L2/98-353

ADDITIONAL CONTROL PICTURES 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 # 2

Tue Nov 3 19:08:48 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

Extensions are proposed to augment Unicode's repertoire of Control Pictures at U+2400 with control pictures for other well-known control sets.

Please refer to the TERMINAL GRAPHICS FOR UNICODE proposal for a discussion of terminal emulation, including motivation for supporting it in Unicode, as well as for acknowledgements to those who helped with this set of proposals.

CONTENTS

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 3.
. Letter-Digit in brackets refers to an Exhibit in Section 4.

For consistency, the References and Exhibits are the same as those in the accompanying, even though most of the items are not referenced here.

1. INTRODUCTION

In the interest of "show[ing] the presence of ... control codes and the SPACE unequivocally when data is displayed" [24,p.6-84], Unicode includes a selection of control pictures. Makers (and supportors, and users) of terminal emulators, PC-based data monitors and protocol analyzers, and most other types of software could use this feature of Unicode to better advantage if it were extended to cover a greater portion of the control space.

Why are Unicode characters needed for this purpose?

a. This was deemed a worthwhile enough concept in the original Unicode design to include a block of control pictures for the C0 set.

b. C1 and EBCDIC control sets are also widely used.

c. Real physical terminals include these glyphs.

d. Debug modes of these terminals (as well as data monitors, etc) show these glyphs in a single fixed-width character cell, of the same size used for regular characters.

e. Since many communications-oriented applications might make use of these glyphs, they should be standardized for interoperability, not only with each other, but also with email, word-processsing, and printing applications to aid in help-desk and documentation procedures.

While this proposal asks that "display-controls" symbols for C1 and EBCDIC control characters be added to Unicode, it does not ask that the corresponding control characters themselves be added.

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.

2. BACKGROUND

Digital VT220 and higher terminals, as well as Televideo, Wyse, HP, Data General, Perkin Elmer, and other models, allow the user (or, in some cases, the host) to select whether control characters are acted upon or displayed graphically. Unicode itself includes its own "control characters" such as line and paragraph separators, directionality controls, etc.

Normally control characters are used to affect the format and presentation of glyphs on the screen. In "display controls", "transparent", or "debug" mode (the terminology varies with the terminal vendor), control characters are shown graphically rather than performing their normal functions; this allows analysis and debugging of the host-terminal data stream using a

terminal, emulator, protocol analyzer, or line monitor. It also allows a more readable form of file dumping and analysis.

A block of control pictures is already found in Unicode at U+2400, but:

a. The illustrations in the Unicode book do not look like the control pictures that are actually used on terminals;

b. They are for C0 only; there is no corresponding set of C1 control pictures;

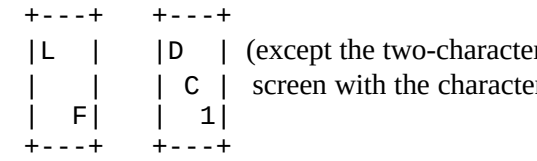
c. There are no pictures for the control characters unique to EBCDIC.

d. Certain other terminal-specific control pictures are missing.

A control picture allows the user to unequivocally determine the identity and position of control characters in the data stream by displaying each control chraracter as a unique (and mnemonic) glyph in a single terminal screen cell.

Terminals do this by arranging the letters (or letter-digit combinations) of the official abbreviation for the control character in diagonally from upper left to lower right, as shown in Figure 5.1.

Figure 2.1: Control Picture Display



The Unicode illustration for control pictures at U+2400, however, depicts the abbreviations horizontally. While the description of this block [24,p.6-84] states that "only the semantic is encoded... a particular application [can] use the graphic representation it prefers," a horizontal arrangement is chosen in the illustration (on p.7-188) for all characters except NL. But if they are implemented this way in a real font, it would be very difficult for the user to discern the boundary between one control picture and the next when several of them appear in a row.

It is suggested, therefore, that that next edition of the Unicode Standard illustrate these characters with the diagonal representation shown in Figure 5.1 (and in ISO 10646 [19]), since it is more likely that Unicode font designers will follow the illustrations in the Unicode Standard than attempt to procure the actual terminals or manuals to see how they do it.

