

TERMINAL GRAPHICS FOR UNICODE

Frank da Cruz  
The Kermit Project  
Columbia University  
New York City USA  
fdc@columbia.edu  
<http://www.columbia.edu/kermit/>

D R A F T # 4

Tue Nov 3 19:08:58 1998

THIS IS A PREFORMATTED PLAIN-TEXT ASCII DOCUMENT. IT IS DESIGNED TO BE VIEWED AS-IS IN A FIXED-PITCH FONT. ITS WIDEST LINE IS 79 COLUMNS. IT CONTAINS NO TABS. IF IT LOOKS MESSY TO YOU, PLEASE FEEL FREE TO PICK UP A CLEAN COPY OF THIS OR THE RELATED PROPOSALS BY ANONYMOUS FTP:

HEX BYTE PICTURES FOR UNICODE (plain text)

<ftp://kermit.columbia.edu/kermit/ucsterminal/hex.txt>

ADDITIONAL CONTROL PICTURES FOR UNICODE (plain text)

<ftp://kermit.columbia.edu/kermit/ucsterminal/control.txt>

TERMINAL GRAPHICS FOR UNICODE (plain text)

<ftp://kermit.columbia.edu/kermit/ucsterminal/ucsterminal.txt>

Glyph Map (PDF, contributed by Michael Everson)

<ftp://kermit.columbia.edu/kermit/ucsterminal/terminal-emulation.pdf>

Clarification of SNI Glyphs (Microsoft Word 7.0)

<ftp://kermit.columbia.edu/kermit/ucsterminal/sni-charsets.doc>

Discussion (plain text)

<ftp://kermit.columbia.edu/kermit/ucsterminal/mail.txt>

(Note, the Exhibits are on paper and not available at the FTP site.)

ABSTRACT

A selection of terminal graphics characters is proposed for Unicode [24] and ISO 10646 [19] to allow Unicode-based terminal emulation software to display glyphs that are found on popular types of terminals but currently are not available in Unicode, and to exchange these characters with other Unicode-based applications.

## CONTENTS

1. Introduction
2. Scope
3. Organization
4. (deleted)
5. 3270 Terminal Operator Status Indicators
6. Math Symbols
7. Line, Box, and Block Characters
8. Unfinished Business
9. Summary of Proposed Additional Characters
10. References
11. Exhibits

### Tables:

- 5.1. 3270 Terminal Operator Status Indicators
- 6.1. Math Symbols for Terminals
- 7.1. Additional Line, Box, and Block Characters
- 9.1. Census of New Characters

### Figures:

- 5.1. Connected Rectangles
- 7.1. "Framus" Glyphs

### Notation:

- . Numbers in (parentheses) are footnote references, keyed to footnotes at the bottom of the section in which they appear.
- . Numbers in [brackets] are keyed to the References in Section 10.
- . Letter-Digit in brackets refers to an Exhibit in Section 11.

Grateful acknowledgements to those whose comments on previous drafts are reflected in this one: Kevin Bracey, Michael Everson, Doug Ewell, Asmus Freytag, Christine Gianone, Tony Harminc, Elliotte Rusty Harold, Edwin Hart, Kent Karlsson, Paul Keinanen, Markus Kuhn, Alain LaBonté, Heinz Lohse, Rick McGowan, Sean O'Leary, Jonathan Rosenne, Otto Stolz, Geoffrey Waigh, Kenneth Whistler. Special thanks to Michael Everson for his rendition of the proposed glyphs.

## 1. INTRODUCTION

Terminal-host communication was the dominant form of interaction between human and computer from about 1974 (when CRTs became affordable)(1) to about 1994 (when the Web and Windows took over the mass market). Terminal-host communication is still widespread, especially in large organizations, and is expected to remain so for decades to come, playing an important part in organizations like universities, hospitals, government agencies, and corporations with central computing facilities, for use in applications ranging from software development and system/network administration, to email and text-based Web access, to data entry and inquiry, to transaction processing, and it is also important to people who use speech or Braille devices and Telecommunications Devices for the Deaf (TDDs).

A text terminal, for purposes of this document, is a device for entry and display of text in a fixed-pitch font on a screen (or on paper) in which graphic characters are displayed as glyph images in rows and columns of "cells" of fixed and uniform size, one glyph image per cell. Text terminals generally display (or otherwise handle) the characters of ASCII [1] or EBCDIC [13], and often also accented or non-Roman letters (or ideograms), and often also "graphics" (2) (non-alphabetic, non-digit, non-punctuation) characters for purposes of line- and box-drawing, mathematics, or other special effects, and they also accept control characters or escape sequences for formatting.

In recent years, physical terminals have largely disappeared from the scene, their functions subsumed into PCs running terminal-emulation software alongside other applications. Unicode (viewed as a process) has effectively met the need for encoding the earth's writing systems, but so far it is not as well suited to terminal emulation as it might be since it lacks some of the required graphics characters.

