This PDF file is an excerpt from *The Unicode Standard, Version 4.0*, issued by the Unicode Consortium and published by Addison-Wesley. The material has been modified slightly for this online edition, however the PDF files have not been modified to reflect the corrections found on the Updates and Errata page (http://www.unicode.org/errata/). For information on more recent versions of the standard, see http://www.unicode.org/standard/versions/enumeratedversions.html.

Dai Kan-Wa Jiten used as the source of reference Kanji codes was written by Tetsuji Morohashi and published by Taishukan Shoten.

Cover and CD-ROM label design: Steve Mehallo, http://www.mehallo.com

The publisher offers discounts on this book when ordered in quantity for bulk purchases and special sales. For more information, customers in the U.S. please contact U.S. Corporate and Government Sales, (800) 382-3419, corpsales@pearsontechgroup.com. For sales outside of the U.S., please contact International Sales, +1 317 581 3793, international@pearsontechgroup.com

Visit Addison-Wesley on the Web: http://www.awprofessional.com

Library of Congress Cataloging-in-Publication Data
The Unicode Standard, Version 4.0 : the Unicode Consortium /Joan Aliprand... [et al.].
p. cm.
Includes bibliographical references and index.
ISBN 0-321-18578-1 (alk. paper)
1. Unicode (Computer character set). I. Aliprand, Joan.

QA268.U545 2004
005.7'2—dc21

2003052158

ISBN 0-321-18578-1
Text printed on recycled paper
1 2 3 4 5 6 7 8 9 10—CRW—0706050403
First printing, August 2003

Chapter 14

Symbols

The universe of symbols is rich and open-ended. The collection of encoded symbols in the Unicode Standard encompasses the following:

- Currency symbols
- Letterlike symbols
- Number forms
- Mathematical symbols
- Mathematical alphabets
- Technical symbols
- Geometrical symbols
- Miscellaneous symbols and dingbats
- Enclosed and square symbols
- Braille patterns
- Western and Byzantine musical symbols

There are other notational systems not covered by the Unicode Standard. Some symbols mark the transition between pictorial items and text elements; because they do not have a well-defined place in plain text, they are not encoded here.

Combining marks may be used with symbols, particularly the set encoded at U+20D0.. U+20FF (see *Section 7.7, Combining Marks*).

Letterlike and currency symbols, as well as number forms including superscripts and subscripts, are typically subject to the same font and style changes as the surrounding text. Some, but not all, of the square and enclosed symbols occur in East Asian contexts and generally follow the prevailing type styles.

Other symbols have an appearance that is independent of type style, or a more limited or altogether different range of type style variation than the regular text surrounding them. For example, mathematical alphanumeric symbols, typically used for mathematical variables, and those letterlike symbols that are part of this set carry semantic information in their type style. This fact restricts, but does not completely eliminate possible style variations. However, symbols such as mathematical operators can be used with any script or independent of any script.

In a bidirectional context (see Unicode Standard Annex #9, "The Bidirectional Algorithm"), symbol characters have no inherent directionality, but resolve according to the Unicode bidirectional algorithm. Where the image of a symbol is not bilaterally symmetric, the mirror image is used when the character is part of the right-to-left text stream (see *Section 4.7, Bidi Mirrored—Normative*).

Dingbats and optical character recognition characters are different from all other characters in the standard in that they are encoded based on their precise appearance.

Braille patterns are a special case, because they can be used to write text. They are included as symbols, as the Unicode Standard encodes only their shapes; the association of letters to patterns is left to other standards. When a character stream is intended primarily to convey text information, it should be coded using one of the scripts. Only when it is intended to convey a particular binding of text to Braille pattern sequence should it be coded using the Braille patterns.

Musical notation, and particularly Western musical notation, is different from ordinary text in the way it is laid out, particularly the representation of pitch and duration in Western musical notation. However, ordinary text commonly refers to the basic graphical elements that are used in musical notation, and it is primarily those symbols that are encoded in the Unicode Standard.

Many symbols encoded in the Unicode Standard are intended to support legacy implementations and obsolescent practices, such as terminal emulation or other character mode user interfaces. Examples include box drawing components and control pictures.

14.1 Currency Symbols

Currency Symbols: U+20A0–U+20CF

This block contains currency symbols not encoded in other blocks. Where the Unicode Standard follows the layout of an existing standard, such as for the ASCII, Latin-1, and Thai blocks, the currency symbols are encoded in those blocks, rather than here.

Unification. The Unicode Standard does not duplicate encodings where more than one currency is expressed with the same symbol. Many currency symbols are overstruck letters. There are therefore many minor variants, such as the U+0024 DOLLAR SIGN $, with one or two vertical bars, or other graphical variation. The Unicode Standard considers these variants to be typographical and provides a single encoding for them.

Claims that glyph variants of a certain currency symbol are used consistently to indicate a particular currency could not be substantiated upon further research. See ISO/IEC 10367, Annex B (informative), for an example of multiple renderings for U+00A3 POUND SIGN.

Lira Sign. A separate currency sign U+20A4 LIRA SIGN is encoded for compatibility with the HP Roman-8 character set, which is still widely implemented in printers. However, in general, U+00A3 POUND SIGN should be used for both the various currencies known as pound (or punt), as well as the various currencies known as lira—for example, the former currency of Italy and the lira still in use in Turkey. Widespread implementation practice in Italian and Turkish systems has long made use of U+00A3 as the currency sign for the lira. As in the case of the dollar sign, the glyphic distinction between single- and double-bar versions of the sign is not indicative of a systematic difference in the currency.

Yen and Yuan. Like the dollar sign and the pound sign, U+00A5 YEN SIGN has been used as the currency sign for more than one currency. While there may be some preferences to use a double-bar glyph for the yen currency of Japan (JPY) and a single-bar glyph for the yuan (renminbi) currency of China (CNY), this distinction is not systematic in all font designs, and there is considerable overlap in usage. There are also a number of CJK characters to represent the words *yen* (or *en*) and *yuan*, and these also tend to overlap in use. In the Unicode Standard, U+00A5 YEN SIGN is intended to be the character for the currency sign for both the yen and yuan, with details of glyphic presentation left to font choice and local preferences.

In general, implementations that are concerned with the *exact* identity of a currency should not depend on an encoded currency sign character, because currency signs vary in shape and are often used for more than one currency. Instead, such implementations should make use of such standards as the ISO 4217 three-letter currency codes, which are *specific* to currencies—for example, USD for U.S. dollar, CAD for Canadian dollar.

Euro Sign. The single currency for member countries of the European Economic and Monetary Union is the euro (EUR). The euro character is encoded in the Unicode Standard as U+20AC EURO SIGN.

Fonts. Currency symbols are commonly designed to display at the same width as a digit (most often a European digit, U+0030..U+0039) to assist in alignment of monetary values in tabular displays. Like letters, they tend to follow the stylistic design features of particular fonts because they are used often and need to harmonize with body text. In particular, even though there may be more or less normative designs for the currency sign per se, as for the euro sign, type designers freely adapt such designs to make them fit the logic of the rest of their fonts. This partly explains why currency signs show more glyph variation than other types of symbols.

Table 14-1 lists common currency symbols encoded in other blocks.

Table 14-1. Other Currency Symbols

Dollar, milreis, escudo, peso	U+0024	DOLLAR SIGN
Cent	U+00A2	CENT SIGN
Pound and lira	U+00A3	POUND SIGN
General currency	U+00A4	CURRENCY SIGN
Yen or yuan	U+00A5	YEN SIGN
Dutch florin	U+0192	LATIN SMALL LETTER F WITH HOOK
Rupee	U+09F2	BENGALI RUPEE MARK
Rupee	U+09F3	BENGALI RUPEE SIGN
Rupee	U+0AF1	GUJARATI RUPEE SIGN
Rupee	U+0BF9	TAMIL RUPEE SIGN
Baht	U+0E3F	THAI CURRENCY SYMBOL BAHT
Riel	U+17DB	KHMER CURRENCY SYMBOL RIEL
(Old) German mark	U+2133	SCRIPT CAPITAL M
Rial	U+FDFC	RIAL SIGN

For additional forms of currency symbols, see Fullwidth Forms (U+FFE0..U+FFE6).

14.2 Letterlike Symbols

Letterlike Symbols: U+2100–U+214F

Letterlike symbols are symbols derived in some way from ordinary letters of an alphabetic script. This block includes three types of symbols based on Latin, Greek, and Hebrew letters. Stylistic variations of single letters are used for semantics in mathematical notation. See "Mathematical Alphanumeric Symbols" in this section for the use of letterlike symbols in mathematical formulas. Some letterforms have given rise to specialized symbols, such as U+211F PRESCRIPTION TAKE.

Numero Sign. U+2116 NUMERO SIGN is provided both for Cyrillic use, where it looks like №, and for compatibility with Asian standards, where it looks like №. The French practice is not to use the symbol character per se, but rather to use an "N" or an "n", according to context, followed by a superscript o (N^{o} or n^{o}; plural N^{os} or n^{os}). Legacy data encoded in ISO/IEC 8859-1 (Latin-1) or other 8-bit character sets may also have represented the *numero sign* by a sequence of "N" followed by the *degree sign* (U+00B0 DEGREE SIGN). Implementations interworking with legacy data should be aware of such alternative representations for the *numero sign* when converting data.

Unit Symbols. Several letterlike symbols are used to indicate units. In most cases, however, such as for SI units (Système International), the use of regular letters or other symbols is preferred. U+2113 SCRIPT SMALL L is commonly used as a non-SI symbol for the *liter*. Official SI usage prefers the regular *lowercase letter l*.

Three letterlike symbols have been given canonical equivalence to regular letters: U+2126 OHM SIGN, U+211A KELVIN SIGN, and U+211B ANGSTROM SIGN. In all three instances the regular letter should be used. In normal use, it is better to represent degrees Celsius "°C" with a sequence of U+00B0 DEGREE SIGN + U+0043 LATIN CAPITAL LETTER C, rather than U+2103 DEGREE CELSIUS. For searching, treat these two sequences as identical. Similarly, the sequence U+00B0 DEGREE SIGN + U+0046 LATIN CAPITAL LETTER F is preferred over U+2109 DEGREE FAHRENHEIT, and those two sequences should be treated as identical for searching.

