents:
3. Information on the user community for the proposed characters (for example: size, demographics, information technology use, or publishing use) is included? *Yes*
   Reference:
4. The context of use for the proposed characters (type of use; common or rare) *Technical*
   Reference:
5. Are the proposed characters in current use by the user community? Yes
   If YES, where? Reference: *STIX, TeX, MathML etc., see attached.*
6. After giving due considerations to the principles in the P&P document must the proposed characters be entirely in the BMP? *Yes*
   If YES, is a rationale provided? *No*
   If YES, reference:
7. Should the proposed characters be kept together in a contiguous range (rather than being scattered)? *N/A*
8. Can any of the proposed characters be considered a presentation form of an existing character or character sequence? *Yes*
   If YES, is a rationale for its inclusion provided? *Yes*
   If YES, reference: *See attached*
9. Can any of the proposed characters be encoded using a composed character sequence of either existing characters or other proposed characters? *No*
   If YES, is a rationale for its inclusion provided?
   If YES, reference:
10. Can any of the proposed character(s) be considered to be similar (in appearance or function) to an existing character? *No*
    If YES, is a rationale for its inclusion provided?
    If YES, reference:
11. Does the proposal include use of combining characters and/or use of composite sequences? *Yes*
    If YES, is a rationale for such use provided? *Yes*
    If YES, reference:
    Is a list of composite sequences and their corresponding glyph images (graphic symbols) provided?
    If YES, reference:
12. Does the proposal contain characters with any special properties such as control function or similar semantics? *Yes*
    If YES, describe in detail (include attachment if necessary)
    See attached
13. Does the proposal contain any Ideographic compatibility character(s)? *No*
    If YES, is the equivalent corresponding unified ideographic character(s) identified?
    If YES, reference: