ISO/IEC JTC1/SC2/WG2 N2336
2001-04-02

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

Doc Type: **Working Group Document**
Title: **Additional Mathematical Symbols**
Source: **Barbara Beeton (AMS), Asmus Freytag (Unicode), and Patrick D. F. Ion (W3C Math WG)**
Status: **Expert contribution**
Action: **For consideration by JTC1/SC2/WG2 and UTC**

The set of mathematical symbols proposed for addition to ISO/IEC 10646-1 and Unicode is based on an extensive search of existing mathematical literature. By its very nature, such a search can never be exhaustive, least of all for a notation that is a living and productive as mathematical notation. In order to have a workable proposal, an initial set was frozen at around the time it was first submitted to WG2.

In the process of refining the initial set, some much needed character candidates were found to have been inadvertently dropped. In parallel to work on refining this initial set, the literature search progressed and generated a list of additional candidate characters.

An ad-hoc group consisting of the authors of this document, with input from Michael Everson, Ken Whistler and Murray Sargent, has narrowed these candidate characters down to the list presented here and provided suggested code locations, names and annotations for use in the Unicode Standard.

Format of this document

While the format of this document follows that of a Unicode character names list, many of the annotations are provided mainly for the readers and reviewers of this document; they are not intended to become part of the standard.

In some cases compatibility ‘mappings’ with novel tags can be found in the names list. These are indications of glyphic or semantic relations between characters, and not suggested formal decompositions. In the final standard they would be replaced by simple cross-references.

Use of Mathematical Symbols with MathML

Since the original proposal that is now part of PDAM1, the W3C has essentially completed its work on Version 2.0 of MathML and many vendors are busy implementing support for it. In moving away from a model that is entirely based on SGML-style entities towards stronger reliance on character codes MathML makes reference to the characters in PDAM1, but also needs to be able to map some of the existing legacy entity sets to character codes. Some of the characters in this proposal are urgently needed for this purpose. Adding these characters to the standard quickly, enables W3C to proceed with MathML without having to add cumbersome temporary work-arounds for these entities.

Proposed code Allocation

Where characters can fill existing holes in the standard (or in PDAM1) the proposed code allocation is indicated directly in the list of characters. Characters that would fit into new blocks are shown with temporary code positions that use private use space.

Three new blocks are to be created to contain the following sequence of characters:

*F540 – F55C Additional Math Operators
*F580 – F597 Supplemental Geometric Shapes
*F570 – F57F Supplementary Arrows-A

The actual location of these blocks is expected to be in the range 2B00 – 2E7F. The code points for the characters proposed here will then have to be adjusted to match the agreed upon location of the blocks.

Notes on proposed characters

2063 INVISIBLE SEPARATOR

The invisible separator (comma) is intended for use in index expressions and other mathematical notation where two adjacent variables form a list and are not implicitly multiplied. In mathematical notation, commas are not always explicitly present, but need to be indicated for symbolic calculation software to help it disambiguate a sequence from a multiplication. For example in the following sequence:

$\sigma^{2\,,\,n}$

is distinct from

σ^{2n}

which could be mis-interpreted as the singly superscripted

$\sigma^{(2n)}$

instead of the functionally equivalent

$\sigma^{2,n}$

which is doubly superscripted as intended. Use of the invisible comma would hint to a math layout program to typeset a small space between the variables.

A fuller discussion of the uses of adjacency in mathematical notation can be found in the forthcoming Unicode Technical Report #28, *Unicode Support for Mathematics.*

This character is currently supported in widely selling mathematical software.

2052 COMMERCIAL MINUS SIGN

The use of the commercial minus sign is attested in commercial or tax related forms or publications in German, where it is pronounced “abzüglich” and is also cited in other sources for the same usage in Scandinavia. The ./. form of this character appears to be something of a fallback representation; while there are citations of both forms in the same publication, they occur with different fonts and without apparent change in meaning. Therefore it is proposed to unify these two forms. If a requirement can be demonstrated for accessing the second form explicitly it can be added as a glyph variation accessible via VS1.

Additional usage: the sign appears in the Finno-Ugric Phonetic Alphabet (FUPA) to mark a structurally related borrowed element of different pronunciation. In Finland, the dingbats ⁒ and ✓ are always used for “correct” and “incorrect” respectively in correcting a student’s paper. This contrasts with e.g. American practice, where ✓ and ✗ might be used for “correct” and “incorrect” respectively in the same context. The symbol may also appear as marginal note in letters, denoting enclosures. One variation replaces the top dot with a digit indicating the number of enclosures.

Since use of this character is not part of standard mathematical notation, it is proposed to encode the character in the General Punctuation block.

F549 MULTIPLICATION ON-LINE

Unlike the period (full stop) this dot is used as an operator, indicating multiplication. The glyph needs different spacing from the glyph for period and the character is explicitly an operator and not sentence terminal punctuation.

Note that use of the full stop in mathematical expressions is not uncommon in its two functions of decimal point and sentence terminal punctuation. The latter use occurs when a displayed expression ends a sentence.

F54A ON-LINE DOT

Unlike F549, this is used in pairs as a fence. Fences have different spacing than other characters of otherwise similar appearance and need to be handled specially in math layout. Glyphs are usually designed with different alignment and spacing.

F576 – F57F Long Arrows

The long forms of arrows are commonly used in mathematical papers. Encoding of these via the VS1 character had been considered, but they are proposed as separate characters here because they represent distinct semantics, rather than mere stylistic glyph differences. One example of a semantic distinction is that the shorter forms are used in connection with limits and the longer with mappings. Their use is so common that they were assigned entities in the ISOAMSA entity set, one of the several mathematical symbol entity sets ISO defined.