Compatibility. Some symbols are composites of several letters. Many of these composite symbols are encoded for compatibility with Asian and other legacy encodings. (See also "CJK Compatibility Ideographs" in *Section 11.1, Han*.) The use of these composite symbols is discouraged where their presence is not required by compatibility. For example, in normal use, the symbols U+2121 TEL TELEPHONE SIGN and U+213B FAX FACSIMILE SIGN are simply spelled out.

In the context of East Asian typography, many letterlike symbols, and in particular composites, form part of a collection of compatibility symbols, the larger part of which is located in the CJK Compatibility block (see *Section 14.8, Enclosed and Square*). When used in this way, these symbols are rendered as "wide" characters occupying a full cell. They remain upright in vertical layout, contrary to the rotated rendering of their regular letter equivalents. See Unicode Standard Annex #11, "East Asian Width," for more information.

Where the letterlike symbols have alphabetic equivalents, they collate in alphabetic sequence; otherwise, they should be treated as neutral symbols. The letterlike symbols may have different directional properties than normal letters. For example, the four transfinite cardinal symbols (U+2135..U+2138) are used in ordinary mathematical text and do not share the strong right-to-left directionality of the Hebrew letters from which they are derived.

Styles. The letterlike symbols include some of the few instances in which the Unicode Standard encodes stylistic variants of letters as distinct characters. For example, there are instances of blackletter (*Fraktur*), double-struck, italic, and script styles for certain Latin letters used as mathematical symbols. The choice of these stylistic variants for encoding reflects their common use as distinct symbols. They form part of the larger set of mathematical alphanumeric symbols. For the complete set and more information on its use, see "Mathematical Alphanumeric Symbols" in this section. These symbols should not be used in ordinary, nonscientific texts.

Despite its name, U+2118 SCRIPT CAPITAL P is neither script nor capital—it is uniquely the Weierstrass elliptic function symbol derived from a calligraphic *lowercase* p. U+2113 SCRIPT SMALL L is derived from a special *italic* form of the *lowercase letter l* and, when it occurs in mathematical notation, is known as the symbol *ell*. Use U+1D4C1 MATHEMATICAL SMALL SCRIPT L as the *lowercase script l* for mathematical notation.

Standards. The Unicode Standard encodes letterlike symbols from many different national standards and corporate collections.

Math Alphanumeric Symbols: U+1D400–U+1D7FF

The Mathematical Alphanumeric Symbols block contains a large extension of letterlike symbols used in mathematical notation, typically for variables. The characters in this block are intended for use only in mathematical or technical notation; they are not intended for use in nontechnical text. When used with markup languages—for example, with Mathematical Markup Language (MathML)—the characters are expected to be used directly, instead of indirectly via entity references or by composing them from base letters and style markup.

Words Used as Variables. In some specialties, whole words are used as variables, not just single letters. For these cases, style markup is preferred because in ordinary mathematical notation the juxtaposition of variables generally implies multiplication, not word formation as in ordinary text. Markup not only provides the necessary scoping in these cases, but also allows the use of a more extended alphabet.

Mathematical Alphabets

Basic Set of Alphanumeric Characters. Mathematical notation uses a basic set of mathematical alphanumeric characters, which consists of the following:

- The set of basic Latin digits (0–9) (U+0030..U+0039)
- The set of basic upper- and lowercase Latin letters (a– z, A–Z)
- The uppercase Greek letters Α–Ω (U+0391..U+03A9), plus the nabla ∇ (U+2207) and the variant of theta ϴ given by U+03F4
- The lowercase Greek letters α–ω (U+03B1..U+03C9), plus the partial differential sign ∂ (U+2202) and the six glyph variants ϵ, ϑ, ϰ, ϕ, ϱ, and ϖ, given by U+03F5, U+03D1, U+03F0, U+03D5, U+03F1, and U+03D6, respectively

Only unaccented forms of the letters are used for mathematical notation, because general accents such as the acute accent would interfere with common mathematical diacritics. Examples of common mathematical diacritics that can interfere with general accents are the circumflex, macron, or the single or double dot above, the latter two of which are used in physics to denote derivatives with respect to the time variable. Mathematical symbols with diacritics are always represented by combining character sequences.

For some characters in the basic set of Greek characters, two variants of the same character are included. This is because they can appear in the same mathematical document with different meanings, even though they would have the same meaning in Greek text. (See "Variant Letterforms" in *Section 7.2, Greek*.)

Additional Characters. In addition to this basic set, mathematical notation uses the four Hebrew-derived characters (U+2135..U+2138). Occasional uses of other alphabetic and numeric characters are known. Examples include U+0428 CYRILLIC CAPITAL LETTER SHA, U+306E HIRAGANA LETTER NO, and Eastern Arabic-Indic digits (U+06F0..U+06F9). However, these characters are used only in their basic forms, rather than in multiple mathematical styles.

Semantic Distinctions. Mathematical notation requires a number of Latin and Greek alphabets that initially appear to be mere font variations of one another. For example, the letter H can appear as plain or upright (H), bold (**H**), italic (*H*), and script. However, in any given document, these characters have distinct, and usually unrelated, mathematical semantics. For example, a normal H represents a different variable from a bold **H**, and so on. If these attributes are dropped in plain text, the distinctions are lost and the meaning of the text is altered. Without the distinctions, the well-known Hamiltonian formula turns into the *integral* equation in the variable H as shown below:

Hamiltonian formula: $\mathcal{H} = \int d\tau(\epsilon E^2 + \mu H^2)$

Integral equation: $H = \int d\tau(\varepsilon E^2 + \mu H^2)$

By encoding a separate set of alphabets, it is possible to preserve such distinctions in plain text.

Mathematical Alphabets. The alphanumeric symbols encountered in mathematics and encoded in the Unicode Standard are given in *Table 14-2*.

Table 14-2. Mathematical Alphanumeric Symbols

Math Style	Characters from Basic Set	Location
plain (upright, serifed)	Latin, Greek, and digits	BMP
bold	Latin, Greek, and digits	Plane 1
italic	Latin and Greek	Plane 1[a]
bold italic	Latin and Greek	Plane 1
script (calligraphic)	Latin	Plane 1[a]
bold script (calligraphic)	Latin	Plane 1
Fraktur	Latin	Plane 1[a]
bold Fraktur	Latin	Plane 1
double-struck	Latin and digits	Plane 1[a]
sans-serif	Latin and digits	Plane 1
sans-serif bold	Latin, Greek, and digits	Plane 1
sans-serif italic	Latin	Plane 1
sans-serif bold italic	Latin and Greek	Plane 1
monospace	Latin and digits	Plane 1

[a] Some of these alphabets have characters in the BMP as noted in the text that follows.

The plain letters have been unified with the existing characters in the Basic Latin and Greek blocks. There are 25 double-struck, italic, Fraktur, and script characters that already exist in the Letterlike Symbols block (U+2100..U+214F). These are explicitly unified with the characters in this block and corresponding holes have been left in the mathematical alphabets.

The alphabets in this block encode only semantic distinction, but not which specific font will be used to supply the actual plain, script, Fraktur, double-struck, sans-serif, or monospace glyphs. Especially the script and double-struck styles can show considerable variation across fonts. Characters from the Mathematical Alphanumeric Symbols block are not to be used for nonmathematical styled text.

Compatibility Decompositions. All mathematical alphanumeric symbols have compatibility decompositions to the base Latin and Greek letters—folding away such distinctions is usually not desirable, however, as it loses the semantic distinctions for which these characters were encoded. See Unicode Standard Annex #15, "Unicode Normalization Forms."

Fonts Used for Mathematical Alphabets

Mathematicians place strict requirements on the *specific* fonts being used to represent mathematical variables. Readers of a mathematical text need to be able to distinguish single-letter variables from each other, even when they do not appear in close proximity. They must be able to recognize the letter itself, whether it is part of the text or is a mathematical variable, and lastly which mathematical alphabet it is from.

Fraktur. The blackletter style is often referred to as *Fraktur* or *Gothic* in various sources. Technically, Fraktur and Gothic typefaces are distinct designs from blackletter, but any of several font styles similar in appearance to the forms shown in the charts can be used. Note that in East Asian typography, the term *Gothic* is commonly used to indicate a sans-serif type style.

Math Italics. Mathematical variables are most commonly set in a form of italics, but not all italic fonts can be used successfully. For example, a math italic font should avoid a "tail" on the lowercase *italic letter z* because it clashes with subscripts. In common text fonts, the *italic letter v* and *Greek letter nu* are not very distinct. A rounded *italic letter v* is therefore preferred in a mathematical font. There are other characters that sometimes have similar shapes and require special attention to avoid ambiguity. Examples are shown in *Figure 14-1*.

Figure 14-1. Easily Confused Shapes for Mathematical Glyphs

italic a	a	α	alpha
italic v (pointed)	v	ν	nu
italic v (rounded)	υ	υ	upsilon
script X	$\mathcal{X}$	χ	chi
plain Y	Y	Υ	Upsilon

Hard-to-Distinguish Letters. Not all sans-serif fonts allow an easy distinction between *lowercase l* and *uppercase I*, and not all monospaced (monowidth) fonts allow a distinction between the *letter l* and the *digit one*. Such fonts are not usable for mathematics. In Fraktur, the letters $\mathfrak{I}$ and $\mathfrak{J}$, in particular, must be made distinguishable. Overburdened blackletter forms are inappropriate for mathematical notation. Similarly, the *digit zero* must be distinct from the *uppercase letter O* for all mathematical alphanumeric sets. Some characters are so similar that even mathematical fonts do not attempt to provide distinct glyphs for them. Their use is normally avoided in mathematical notation unless no confusion is possible in a given context—for example, *uppercase A* and *uppercase Alpha*.

