L2/15-083

Proposal for addition of Group Mark symbol

Ken Shirriff
github.com/shirriff/groupmark
Feb 14, 2015

Abstract

The group mark symbol was part of the character set of important computers of the 1960s and 1970s, such as the IBM 705, 7070, 1401 and 1620. Books about these computers and manuals for them used the symbol ǂ to represent the group mark in running text. Unfortunately, this symbol does not exist in Unicode, making it difficult to write about technical details of these historical computers.

Introduction

The group mark was introduced in the 1950s as a separator character for I/O operations. In written text, the group mark is indicated by the symbol: ǂ. Unicode doesn't include this symbol, which is inconvenient when writing about the group mark or the character set of these computers.

The group mark became part of IBM's Standard BCD Interchange Code (BCDIC) in 1962. [1, page 20 and figure 56]. This standard was used by the IBM 1401, 1440, 1410, 1460, 7010, 7040, and 7044 data processing systems [2]. The BCDIC standard provided consistent definitions of codes and the relation between these codes and printed symbols, including uniform graphics for publications.[2, page A-2]. Unicode can represent all the characters in BCDIC, except for the group mark.

This proposal is for a Unicode code point for the group mark for use in running text with an associated glyph ǂ. This is orthogonal to the use of the group mark as a separator in data files. While the proposed group mark could be used in data files, that is not the focus of this proposal. In other words, this proposal is for a text symbol, not a control character.

1 Eric Fischer, *The Evolution of Character Codes, 1874-1968*.
2 Jack Melnick, *IBM 1401, 1440 and 1460 Programming Techniques*, 1964.

Examples of the group mark in text

The group mark symbol ǂ has been used in running text from the 1950s up to the present, in documentation, books and on the web. This section gives examples of the use of the group mark symbol from a variety of sources.

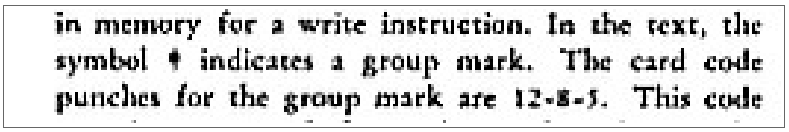

in memory for a write instruction. In the text, the
symbol ǂ indicates a group mark. The card code
punches for the group mark are 12-8-5. This code

Figure 1. Example of group mark symbol in use in text from 1956, in a manual for the IBM 705 data processing system. While the resolution of the symbol is bad, this example illustrates an early use of the group mark. [3, p20]

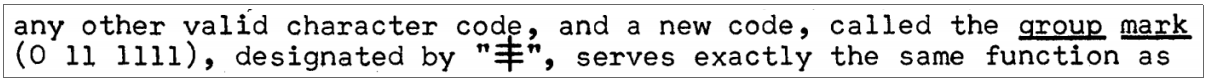

any other valid character code, and a new code, called the group mark
(0 11 1111), designated by "ǂ", serves exactly the same function as

Figure 2. Example of group mark symbol in use in text from a Univac manual from 1957. [4, p1103.6]

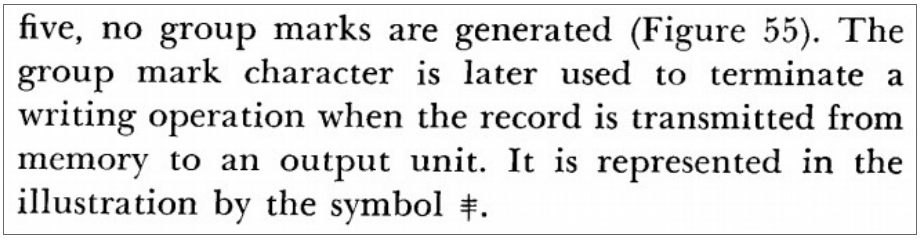

five, no group marks are generated (Figure 55). The
group mark character is later used to terminate a
writing operation when the record is transmitted from
memory to an output unit. It is represented in the
illustration by the symbol ǂ.

Figure 3. Group mark explained and symbol used in running text in General Information Manual for the IBM 705. [5, p34]

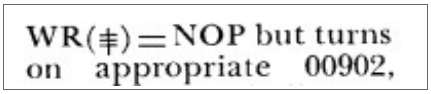

WR(ǂ) = NOP but turns
on appropriate 00902,

Figure 4. Group mark symbol in use in text from 1959, in a reference manual for the 705 data processing system [6, p157]

3 *IBM electronic data-processing machines, type 705: Preliminary manual of operation*, 1956.
4 *Univac II Marketing Manual*, 1957.
5 IBM, *General Information Manual, 705 Data Processing System*, 1959.
6 IBM, *Reference Manual, 705 Data Processing System*, 1959.