MathML is an XML application, and MathML is intended to support the full legacy collection of the ISO math entity sets. This is in part because publishers who use forms of the ISO 12087 standards for mathematics would reasonably expect it to. The STIX project of the STIPUB group of publishers included those ISO sets in its collection of mathematical characters that formed the source list submitted to the UTC for standardization.

These characters are needed to complete the mapping to the entity set in question.

29D8 LEFT WIGGLY FENCE

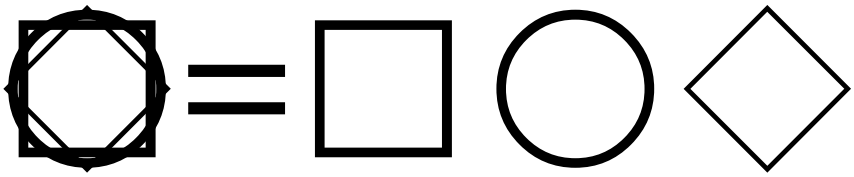
The proposed character looks superficially similar to FE34 PRESENTATION FORM FOR VERTICAL LOW LINE, which is intended for legacy support as an ‘underlining’ character in vertical context. It has a compatibility mapping to 005F in Unicode. This represents a very different use and properties from the standard use of fence characters in mathematical notation — therefore the characters cannot be unified.

29E6 LARGE SQUARE WITH BLACK MEDIUM SMALL CIRCLE

This character looks superficially like a larger version of 22A1 SQUARED DOT OPERATOR – but the latter is unambiguously an operator, whereas the proposed LARGE SQUARE WITH BLACK MEDIUM SMALL CIRCLE is an object and larger.

Geometrical Shapes

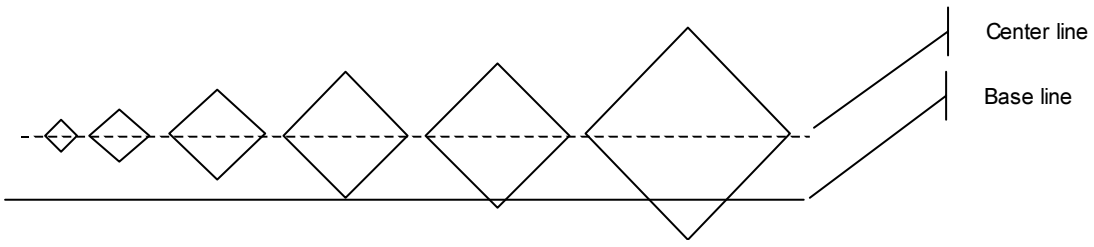
Ideal sizes. Mathematical usage requires at least four distinct sizes of simple shapes, and sometimes more. The size gradation must allow each size to be recognized, even when it occurs in isolation. In other words shapes of the *same* size should ideally have roughly the same visual "impact" as opposed to same nominal height or width. For mathematical usage simple shapes ideally share a common center. The following diagram shows which size relationship *across* shapes of the same nominal size is considered ideal.



Please note that neither the current set of glyphs in the standard nor the glyphs from many commonly available non-mathematical fonts show this kind of size relation.

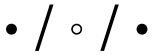
Actual sizes. The sizes of existing characters and their names are not always consistent. For mathematical usage, therefore, the MEDIUM SMALL SQUARE should be used together with the MEDIUM size of the other basic shapes, and correspondingly for the other sizes. The basic shapes from the ZapfDingbats font match the unmarked size for triangle, diamond and circle and the MEDIUM size for the square. To achieve the correct size relation, mathematical fonts may need to deviate in minor amounts from the sizes shown in the character charts.

Positioning: For a mathematical font, the centerline should go through the middle of a parenthesis, which should go from bottom of descender to top of ascender. This is the same level as the minus or the middle of the plus and equal signs. For correct positioning, the glyph will descend below the baseline for the larger sizes of the basic shapes as in the following schematic diagram:



The standard triangles used for mathematics are also center aligned. This is different from the positioning for the reference glyphs of existing characters shown in the charts. Mathematical fonts may need to deviate in positioning of these triangles.

Relative sizes of simple geometrical shapes

Size	Circles	Squares	Diamonds	Lozenges	Triangles
Large					
Unmarked					
Medium					
Med.Small					
Small					
Very Small					
Tiny					

Size	Circles	Squares*	Diamonds	Lozenges	Triangles
Large	**25EF**	**2588**	F588	F591	
Unmarked	**25CF**	**25A0**	**25C6**	**29EB / 25CA**	**25B2**
Medium	F59A	**25FC**	F58B	F594	F580
Med. Small	**2981**	**25FE**	F58D	F596	
Small	**2022 / 25E6**	**25AA**	F58F	**22C4*** / F598	25B4
Very Small	**2219 / 2218/ 00B7***	F59C			
Tiny	**22C5***				