Font Support for Combining Diacritics. Mathematical equations require that characters be combined with diacritics (dots, tilde, circumflex, or arrows above are common), as well as followed or preceded by super- or subscripted letters or numbers. This requirement leads to designs for *italic* styles that are less inclined, and *script* styles that have smaller overhangs and less slant than equivalent styles commonly used for text such as wedding invitations.

Type Style for Script Characters. In some instances, a deliberate unification with a non-mathematical symbol has been undertaken; for example, U+2133 is unified with the pre-1949 symbol for the German currency unit *Mark*. This unification restricts the range of glyphs that can be used for this character in the charts. Therefore the font used for the representative glyphs in the code charts uses a simplified "English Script" style, as per recommendation by the American Mathematical Society. For consistency, other script characters in the Letterlike Symbols block are now shown in the same type style.

Double-Struck Characters. The double-struck glyphs shown in earlier editions of the standard attempted to match the design used for all the other Latin characters in the standard, which is based on Times. The current set of fonts was prepared in consultation with the American Mathematical Society and leading mathematical publishers, and shows much simpler forms that are derived from the forms written on a blackboard. However, both serifed and non-serifed forms can be used in mathematical texts, and inline fonts are found in works published by certain publishers.

14.3 Number Forms

Number Forms: U+2150–U+218F

Number form characters are encoded solely for compatibility with existing standards. The same considerations with respect to compatibility apply as noted in the discussion of letterlike symbols.

Fractions. The Number Forms block contains a series of vulgar fraction characters, encoded for compatibility with legacy character encoding standards. These characters are intended to represent both of the common forms of vulgar fractions: forms with a right-slanted division slash, such as ¾, as shown in the code charts, and forms with a horizontal division line, such as $\frac{3}{4}$, which are considered to be alternative glyphs for the same fractions. A few other vulgar fraction characters are located in the Latin-1 block in the range U+00BC..U+00BE.

The vulgar fraction characters are given compatibility decompositions using U+2044 "/" FRACTION SLASH. Use of the *fraction slash* is the more generic way to represent fractions in text; it can be used to construct fractional number forms that are not included in the collections of vulgar fraction characters. For more information on the *fraction slash*, see "Other Punctuation" in *Section 6.2, General Punctuation*.

Roman Numerals. The Roman numerals can be composed of sequences of the appropriate Latin letters. Upper- and lowercase variants of the Roman numerals through 12, plus L, C, D, and M, have been encoded for compatibility with East Asian standards.

U+2180 ROMAN NUMERAL ONE THOUSAND C D and U+216F ROMAN NUMERAL ONE THOUSAND can be considered to be glyphic variants of the same Roman numeral, but are distinguished because they are not generally interchangeable, and because U+2180 cannot be considered to be a compatibility equivalent to the Latin letter M. U+2181 ROMAN NUMERAL FIVE THOUSAND and U+2182 ROMAN NUMERAL TEN THOUSAND are distinct characters used in Roman numerals; they do not have compatibility decompositions in the Unicode Standard. U+2183 ROMAN NUMERAL REVERSED ONE HUNDRED is a form used in combinations with C and/or I to form large numbers—some of which vary with single character number forms such as D, M, U+2181, or others.

Suzhou-Style Numerals. The Suzhou-style numerals (Mandarin *su1zhou1ma3zi*) are CJK ideographic number forms encoded in the CJK Symbols and Punctuation block in the ranges U+3021..U+3029 and U+3038..U+303A.

The Suzhou-style numerals are modified forms of CJK ideographic numerals that are used by shopkeepers in China to mark prices. They are also known as "commercial forms," "shop units," or "grass numbers." They are encoded for compatibility with the CNS 11643-1992 and Big Five standards. The forms for ten, twenty, and thirty, encoded at U+3038..U+303A, are also encoded as CJK unified ideographs: U+5341, U+5344, and U+5345, respectively. (For twenty, see also U+5EFE and U+5EFF.)

These commercial forms of Chinese numerals should be distinguished from the use of other CJK unified ideographs as accounting numbers to deter fraud. See *Table 4-4* in *Section 4.6, Numeric Value—Normative*, for a list of ideographs used as accounting numbers.

Why are the Suzhou numbers called Hangzhou numerals in the Unicode names? No one has been able to trace this back. Hangzhou is a district in China that is near the Suzhou district, but the name "Hangzhou" does not occur in other sources that discuss these number forms.

Superscripts and Subscripts: U+2070–U+209F

In general, the Unicode Standard does not attempt to describe the positioning of a character above or below the baseline in typographical layout. Therefore, the preferred means to encode superscripted letters or digits, such as "1^{st}" or "$DC00_{16}$", is by style or markup in rich text. However, in some instances superscript or subscript letters are used as part of the plain text content of specialized phonetic alphabets, such as the Uralic Phonetic Alphabet. These superscript and subscript letters are mostly from the Latin or Greek scripts. These characters are encoded in other character blocks, along with other modifier letters or phonetic letters. In addition, superscript digits are used to indicate tone in transliteration of many languages. The use of *superscript two* and *superscript three* is common legacy practice when referring to units of area and volume in general texts.

A certain number of additional superscript and subscript characters are needed for round-trip conversions to other standards and legacy code pages. Most such characters are encoded in this block and are considered compatibility characters. In the Unicode Character Database, superscript and subscript digits have not been given the General Category property of decimal digit (Nd), so as to prevent expressions like 2^3 from being treated like 23 by simplistic parsers.

Standards. Many of the characters in the Superscripts and Subscripts block are from character sets registered in the ISO International Register of Coded Character Sets to be Used With Escape Sequences, under the registration standard ISO/IEC 2375, for use with ISO/IEC 2022. Two MARC 21 character sets used by libraries include the digits, plus signs, minus signs, and parentheses.

Superscripts and Subscripts in Other Blocks. The superscript digits one, two, and three are coded in the Latin-1 Supplement block to provide code point compatibility with ISO/IEC 8859-1. For a discussion of U+00AA FEMININE ORDINAL INDICATOR and U+00BA MASCULINE ORDINAL INDICATOR, see "Letters of the Latin-1 Supplement" in *Section 7.1, Latin*. U+2120 SERVICE MARK and U+2122 TRADE MARK SIGN are commonly used symbols, encoded in the Letterlike Symbols block (U+2100..U+214F); they consist of sequences of two superscripted letters each.

For phonetic usage, there are a small number of superscript letters located in the Spacing Modifier Letters block (U+02B0..U+02FF) and a large number of superscript or subscript letters in the Phonetic Extensions block (U+1D00..U+1D7F). The superscripted letters do not contain the word "superscript" in their character names, but are simply called modifier letters. Finally, a small set of superscripted CJK ideographs, used for the Japanese system of syntactic markup of Classical Chinese text for reading, is located in the Kanbun block (U+3190..U+319F).

14.4 Mathematical Symbols

Mathematical Operators: U+2200–U+22FF

The Mathematical Operators block includes character encodings for operators, relations, geometric symbols, and a few other symbols with special usages confined largely to mathematical contexts.

In addition to the characters in this block, mathematical operators are found in the Basic Latin (ASCII) and Latin-1 Supplement blocks. A few of the symbols from the Miscellaneous Technical block and characters from General Punctuation are also used in mathematical notation. For Latin and Greek letters in special font styles that are used as mathematical variables, such as U+210B $\mathscr{H}$ SCRIPT CAPITAL H , as well as the Hebrew letter *alef* used as the first transfinite cardinal symbol encoded by U+2135 ℵ ALEF SYMBOL, see "Letterlike Symbols" and "Mathematical Alphanumeric Symbols" in *Section 14.2, Letterlike Symbols*.

Standards. Many national standards' mathematical operators are covered by the characters encoded in this block. These standards include such special collections as ANSI Y10.20, ISO 6862, ISO 8879, and portions of the collection of the American Mathematical Society, as well as the original repertoire of TEX.

Encoding Principles. Mathematical operators often have more than one meaning. Therefore the encoding of this block is intentionally rather shape-based, with numerous instances in which several semantic values can be attributed to the same Unicode code point. For example, U+2218 ∘ RING OPERATOR may be the equivalent of *white small circle* or *composite function* or *apl jot*. The Unicode Standard does not attempt to distinguish all possible semantic values that may be applied to mathematical operators or relation symbols.

The Unicode Standard does include many characters that appear to be quite similar to one another, but that may well convey different meanings in a given context. On the other hand, mathematical operators, and especially relation symbols, may appear in various standards, handbooks, and fonts with a large number of purely graphical variants. Where variants were recognizable as such from the sources, they were not encoded separately. For relation symbols, the choice of a vertical or forward-slanting stroke typically seems to be an aesthetic one, but both slants might appear in a given context. However, a back-slanted stroke has almost always a distinct meaning compared to the forward-slanted stroke. See *Section 15.6, Variation Selectors*, for more information on some particular variants.

Unifications. Mathematical operators such as *implies* ⇒ and *if and only if* ↔ have been unified with the corresponding arrows (U+21D2 RIGHTWARDS DOUBLE ARROW and U+2194 LEFT RIGHT ARROW, respectively) in the Arrows block.

The operator U+2208 ELEMENT OF is occasionally rendered with a taller shape than shown in the code charts. Mathematical handbooks and standards consulted treat these characters as variants of the same glyph. U+220A SMALL ELEMENT OF is a distinctively small version of the *element of* that originates in mathematical pi fonts.

The operators U+226B MUCH GREATER-THAN and U+226A MUCH LESS-THAN are sometimes rendered in a nested shape. The nested shapes are encoded separately as U+2AA2 DOUBLE NESTED GREATER-THAN and U+2AA1 DOUBLE NESTED LESS-THAN.

A large class of unifications applies to variants of relation symbols involving negation. Variants involving vertical or slanted *negation slashes* and *negation slashes* of different lengths

are not separately encoded. Thus, for example, U+2288 NEITHER A SUBSET OF NOR EQUAL TO is the archetype for several different glyph variants noted in various collections.