The following sections discuss the different control sets, and propose a new set of control picture glyphs for each set except the C0 set. Each subsection is to be considered separately except insofar as they overlap.

Control picture characters 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
Mathematical:    No
```

3. C0 CONTROL PICTURES

Table 3.1 lists the C0 Control Characters from the ASCII Standard [1] (and also in ISO 646 and ISO 6429). Each C0 control character has an official designator (from the appropriate ANSI [1] or ISO [18] standard): a 2- or 3-character sequence of (ASCII) alphanumeric characters.

In some terminals, such as the DEC VT320 and above [B1,B2,C1], the control picture shows the designation in full. In most others, such as the VT220 and 240 [A1-A2], Data General [D1], Televideo [M1], HP [K1], and Perkin Elmer [20], each 3-character designator is replaced by a 2-character short form, referred to in this document as the "2X" form. For example, the character called DELETE has an official abbreviation DEL and a 2X form DT.

The columns of Table 3.1 are as follows:

```
Code: The Unicode value in hexadecimal.
Val:  The value of the control character's code in hexadecimal.
Name: The full ASCII abbreviation for the control character's name.
2X:   The 2-character abbreviation used on Televideo, Wyse, HP, etc.
Description: "Symbol for" followed by the character's standard name.
```

Table 3.1: C0 Control Characters

Code	Val	Name	2X	Description
2400	00	NUL	NU	Symbol for Null
2401	01	SOH	SH	Symbol for Start of Heading
2402	02	STX	SX	Symbol for Start of Text
2403	03	ETX	EX	Symbol for End of Text
2404	04	EOT	ET	Symbol for End of Transmission
2405	05	ENQ	EQ	Symbol for Enquiry
2406	06	ACK	AK	Symbol for Acknowledge
2407	07	BEL	BL	Symbol for Bell
2409	09	BS	BS	Symbol for Backspace
2409	09	HT	HT	Symbol for Horizontal Tab (1)
240A	0A	LF	LF	Symbol for Line Feed (1)
240B	0B	VT	VT	Symbol for Vertical Tab (1)
240C	0C	FF	FF	Symbol for Form Feed (2)
240D	0D	CR	CR	Symbol for Carriage Return (1)
240E	0E	SO	SO	Symbol for Shift Out