Without a standard encoding for the missing glyphs, each maker of terminal emulation software must create or contract for custom fonts with private encodings. Such fonts are not compatible with other (otherwise compatible) fonts on the same platform (e.g. when copying from a terminal window and pasting to a word processor), nor with each other. Furthermore, should Unicode printers become standard equipment on PCs, terminal graphics characters will not print correctly on them (e.g. when used with the terminal's transparent printing, autoprinting, or dump-screen features).

This document proposes a modest repertoire of terminal graphics characters to be added to Unicode and ISO 10646, to supplement those already there (e.g. the line and box drawing characters at U+2500) to which all makers of fonts, code pages, and printers can refer when designing their products, and upon which all makers of terminal emulation and/or debugging software can base their screen displays.

To state the motivation for this and the companion proposals (see Section 2.2.) as clearly as I can:

1. There are numerous terminal emulation products on the market, with a user base numbering in the millions.
2. Increasingly, these products are designed for and used on systems -- like Windows NT -- that have Unicode fonts.
3. Many terminal based applications take full advantage of the features and glyph repertoires of the terminals they are designed for (far beyond the simple models supported, e.g. by termcap/terminfo).
4. The glyph repertoire of many common terminals -- VT100/VT220, Wyse, Siemens Nixdorf, Data General, etc, include glyphs that are not presently in Unicode.
5. Customers of terminal emulation products often demand complete and accurate emulation.
6. In order to succeed, makers of terminal emulation software must create private fonts containing the missing glyphs (which, as an aside, unnecessarily drives up the cost of the product for the end user) in the Private Use area.
7. Because of the closed and proprietary nature of this process, each terminal emulation product potentially (and in fact) encodes the same characters at different places.
8. Other applications use the Private Use Area for other purposes (and other glyphs).
9. The result is that terminal emulation products do not interoperate with each other or with other applications on the same platform.

For example, a VT100 or HP forms-based screen can not be pasted into a word processing document without changing the forms borders (etc, depending on exactly how they are encoded) into whatever other glyphs happen to be defined at the same code points in the font used by the other application. Ditto for mathematical formulae displayed on DEC or Siemens Nixdorf screens. Ditto for character-cell illustrations or tables in numerous online texts intended for display on any of the widespread terminals.

Notes:

- (1) Strictly speaking, terminals predate electronic computers by some decades; the Teletype (used as the control terminal on many mainframes and most minicomputers in the 1950s through 1970s) dates back to 1929.
- (2) Note the distinction between "graphic" meaning "printing" (as in "ISO 8859-1 is a graphic character set") versus "graphics" meaning having something to do with pictures. Note that graphics terminals (such as the Tektronix 4010) also exist, but are not relevant to this proposal.

## 2. SCOPE

This document represents a survey of the following terminals:

Data General D210,215,217,413,463 [2]  
Digital Equipment Corporation VT100 through VT520 [3-9]  
Heath / Zenith 19 [10]  
Hewlett Packard HP-2621 and HP-2648 [11,12]  
IBM 3164 and 3270 [15,16,27]  
Siemens Nixdorf 97801 [21]  
Televideo 922 and 965 [22,23]  
Wyse 60 and 370 [25,26]

as well as:

IBM PC code page 437 [14]

which is the basis for numerous PC-oriented so-called ANSI emulations.

### 2.1. Problems

Even within this fairly narrow scope, arriving at a sufficient set of character-cell terminal graphics for Unicode is complicated by the well-known problems that affect other preexisting character sets to varying degrees:

1. Lack of official names for the characters of some of the sets.
2. Lack of definitive, high-quality pictures of the glyphs in some cases.
3. Lack of descriptions of the purpose and intended use of the glyphs.
4. Lack of a current registration authority or owner in some cases.
5. Questions of unification of glyphs from different terminal makers.
6. End-user demand for specific characters or sets.

The issue of unification is complicated by the fact that some of the terminal graphics characters are designed to join at cell boundaries to form "pictures" (such as boxes or forms to be filled out) or large characters (such as big math symbols) spanning multiple rows and/or columns. The relationship of similar-looking glyphs for different terminals is difficult to determine -- e.g. exactly where does a line touch an edge, and at what angle, and does it make a difference?

The question of unification should be considered not only in the GUI environment but also for platforms where only one font is available -- a fixed-pitch "console" font -- and in "DOS"-like windows or fullscreen sessions, where only one fixed-pitch font may be used; this sort of environment is often host to terminal applications. Examples: a full-screen Windows NT session; the new Unicode-based Linux console driver and font.