In two instances in this block, essentially stylistic variants are separately encoded: U+2265 GREATER-THAN OR EQUAL TO is distinguished from U+2267 GREATER-THAN OVER EQUAL TO; the same distinction applies to U+2264 LESS-THAN OR EQUAL TO and U+2266 LESS-THAN OVER EQUAL TO. Further instances of the encoding of such stylistic variants can be found in the supplemental blocks of mathematical operators.

Greek-Derived Symbols. Several mathematical operators derived from Greek characters have been given separate encodings because they are used differently from the corresponding letters. These operators may occasionally occur in context with Greek-letter variables. They include U+2206 Δ INCREMENT , U+220F ∏ N-ARY PRODUCT, and U+2211 ∑ N-ARY SUMMATION. The latter two are large operators that take limits.

Other duplicated Greek characters are those for U+00B5 µ MICRO SIGN in the Latin-1 Supplement block, U+2126 Ω OHM SIGN in Letterlike Symbols, and several characters among the APL functional symbols in the Miscellaneous Technical block. Most other Greek characters with special mathematical semantics are found in the Greek block because duplicates were not required for compatibility. Additional sets of mathematical-style Greek alphabets are found in the Mathematical Alphanumeric Symbols block.

N-ary Operators. N-ary operators are distinguished from binary operators by their larger size and the fact that in mathematical layout, they take limit expressions.

Invisible Operators. In mathematics some operators or punctuation are often implied but not displayed. For a set of invisible operators that can be used to mark these implied operators in the text, see *Section 15.3, Invisible Operators*.

Minus Sign. U+2212 "−" MINUS SIGN is a mathematical operator, to be distinguished from the ASCII-derived U+002D "-" HYPHEN-MINUS, which may look the same as a minus sign, or may be shorter in length. (For a complete list of dashes in the Unicode Standard, see *Table 6-3*.) U+22EE..U+22F1 are a set of ellipses used in matrix notation. U+2052 "⁒" COMMERCIAL MINUS SIGN is a specialized form of the minus sign. Its use is described in *Section 6.2, General Punctuation*.

Mathematical Property. The mathematical (*math*) property is an informative property of characters that are used as operators in mathematical formulas. The mathematical property may be useful in identifying characters commonly used in mathematical text and formulas. However, a number of these characters have multiple usages and may occur with nonmathematical semantics. For example, U+002D HYPHEN-MINUS may also be used as a hyphen—and not as a mathematical minus sign. Other characters, including some alphabetic, numeric, punctuation, spaces, arrows, and geometric shapes, are used in mathematical expressions as well, but are even more dependent on the context for their identification. A list of characters with the mathematical property is provided in the Unicode Character Database.

In addition to the symbols in these blocks, mathematical and scientific notation makes frequent use of arrows, punctuation characters, letterlike symbols, geometrical shapes, and other miscellaneous and technical symbols.

For an extensive discussion of mathematical alphanumeric symbols, see *Section 14.2, Letterlike Symbols*. For additional information on all the mathematical operators and other symbols, see Unicode Technical Report #25, "Unicode Support for Mathematics."

Supplements to Mathematical Symbols and Arrows

The Unicode Standard defines a number of additional blocks to supplement the repertoire of mathematical operators and arrows. These additions are intended to extend the Unicode repertoire sufficiently to cover the needs of such applications as MathML, modern mathematical formula editing and presentation software, and symbolic algebra systems.

Standards. MathML, an XML application, is intended to support the full legacy collection of the ISO mathematical entity sets. Accordingly, the repertoire of mathematical symbols for the Unicode Standard has been supplemented by the full list of mathematical entity sets in ISO TR 9573-13, *Public entity sets for mathematics and science*. An additional repertoire was provided from the amalgamated collection of the STIX Project (Scientific and Technical Information Exchange). That collection includes, but is not limited to, symbols gleaned from mathematical publications by experts of the American Mathematical Society and symbol sets provided by Elsevier Publishing and by the American Physical Society.

Semantics. The same mathematical symbol may have different meanings in different subdisciplines or different contexts. The Unicode Standard encodes only a single character for a single symbolic form. For example, the "+" symbol normally denotes addition in a mathematical context, but might refer to concatenation in a computer science context dealing with strings, or incrementation, or have any number of other functions in given contexts. It is up to the application to distinguish such meanings according to the appropriate context. Where information is available about the usage (or usages) of particular symbols, it has been indicated in the character annotations in *Chapter 16, Code Charts*.

Supplemental Math Operators: U+2A00–U+2AFF

The Supplemental Mathematical Operators block contains many additional symbols to supplement the collection of mathematical operators.

Miscellaneous Math Symbols-A: U+27C0–U+27EF

The Miscellaneous Mathematical Symbols-A block contains symbols used mostly as operators or delimiters in mathematical notation.

Mathematical Brackets. The mathematical white square brackets, angle brackets, and double angle brackets encoded at U+27E6..U+27EB are intended for ordinary mathematical use of these particular bracket types. They are unambiguously narrow, for use in mathematical and scientific notation, and should be distinguished from the corresponding wide forms of white square brackets, angle brackets, and double angle brackets used in CJK typography. (See the discussion of the CJK Symbols and Punctuation block in *Section 6.2, General Punctuation*.) Note especially that the "bra" and "ket" angle brackets, U+2329 LEFT-POINTING ANGLE BRACKET and U+232A RIGHT-POINTING ANGLE BRACKET, are now deprecated for use with mathematics because of their canonical equivalence to CJK angle brackets, which is likely to result in unintended spacing problems if used in mathematical formulae.

Miscellaneous Math Symbols-B: U+2980–U+29FF

The Miscellaneous Mathematical Symbols-B block contains miscellaneous symbols used for mathematical notation, including fences and other delimiters. Some of the symbols in this block may also be used as operators in some contexts.

Wiggly Fence. U+29DB LEFT WIGGLY FENCE has a superficial similarity to U+FE34 PRESENTATION FORM FOR VERTICAL LOW LINE. The latter is a wiggly sidebar character, intended for legacy support as a style of underlining character in a vertical text layout context; it has a compatibility mapping to U+005F LOW LINE. This represents a very different usage from the standard use of fence characters in mathematical notation.

Arrows: U+2190–U+21FF

Arrows are used for a variety of purposes: to imply directional relation, to show logical derivation or implication, and to represent the cursor control keys.

Accordingly, the Unicode Standard includes a fairly extensive set of generic arrow shapes, especially those for which there are established usages with well-defined semantics. It does not attempt to encode every possible stylistic variant of arrows separately, especially where their use is mainly decorative. For most arrow variants, the Unicode Standard provides encodings in the two horizontal directions, often in the four cardinal directions. For the single and double arrows, the Unicode Standard provides encodings in eight directions.

Standards. The Unicode Standard encodes arrows from many different international and national standards, as well as corporate collections.

Unifications. Arrows expressing mathematical relations have been encoded in the Arrows block, as well as in the Supplemental Arrows-A and the Supplemental Arrows-B blocks. An example is U+21D2 ⇒ RIGHTWARDS DOUBLE ARROW, which may be used to denote *implies*. Where available, such usage information is indicated in the annotations to individual characters in *Chapter 16, Code Charts*. However, because the arrows have such a wide variety of applications, there may be several semantic values for the same Unicode character value.

Supplemental Arrows

The Supplemental Arrows-A (U+27F0..U+27FF), Supplemental Arrows-B (U+2900..U+297F), and Miscellaneous Symbols and Arrows (U+2B00..U+2BFF) blocks contain a large repertoire of arrows to supplement the main set in the Arrows block.

Long Arrows. The long arrows encoded in the range U+27F5..U+27FF map to standard SGML entity sets supported by MathML. Long arrows represent distinct semantics from their short counterparts, rather than mere stylistic glyph differences. For example, the shorter forms of arrows are often used in connection with limits, whereas the longer ones are associated with mappings. The use of the long arrows is so common that they were assigned entity names in the ISOAMSA entity set, one of the suite of mathematical symbol entity sets covered by the Unicode Standard.

Standardized Variants of Mathematical Symbols

These mathematical variants are all produced with the addition of U+FE00 VARIATION SELECTOR-1 (VS1) to mathematical operator base characters. The valid combinations are listed in the file StandardizedVariants.txt in the Unicode Character Database. All combinations not listed there are unspecified and are reserved for future standardization; no conformant process may interpret them as standardized variants.

Change in Representative Glyphs for U+2278 and U+2279. In Version 3.2 of the Unicode Standard, the representative glyphs for U+2278 NEITHER LESS-THAN NOR GREATER-THAN

and U+2279 NEITHER GREATER-THAN NOR LESS-THAN were changed from using a vertical cancellation to using a slanted cancellation. This change was made to match the long-standing canonical decompositions for these characters, which use U+0338 COMBINING LONG SOLIDUS OVERLAY. The symmetric forms using the vertical stroke continue to be acceptable glyph variants. Using U+2278 or U+2279 with VS1 will request these variants explicitly, as will using U+2276 LESS-THAN OR GREATER-THAN or U+2277 GREATER-THAN OR LESS-THAN with U+20D2 COMBINING LONG VERTICAL LINE OVERLAY. Unless fonts are created with the intention to add support for both forms (via VS1 for the upright forms), there is no need to revise the glyphs in existing fonts; the glyphic range implied by using the base character code alone encompasses both shapes. For more information, see *Section 15.6, Variation Selectors*.

14.5 Technical Symbols

Control Pictures: U+2400–U+243F

The need to show the presence of the C0 control codes and the SPACE unequivocally when data are displayed has led to conventional representations for these nongraphic characters.