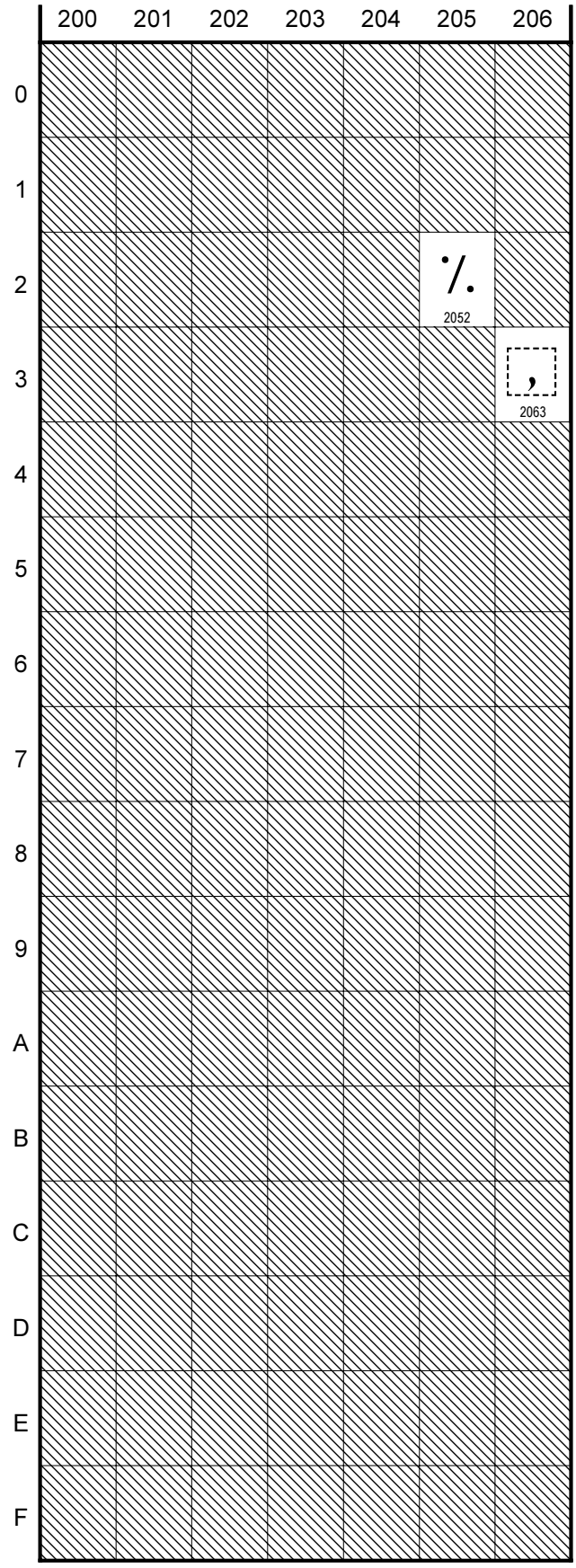
Characters already coded or in the PDAM are shown in black in the upper table with bold code points in the lower table. Where a white symbol is shown the black symbol is not currently encoded, or the relation between white and black symbol is not obvious from the name. Symbols in blue correspond to the current proposal.

The light green comparison glyphs are from the ZapfDingbats font.

The left column in both tables gives a size indication. Where possible it matches the size indication in the character name. Sometimes the names do not indicate a size, in which case, the characters have been inter-filed according to relative sizes of their glyph (e.g. as done for the characters that consist of smaller circles).

Notes:

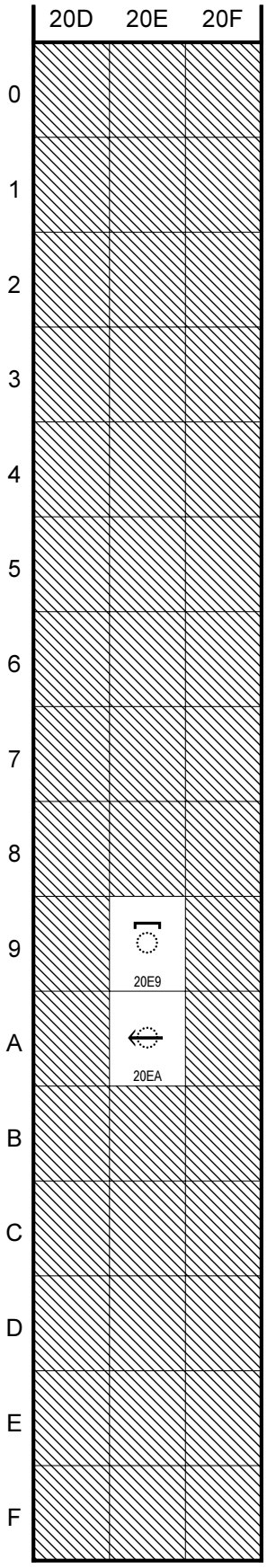
- DIAMOND OPERATOR is a misnomer for *SMALL LOZENGE
- The squares are consistently one size larger than their nominal size
- Circles and dots occupy the same series. 00B7 may also be rendered in tiny size


200
201
202
203
204
205
206
0
1
2
3
4
5
6
7
8
9
A
B
C
D
E
F
⁒
2052
,
2063

2052 ⁒ COMMERCIAL MINUS SIGN
= abzüglich (German), med avdrag av (Swedish), piska (Swedish, "whip")
• a common glyph variant and fallback representation looks like ./.
• may also be used as a dingbat to indicate correctness
• used in Finno-Ugric Phonetic Alphabet to indicate a related borrowed form with different sound
→ 0025 % percent sign
→ 066A ٪ arabic percent sign