```
  240F   0F   SI    SI   Symbol for Shift In
  2410   10   DLE   DL   Symbol for Data Link Escape
  2411   11   DC1   D1   Symbol for Device Control 1 (2)
  2412   12   DC2   D2   Symbol for Device Control 2 (2)
  2413   13   DC3   D3   Symbol for Device Control 3 (2)
  2414   14   DC4   D4   Symbol for Device Control 4 (2)
  2415   15   NAK   NK   Symbol for Negative Acknowledge
  2416   16   SYN   SY   Symbol for Synchronous Idle
  2417   17   ETB   EB   Symbol for End of Transmission Block
  2418   18   CAN   CN   Symbol for Cancel
  2419   19   EM    EM   Symbol for End of Medium
  241A   1A   SUB   SU   Symbol for Substitute
  241B   1B   ESC   EC   Symbol for Escape
  241C   1C   FS    FS   Symbol for Field Separator (3)
  241D   1D   GS    GS   Symbol for Group Separator (3)
  241E   1E   RS    RS   Symbol for Record Separator (3)
  241F   1F   US    US   Symbol for Unit Separator (3)
  2420   20   SP    SP   Symbol for Space (4)
  2421   7F   DEL   DT   Symbol for Delete (4)
```

Notes:
(1) This symbol is also used in the DEC Special Graphics Set.
(2) Note the conflict/coincidence of these 2-character forms with hex bytes; see Note (3) in Section 4.
(3) These C0 controls have alternative names, listed in Section 7.
(4) Not, strictly speaking, a control character, but not a visible one either.

Summary and Status:
No new characters, but it is recommended that C0 control pictures be illustrated diagonally in the Unicode Standard, and that the "2X" forms be listed as alternatives for font designers, especially for low resolutions or small point sizes.

4. C1 CONTROL PICTURES

Since Unicode is used as the internal character set in applications (such as terminal emulators) that deal with non-Unicode character sets externally -- e.g. on network or modem connections -- the other widely-used control sets should also have control-picture glyphs, just as the C0 set does now.

C1 Control characters are specified in ISO 6429 [18] (ISO Registration Number 77 [28]) and used, among other places, in the VT220 family of terminals [5-9], Data General terminals [2], and the Wyse 370 [26], where they are represented in the right half of the "display controls" font as shown in Table 4.1 (DEC VT320 and higher terminals use the full name [B1-B2], Wyse terminals use the 2X name [G1-G4]; the DEC VT220 puts the hex value in a single character cell [A1,A2]). As with C0 controls, the "name" is displayed diagonally within the character cell in all these terminals. Unicode presently includes no C1 control pictures.

The "Code" column in the table shows the temporary Unicode value for reference within this document only; actual code assignments should be outside the

Private Use area. The other columns are labeled as in Table 3.1.

Table 4.1: C1 Control Characters

Code	Val	Name	2X	Description
	80	80	80	(1)
	81	81	81	(1)
E022	82	BPH	82	Symbol for Break Permitted Here (2)
E023	83	NBH	83	Symbol for No Break Here (2)
E024	84	IND	IN	Symbol for Index (3)
E025	85	NEL	NL	Symbol for Next Line (4)
E026	86	SSA	SS	Symbol for Start Selected Area
E027	87	ESA	ES	Symbol for End Selected Area
E028	88	HTS	HS	Symbol for Character Tabulation Set
E029	89	HTJ	HJ	Symbol for Character Tabulation with Justification
E02A	8A	VTS	VS	Symbol for Line Tabulation Set
E02B	8B	PLD	PD	Symbol for Partial Line Forward
E02C	8C	PLU	PU	Symbol for Partial Line Backward
E02D	8D	RI	RI	Symbol for Reverse Line Feed
E02E	8E	SS2	S2	Symbol for Single Shift 2
E02F	8F	SS3	S3	Symbol for Single Shift 3
E030	90	DCS	DC	Symbol for Device Control String
E031	91	PU1	P1	Symbol for Private Use 1
E032	92	PU2	P2	Symbol for Private Use 2
E033	93	STS	SE	Symbol for Set Transmit State
E034	94	CCH	CC	Symbol for Cancel Character
E035	95	MW	MW	Symbol for Message Waiting
E036	96	SPA	SP	Symbol for Start Protected (Guarded) Area
E037	97	EPA	EP	Symbol for End Protected (Guarded) Area
E038	98	SOS	98	Symbol for Start of String (2)
	99		99	(1)
E03A	9A	SCI	9A	Symbol for Single Character Introducer (2)
E03B	9B	CSI	CS	Symbol for Control Sequence Introducer (5)
E03C	9C	ST	ST	Symbol for String Terminator
E03D	9D	OSC	OS	Symbol for Operating System Command
E03E	9E	PM	PM	Symbol for Privacy Message
E03F	9F	APC	AP	Symbol for Application Program Command

Notes;
(1) Undefined in ISO-6429, shown on VT320/WY370 terminal by hex byte symbols (see text just below these notes).
(2) Defined in ISO-6429, but shown on VT320/WY370 terminal by hex value.
(3) Removed from ISO-6429 in the third edition, but shown as indicated on VT320 and WY370 terminals. Data General terminals show "ID" rather than "IN" [D7].
(4) Note the unfortunate coincidence of the 2X form of this character, "NL", with the EBCDIC Newline (NL) control. Data General Terminals show "NE" rather than "NL" [D7]. Also see notes in Section 5.
(5) Data General terminals show "CI" rather than "CS" [D7].

As the table indicates, three of the C1 control pictures are unassigned (the ones marked by "(1)", that would be at U+E020, U+E021, and U+E039 if these were assigned). These positions should be left vacant in case names are assigned to these characters in a future revision of ISO 6429, or terminals

are discovered with control pictures for these codes. In the meantime, hex bytes are used (because this is what the real terminals do); if a hex-byte block (separate proposal) is defined, they can be taken from that block; otherwise, the particular values shown here (80, 81, and 99, and possibly also 98 and 9A) must be defined for this block.

As with C0 controls, it is a matter for the font designer to choose the full designator from the Name column, or the 2-character alternatives from the 2X column.

Summary:
29 New characters (if hex bytes are also approved) or 32 (if they are not).

Status:
Needed to replicate the debugging functions of (at least) VT320/420/520 and WY370 terminals, and for debugging any data stream that contains ISO 6429 C1 controls.

5. EBCDIC CONTROL PICTURES

The EBCDIC family of character sets [13,14,29] includes its own repertoire of control characters. Many of them, like NUL, SOH, FF, SO, SI, and so on, are coincident with ASCII C0 controls in name and semantics, and sometimes also in encoding. Others are unique to EBCDIC.

Table 5.1 shows the EBCDIC control characters [29], in EBCDIC order. The Code column shows the Unicode value; those starting with 24 are already in Unicode block U+2400; those starting with E need to be added (these are also marked with "+" for emphasis). The Val column shows the EBCDIC value (hex). The Name column shows the EBCDIC abbreviation for the code, and the description lists "Symbol for" plus the EBCDIC name. No known "2X" forms exist.

Table 5.1: EBCDIC Control Characters

	Code	Val	Name	Description
	2400	00	NUL	Symbol for Null
	2401	01	SOH	Symbol for Start of Heading
	2402	02	STX	Symbol for Start of Text
	2403	03	ETX	Symbol for End of Text
+	E040	04	SEL	Symbol for Select (6)
	2409	05	HT	Symbol for Horizontal Tab
+	E041	06	RNL	Symbol for Required New Line (6)
	2421	07	DEL	Symbol for Delete
+	E042	08	GE	Symbol for Graphic Escape
+	E043	09	SPS	Symbol for Superscript
+	E044	0A	RPT	Symbol for Repeat (6)
	240B	0B	VT	Symbol for Vertical Tab
	240C	0C	FF	Symbol for Form Feed (1)
	240D	0D	CR	Symbol for Carriage Return
	240E	0E	SO	Symbol for Shift Out
	240F	0F	SI	Symbol for Shift In
	2410	10	DLE	Symbol for Data Link Escape