Code Points for Pictures for Control Codes. By definition, control codes themselves are manifested only by their action. However, it is sometimes necessary to show the position of a control code within a data stream. Conventional illustrations for the ASCII C0 control codes have been developed—but the characters U+2400..U+241F and U+2424 are intended for use as unspecified graphics for the corresponding control codes. This choice allows a particular application to use *any* desired pictorial representation of the given control code. It assumes that the particular pictures used to represent control codes are often specific to different systems, and are rarely the subject of text interchange between systems.

Pictures for ASCII Space. By definition, the SPACE is a blank graphic. Conventions have also been established for the visible representation of the space. Two specific characters are provided that may be used to visually represent the ASCII space character, U+2420 SYMBOL FOR SPACE and U+2422 BLANK SYMBOL.

Standards. The CNS 11643 standard encodes characters for pictures of control codes. Standard representations for control characters have been defined—for example, in ANSI X3.32 and ISO 2047. If desired, characters U+2400..U+241F may be used for these representations.

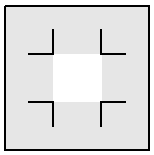
Miscellaneous Technical: U+2300–U+23FF

This block encodes technical symbols, including keytop labels such as U+232B ERASE TO THE LEFT. Excluded from consideration were symbols that are not normally used in one-dimensional text but are intended for two-dimensional diagrammatic use, such as symbols for electronic circuits.

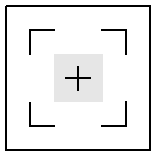
Keytop Labels. Where possible, keytop labels have been unified with other symbols of like appearance—for example, U+21E7 UPWARDS WHITE ARROW to indicate the Shift key. While symbols such as U+2318 PLACE OF INTEREST SIGN and U+2388 HELM SYMBOL are generic symbols that have been adapted to use on keytops, other symbols specifically follow ISO/IEC 9995-7.

Corner Brackets. Applications that need corner brackets should use the floor and ceiling symbols encoded at U+2308..U+230B. These should not be confused with the CJK corner brackets at U+300C and U+300D, which are used as quotation marks.

Crops and Quine Corners. Crops and quine corners are most properly used in two-dimensional layout but may be referred to in plain text. The usage of crops and quine corners is as indicated in the following diagram:



Use of crops



Use of quine corners

Angle Brackets. U+2329 LEFT-POINTING ANGLE BRACKET and U+232A RIGHT-POINTING ANGLE BRACKET have long been canonically equivalent to the CJK punctuation characters, U+3008 LEFT ANGLE BRACKET and U+3009 RIGHT ANGLE BRACKET, respectively. This canonical equivalence implies that the use of the latter (CJK) code points is preferred, and that U+2329 and U+232A are also "wide" characters. (See Unicode Standard Annex #11, "East Asian Width," for the definition of the East Asian wide property.) For this reason, the use of U+2329 and U+232A is deprecated for mathematics and for technical publication, where the wide property of the characters has the potential to interfere with the proper formatting of mathematical formulae. The angle brackets specifically provided for mathematics, U+27E8 MATHEMATICAL LEFT ANGLE BRACKET and U+27E9 MATHEMATICAL RIGHT ANGLE BRACKET, should be used instead. See *Section 14.4, Mathematical Symbols*.

APL Functional Symbols. APL (A Programming Language) makes extensive use of functional symbols constructed by composition with other, more primitive functional symbols. It used backspace and overstrike mechanisms in early computer implementations. In principle, functional composition is productive in APL; in practice, a relatively small number of composed functional symbols have become standard operators in APL. This relatively small set is encoded in its entirety in this block. All other APL extensions can be encoded by composition of other Unicode characters. For example, the APL symbol *a underbar* can be represented by U+0061 LATIN SMALL LETTER A + U+0332 COMBINING LOW LINE.

Symbol Pieces. The characters in the range U+239B..U+23B3, plus U+23B7, constitute a set of bracket and other symbol fragments for use in mathematical typesetting. These pieces originated in older font standards, but have been used in past mathematical processing as characters in their own right to make up extra-tall glyphs for enclosing multiline mathematical formulae. Mathematical fences are ordinarily sized to the content that they enclose. However, in creating a large fence, the glyph is not scaled proportionally; in particular, the displayed stem weights must remain compatible with the accompanying smaller characters. Thus, simple scaling of font outlines cannot be used to create tall brackets. Instead, a common technique is to build up the symbol from pieces. In particular, the characters U+239B LEFT PARENTHESIS UPPER HOOK through U+23B3 SUMMATION BOTTOM represent a set of glyph pieces for building up large versions of the fences (,), [,], {, and }, and of the large operators ∑ and ∫. These brace and operator pieces are compatibility characters. They should not be used in stored mathematical text, but are often used in the data stream created by display and print drivers.

Table 14-3 shows which pieces are intended to be used together to create specific symbols.

Table 14-3. Use of Symbol Pieces

	Two-Row	Three-Row	Five-Row
Summation	23B2, 23B3		
Integral	2320, 2321	2320, 23AE, 2321	2320, 3×23AE, 2321
Left parenthesis	239B, 239D	239B, 239C, 239D	239B, 3×239C, 239D
Right parenthesis	239E, 23A0	239E, 239F, 23A0	239E, 3×239F, 23A0
Left bracket	23A1, 23A3	23A1, 23A2, 23A3	23A1, 3×23A2, 23A3
Right bracket	23A4, 23A6	23A4, 23A5, 23A6	23A4, 3×23A5, 23A6
Left brace	23B0, 23B1	23A7, 23A8, 2389	23A7, 23AA, 23A8, 23AA, 2389
Right brace	23B1, 23B0	23AB, 23AC, 23AD	23AB, 23AA, 23AC, 23AA, 23AD

For example, an instance of U+239B can be positioned relative to instances of U+239C and U+239D to form an extra-tall (three or more line) left parenthesis. The center sections encoded here are meant to be used only with the top and bottom pieces encoded adjacent to them because the segments are usually graphically constructed within the fonts so that they match perfectly when positioned at the same *x* coordinates.

Vertical Square Brackets. The vertical square brackets, U+23B4 TOP SQUARE BRACKET and U+23B5 BOTTOM SQUARE BRACKET, are compatibility characters for legacy applications emulating certain terminals. They are intended for those terminal applications only, for limited use in vertically oriented bracketed expressions. U+23B6 BOTTOM SQUARE BRACKET OVER TOP SQUARE BRACKET is used when a single character cell is both the end of one such expression and the start of another. These compatibility characters should not be confused with the general need for rotated *glyphs* for parentheses, brackets, braces, and quotation marks for vertically rendered CJK text. Such rotations should be handled by fonts and rendering software, rather than by separate encoding of each rotated glyph as a character. See further discussion in *Section 6.2, General Punctuation*.

Terminal Graphics Characters. In addition to the box-drawing characters in the Box Drawing block, a small number of vertical or horizontal line characters are encoded in the Miscellaneous Technical symbols block to complete the set of compatibility characters needed for applications that need to emulate various old terminals. The horizontal scan line characters, U+23BA HORIZONTAL SCAN LINE-1 through U+23BD HORIZONTAL SCAN LINE-9, in particular, represent characters that were encoded in character ROM for use with 9-line character graphic cells. Horizontal scan line characters are encoded for scan lines 1, 3, 7, and 9. The horizontal scan line character for scan line 5 is unified with U+2500 BOX DRAWINGS LIGHT HORIZONTAL.

Dental Symbols. The set of symbols from U+23BE to U+23CC form a set of symbols from JIS X 0213 for use in dental notation.

Standards. This block contains a large number of symbols from ISO/IEC 9995-7:1994, *Information technology—Keyboard layouts for text and office systems—Part 7: Symbols used to represent functions.*

ISO/IEC 9995-7 contains many symbols that have been unified with existing and closely related symbols in Unicode. These symbols are shown with their ordinary shapes in the code charts, not with the particular glyph variation required by conformance to ISO/IEC 9995-7. Implementations wishing to be conformant to ISO/IEC 9995-7 in the depiction of these symbols should make use of a suitable font.

Optical Character Recognition: U+2440–U+245F

This block includes those symbolic characters of the OCR-A character set that do not correspond to ASCII characters as well as magnetic ink character recognition (MICR) symbols used in check processing.

Standards. Both sets of symbols are specified in ISO 2033.

14.6 Geometrical Symbols

Box Drawing: U+2500–U+257F

The characters in the Box Drawing block are encoded solely to facilitate the support of legacy implementations, such as terminal emulation.

Standards. GB 2312, KS X 1001, and industry standards were used to develop this block.

Block Elements: U+2580–U+259F

The Block Elements block represents a graphic compatibility zone in the Unicode Standard. A number of existing national and vendor standards, including IBM PC Code Page 437, contain a number of characters intended to enable a simple kind of display cell graphics by filling some fraction of each cell, or by filling each display cell with some degree of shading. The Unicode Standard does not encourage this kind of character-based graphics model but includes a minimal set of such characters for backward compatibility with the existing standards.

Half-block fill characters are included for each half of a display cell, plus a graduated series of vertical and horizontal fractional fills based on one-eighth parts. Also included is a series of shades based on one-quarter shadings. The fractional fills do not form a logically complete set but are intended only for backward compatibility.

Geometric Shapes: U+25A0–U+25FF

The Geometric Shapes are a collection of characters intended to encode prototypes for various commonly used geometrical shapes—mostly squares, triangles, and circles. The collection is somewhat arbitrary in scope; it is a compendium of shapes from various character and glyph standards. The typical distinctions more systematically encoded include black versus white, large versus small, basic shape (square versus triangle versus circle), orientation, and top versus bottom or left versus right part.

The hatched and cross-hatched squares at U+25A4..U+25A9 are derived from the Korean national standard (KS X 1001), in which they were probably intended as representations of fill patterns. Because the semantics of those characters is insufficiently defined in that standard, the Unicode character encoding simply carries the glyphs themselves as geometric shapes to provide a mapping for the Korean standard.

U+25CA ◊ LOZENGE is a typographical symbol seen in PostScript and in the Macintosh character set. It should be distinguished from both the generic U+25C7 WHITE DIAMOND and the U+2662 WHITE DIAMOND SUIT, as well as from another character sometimes called a lozenge, U+2311 SQUARE LOZENGE.