Invisible Operator

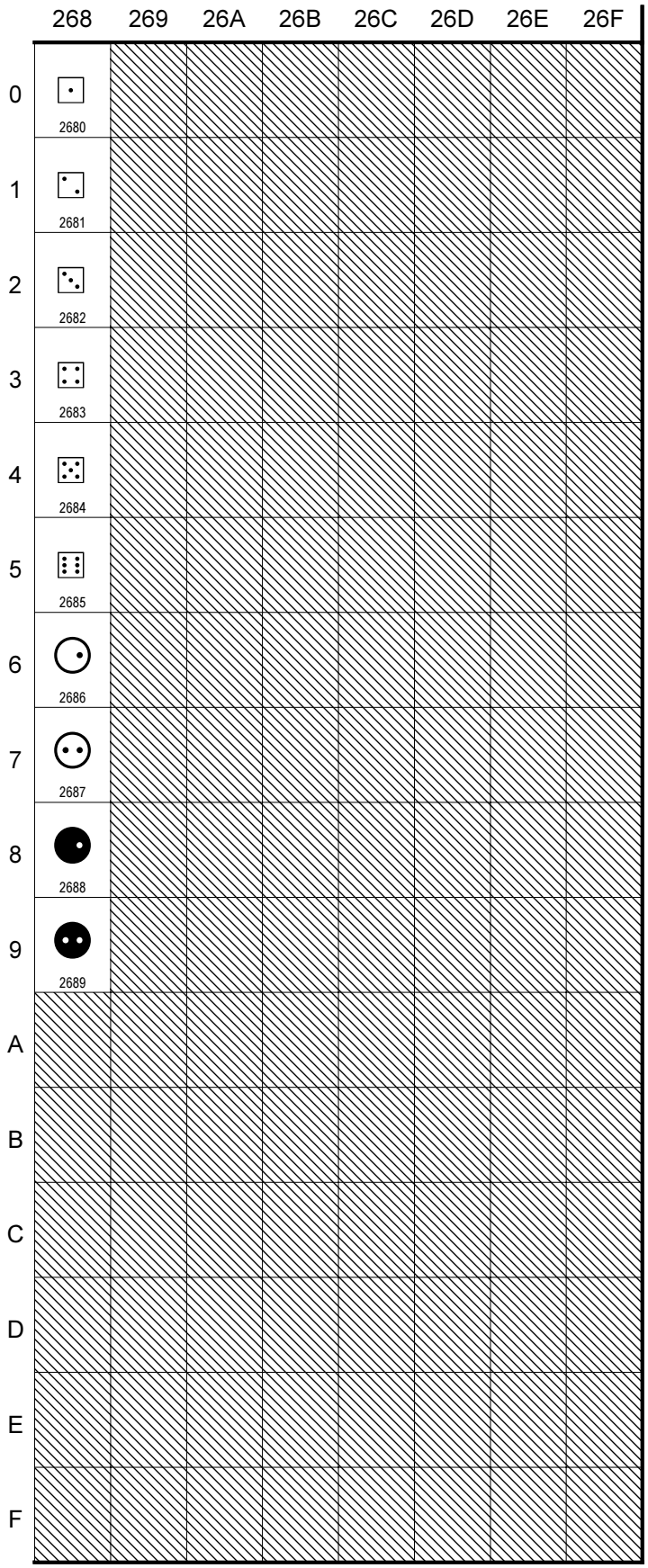
2063 ⁣ INVISIBLE SEPARATOR
= invisible comma
• contiguity operator indicating that adjacent mathematical symbols form a list, e.g. when no comma is used between multiple indices


20D
20E
20F
0
1
2
3
4
5
6
7
8
9
A
B
C
D
E
F
20E9
20EA

Combining characters

20E9 ◌⃩ COMBINING WIDE BRIDGE ABOVE
= contraction operator
• this character extends the full width of the base character
→ 0346 ◌͆ combining bridge above