Now suppose a particular terminal had a special glyph for "Superscript small Latin letter i". In the GUI environment, one would argue that the rendering software should change the size and baseline of the regular small "i" at

U+0069 to achieve the desired effect, and therefore a new character would not be needed. But this could not be done in a fixed-pitch font, and thus a new character would be needed after all if the terminal were to be emulated accurately and the meaning of the display not be altered (is it "3i" or 3 to the ith power?).

## 2.2. What This Proposal Does Not Propose

This proposal does not require any action for well-known terminal presentation forms such as double-high and/or double-wide characters, bold, blinking, inverse, italic, underlining, color, etc, since these are not encoding issues. In particular, no special code points are needed for double-high or double-wide characters, such as those seen on the DEC VT100 family of terminals, nor for compressed characters as seen on Data General and DEC terminals.

This proposal also does not cover true graphics terminals, such as Tektronix vector graphics units, DEC ReGIS or Sixel graphics, BBN Bitgraph, etc, since these graphics regimes are not character-cell based.

No attempt was made to account for the many Viewdata, Videotex, Minitel, NAPLPS, or other mosaic graphics character sets. These should be tackled, if at all, by someone who knows something about them.

Note that the graphic characters listed in this proposal rarely, if ever, appear on keyboard key labels. In general, these characters are never typed, not even on real terminals, but are displayed when the terminal is commanded into a special mode by the host; for example, with ISO 2022 [17] character-set designation and invocation escape sequences.

## 2.2. Related Proposals

This proposal contains only glyphs that appear on the screens of popular terminals during normal modes of operation, and not during debugging or "show invisibles" modes.

Terminals as well as special-purpose data monitors and protocol analyzers include debugging capabilities to allow otherwise invisible, illegal, or unknown characters to be displayed visually, either mnemonically (e.g. by control-character abbreviation) or in hexadecimal. Glyphs for these purposes are proposed separately in documents entitled "ADDITIONAL CONTROL PICTURES FOR UNICODE" and "HEX BYTE PICTURES FOR UNICODE".

## 3. ORGANIZATION

The following character categories are presented in sections 5 through 7:

### 3270 Terminal Operator Status Indicators

These glyphs are shown on IBM 3270 terminal [15] Operator Information Area. Section 5.

## Math Symbols

Although most math symbols found on terminals are already in Unicode, certain terminal-based applications rely on the ability to construct large symbols (integral and summation signs, braces, brackets) from smaller character-cell-sized pieces. This category also can apply to mathematical typesetting systems such as TeX [30]. Section 6.

## Line, Box, and Block Drawing

Used for data entry, transaction processing, forms filling, etc, in markets ranging from car rental and airline reservations, to 911 operators, to medical information systems, to online library catalogs. Although Unicode does include a basic set (mainly those as U+2500), some others are missing. Section 7.

### 3.1. Temporary Reference Code Assignments

The characters proposed in this document are assigned temporary Unicode values from the Private Use area, strictly for reference within (or to) this document only. Final values should be assigned outside of the Private Use range. The temporary allocations are:

E080-E086	3270	Status Symbols	8
E0A0-E0BF		Math Symbols	19
E0D0-E0EF		Line and Box Drawing	32

For a total of 59 positions. Obviously the final counts, code values, and block allocations, including reserved positions, are likely to change as this proposal evolves.

### 3.2. Character Properties

All new characters proposed in this document should be precomposed, since no terminals (with the exception of certain APL and ALA terminals) are capable of composing characters on the fly from nonspacing diacritics or by overstriking. All proposed characters have Combining Class 0 (but note that some of the corresponding glyphs are designed to "combine" (connect) with other glyphs in adjacent display cells).

All new characters proposed in this document that are approved should be defined in the Basic Multilingual Plane (Plane 0), since otherwise they would be of no use in operating systems such as Windows 95, which dominates the market and which does not and never will support the "astral planes."

No "Letter" characters are proposed, therefore none of the proposed additions has the Case property. All proposed characters are "Other Neutrals" (ON) as to directionality, since they fall into the "Punctuation, Symbols" category [24,Table 4-4]. None of the proposed characters has the Numeric Value Property.

Some of the proposed box-drawing and math-technical characters have the

Mirrored Property; this should be rather obvious when its name or description contains the word "left", or "right".

I would venture that the proposed math symbols (Section 6) would have the Mathematical Property, including the extensible ones, since the current Integral Top and Bottom at U+2320, U+2321 have this property [24, Section 1.9].

Summary: The characters in this proposal should have the following properties:

Case:	No
Combining Class:	0
Combining Jamo:	No
Directionality:	Other Neutral (ON)
Jamo Short Name:	No
Numeric Value:	No
Private Use:	No
Surrogate:	No
Mirrored:	No, except where noted in Section 6.
Mathematical:	Those in Section 6, Yes; others, No.

#### 4. (DELETED)

(This section moved to a separate proposal.)