The squares and triangles at U+25E7..U+25EE are derived from the Linotype font collection. U+25EF LARGE CIRCLE is included for compatibility with the JIS X 0208-1990 Japanese standard.

Many geometrical shapes are used in mathematics. When used for this purpose, the centerpoints of the glyphs representing geometrical shapes should line up at the centerline of the mathematical font. This differs from the alignment used for some of the representative glyphs in the code charts.

For several simple geometrical shapes—circle, square, triangle, diamond, and lozenge—differences in size carry semantic distinctions in mathematical notation, such as the difference between use of the symbol as a variable or as one of a variety of operator types. In some cases, other blocks, such as Mathematical Operators or General Punctuation, contain these other sizes of geometrical symbols.

For more details on the use of geometrical shapes in mathematics, see Unicode Technical Report #25, "Unicode Support for Mathematics."

Standards. The Geometric Shapes are derived from a large range of national and vendor character standards.

14.7 Miscellaneous Symbols and Dingbats

Miscellaneous Symbols: U+2600–U+26FF

The Miscellaneous Symbols block consists of a very heterogeneous collection of symbols that do not fit in any other Unicode character block and that tend to be rather pictographic in nature. These symbols are typically used for text decorations, but they may also be treated as normal text characters in applications such as typesetting chess books, card game manuals, and horoscopes.

Characters in the Miscellaneous Symbols block may be rendered in more than one way, unlike characters in the Dingbats block, in which characters generally correspond to an explicit glyph. For example, both U+2641 EARTH and U+2645 URANUS have common alternative glyphs. EARTH can be rendered as ♁ or ⊕, and URANUS can be rendered as ♅ or ⛢.

The order of the Miscellaneous Symbols is completely arbitrary, but an attempt has been made to keep like symbols together and to group subsets of them into meaningful orders. Some of these subsets include weather and astronomical symbols, pointing hands, religious and ideological symbols, the Yijing (I Ching) trigrams, planet and zodiacal symbols, chess pieces, card suits, musical dingbats, and recycling symbols. (For other moon phases, see the circle-based shapes in the Geometric Shapes block.)

Corporate logos and collections of pictures of animals, vehicles, foods, and so on are not included because they tend either to be very specific in usage (logos, political party symbols) or nonconventional in appearance and semantic interpretation (pictures of cows or of cats; fizzing champagne bottles), and hence are inappropriate for encoding as characters. The Unicode Standard recommends that such items be incorporated in text via higher protocols that allow intermixing of graphic images with text, rather than by indefinite extension of the number of Miscellaneous Symbols encoded as characters. However, a large unassigned space has been set aside in this block with the expectation that other conventional sets of such symbols may be found appropriate for character encoding in the future.

Plastic Bottle Material Code System. The seven numbered logos encoded from U+2673 to U+2679, ♳♴♵♶♷♸♹, are from "The Plastic Bottle Material Code System," introduced in 1988 by the Society of the Plastics Industry (SPI). This set consistently uses thin, two-dimensional curved arrows suitable for use in plastics molding. In actual use, the symbols often are combined with an abbreviation of the material class below the triangle. Such abbreviations are not universal; therefore, they are not present in the representative glyphs in *Chapter 16, Code Charts*.

Recycling Symbol for Generic Materials. An unnumbered plastic resin code symbol U+267A ♺ RECYCLING SYMBOL FOR GENERIC MATERIALS is not formally part of the SPI system, but is found in many fonts. Occasional use of this symbol as a generic materials code symbol can be found in the field, usually with a text legend below, but sometimes also surrounding or overlaid by other text or symbols. Sometimes, the UNIVERSAL RECYCLING SYMBOL is substituted for the generic symbol in this context.

Universal Recycling Symbol. Unicode encodes two common glyph variants of this symbol: U+2672 ♲ UNIVERSAL RECYCLING SYMBOL and U+267B ♻ BLACK UNIVERSAL RECYCLING SYMBOL. Both are used to indicate that the material is recyclable. The white form is the traditional version of the symbol, but the black form is sometimes substituted, presumably because the thin outlines of the white form do not always reproduce well.

Paper Recycling Symbols. The two paper recycling symbols, U+267C ♼ RECYCLED PAPER SYMBOL and U+267D ♽ PARTIALLY-RECYCLED PAPER SYMBOL, can be used to distinguish between fully and partially recycled fiber content in paper products or packaging. They are usually accompanied by additional text.

Standards. The Miscellaneous Symbols are derived from a large range of national and vendor character standards.

Dingbats: U+2700–U+27BF

The Dingbats are derived from a well-established set of glyphs, the ITC Zapf Dingbats series 100, which constitutes the industry standard "Zapf Dingbat" font currently available in most laser printers. Other series of dingbat glyphs also exist, but are not encoded in the Unicode Standard because they are not widely implemented in existing hardware and software as character-encoded fonts. The order of the Dingbats block basically follows the PostScript encoding.

Unifications. Where a dingbat from the ITC Zapf Dingbats series 100 could be unified with a generic symbol widely used in other contexts, only the generic symbol was encoded. This accounts for the encoding gaps in the Dingbats block. Examples of such unifications include card suits, BLACK STAR, BLACK TELEPHONE, and BLACK RIGHT-POINTING INDEX (see "Miscellaneous Symbols"); BLACK CIRCLE and BLACK SQUARE (see "Geometric Shapes"); white encircled numbers 1 to 10 (see "Enclosed Alphanumerics"); and several generic arrows (see "Arrows"). Those four entries appear elsewhere in this chapter.

In other instances, other glyphs from the ITC Zapf Dingbats series 100 glyphs have come to be recognized as having applicability as generic symbols, despite having originally been encoded in the Dingbats block. For example, the series of negative (black) circled numbers 1 to 10 are now treated as generic symbols for this sequence, the continuation of which can be found in "Enclosed Alphanumerics." Other examples include U+2708 AIRPLANE and U+2709 ENVELOPE, which have definite semantics independent of the specific glyph shape, and which therefore should be considered generic symbols, rather than as symbols representing only the Zapf Dingbat glyph shapes.

For many of the remaining characters in the Dingbat block, their semantic value is primarily their shape; unlike characters that represent letters from a script, there is no well-established range of typeface variations for a dingbat that will retain its identity and therefore its semantics. It would be incorrect to arbitrarily replace U+279D TRIANGLE-HEADED RIGHTWARDS ARROW with any other right arrow dingbat or with any of the generic arrows from the Arrows block (U+2190..U+21FF). However, exact shape retention for the glyphs is not always required to maintain the relevant distinctions. For example, ornamental characters such as U+2741 EIGHT PETALLED OUTLINE BLACK FLORETTE have been successfully implemented in font faces other than Zapf Dingbats with glyph shapes that are similar, but not identical to the ITC Zapf Dingbats series 100.

The following guidelines are provided for font developers wishing to support this block of characters. Characters showing large sets of contrastive glyph shapes in the Dingbats block, and in particular the various arrow shapes at U+2794..U+27BE, should have glyphs that are closely modeled on the ITC Zapf Dingbats series 100, which are shown as representative glyphs in the code charts. The same applies to the various stars, asterisks, snowflakes, dropshadowed squares, checkmarks, and x's, many of which are ornamental and have elaborate names describing their glyphs.

Where the above does not apply, or where dingbats have more generic applicability as symbols, their glyphs do not need to match the representative glyphs in the code charts in every detail.

Ornamental Brackets. The 14 ornamental brackets encoded at U+2768..U+2775 are part of the set of Zapf Dingbats. Although they have always been included in Zapf Dingbats fonts, they were unencoded in PostScript versions of the fonts on some platforms. The Unicode Standard treats these brackets as punctuation characters.

Yijing Hexagram Symbols: U+4DC0–U+4DFF

Usage of the Yijing Hexagram Symbols in China begins with a text called《周易》*Zhou Yi,* ("the Zhou Dynasty classic of change"), said to have originated circa 1000 BCE. This text is now popularly known as the *Yijing*, *I Ching*, or *Book of Changes*. These symbols represent a primary level of notation in this ancient philosophical text, which is traditionally considered the first and most important of the Chinese classics. Today, these symbols appear in many print and electronic publications, produced in Asia and all over the world. The important Chinese character lexicon *Hanyu Da Zidian*, for example, makes use of these symbols in running text. These symbols are semantically distinct written signs associated with specific words. Each of the 64 hexagrams has a unique one- or two-syllable name. Each hexagram name is intimately connected with interpretation of the six lines. Related characters are Monogram and Digram Symbols (U+268A..U+268F), Yijing Trigram Symbols (U+2630..U+2637), and Tai Xuan Jing Symbols (U+1D300..U+1D356).

Tai Xuan Jing Symbols: U+1D300–U+1D356

Usage of these symbols in China begins with a text called《太玄經》*Tai Xuan Jing* (literally, "the exceedingly arcane classic"). Composed by a man named 楊雄 Yang Xiong (53 BCE–18 CE), the first draft of this work was completed in 2 BCE, in the decade before the fall of the Western Han Dynasty. This text is popularly known in the West under several titles, including *The Alternative I Ching* and *The Elemental Changes*. A number of annotated editions of *Tai Xuan Jing* have been published and reprinted in the 2,000 years since the original work appeared.

These symbols represent a primary level of notation in the original ancient text, following and expanding upon the traditions of the Chinese classic *Yijing*. The tetragram signs are less well known and less widely used than the hexagram signs. For this reason they were encoded on Plane 1 rather than the BMP. These tetragram symbols are semantically distinct written signs associated with specific words. Each of the 81 tetragrams has a unique monosyllabic name, and each tetragram name is intimately connected with interpretation of the four lines. The 81 tetragram symbols (U+1D306..U+1D356) encoded on Plane 1 constitute a complete set. Within this set of 81 signs, a subset of 16 signs known as the Yijing tetragrams is of importance to Yijing scholarship. These are used in the study of the "nuclear trigrams." Related characters are Monogram and Digram Symbols (U+268A..U+268F), Yijing Trigram Symbols (U+2630..U+2637), and Yijing Hexagram Symbols (U+4DC0..U+4DFF).