20EA ◌⃪ COMBINING LEFTWARD ARROW OVERLAY

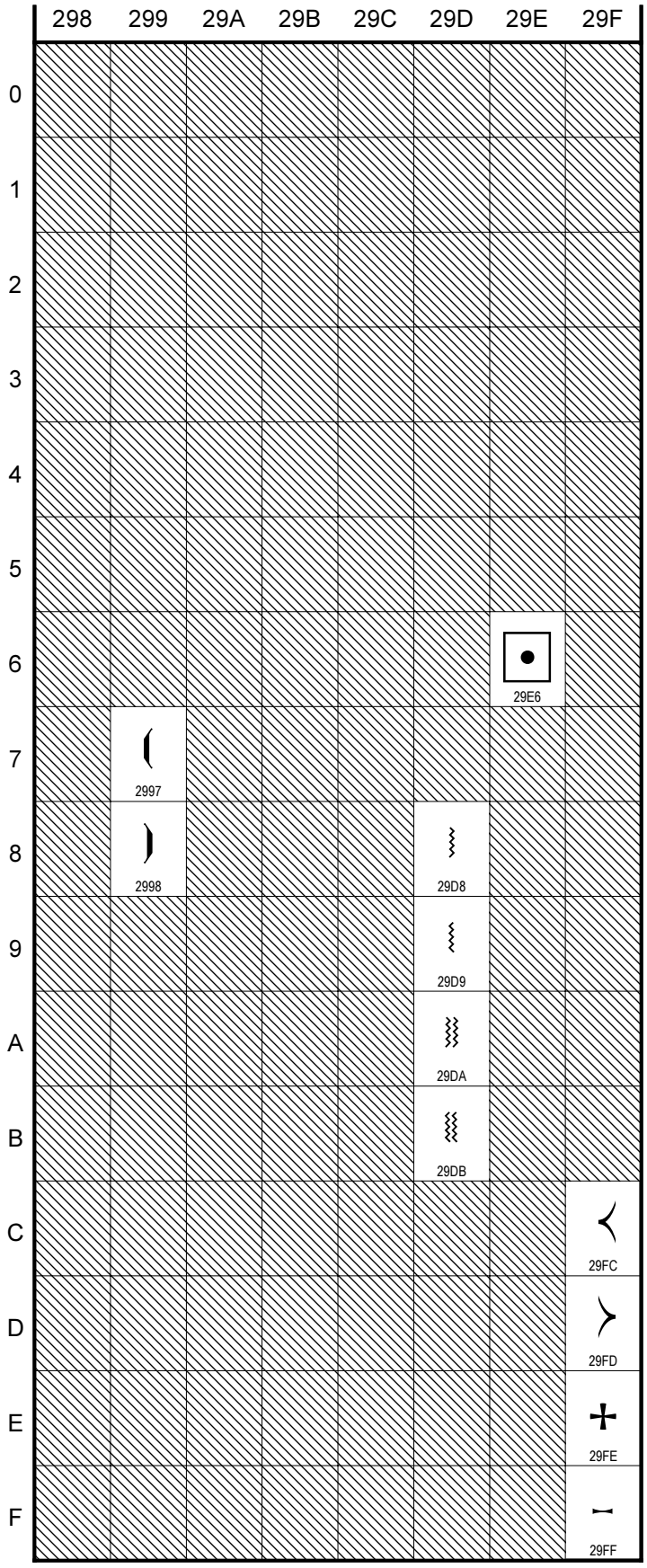

268
269
26A
26B
26C
26D
26E
26F
0
1
2
3
4
5
6
7
8
9
A
B
C
D
E
F
2680
2681
2682
2683
2684
2685
2686
2687
2688
2689

Dice

2680 ⚀ DIE FACE-1
2681 ⚁ DIE FACE-2
2682 ⚂ DIE FACE-3
2683 ⚃ DIE FACE-4
2684 ⚄ DIE FACE-5
2685 ⚅ DIE FACE-6

Go markers

2686 ⊙ WHITE CIRCLE WITH DOT RIGHT
2687 ⊚ WHITE CIRCLE WITH TWO DOTS
2688 ⚈ BLACK CIRCLE WITH WHITE DOT RIGHT
2689 ⚉ BLACK CIRCLE WITH TWO WHITE DOTS


298
299
29A
29B
29C
29D
29E
29F
0
1
2
3
4
5
6
7
8
9
A
B
C
D
E
F
29E6
2997
2998
29D8
29D9
29DA
29DB
29FC
29FD
29FE
29FF

Brackets

2997 ⦗ LEFT BLACK TORTOISE SHELL BRACKET
≈ <black> 3014 〔 left tortoise shell bracket

2998 ⦘ RIGHT BLACK TORTOISE SHELL BRACKET
≈ <black> 3015 〕 right tortoise shell bracket

Fences

29D8 ⧘ LEFT WIGGLY FENCE
→ FE34 ﴾ presentation form for vertical low line

29D9 ⧙ RIGHT WIGGLY FENCE
→ FE34 ﴿ presentation form for vertical low line

29DA ⧚ LEFT DOUBLE WIGGLY FENCE

29DB ⧛ RIGHT DOUBLE WIGGLY FENCE

Other Symbol

29E6 ⧦ LARGE SQUARE WITH BLACK MEDIUM SMALL CIRCLE
→ 29C7 ⧇ squared small circle
→ 22A1 ⊡ squared dot operator

Brackets

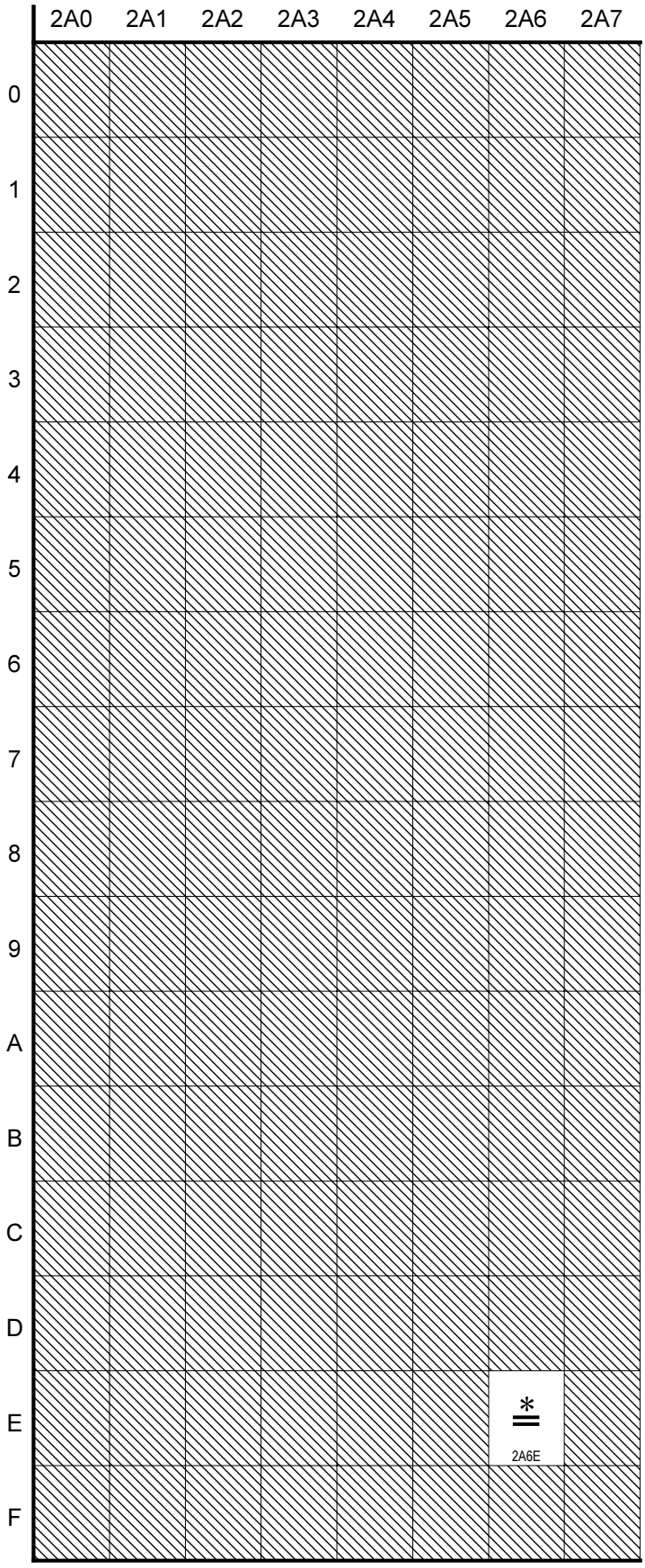
29FC ⧼ LEFT POINTING CURVED ANGLE BRACKET
→ 2329 〈 left pointing angle bracket