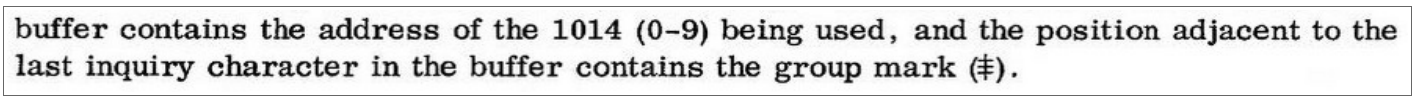
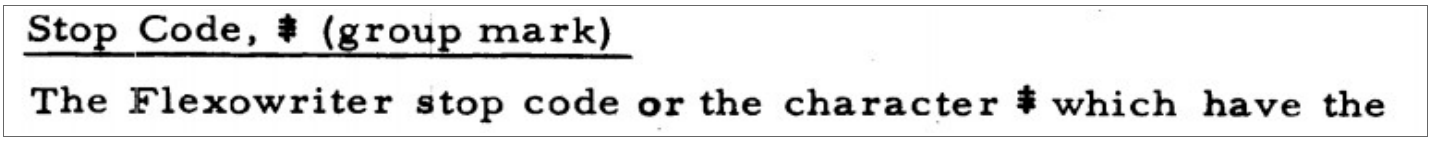

buffer contains the address of the 1014 (0-9) being used, and the position adjacent to the last inquiry character in the buffer contains the group mark (≢).

Figure 5. Group mark in the manual for an IBM I/O system. [7, p4]


Stop Code, ≢ (group mark)

The Flexowriter stop code or the character ≢ which have the

Figure 6. The group mark used in a programming document from Scientific Data Systems, Inc. [8, p3]

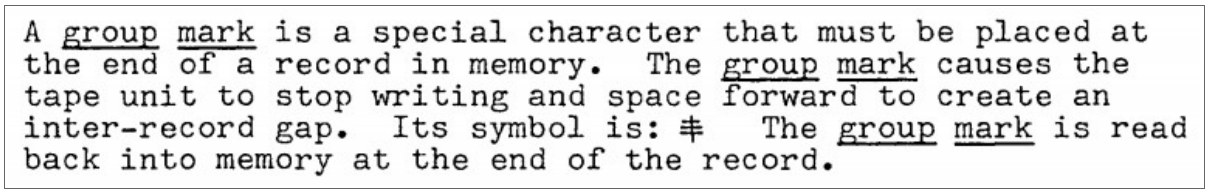

A group mark is a special character that must be placed at the end of a record in memory. The group mark causes the tape unit to stop writing and space forward to create an inter-record gap. Its symbol is: ≢ The group mark is read back into memory at the end of the record.

Figure 7. The group mark used in a book from 1962 about the IBM 1401. [9, p146]

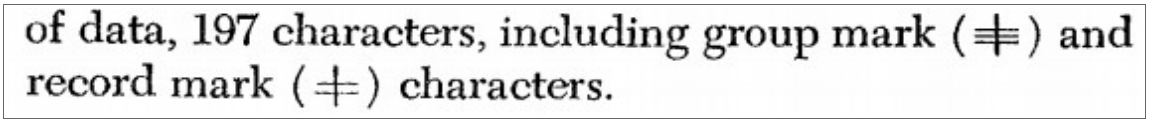

of data, 197 characters, including group mark (≢) and record mark (≠) characters.

Figure 8. Group mark from a 1964 printer manual. Note the similar record mark character, with two horizontal lines instead of three. [10, p1]

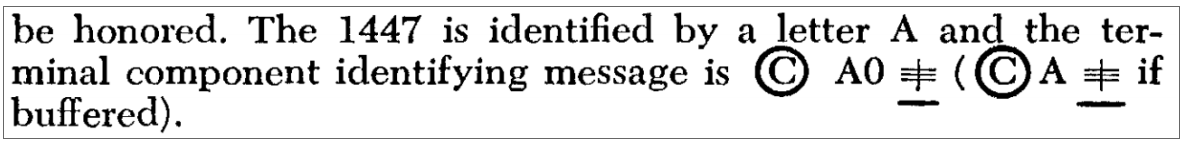
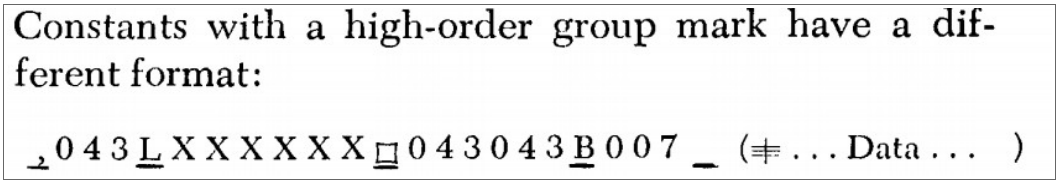

be honored. The 1447 is identified by a letter A and the terminal component identifying message is © A0 ≢ (©A ≢ if buffered).

Figure 9. Group mark symbol in text in an IBM reference document from 1964 [11, p26]


Constants with a high-order group mark have a different format:

␣043LXXXXXX□043043B007_ (≢ . . . Data . . .)