14.8 Enclosed and Square

Enclosed Alphanumerics: U+2460–U+24FF

The enclosed numbers and Latin letters of this block come from several sources, chiefly East Asian standards, and are provided for compatibility with them.

Standards. Enclosed letters and numbers occur in the Korean national standard, KS X 1001:1998, and in the Chinese national standard, GB 2312, as well as in various East Asian industry standards.

The Zapf Dingbat character set in widespread industry use contains four sets of encircled numbers (including encircled zero). The black-on-white set that has numbers with serifs is encoded here (U+2460..U+2468, and U+24EA). The other three sets are encoded in the range U+2776..U+2793 in the Dingbats block.

Decompositions. The parenthesized letters or numbers have compatibility decompositions to a sequence of opening parenthesis, letter or digit(s), closing parenthesis. The numbers with a period may be decomposed to digit(s), followed by a period. The encircled letters and single-digit numbers may be decomposed to a letter or digit followed by U+20DD ◌⃝ COMBINING ENCLOSING CIRCLE. Decompositions for the encircled numbers 10 through 20 are not supported in Unicode plain text.

Enclosed CJK Letters and Months: U+3200–U+32FF

Standards. This block provides mapping for all the enclosed Hangul elements from Korean standard KS X 1001:1998 as well as parenthesized ideographic characters from the JIS X 0208-1990 standard, CNS 11643, and several corporate registries.

CJK Compatibility: U+3300–U+33FF

CJK squared Katakana words are Katakana-spelled words that fill a single display cell (em-square) when intermixed with CJK ideographs. Likewise, squared Latin abbreviation symbols are designed to fill a single character position when mixed with CJK ideographs.

These characters are provided solely for compatibility with existing character encoding standards. Modern software can supply an infinite repertoire of Kana-spelled words or squared abbreviations on the fly.

Standards. CJK Compatibility characters are derived from the KS X 1001:1998 and CNS 11643 national standards, and from various company registries.

Japanese Era Names. The Japanese era names refer to the dates given in *Table 14-4*.

Table 14-4. Japanese Era Names

U+337B SQUARE ERA NAME HEISEI	1989-01-07 to present day
U+337C SQUARE ERA NAME SYOUWA	1926-12-24 to 1989-01-06
U+337D SQUARE ERA NAME TAISYOU	1912-07-29 to 1926-12-23
U+337E SQUARE ERA NAME MEIZI	1867 to 1912-07-28

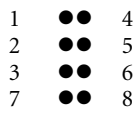
14.9 Braille

Braille Patterns: U+2800–U+28FF

Braille is a writing system used by blind people worldwide. It uses a system of six or eight raised dots, arranged in two vertical rows of three or four dots respectively. Eight-dot systems build on six-dot systems by adding two extra dots above or below the core matrix. Six-dot Braille allows 64 possible combinations, and eight-dot Braille allows 256 possible patterns of dot combinations. There is no fixed correspondence between a dot pattern and a character or symbol of any given script. Dot pattern assignments are dependent on context and user community. A single pattern can represent an abbreviation or a frequently occurring short word. For a number of contexts and user communities, the series of ISO technical reports starting with ISO/TR 11548-1 provide standardized correspondence tables as well as invocation sequences to indicate a context switch.

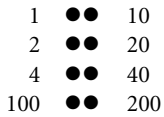
The Unicode Standard encodes a single complete set of 256 eight-dot patterns. This set includes the 64 dot patterns needed for six-dot Braille.

The character names for Braille patterns are based on the assignments of the dots of the Braille pattern to digits 1 to 8 as follows:



The designation of dots 1 to 6 corresponds to that of 6-dot Braille. The additional dots 7 and 8 are added beneath. The character name for a Braille pattern consists of BRAILLE PATTERN DOTS-12345678, where only those digits corresponding to dots in the pattern are included. The name for the empty pattern is BRAILLE PATTERN BLANK.

The 256 Braille patterns are arranged in the same sequence as in ISO/TR 11548-1, which is based on an octal number generated from the pattern arrangement. Octal numbers are associated with each dot of a Braille pattern in the following way:

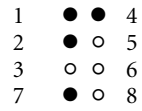


The octal number is obtained by adding the values corresponding to the dots present in the pattern. Octal numbers smaller than 100 are expanded to three digits by inserting leading zeroes. For example, the dots of BRAILLE PATTERN DOTS-1247 are assigned to the octal values of 1_8, 2_8, 10_8, and 100_8. The octal number representing the sum of these values is 113_8.

The assignment of meanings to Braille patterns is outside the scope of this standard.

Example. According to ISO/TR 11548-2, the character LATIN CAPITAL LETTER F can be represented in eight-dot Braille by the combination of the dots 1, 2, 4, and 7 (BRAILLE PATTERN

DOTS-1247). A full circle corresponds to a tangible (set) dot, and empty circles serve as position indicators for dots not set within the dot matrix:



Usage Model. The eight-dot Braille patterns in the Unicode Standard are intended to be used with either style of eight-dot Braille system, whether the additional two dots are considered to be in the top row or in the bottom row. These two systems are never intermixed in the same context, so their distinction is a matter of convention. The intent of encoding the 256 Braille patterns in the Unicode Standard is to allow input and output devices to be implemented that can interchange Braille data without having to go through a context-dependent conversion from semantic values to patterns, or vice versa. In this manner, final form documents can be exchanged and faithfully rendered. On the other hand, processing of textual data that require semantic support is intended to take place using the regular character assignments in the Unicode Standard.

Imaging. When output on a Braille device, dots shown as black are intended to be rendered as tangible. Dots shown in the standard as open circles are blank (not rendered as tangible). The Unicode Standard does not specify any physical dimension of Braille characters.

In the absence of a higher-level protocol, Braille patterns are output from left to right. When used to render final form (tangible) documents, Braille patterns are normally not intermixed with any other Unicode characters except control codes.

14.10 Byzantine Musical Symbols

Byzantine Musical Symbols: U+1D000–U+1D0FF

Byzantine musical notation first appeared in the seventh or eighth century CE, developing more fully by the tenth century. Byzantine Musical Symbols are chiefly used to write the religious music and hymns of the Christian Orthodox Church, although folk music manuscripts are also known. In 1881, the Orthodox Patriarchy Musical Committee redefined some of the signs and established the New Analytical Byzantine Musical Notation System, which is in use today. About 95 percent of the more than 7,000 musical manuscripts using this system are in Greek. Other manuscripts are in Russian, Bulgarian, Romanian, and Arabic.

Processing. Computer representation of Byzantine Musical Symbols is quite recent, although typographic publication of religious music books began in 1820. Two kinds of applications have been developed: applications to enable musicians to write the books they use, and applications that compare or convert this musical notation system to the standard Western system. (See *Section 14.11, Western Musical Symbols.*)

Byzantine Musical Symbols are divided into 15 classes according to function. Characters interact with one another in the horizontal and vertical dimension. There are three horizontal "stripes" in which various classes generally appear and rules as to how other characters interact within them. These rules are still being specified, and at present the plain text manipulation of Byzantine musical symbols, like that of Western musical symbols, is outside the scope of the Unicode Standard.

14.11 Western Musical Symbols

Musical Symbols: U+1D100–U+1D1FF

The musical symbols encoded in the Unicode Standard are intended to cover basic Western musical notation and its antecedents: mensural notation and plainsong (or Gregorian) notation. The most comprehensive coded language in regular use for representing sound is the common musical notation (CMN) of the Western world. Western musical notation is a system of symbols that is relatively, but not completely, self-consistent and relatively stable but still, like music itself, evolving. This open-ended system has survived over time partly because of its flexibility and extensibility. In the Unicode Standard, musical symbols have been drawn primarily from CMN. Commonly recognized additions to the CMN repertoire, such as quarter-tone accidentals, cluster noteheads, and shape-note noteheads, have also been included.

Graphical score elements are not included in the Musical Symbols block. These are pictographs usually created for a specific repertoire or sometimes even a single piece. Characters that have some specialized meaning in music but that are found in other character blocks are not included. They include numbers for time signatures and figured basses, letters for section labels and Roman numeral harmonic analysis, and so on.

Musical symbols are used worldwide in a more or less standard manner by a very large group of users. The symbols frequently occur in running text and may be treated as simple spacing characters with no special properties, with a few exceptions. Musical symbols are used in contexts such as theoretical works, pedagogical texts, terminological dictionaries, bibliographic databases, thematic catalogs, and databases of musical data. The musical symbol characters are also intended to be used within higher-level protocols, such as music description languages and file formats for the representation of musical data and musical scores.

Because of the complexities of layout and of pitch representation in general, the encoding of musical pitch is intentionally outside the scope of the Unicode Standard. The Musical Symbols block provides a common set of elements for interchange and processing. Encoding of pitch, and layout of the resulting musical structure, involves not only specifications for the vertical relationship between multiple notes simultaneously, but also in multiple staves, between instrumental parts, and so forth. These musical features are expected to be handled entirely in higher-level protocols making use of the graphical elements provided. Lack of pitch encoding is not a shortcoming, but is a necessary feature of the encoding.

The glyphs used in the accompanying charts are representative of typical cases; however, note in particular that the stem direction is not specified by the Unicode Standard and can only be determined in context. For a font that is intended to provide musical symbols in running text, either stem direction is acceptable. In some contexts—particularly for applications in early music—note heads, stems, flags, and other associated symbols may need to be rendered in different colors—for example, red.