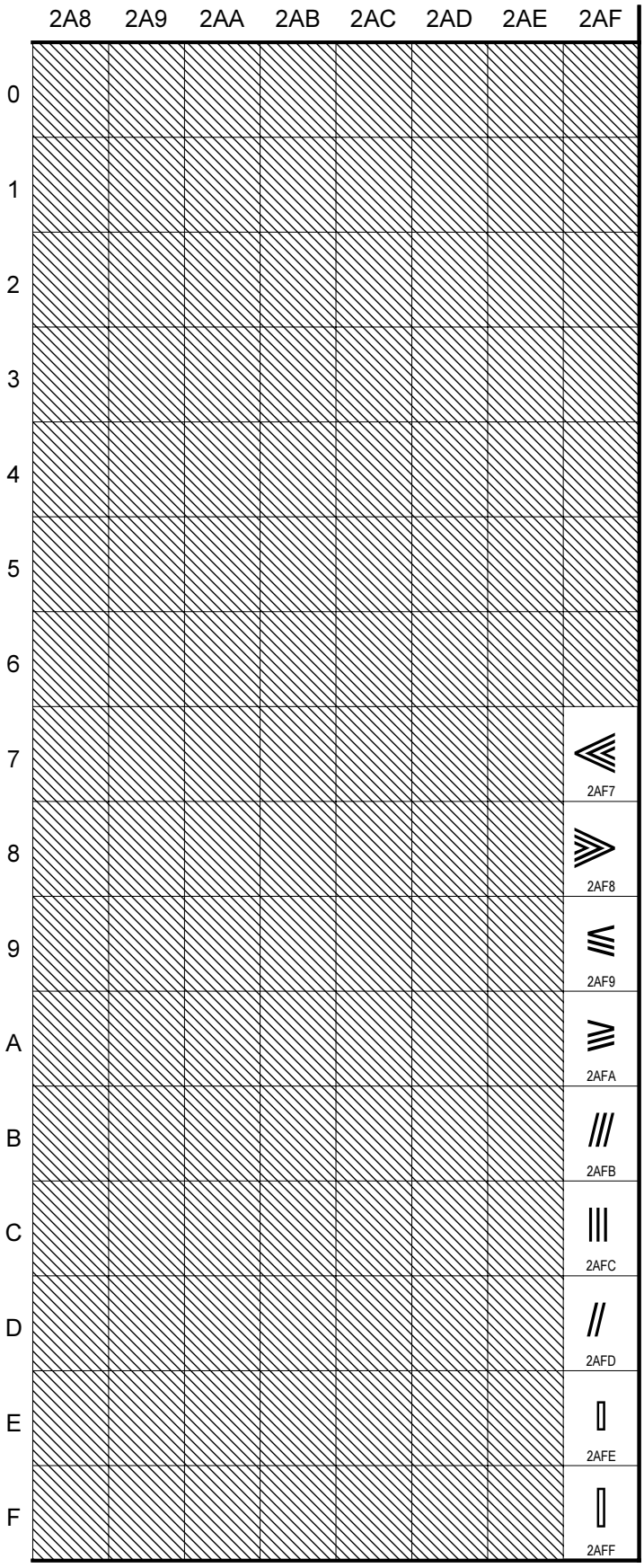
29FD ⧽ RIGHT POINTING CURVED ANGLE BRACKET
→ 232A 〉 right pointing angle bracket

Other Symbols

29FE ⧾ TINY

29FF ⧿ MINY


2A0
2A1
2A2
2A3
2A4
2A5
2A6
2A7
0
1
2
3
4
5
6
7
8
9
A
B
C
D
E
F
≠
2A6E


2A8
2A9
2AA
2AB
2AC
2AD
2AE
2AF
0
1
2
3
4
5
6
7
8
9
A
B
C
D
E
F
2AF7
2AF8
2AF9
2AFA
2AFB
2AFC
2AFD
2AFE
2AFF

2A6E ≛ EQUALS WITH ASTERISK

Relations

2AF7 ⫷ STACKED VERY MUCH LESS-THAN
≈ <stacked> 22D8 ⋘ very much less-than

2AF8 ⫸ STACKED VERY MUCH GREATER-THAN
≈ <stacked> 22D9 ⋙ very much greater-than

2AF9 ⩽ VARIANT LESS-THAN OVER EQUAL TO
≈ <variant> 2266 ≦ less-than over equal to

2AFA ⩾ VARIANT GREATER-THAN OVER EQUAL TO
≈ <variant> 2267 ≧ greater-than over equal to

2AFB ⫻ TRIPLE SOLIDUS BINARY RELATION
= triple slash binary relation
→ 2AF4 ⫼ triple vertical bar binary relation