#### 5. 3270 TERMINAL OPERATOR STATUS INDICATORS

The IBM 3270 terminal shows a variety of unique glyphs in its Operator Information Area [15, Figure A-4]. Although they are not encoded in any IBM known character set, or assigned IBM Graphic Character Global Identifiers (GCGIDs) in [29, as updated], they nevertheless appear on the screen, and are therefore required for accurate terminal emulation. These glyphs are listed in Table 5.1 and illustrated in Exhibit [H1].

Table 5.1: 3270 Terminal Operator Status Indicators

Code	Description
E080	Human stick figure (1)
E081	Human stick figure in box
E082	Clock at 6:10 (or 2:30) (1)
E083	White rectangle with stroke (2)
E084	Black rectangle with stroke (3)
E085	Lighting with stroke (4)
E086	Security key (5)
E087	Black and White Right-Pointing Triangles (6)

Notes:

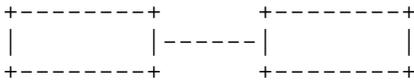
- (0) 0 = Combining Class 0; N = Neutral directionality;
- (1) Human stick figures are also found in several other terminal character sets, such as SNI Facet [21]. A clock (but with the hands at 3:00

rather than 6:10) is also found in SNI Klammern [E2].

- (2) A rectangle like the one at U+25AD with an oblique stroke through it.  
Note that "white" and "black" are used in the sense of the Unicode standard, and do not imply any particular colors or measure of goodness.
- (3) A rectangle like the one at U+25AC with an oblique stroke through it.
- (4) A horizontal lightning symbol with an oblique stroke through it.
- (5) A picture of a key (indicating the keyboard is locked). Reportedly, other proposals include similar glyphs for related, but different, purposes, such as a Shift Lock indicator. Care should be taken not to unify the IBM Security Key (a simple key) with other such symbols, such as padlocks with or without keys in them.
- (6) Like U+25B8 and U+25B9 in the same cell, arranged horizontally, left to right, like a double right-pointing arrowhead, used as a supplementary indicator.

In many cases, black and/or white rectangles (U+25AD, U+25AC, U+E083, U+E084) are connected with a centered horizontal line such as the one at U+2500; two rectangles connected this way generally symbolize a 3270 terminal with a printer attached. Figure 5.1 shows an example; also see Exhibit [H1]. The font designer must ensure that a sequence: rectangle, line, rectangle, results in a pair of connected rectangles.

Figure 5.1: Connected Rectangles



Summary:

8 new characters, E080-E087

Status:

Needed for proper emulation of IBM 3270 screens. This block of characters is separate and distinct from, and independent of, all other blocks in this proposal.

## 6. MATH SYMBOLS

Unicode has a generous supply of math symbols, and no doubt more are in the works. And of course it also includes the Latin, Greek, Fraktur, Hebrew, and other letters used in mathematical notation.

However, terminal emulators also need special glyphs designed to be joined together in adjacent character cells, vertically or horizontally, to form large math symbols such as integrals, summation signs, braces, or brackets, such as the integral top and bottom that already exist at U+2320 and U+2321. Several other single-cell characters are also missing, including the small radical sign from the DEC Technical character set.

Extensible math characters also appear in the TeX Standard Extension

Font [30,p.175], and in the widely used Apple Symbol font, also used in Adobe PostScript.

Table 6.1 lists the needed characters, along with suggested temporary codes for them. At least one real terminal reference is shown for each character, in column/row notation, and/or an IBM Graphic Character Global Identifier (GCGID) [14]. All characters in this table have the Mathematical Property; those marked by "\*" also have the Mirrored Property.

Legend:

- SB = Square Bracket
- UL = Upper Left
- LL = Lower Left
- UR = Upper Right
- LR = Lower Right

Table 6.1: Math Symbols for Terminals

Code	Description	Reference
* E0A0	Extensible left brace middle	DEC Tech 02/15 (1)
* E0A1	Extensible left parenthesis bottom	DEC Tech 02/12,IBM SS210000
* E0A2	Extensible left parenthesis top	DEC Tech 02/11,IBM SS200000
* E0A3	Extensible left SB bottom	DEC Tech 02/08
* E0A4	Extensible left SB top	DEC Tech 02/07
* E0A5	Extensible right brace middle	DEC Tech 03/00 (1)
* E0A6	Extensible UR or LL brace section	IBM SS240000
* E0A7	Extensible LR or UL brace section	IBM SS250000
* E0A8	Extensible right parenthesis bottom	DEC Tech 02/14,IBM SS230000
* E0A9	Extensible right parenthesis top	DEC Tech 02/13,IBM SS220000
* E0AA	Extensible right SB bottom	DEC Tech 02/10
* E0AB	Extensible right SB top	DEC Tech 02/08
E0AC	Summation symbol bottom	DEC Tech 03/02,DG Math 01/09(2)
E0AD	Summation symbol top	DEC Tech 03/01,DG Math 01/08(2)
* E0AE	Right ceiling corner	DEC Tech 03/05
* E0AF	Right floor corner	DEC Tech 03/06
E0B0	Radical symbol, small	DEC Tech 00/01
E0B1	Radical symbol with stroke	DG Math 01/13
E0B2	Superscript Latin small letter i	SNI Math 03/00