```
  2411   11   DC1   Symbol for Device Control 1
  2412   12   DC2   Symbol for Device Control 2
  2413   13   DC3   Symbol for Device Control 3 (6)
+ E045   14   RES   Symbol for Restore
  2424   15   NL    Symbol for New Line (2)
  2409   16   BS    Symbol for Backspace
+ E046   17   POC   Symbol for Program Operator Communication (6)
  2418   18   CAN   Symbol for Cancel
  2419   19   EM    Symbol for End of Medium
+ E047   1A   UBS   Symbol for Unit Back Space
+ E048   1B   CU1   Symbol for Customer Use 1
+ E049   1C   IFS   Symbol for Interchange File Separator
+ E04A   1D   IGS   Symbol for Interchange Group Separator
+ E04B   1E   IRS   Symbol for Interchange Record Separator
+ E04C   1F   IUS   Symbol for Interchange Unit Separator (3)
+ E04D   20   DS    Symbol for Digit Select
+ E04E   21   SOS   Symbol for Start of Significance
  241C   22   FS    Symbol for Field Separator
+ E04F   23   WUS   Symbol for Word Underscore
+ E050   24   BYP   Symbol for Bypass
  240A   25   LF    Symbol for Line Feed
  2417   26   ETB   Symbol for End of Transmission Block
  241B   27   ESC   Symbol for Escape
+ E051   28   SA    Symbol for Set Attribute
+ E052   29   SFE   Symbol for Start Field Extended
+ E053   2A   SM    Symbol for Set Mode (4)
+ E054   2B   CSP   Symbol for Control Sequence Prefix (6)
+ E055   2C   MFA   Symbol for Modify Field Attribute
  2405   2D   ENQ   Symbol for Enquiry
  2406   2E   ACK   Symbol for Acknowledge
  2407   2F   BEL   Symbol for Bell
+ E056   30         (Reserved by IBM for future use)
+ E057   31         (Reserved by IBM for future use)
  2416   32   SYN   Symbol for Synchronous Idle
+ E058   33   IR    Symbol for Index Return
+ E059   34   PP    Symbol for Presentation Position (6)
+ E05A   35   TRN   Symbol for Transparent (6)
+ E05B   36   NBS   Symbol for Numeric Backspace (6)
  2404   37   EOT   Symbol for End of Transmission
+ E05C   38   SBS   Symbol for Subscript
+ E05D   39   IT    Symbol for Indent Tabulation
+ E05E   3A   RFF   Symbol for Reverse Form Feed
+ E05F   3B   CU3   Symbol for Customer Use 3 (5)
  2414   3C   DC4   Symbol for Device Control 4
  2415   3D   NAK   Symbol for Negative Acknowledge
+ E060   3E         (Reserved by IBM for future use)
  241A   3F   SUB   Symbol for Substitute
```

Notes:

(1) Conflict/coincidence with a hex byte.

(2) Conflict/coincidence with C1 2X form; see text just below these notes. Also note that the NL glyph is part of the DEC Special Graphics character set [3-9].

(3) The IUS control is sometimes also labeled ITB.

(4) The SM control is sometimes also labeled SW (= Switch).

(5) Note: There is no longer a Customer Use 2 (see Table 5.2).

(6) Supersedes old name from Table 5.2.

The fact that the EBCDIC control character name "NL" is the same as one of the 2X forms of the C1 control character name "NEL" (the form used by DG terminals is "NE", not "NL"), together with the fact that the semantics of these two control characters are similar (though not identical) in their respective domains, does not necessarily make them candidates for unification, since the purpose of these sections is to encode the names of the controls in each domain (ASCII/ISO, EBCDIC, Unicode), not the controls themselves. If NEL and NL can be unified, then by this logic, so could numerous other C0, C1, EBCDIC, and Unicode controls whose names were less similar, e.g. C1 CSI (Control Sequence Introducer) and EBCDIC CSP (Control Sequence Prefix), or C1 BHP (Break Permitted Here) and Unicode ZWS (Zero Width Space), and this would defeat the advantage of encoding glyphs for the names used in each control-character domain, namely that the glyphs would contain names that are familiar to users of that domain.

Summary:
33 new characters, E040-E060, including 3 reserved.

Status:
Needed for debugging EBCDIC data streams. This block of characters is separate and distinct from, and independent of, all other blocks in this proposal. In particular, it is independent of the C1 controls.

For reference, Table 5.2 shows the original names for EBCDIC control characters [13] that have been superseded by the names shown in Table 5.1. This proposal does not advocate additional glyphs for these names.

Table 5.2: Obsolete EBCDIC Control Characters

Val	Name	Description	Replaced By
04	PF	Punch Off	SEL
06	LC	Lower Case	RNL
0A	SMM	Start of Manual Message	RPT
13	TM	Tape Mark	DC3
17	IL	Idle	POC
1A	CC	Cursor Control	UBX
2B	CU2	Customer Use 2	CSP
34	PN	Punch On	PP
35	RS	Record Separator	TRN
36	UC	Upper Case	NBS

6. IBM 3270 TERMINAL ORDERS AND CONTROLS

Names for IBM 3270(1) terminal orders and controls [27] that are not already listed in Tables 3.1-5.1 are shown in Table 6.1, to be used in debugging 3270 data streams. Columns are as in the previous tables, except the Type column, in which:

```
O = 3270 Terminal Order [27,Table 4-1]
```

```
D = 3270 Terminal Order in normal display [27,p.E-3]
L = LU 1 SCS Control Codes [27,Table 8-2]
F = 3270 Format Control Order [27,Table 4-3]
```

Notes:
(1) "3270" refers to the IBM 3270 terminal architecture, and not to any specific 3270 terminal model, such as 3277, 3278, etc.

Table 6.1: 3270 Control Characters

Code	Val	Name	Type	Description
E070	1D	SF	O	Symbol for Start Field
E071	11	SBA	O	Symbol for Set Buffer Address
E072	2C	MF	O	Symbol for Modify Field
E073	13	IC	O	Symbol for Insert Cursor
E074	05	PT	O	Symbol for Program Tab
E075	3C	RA	O	Symbol for Repeat to Address
E076	12	EUA	O	Symbol for Erase to Unprotected Address
E077	04	VCS	L	Symbol for Vertical Channel Select
E078	14	ENP	L	Symbol for Enable Presentation
E079	24	INP	L	Symbol for Inhibit Presentation
E07A	2B	FMT	L	Symbol for Format
E07B	1C	DUP	F	Symbol for Duplicate
E07C	1C	DUP	D	Overscore asterisk (1)
E07D	1E	FM	F	Symbol for Field Mark
E07E	1E	FM	D	Overscore semicolon (1)
E07F	FF	EO	F	Symbol for Eight Ones

Notes:
(1) When displayed by an actual 327x terminal, as opposed to an emulator in "display controls" mode.

Summary:
16 new characters, E070-E07F.

Status:
Needed for debugging IBM 3270 data streams. This block of characters is supplementary to the one in Section 5, and should not be approved unless the EBCDIC control picture glyphs are also approved.

7. ADDITIONAL CONTROL-LIKE PICTURES

Table 7.1 shows additional characters included in "display controls" mode on various terminals.

Table 7.1: Additional Control-Like Pictures

Code	Name	Description
E090	LS1	Symbol for Locking Shift 1 (1)
E091	LS0	Symbol for Locking Shift 0 (2)
E092	CEX	Symbol for Control Extension (3)
E093	IS4	Symbol for Information Separator 4 (4)