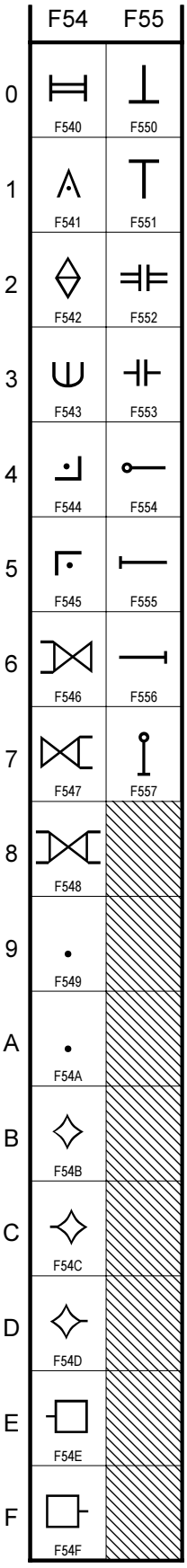
Operators

2AFC ⫼ LARGE TRIPLE VERTICAL BAR OPERATOR
often n-ary
→ 2AF4 ⫼ triple vertical bar binary relation
→ 2980 ⫼ triple vertical bar delimiter

2AFD ⫽ DOUBLE SLASH OPERATOR
= tangential to

2AFE ⫾ WHITE VERTICAL BAR
= Dijkstra choice

2AFF ⫿ N-ARY WHITE VERTICAL BAR
= N-ary Dijkstra choice


F54
F55
0
F540
F550
1
F541
F551
2
F542
F552
3
F543
F553
4
F544
F554
5
F545
F555
6
F546
F556
7
F547
F557
8
F548
9
F549
A
F54A
B
F54B
C
F54C
D
F54D
E
F54E
F
F54F

Relation

F540 ⧧ GLEICH STARK
= tautological equivalent

Operators

F541 ⩑ AND WITH DOT
F542 ⧫ LOZENGE DIVIDED BY HORIZONTAL RULE
F543 ⩏ ELEMENT OF OPENING UPWARDS
→ 2AD9 ⩍ element of opening downwards
F544 ⌟ LOWER RIGHT CORNER WITH DOT
= pullback
F545 ⌜ UPPER LEFT CORNER WITH DOT
= pushout
F546 ⟕ LEFT OUTER JOIN
F547 ⟖ RIGHT OUTER JOIN
F548 ⟗ FULL OUTER JOIN

Operator

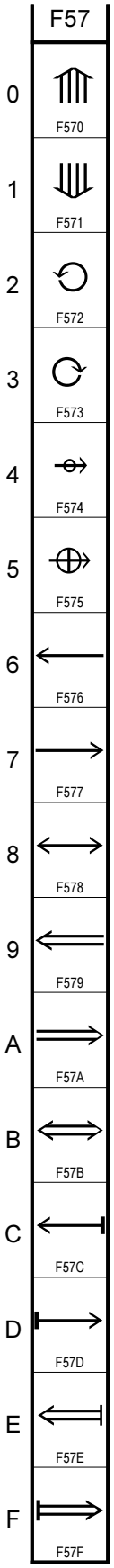
F549 . MULTIPLICATION ON-LINE
= often omitted
→ 2062 ⁢ invisible times
→ 22C5 ⋅ dot operator
→ 2024 . one dot leader
→ 002E . full stop
→ F54A . on-line dot

Fence

F54A . ON-LINE DOT
= paired, used as fence
→ F549 . multiplication on-line

Operators

F54B ✧ WHITE CONCAVE-SIDED DIAMOND
= never (modal operator)
F54C ✧ WHITE CONCAVE-SIDED DIAMOND WITH LEFTWARDS TICK
= was never (modal operator)
F54D ✧ WHITE CONCAVE-SIDED DIAMOND WITH RIGHTWARDS TICK
= will never be (modal operator)
F54E ⟡ WHITE SQUARE WITH LEFTWARDS TICK
= was always (modal operator)
F54F ⟡ WHITE SQUARE WITH RIGHTWARDS TICK
= will always be (modal operator)
F550 ⟘ LARGE UP TACK
≈ <large> 22A5 ⊥ up tack
F551 ⟙ LARGE DOWN TACK
≈ <large> 22A4 ⊤ down tack
F552 ⥡ LEFT AND RIGHT DOUBLE TURNSTILE
F553 ⥢ LEFT AND RIGHT TACK
F554 ⟜ LEFT MULTIMAP
→ 22B8 ⊸ multimap
F555 ⟝ LONG RIGHT TACK
= discrete Fourier transform
F556 ⟞ LONG LEFT TACK
F557 ⫯ UP TACK WITH CIRCLE ABOVE
= radial component
→ 2AF1 ⫱ down tack with circle below