References:

- DEC Tech = Digital Equipment Corporation Technical Character Set [C2], VT320 and later.
- SNI Math = Siemens Nixdorf Mathematisch [E5], SNI 97801.
- SNI IBM = Siemens Nixdorf IBM [E4], SNI 97801.
- DG Math = Data General Word-Processing, Greek, and Math Character Set [D2]
- IBM = IBM Graphic Character Global Identifier (GCGID) [14]

Notes:

- (1) Also found in the Microsoft and Apple/Adobe Symbol fonts.
- (2) Also GCGID SS280000 and SS290000.

Summary: 24 new characters, E0A0-E0B7.

Status:

These symbols are needed for accurate emulation of DEC, DG, SNI and other terminals, and very likely also by mathematical typesetting systems. This block of characters is separate and distinct from, and independent of, all other blocks in this proposal.

## 7. LINE, BOX, AND BLOCK CHARACTERS

A particular need addressed by this proposal is the continued ability to support (sometimes mission-critical) terminal-based forms-filling applications that also require entry and display of international characters, as terminals are replaced by PCs. So far, Unicode has provided the international characters, but not necessarily all the needed character-cell based forms-drawing capabilities.

Some terminals have vertical and horizontal lines that are not centered within the character cell, and currently not found in Unicode. Others have black rectangles or other shapes not found in the U+2580 block.

Table 7.1 lists the additional line, box, and block characters needed to emulate the target terminals.

Abbreviations:

V = Vertical  
H = Horizontal  
L = Left  
R = Right  
LL = Lower Left  
LR = Lower Right  
UL = Upper Left  
UR = Upper Right

Terminology:

Quadrant

A black rectangle filling one quarter of a cell, with one corner in the center and the opposite corner at a corner of the cell. So "Quadrant UL" is the upper left quadrant; "Quadrant UL and UR" is the top half of the cell (which happens to be coincident with U+2580 and so is not included here).

Line

Refers to a line that extends all the way to opposite edge(s) of a cell, designed to be joined to (a) line(s) in the adjacent cell(s).

Bar

Refers to a horizontal line that does not touch any cell edges.

Wedge

Refers to a character cell with a diagonal line connecting opposite

corners, dividing it into two triangles; one black, the other white; the wedge is the black part. Thus an UL Wedge is similar to U+25E9, except it fills the entire character cell.

### Framus

(Pick a better word!) is a shape composed of two triangles with their points meeting at the center of the cell to form an X with bars across the top and bottom, closing the open ends. A black framus has the two triangles filled in; a white one is in outline form. A framus with center bar has a horizontal line through the center of the cell.

Figure 7.1: "Framus" Glyphs

```

White      Black      With Bar
*****    *****    *****
*  *      *  *      *  *
* *      * *      * *
 *       *       * *
* *      * *      *****
* *      * *      * *
*  *      *  *      *  *
*****    *****    *****

```

Table 7.1: Additional Line, Box, and Block Characters

Code	Description	References
E0D0	L V box line, extensible	H19 07/12 (1)
E0D1	R V box line, extensible	H19 07/13 (1)
E0D2	UL Wedge	H19 07/02, IBM SF870000
E0D3	UR Wedge	H19 05/14, IBM SF860000
E0D4	LL Wedge	IBM SF850000
E0D5	LR Wedge	IBM SF840000
E0D6	H line - Scan 1	DSG 06/15, H19 07/10, WG3 05/00, TVI 09/00
E0D7	H line - Scan 3	DSG 07/00, Wyse ANSI 01/01, WG3 05/00
E0D8	H line - Scan 5	DSG 07/01, Wyse ANSI 02/02 (2)
E0D9	H line - Scan 7	DSG 07/02, Wyse ANSI 01/03, WG3 05/01
E0DA	H line - Scan 9	DSG 07/03, H19 07/11, WG3 05/01, TVI 09/01
E0DB	Quadrant LL	H19 06/13, WG3 05/05, TVI 09/05
E0DC	Quadrant LR	H19 06/12, WG3 05/04, TVI 09/04
E0DD	Quadrant UL	H19 06/14, WG3 05/06, TVI 09/06
E0DE	Quadrant UL and LL and LR	WG3 05/11, TVI 09/11
E0DF	Quadrant UL and LR	H19 06/10 (3)
E0E0	Quadrant UL and UR and LL	WG3 05/12, TVI 09/12
E0E1	Quadrant UL and UR and LR	WG3 05/13, TVI 09/13
E0E2	Quadrant UR	H19 111, WG3 83, TVI 09/03
E0E3	Quadrant UR and LL	(for completeness)
E0E4	Quadrant UR and LL and LR	WG3 05/14, TVI 09/14
E0E5	Full black diamond	TVI 09/02 (4)
E0E6	Black framus	DGM 06/08
E0E7	Black framus + H center bar	DGM 06/09
E0E8	White framus	DGM 06/10
E0E9	White framus + H center bar	DGM 06/11
E0EA	R & L arrow to V center bar	DGM 03/13
E0EB	Up arrow to H center line	DGL 02/12
E0EC	R arrow to V center line	DGL 02/13
E0ED	L arrow to V center line	DGL 02/14