```
E094  IS3   Symbol for Information Separator 3 (5)
E095  IS2   Symbol for Information Separator 2 (6)
E096  IS1   Symbol for Information Separator 1 (7)
E097        Picture of Bell (8)
E098  BP    Word Processing Symbol BP (9)
E099  BE    Word Processing Symbol BE (9,10)
E09A  FN    Word Processing Symbol FN (9)
E09B  FE    Word Processing Symbol FE (9,10)
E09C  HF    Word Processing Symbol BP (9)
2426        Symbol for Substitute Form Two (Reverse Question Mark) (11)
```

Notes:
(1) ISO name for SO [18].
(2) ISO name for SI [18].
(3) From JIS C 6225-1979 / ISO # 74 [28].
(4) ISO Name for FS [18].
(5) ISO Name for GS [18].
(6) ISO Name for RS [18].
(7) ISO Name for US [18].
(8) Used on HP terminals in place of Symbol for BEL (U+2407) [K1].
(9) From the Data General Word Processing Set [2].
(10) Conflict/Coincidence with Hex Byte; see Note (3) in Section 4.

(11) The upright reverse question mark is used by DEC VT terminals to indicate that an invalid code was received. It also stands for SUB and/or RS in Wyse 370 [G2] and VT220 [A1] display controls mode, and is a glyph in its own right in the DEC Technical Character Set [C2], the DG Special Graphics Character Set [D4], and several others. This one is not in Unicode at present, but is encoded in Amendment 18 to ISO 10646 at the code point shown, with the requisite shape of reverse upright question mark.

Note that several other C0 controls have distinctive ISO names, such as TC1 for SOH, TC2 for STX, TC3 for ETX...; FE0 for BS, FE1 for HT, FE2 for LF, etc [28, Registration #001, the ISO 646 Control Set], but I have never seen these used outside the standard itself.

Summary:
13 characters, E090-E09C.

Status:
The ISO names LS1, LS0, IS4, IS3, IS2, IS1 are suggested for standards compliance; these might be suggested as glyph variants for SO, SI, FS, GS, RS, and US rather than encoded separately. However, the HP and DG symbols, as well as the reverse question mark, are are needed by terminal emulators.

8. UNICODE CONTROL PICTURES

Table 8.1 lists the nonprinting Unicode characters used for spacing, directionality control, and general formatting. These characters are in

the U+2000 block, and are indicated by mnemonics inside broken-line squares.

The Code column contains the temporary code value for the proposed symbol. The Val column contains the Unicode value of the character for which the symbolic representation is proposed. The Name column contains the desginator shown in the broken-line square in the Unicode code table, with a space standing for a line break (but see Note 2).

The suggested glyphs are those shown in the Unicode Standard.

Table 8.1: Unicode Control Characters

Code	Val	Name	Description
E000	2000	NQ SP	Symbol for En Quad
E001	2001	MQ SP	Symbol for Em Quad
E002	2002	EN SP	Symbol for En Space
E003	2003	EM SP	Symbol for Em Space
E004	2004	3/M SP	Symbol for Three-Per-Em-Space
E005	2005	4/M SP	Symbol for Four-Per-Em-Space
E006	2006	6/M SP	Symbol for Six-Per-Em-Space
E007	2007	F SP	Symbol for Figure Space
E008	2008	P SP	Symbol for Punctuation Space
E009	2009	TH SP	Symbol for Thin Space
E00A	200A	H SP	Symbol for Hair Space
E00B	200B	ZW SP	Symbol for Zero-Width Space
E00C	200C	ZW NJ	Symbol for Zero-Width Non-Joiner
E00D	200D	ZW J	Symbol for Zero-Width Joiner
E00E	200E	LRM	Symbol for Left-to-Right Mark
E00F	200F	RLM	Symbol for Right-to-Left Mark
E010	2028	L SEP	Symbol for Line Separator
E011	2029	P SEP	Symbol for Paragraph Separator
E012	202A	LRE	Symbol for Left-to-Right Embedding
E013	202B	RLE	Symbol for Right-to-Left Embedding
E014	202C	PDF	Symbol for Pop Directional Formatting
E015	202D	LRO	Symbol for Left-to-Right Override
E016	202E	RLO	Symbol for Right-to-Left Override
E017	206A	I SS	Symbol for Inhibit Symmetric Swapping
E018	206B	A SS	Symbol for Activate Symmetric Swapping
E019	206C	I AFS	Symbol for Inhibit Arabic Form Shaping
E01A	206D	A AFS	Symbol for Activate Arabic Form Shaping
E01B	206E	NA DS	Symbol for National Digit Shapes
E01C	206F	NO DS	Symbol for Nominal Digit Shapes