U+266D MUSIC FLAT SIGN, U+266E MUSIC NATURAL SIGN, and U+266F MUSIC SHARP SIGN—three characters that occur frequently in music notation—are encoded in the Miscellaneous Symbols block (U+2600..U+267F). However, four characters also encoded in that block are to be interpreted merely as dingbats or miscellaneous symbols, not as representing actual musical notes:

U+2669 QUARTER NOTE

U+266A EIGHTH NOTE

U+266B BEAMED EIGHTH NOTES

U+266C BEAMED SIXTEENTH NOTES

The *punctum*, or Gregorian *brevis*, a square shape, is unified with U+1D147 MUSICAL SYMBOL SQUARE NOTEHEAD BLACK. The Gregorian *semibrevis*, a diamond or lozenge shape, is unified with U+1D1BA MUSICAL SYMBOL SEMIBREVIS BLACK. Thus, Gregorian notation, medieval notation, and modern notation require either separate fonts in practice or need font features to make subtle differentiations between shapes where required.

Processing. Most musical symbols can be thought of as simple spacing characters when used inline within texts and examples, even though they behave in a more complex manner in full musical layout. Some characters are meant only to be combined with others to produce combined character sequences, representing musical notes and their particular articulations. Musical symbols can be input, processed, and displayed in a manner similar to mathematical symbols. When embedded in text, most of the symbols are simple spacing characters with no special properties. A few characters have format control functions, as described below.

Input Methods. Musical symbols can be entered via standard alphanumeric keyboard, via piano keyboard or other device, or by a graphical method. Keyboard input of the musical symbols may make use of techniques similar to those used for Chinese, Japanese, and Korean. In addition, input methods utilizing pointing devices or piano keyboards could be developed similar to those in existing musical layout systems. For example, within a graphical user interface, the user could choose symbols from a palette-style menu.

Directionality. There are no known bidirectional implications for Musical Symbols. When combined with right-to-left texts—in Hebrew or Arabic, for example—the music notation is still written from left to right as usual. The words are divided into syllables and placed under or above the notes in the same fashion as for Latin scripts. The individual words or syllables corresponding to each note, however, are written in the dominant direction of the script.

Format Characters. Extensive ligature-like beams are used frequently in music notation between groups of notes having short values. The practice is widespread and very predictable, and is therefore amenable to algorithmic handling. The format characters U+1D173 MUSICAL SYMBOL BEGIN BEAM and U+1D174 MUSICAL SYMBOL END BEAM can be used to indicate the extents of beam groupings. In some exceptional cases, beams are left unclosed on one end. This can be indicated with a U+1D159 MUSICAL SYMBOL NULL NOTEHEAD character if no stem is to appear at the end of the beam.

Similarly, format characters have been provided for other connecting structures. The characters U+1D175 MUSICAL SYMBOL BEGIN TIE, U+1D176 MUSICAL SYMBOL END TIE, U+1D177 MUSICAL SYMBOL BEGIN SLUR, U+1D178 MUSICAL SYMBOL END SLUR, U+1D179 MUSICAL SYMBOL BEGIN PHRASE, and U+1D17A MUSICAL SYMBOL END PHRASE indicate the extent of these features. Like beaming, these features are easily handled in an algorithmic fashion.

These pairs of characters modify the layout and grouping of notes and phrases in full music notation. When musical examples are written or rendered in plain text without special software, the start/end format characters may be rendered as brackets or left uninterpreted. To the extent possible, more sophisticated software that renders musical examples inline with natural-language text might interpret them in their actual format control capacity, rendering slurs, beams, and so forth, as appropriate.

Precomposed Note Characters. For maximum flexibility, the character set includes both precomposed note values and primitives from which complete notes may be constructed. The precomposed versions are provided mainly for convenience. However, if any normalization form is applied, including NFC, the characters will be decomposed. For further information, see Unicode Standard Annex #15, "Unicode Normalization Forms." The canonical equivalents for these characters are given in the Unicode Character Database and illustrated in *Table 14-5*. In this table and subsequent examples, the names of the Unicode musical symbol characters are abbreviated by omitting the phrases MUSICAL SYMBOL or MUSICAL SYMBOL ORNAMENT. In *Table 14-5*, the "Equivalent to" column shows the full decomposition, which matches the expanded diagram to the right of the equals sign in the "Depiction" column.

Table 14-5. Precomposed Note Characters

Depiction	Precomposed Note	Equivalent to
𝅗𝅥 = 𝅗 + ◌𝅥	1D15E HALF NOTE	1D157 VOID NOTEHEAD + 1D165 COMBINING STEM
♩ = ♩ + ◌𝅥	1D15F QUARTER NOTE	1D158 NOTEHEAD BLACK + 1D165 COMBINING STEM
♪ = ♩ + ◌𝅥 + ◌𝅘	1D160 EIGHTH NOTE	1D158 NOTEHEAD BLACK + 1D165 COMBINING STEM + 1D16E COMBINING FLAG-1
𝅘𝅥𝅯 = ♩ + ◌𝅥 + ◌𝅘𝅥𝅯	1D161 SIXTEENTH NOTE	1D158 NOTEHEAD BLACK + 1D165 COMBINING STEM + 1D16F COMBINING FLAG-2
𝅘𝅥𝅰 = ♩ + ◌𝅥 + ◌𝅘𝅥𝅰	1D162 THIRTY-SECOND NOTE	1D158 NOTEHEAD BLACK + 1D165 COMBINING STEM + 1D170 COMBINING FLAG-3
𝅘𝅥𝅱 = ♩ + ◌𝅥 + ◌𝅘𝅥𝅱	1D163 SIXTY-FOURTH NOTE	1D158 NOTEHEAD BLACK + 1D165 COMBINING STEM + 1D171 COMBINING FLAG-4
𝅘𝅥𝅲 = ♩ + ◌𝅥 + ◌𝅘𝅥𝅲	1D164 ONE HUNDRED TWENTY-EIGHTH NOTE	1D158 NOTEHEAD BLACK + 1D165 COMBINING STEM + 1D172 COMBINING FLAG-5

Alternative Noteheads. More complex notes built up from alternative noteheads, stems, flags, and articulation symbols are necessary for complete implementations and complex scores. Examples of their use include American shape-note and modern percussion notations, as shown in *Table 14-6*.

Table 14-6. Alternative Noteheads

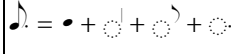
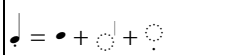
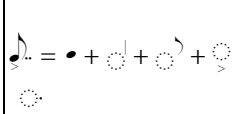
▪ = ▪ + ◌𝅥 1D147 SQUARE NOTEHEAD BLACK + 1D165 COMBINING STEM

× = × + ◌𝅥 1D143 X NOTEHEAD + 1D165 COMBINING STEM

Augmentation Dots and Articulation Symbols. Augmentation dots and articulation symbols may be appended to either the precomposed or built-up notes. In addition, augmentation dots and articulation symbols may be repeated as necessary to build a complete note symbol. Examples of the use of augmentation dots with both the NFC and NFD normalized forms are shown in *Table 14-7*. The full decomposition (NFD) matches the expanded diagram to the right of the equals sign in the "Depiction" column.

Note that in Normalization Form D, the canonically reordered combining marks may occur in a different order from that expected from the application order of augmentation dots and articulation symbols.

Table 14-7. Augmentation Dots and Articulation Symbols

Depiction	NFC Form	NFD Form
	1D160 EIGHTH NOTE + 1D16D COMBINING AUGMENTATION DOT	1D158 NOTEHEAD BLACK + 1D165 COMBINING STEM + 1D16E COMBINING FLAG-1 + 1D16D COMBINING AUGMENTATION DOT
	1D15F QUARTER NOTE + 1D17C COMBINING STACCATO	1D158 NOTEHEAD BLACK + 1D165 COMBINING STEM + 1D17C COMBINING STACCATO
	1D160 EIGHTH NOTE + 1D17B COMBINING ACCENT + 1D16D COMBINING AUGMENTATION DOT + 1D16D COMBINING AUGMENTATION DOT	1D158 NOTEHEAD BLACK + 1D165 COMBINING STEM + 1D16E COMBINING FLAG-1 + 1D17B COMBINING ACCENT + 1D16D COMBINING AUGMENTATION DOT + 1D16D COMBINING AUGMENTATION DOT

Ornamentation. *Table 14-8* lists common eighteenth-century ornaments and the combining sequences of characters from which they can be generated.

Table 14-8. Examples of Ornamentation

	1D19C STROKE-2 + 1D19D STROKE-3
	1D19C STROKE-2 + 1D1A0 STROKE-6 + 1D19D STROKE-3
	1D1A0 STROKE-6 + 1D19C STROKE-2 + 1D19C STROKE-2 + 1D19D STROKE-3
	1D19C STROKE-2 + 1D19C STROKE-2 + 1D1A0 STROKE-6 + 1D19D STROKE-3
	1D19C STROKE-2 + 1D19C STROKE-2 + 1D1A3 STROKE-9
	1D1A1 STROKE-7 + 1D19C STROKE-2 + 1D19C STROKE-2 + 1D19D STROKE-3
	1D1A2 STROKE-8 + 1D19C STROKE-2 + 1D19C STROKE-2 + 1D19D STROKE-3
	1D19C STROKE-2 + 1D19C STROKE-2 + 1D19D STROKE-3 + 1D19F STROKE-5
	1D1A1 STROKE-7 + 1D19C STROKE-2 + 1D19C STROKE-2 + 1D1A0 STROKE-6 + 1D19D STROKE-3
	1D1A1 STROKE-7 + 1D19C STROKE-2 + 1D19C STROKE-2 + 1D19D STROKE-3 + 1D19F STROKE-5
	1D1A2 STROKE-8 + 1D19C STROKE-2 + 1D19C STROKE-2 + 1D1A0 STROKE-6 + 1D19D STROKE-3
	1D19B STROKE-1 + 1D19C STROKE-2 + 1D19C STROKE-2 + 1D19D STROKE-3
	1D19B STROKE-1 + 1D19C STROKE-2 + 1D19C STROKE-2 + 1D19D STROKE-3 + 1D19E STROKE-4
	1D19C STROKE-2 + 1D19D STROKE-3 + 1D19E STROKE-4