E0EE Down arrow to H center line DGL 02/12  
E0EF Box drawing double dash H DGL 03/12 (5)

#### References:

DGM = Data General Word-Processing, Greek, and Math Character Set [D2]  
DGL = Data General Line Drawing Character Set [D3]  
DSG = The DEC Special Graphics Character Set [A3]  
H19 = The Heath/Zenith 19 Graphics Character Set [L1]  
WG3 = The Wyse Graphics 3 Character Set [F2]  
TVI = The Televideo 965 Multinational Character Set [23]  
IBM = Graphic Character Global Identifier (GCGID) [14]  
Wyse ANSI = Wyse 60 "Standard ANSI", "UK ANSI", and "ANSI Graphics" [F3]

#### Notes:

- (1) The vertical box lines are near, but not touching, the left and right edges of the cell, respectively, and are two pixels thick on the H19 screen. Similar to IBM GCID SF640000 and SF650000, respectively.
- (2) A centered horizontal line is already in Unicode U+2500, but this one might need to be encoded separately if the existing one does not mesh well with other line and box characters (which is not known, since no semantics are stated for it).
- (3) Only on Zenith models, not original Heathkits.
- (4) Full black diamond, with points touching midpoint of each cell wall.
- (5) Similar to U+2504 but double rather than triple.

Also note that Quadrants UL+UR, UR+LR, LL+LR, UL+LL (half blocks) are already encoded at Unicode block U+2580.

#### Summary:

32 New glyphs, Range E0D0 to E0EF.

#### Status:

These symbols are needed for accurate emulation of DEC, DG, SNI, Heath, Wyse, Televideo, and other terminals. This block of characters is separate and distinct from, and independent of, all other blocks in this proposal.

## 8. UNFINISHED BUSINESS

The selection of characters presented in this draft is far from comprehensive. Hundreds of other terminals from the past 30+ years are likely to have glyphs or entire character sets covered neither here nor in Unicode, and these might or might not be important in some application somewhere. Readers of this draft are invited, therefore, to suggest any needed additions, bearing in mind that Unicode code space is not unlimited.

Several character sets found in the references consulted are ignored here, fully or in part, due to lack of motivation (nobody has ever asked us, in our role of terminal emulator maker, to support them). Obviously these, and any other missing sets (such as the many Videotex/Viewdata/etc mosaic sets),

can be considered if there is a demand.

#### Siemens Nixdorf Facet [E3]

A set of 95 mosaic graphics, but not resembling any of the ISO Videotex mosaic sets; difficult to describe.

#### Hewlett Packard Line Drawing

Mostly coincident with Unicode box-drawing set at U+2500, but with a handful of unique characters, such as single-to-triple box intersections, single-to-double intersections with wide spacing, etc. These should be mappable to existing U+25xx glyphs without causing riots in the streets.

#### Hewlett Packard Big Character Pieces

Thick line segments specially designed for drawing large digits and letters, used on the HP-2648.

## 9. SUMMARY OF PROPOSED NEW CHARACTERS

If all the proposed new characters are added to the UCS, this will enable terminal emulators to fully handle at least the following terminal character sets, which were not previously covered in full:

- DEC Technical (VT320 and later)
- DEC Special Graphics (VT100 and later)
- Data General Word-Processing, Greek, and Math
- Data General Line Drawing (1)
- Heath/Zenith 19 Graphics
- Hewlett Packard 2621 and HPTERM
- Siemens Nixdorf's "IBM" set (plus parts of its Klammern and Facet sets)
- Televideo Multinational
- Wyse Graphics 3 (Graphics 1 and 2 were already covered)
- Wyse "Standard ANSI", "UK ANSI", and "ANSI Graphics"

#### Notes:

- (1) Except the DG logo character, which, like other corporate logos, is off limits.

Terminals supporting these character sets are numerous indeed. An incomplete list includes: DEC VT100, VT102, VT220/240, VT320/330/340, VT420, VT520/525; Data General 210, 215, 217, 413, and 463; the Heath / Zenith 19; the Perkin Elmer 550 and 1100; and numerous Televideo and Wyse models.