E01D	FEFF	ZWN BSP	Symbol for Zero Width No Break Space
E01E	FFFE	FF FE	Symbol for Not A Character (Byte Order) (1)
E01F	FFFF	FF FF	Symbol for Not A Character (1)

Notes:
(1) No mnemonic or abbreviation is given for the "not-a-character" characters in the Unicode Standard. A glyph is suggested for this character to allow Unicode-based debugging software or monitors to be able to unambiguously indicate its presence in the data stream.

Summary:
32 characters, E0000-E01F.

Status:

Controversial. Unicode control pictures are not needed for terminal emulation (at least not unless and until a Unicode-based terminal is defined), but are included for symmetry with the situation for C0 controls, and for completeness and reference. Makers of word processors, Web browsers, and other Unicode-based applications might find it desirable to add debugging features to their products using these glyphs.

9. SUMMARY OF PROPOSED ADDITIONAL CHARACTERS

The following control pictures are proposed:

Unicode Controls: 32 new characters, E000-E01F
C0 Controls: 0 new characters
C1 Controls: 32 new characters, E020-E03F
EBCDIC Controls: 33 new characters, E040-E060
3270 Controls: 16 new characters, E070-E07F
Misc Controls: 13 new characters, E090-E09C

Total Control Pics: 126
Without Unicode: 94

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:

ASCII/ISO Display Controls for DEC, Hewlett Packard, Wyse, Televideo, and others.
EBCDIC Display Controls for the IBM 3270

Table 9.1: Census of New Characters

Code	Description
E000	Symbol for En Quad
E001	Symbol for Em Quad
E002	Symbol for En Space
E003	Symbol for Em Space
E004	Symbol for Three-Per-Em-Space
E005	Symbol for Four-Per-Em-Space
E006	Symbol for Six-Per-Em-Space
E007	Symbol for Figure Space
E008	Symbol for Punctuation Space
E009	Symbol for Thin Space
E00A	Symbol for Hair Space
E00B	Symbol for Zero-Width Space
E00C	Symbol for Zero-Width Non-Joiner
E00D	Symbol for Zero-Width Joiner
E00E	Symbol for Left-to-Right Mark
E00F	Symbol for Right-to-Left Mark
E010	Symbol for Line Separator
E011	Symbol for Paragraph Separator

E012 Symbol for Left-to-Right Embedding
E013 Symbol for Right-to-Left Embedding
E014 Symbol for Pop Directional Formatting
E015 Symbol for Left-to-Right Override
E016 Symbol for Right-to-Left Override
E017 Symbol for Inhibit Symmetric Swapping
E018 Symbol for Activate Symmetric Swapping
E019 Symbol for Inhibit Arabic Form Shaping
E01A Symbol for Activate Arabic Form Shaping
E01B Symbol for National Digit Shapes
E01C Symbol for Nominal Digit Shapes
E01D Symbol for Zero Width No Break Space
E01E Symbol for Not A Character (Byte Order)
E01F Symbol for Not A Character

E020 (Reserved)
E021 (Reserved)
E022 Symbol for Break Permitted Here
E023 Symbol for No Break Here
E024 Symbol for Index
E025 Symbol for Next Line
E026 Symbol for Start Selected Area
E027 Symbol for End Selected Area
E028 Symbol for Character Tabulation Set
E029 Symbol for Character Tabulation with Justification
E02A Symbol for Line Tabulation Set
E02B Symbol for Partial Line Forward
E02C Symbol for Partial Line Backward
E02D Symbol for Reverse Line Feed
E02E Symbol for Single Shift 2
E02F Symbol for Single Shift 3
E030 Symbol for Device Control String
E031 Symbol for Private Use 1
E032 Symbol for Private Use 2
E033 Symbol for Set Transmit State
E034 Symbol for Cancel Character
E035 Symbol for Message Waiting
E036 Symbol for Start Protected (Guarded) Area
E037 Symbol for End Protected (Guarded) Area
E038 Symbol for Start of String
E039 (Reserved)
E03A Symbol for Single Character Introducer
E03B Symbol for Control Sequence Introducer
E03C Symbol for String Terminator
E03D Symbol for Operating System Command
E03E Symbol for Privacy Message
E03F Symbol for Application Program Command

E040 Symbol for Select
E041 Symbol for Required New Line
E042 Symbol for Graphic Escape
E043 Symbol for Superscript
E044 Symbol for Repeat
E045 Symbol for Restore
E046 Symbol for Program Operator Communication
E047 Symbol for Unit Back Space
E048 Symbol for Customer Use 1