F57
0
F570
1
F571
2
F572
3
F573
4
F574
5
F575
6
F576
7
F577
8
F578
9
F579
A
F57A
B
F57B
C
F57C
D
F57D
E
F57E
F
F57F

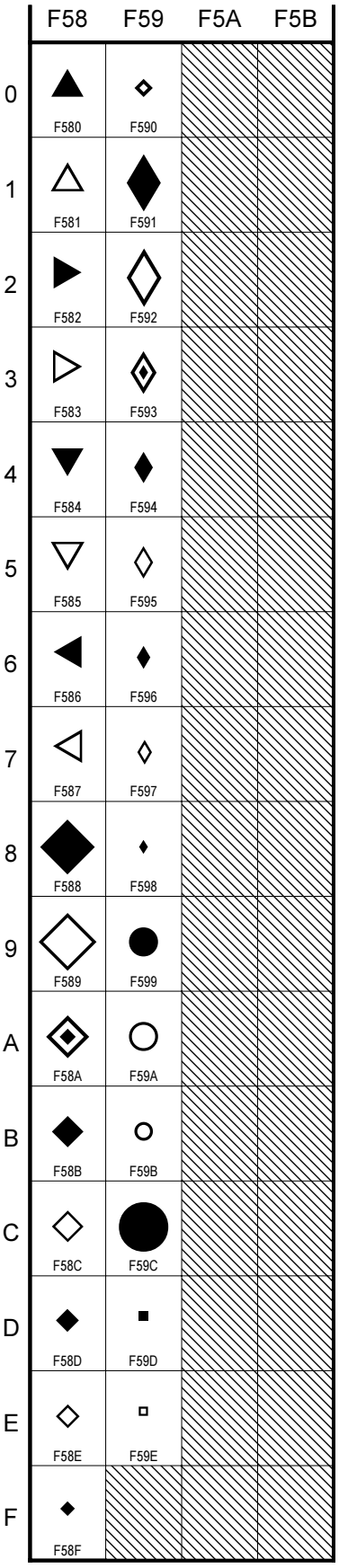
Arrows

F570 ⟰ UPWARDS QUADRUPLE ARROW
F571 ⟱ DOWNWARDS QUADRUPLE ARROW
F572 ⟲ ANTICLOCKWISE GAPPED CIRCLE ARROW
→ 21BA ↺ anticlockwise open circle arrow
F573 ⟳ CLOCKWISE GAPPED CIRCLE ARROW
→ 21BB ↻ clockwise open circle arrow
F574 ⟴ RIGHT ARROW WITH SMALL CIRCLE
F575 ⟴ RIGHT ARROW WITH CIRCLED PLUS

Long Arrows

These are used for mapping whereas the short forms would be used in limits. They are also needed for MathML to complete mapping to the ASMA sets.

F576 ⟵ LONG LEFTWARDS ARROW
F577 ⟶ LONG RIGHTWARDS ARROW
F578 ⟷ LONG LEFT RIGHT ARROW
F579 ⟸ LONG LEFTWARDS DOUBLE ARROW
F57A ⟹ LONG RIGHTWARDS DOUBLE ARROW
F57B ⟺ LONG LEFT RIGHT DOUBLE ARROW
F57C ⟻ LONG LEFTWARDS ARROW FROM BAR
= maps from
F57D ⟼ LONG RIGHTWARDS ARROW FROM BAR
= maps to
F57E ⟽ LONG LEFTWARDS DOUBLE ARROW FROM BAR
F57F ⟾ LONG RIGHTWARDS DOUBLE ARROW FROM BAR


F58
F59
F5A
F5B
0
F580
F590
1
F581
F591
2
F582
F592
3
F583
F593
4
F584
F594
5
F585
F595
6
F586
F596
7
F587
F597
8
F588
F598
9
F589
F599
A
F58A
F59A
B
F58B
F59B
C
F58C
F59C
D
F58D
F59D
E
F58E
F59E
F
F58F

Triangles

Triangles exist in regular and small sizes. An additional intermediate size is needed.

F580 ▲ BLACK UP-POINTING MEDIUM TRIANGLE
F581 △ WHITE UP-POINTING MEDIUM TRIANGLE
F582 ▶ BLACK RIGHT-POINTING MEDIUM TRIANGLE
F583 ▷ WHITE RIGHT-POINTING MEDIUM TRIANGLE
F584 ▼ BLACK DOWN-POINTING MEDIUM TRIANGLE
F585 ▽ WHITE DOWN-POINTING MEDIUM TRIANGLE
F586 ◀ BLACK LEFT-POINTING MEDIUM TRIANGLE
F587 ◁ WHITE LEFT-POINTING MEDIUM TRIANGLE

Diamonds

Diamonds are needed in several sizes, including one larger size. The sizes here match the sizes of the lozenges. Widely available fonts contain black diamonds in six sizes down to tiny.

F588 ◆ BLACK LARGE DIAMOND
F589 ◇ WHITE LARGE DIAMOND
F58A ◈ WHITE DIAMOND CONTAINING BLACK SMALL DIAMOND
F58B ◆ BLACK MEDIUM DIAMOND
F58C ◇ WHITE MEDIUM DIAMOND
F58D ◆ BLACK MEDIUM SMALL DIAMOND
F58E ◇ WHITE MEDIUM SMALL DIAMOND
F58F ◆ BLACK SMALL DIAMOND
F590 ◇ WHITE SMALL DIAMOND

Lozenges

Lozenges are needed in several sizes, including one larger size. The sizes here are chosen to interpolate between the two already coded sizes. Widely available fonts contain black diamonds in six sizes down to tiny.

F591 ⧫ BLACK LARGE LOZENGE
F592 ◊ WHITE LARGE LOZENGE
→ 25CA ◊ lozenge
→ 29EB ⧫ black lozenge
F593 ◈ WHITE LOZENGE CONTAINING BLACK SMALL LOZENGE
F594 ⧫ BLACK MEDIUM LOZENGE
F595 ◊ WHITE MEDIUM LOZENGE
F596 ⧫ BLACK MEDIUM SMALL LOZENGE
F597 ◊ WHITE MEDIUM SMALL LOZENGE
F598 ◆ BLACK SMALL LOZENGE
→ 25C4 ◄ diamond operator

Circles

Circles exist in many sizes, but there is one gap corresponding to medium size. The medium size here corresponds to the size of the inner circles in 25C9 and 25CE. Medium small and large circles only exist in one color.

F599 ● BLACK MEDIUM CIRCLE
F59A ○ WHITE MEDIUM CIRCLE
F59B ◦ WHITE MEDIUM SMALL CIRCLE
→ 2981 • z notation spot
F59C ⬤ LARGE BLACK CIRCLE
→ 25EF ◯ large circle

Squares

For the same size designation, the size of encoded squares are one step larger than those of other shapes. A very small size is needed to match up with the small sizes for lozenge, diamond, triangle and circle.

F59D ▪ BLACK VERY SMALL SQUARE
F59E ▫ WHITE VERY SMALL SQUARE