The new characters proposed in this document are listed in Table 10.1.

Table 9.1: Census of New Characters

Code	Description
E080	Human stick figure
E081	Human stick figure in box
E082	Clock at 6:10 (or 1:30)

E083 White rectangle with stroke  
E084 Black rectangle with stroke  
E085 Lighting with stroke  
E086 Security key  
E087 Black and White Right-Pointing Triangles

E0A0 Extensible left brace middle  
E0A1 Extensible left parenthesis bottom  
E0A2 Extensible left parenthesis top  
E0A3 Extensible left SB bottom  
E0A4 Extensible left SB top  
E0A5 Extensible right brace middle  
E0A6 Extensible UR or LL brace section  
E0A7 Extensible LR or UL brace section  
E0A8 Extensible right parenthesis bottom  
E0A9 Extensible right parenthesis top  
E0AA Extensible right SB bottom  
E0AB Extensible right SB top  
E0AC Summation symbol bottom  
E0AD Summation symbol top  
E0AE Right ceiling corner  
E0AF Right floor corner  
E0B0 Radical symbol, small  
E0B1 Radical symbol with stroke  
E0B2 Superscript Latin small letter i

E0D0 L V box line, extensible  
E0D1 R V box line, extensible  
E0D2 UL Wedge  
E0D3 UR Wedge  
E0D4 LL Wedge  
E0D5 LR Wedge  
E0D6 H line - Scan 1  
E0D7 H line - Scan 3  
E0D8 H line - Scan 5  
E0D9 H line - Scan 7  
E0DA H line - Scan 9  
E0DB Quadrant LL  
E0DC Quadrant LR  
E0DD Quadrant UL  
E0DE Quadrant UL and LL and LR  
E0DF Quadrant UL and LR  
E0E0 Quadrant UL and UR and LL  
E0E1 Quadrant UL and UR and LR  
E0E2 Quadrant UR  
E0E3 Quadrant UR and LL  
E0E4 Quadrant UR and LL and LR  
E0E5 Full black diamond  
E0E6 Black framus  
E0E7 Black framus + H center bar  
E0E8 White framus  
E0E9 White framus + H center bar  
E0EA R & L arrow to V center bar  
E0EB Up arrow to H center line  
E0EC R arrow to V center line  
E0ED L arrow to V center line  
E0EE Down arrow to H center line

E0EF Box drawing double dash H

Summary:

3270 Symbols:	8
Math Symbols:	19
Line/Box/Block:	32

Total: 59

## 10. REFERENCES

- [1] American National Standards Institute, ANSI X3.4-1986, Code for Information Interchange (ASCII), 1986.
- [2] Data General, Programming the Display Terminal: Models D217, D413, and D463, Westboro, MA, 1991.
- [3] Digital Equipment Corporation, VT100 User Guide, EK-VT100-UG-002, Maynard, MA, 1979.
- [4] Digital Equipment Corporation, VT102 Video Terminal User Guide, EK-VT102-UG-003, Maynard, MA, 1982.
- [5] Digital Equipment Corporation, VT220 Owner's Manual, EK-VT220-UG-003, Maynard, MA, 1984.
- [6] Digital Equipment Corporation, VT220 Series Programmer Reference Manual, EK-VT240-RM-002, Maynard, MA, 1984.
- [7] Digital Equipment Corporation, VT330/VT340 Programmer Reference Manual, Volume 1: Text Programming, ED-VT3XX-TP-002, Maynard, MA, 1988.
- [8] Digital Equipment Corporation, Installing and Using the VT420 Video Terminal EK-VT420-UG.002, Maynard, MA, 1988.
- [9] Digital Equipment Corporation, VT520/VT525 Video Terminal Programmer Information, EK-VT520-RM.A01, Maynard, MA, 1994.
- [10] Heathkit Manual for the Video Terminal Model H19, The Heath Company, Benton Harbor, MI, 1979.
- [11] Hewlett Packard 2621A/P Interactive Terminal Owner's Manual, 1978.
- [12] Hewlett Packard 2648A Graphics Terminal Reference Manual, 1977.
- [13] IBM System/360 Principles of Operation, GA22-6821-8, Poughkeepsie, NY, 1970.
- [14] IBM National Language Design Guide, Volume 2: National Language Support Reference Manual, 4th Edition, SE09-8002-03, North York