```
E049  Symbol for Interchange File Separator
E04A  Symbol for Interchange Group Separator
E04B  Symbol for Interchange Record Separator
E04C  Symbol for Interchange Unit Separator
E04D  Symbol for Digit Select
E04E  Symbol for Start of Significance
E04F  Symbol for Word Underscore
E050  Symbol for Bypass
E051  Symbol for Set Attribute
E052  Symbol for Start Field Extended
E053  Symbol for Set Mode
E054  Symbol for Control Sequence Prefix
E055  Symbol for Modify Field Attribute
E056  (Reserved)
E057  (Reserved)
E058  Symbol for Index Return
E059  Symbol for Presentation Position
E05A  Symbol for Transparent
E05B  Symbol for Numeric Backspace
E05C  Symbol for Subscript
E05D  Symbol for Indent Tabulation
E05E  Symbol for Reverse Form Feed
E05F  Symbol for Customer Use 3
E060  (Reserved)
E070  Symbol for Start Field
E071  Symbol for Set Buffer Address
E072  Symbol for Modify Field
E073  Symbol for Insert Cursor
E074  Symbol for Program Tab
E075  Symbol for Repeat to Address
E076  Symbol for Erase to Unprotected Address
E077  Symbol for Vertical Channel Select
E078  Symbol for Enable Presentation
E079  Symbol for Inhibit Presentation
E07A  Symbol for Format
E07B  Symbol for Duplicate
E07C  Overscore asterisk
E07D  Symbol for Field Mark
E07E  Overscore semicolon
E07F  Symbol for Eight Ones

E090  Symbol for Locking Shift 1
E091  Symbol for Locking Shift 0
E092  Symbol for Control Extension
E093  Symbol for Information Separator 4
E094  Symbol for Information Separator 3
E095  Symbol for Information Separator 2
E096  Symbol for Information Separator 1
E097  Picture of Bell
E098  Word Processing Symbol BP
E099  Word Processing Symbol BE
E09A  Word Processing Symbol FN
E09B  Word Processing Symbol FE
E09C  Word Processing Symbol BP
```

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 Inforamtion, 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, Bildschirmeinheit 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)