- ON, 1994.
- [15] IBM 3270 Information Display System, Component Description, GA27-2749-10, 1980.
- [16] IBM 3164 ASCII Color Display Station Description, GA18-2317-1, 1986.
- [17] ISO International Standard 2022, Information processing -- ISO 7-bit and 8-bit coded character sets -- Code extension techniques, Third Edition, Geneva, 1986.
- [18] ISO/IEC International Standard 6429, Information technology -- Control functions for coded character sets, Third Edition, Geneva, 1992.
- [19] ISO/IEC 10646-1, International Standard 10646, Information Processing -- Multiple-Octet Coded Character Set, 1993-now.
- [20] Perkin Elmer Model 1100 User's Manual, Randolph, NJ, 1978.
- [21] Siemens Nixdorf, Bildschirmereinheit 97801-5xx Schnittstellen, Benutzerhandbuch, München, 1991.
- [22] Televideo 922 Video Terminal Display Operator's Manual, Sunnyvale, CA, 1984.
- [23] Televideo 965 Video Terminal Display Operator's Manual, Sunnyvale, CA, 1988.
- [24] The Unicode Standard, Version 2.0, Addison-Wesley Developers Press, 1996.
- [25] Wyse WY-60 Programmer's Guide, Wyse Technology, San Jose, CA, 1987.
- [26] Wyse WY-370 Programmer's Guide, Wyse Technology, San Jose, CA, 1990.
- [27] IBM 3270 Information Display System, Data Stream Programmer's Reference, GA23-0059-06, 1991.
- [28] ISO International Register of Coded Characters to Be Used with Escape Sequences, European Computer Manufacturers Association (ECMA), Geneva, 1985-present.
- [29] IBM Character Data Representation Architecture, Level 1 Registry, IBM Canada Ltd., National Language Technical Centre, Ontario, SC09-1391-00, 1990 (superseded by: IBM Character Data Representation Architecture, Registration and Registry, IBM Canada Ltd., Toronto, SC09-2190-00, 1995).
- [30] Knuth, Donald, "TeX and METAFONT, New Directions in Typesetting", American Mathematical Society / Digital Press, Bedford MA, 1979.

- [31] Apple Computer Corporation, Inside Macintosh, 1984.
- [32] HDS-3200 Terminal Series Owner's Manual, Philadelphia PA, 1987.
- [33] Zenith Data Systems Video Terminal Z-19-CN Operation Manual, Saint Joseph, MI, 1981.
- [34] Interview 30A/40A Operator's Field Reference Guide, Atlantic Research Corporation, ATLC-107-919-101, Alexandria, VA, 1982.

## 11. EXHIBITS

The following exhibits, available only on paper, are reproduced from the terminal manuals indicated by the numeric reference number. Each exhibit is 1 page unless otherwise indicated.

- [A1] VT220 Display Controls Font (Left Half) [5].
- [A2] VT220 Display Controls Font (Right Half) [5].
- [A3] VT220 DEC Special Graphics Character Set [5].
- [B1] VT320 Display Controls Font (Left Half) [7].
- [B2] VT320 Display Controls Font (Right Half) [7].
- [C1] VT420 Display Controls Font (Both Halves) [8].
- [C2] VT420 DEC Technical Character Set [8].
- [C3] HDS-3200 DEC Technical Character Set [32].
- [D1] Data General US ASCII Character Set [2].
- [D2] Data General Word-Processing, Greek, and Math Character Set [2].
- [D3] Data General Line Drawing Character Set [2].
- [D4] Data General Special Graphics Character Set [2].
- [D5] Data General VT Multinational Character Set [2].
- [D6] Data General VT Special Graphics Character Set [2].
- [D7] Data General ISO 8859/1.2 Character Set [2].
- [E1] Siemens Nixdorf 97801 ISO 8859-1 Character Set [21].
- [E2] Siemens Nixdorf 97801 Klammern (Brackets) Character Set [21].

[E3] Siemens Nixdorf 97801 Facet Character Set [21].

[E4] Siemens Nixdorf 97801 IBM Character Set [21].

[E5] Siemens Nixdorf 97801 Math Character Set [21].

[E6] Siemens Nixdorf 97801 Character Generator (8 pages) [21].

[F1] Wyse 60 Native, Multinational, PC, and ASCII Character Sets [25].

[F2] Wyse 60 Graphics 1, 2, and 3 Character Sets [25].

[F3] Wyse 60 Standard ANSI, ANSI Graphics, and UK ANSI Character Sets [25].

[G1] Wyse 370 Controls Display Mode (74Hz) [26].

[G2] Wyse 370 Controls Display Mode (60Hz) [26].

[G3] Wyse 370 C0, ASCII, and Special Graphics Character Sets [26].

[G4] Wyse 370 C1, Multinational, and Latin-1 Character Sets [26].

[H1] IBM 3270 Operator Information Area Symbols (10 pages) [15].

[I1] TeX Standard Extension Font [30].

[J1] Apple Symbol Font (2 pages) [31].

[K1] Hewlett Packard 2621A/P National Terminal Character Set [11].

[L1] Heath/Zenith-19 Graphic Symbols (2 pages) [33].

[M1] Televideo 922 ASCII, Supplemental, Special Character Sets (4 pages) [22].

[N1] Sample screen from a data analyzer showing hex display [34].

(End)