Figure 10. Group mark used in text in another IBM manual. [12, p51]

---

7 *IBM 7070-7074 Data Processing System Bulletin*, 1962.
8 Scientific Data Systems, *SDS 900 Series Program Library*, Catalog No. 000011, 1962.
9 James A. Saxon and William S. Plette, *Programming the 1401*, Prentice Hall, 1962.
10 *IBM 1443 Printer for 1620/1710 Systems*, 1964.
11 IBM Systems Reference Library, *IBM 1447 Console*, 1964.
12 *Autocoder (on Tape) Language Specifications and Operating Procedures IBM 1401 and 1460*,

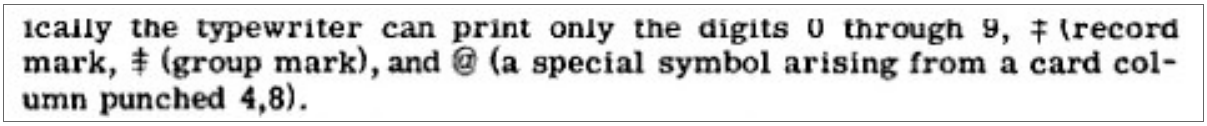

ically the typewriter can print only the digits 0 through 9, ‡ (record mark, ǂ (group mark), and @ (a special symbol arising from a card column punched 4,8).

Figure 11. Group mark in a programming book for the IBM 1620 from 1965 [13, p130]

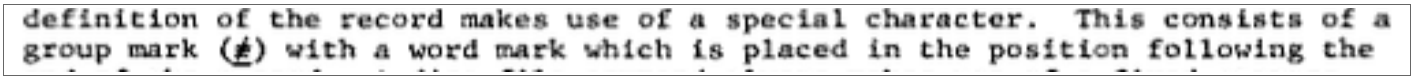

definition of the record makes use of a special character. This consists of a group mark (≠) with a word mark which is placed in the position following the

Figure 12. The group mark was also in use on the IBM 7070. This hand-drawn group mark has an slightly slanted vertical bar. The underline indicates the character has been combined with a word mark. [14, p84]

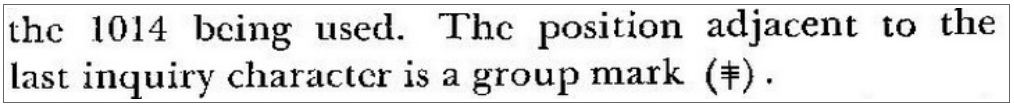

thc 1014 bcing used. Thc position adjacent to the last inquiry character is a group mark (≢).

Figure 13. The group mark in an IBM 7094 manual from 1966. [15, p96]

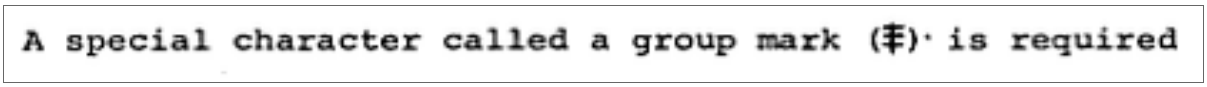

A special character called a group mark (≢)· is required

Figure 14. Group mark in a programming book from 1968 [16 ,p253]

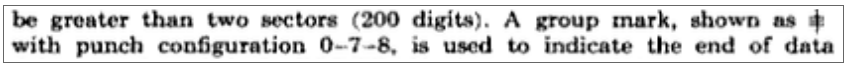

be greater than two sectors (200 digits). A group mark, shown as ǂ with punch configuration 0-7-8, is used to indicate the end of data

Figure 15. Group mark in a FORTRAN book from 1969. [17, 79]


△ mode change +++ segment mark
ǂ group mark
Fig. 23-21. The IBM BCD interchange code.

Figure 16. Group mark in an 1982 engineering handbook showing its use in BCD Interchange Code. [18, p23-17] This shows that books were using the group mark symbol into the 1980s.

Additional evidence that the group mark is considered a typographical character for use in text is found in a catalog for the Selectric typewriter. The group mark is listed as an available character and is

---

1964.

13 Eric A Weiss, *Programming the IBM 1620: The Hands-on Approach,* McGraw-Hill, 1965.

14 Gordon Bitter Davis, *An introduction to the IBM system/1401 computer*, McGraw-Hill, 1965.

15 *IBM 7094 Principles of Operation*, 1966.

16 William J. Claffey, *Principles of programming the IBM 1620 computer,* Dickenson Pub. Co., 1968.

17 D. K. Carver, *Introduction to Fortran II and Fortran IV programming,* Wiley, 1969.

18 Electronics engineers' handbook. Donald G. Fink, Donald Christiansen, McGraw-Hill, 1982.

available on a “88 character universal” type ball [19, p11,12]. In addition, page 114 shows the group mark character available for printer type bars. Thus, the group mark is intended to be usable by typewriters and printers in text.

The group mark has also been used in many tables (character sets, collation sequences, etc) and diagrams in books and documentation. These examples have not been included here, but are available if needed.

The group mark symbol is in current use on the web, as seen in several Wikipedia pages and other pages. Because the symbol is missing from Unicode, the pages need to use a bitmap image for the group mark.

BCD文字	Print-A	Print-H	カード	BCD	操作	定義その他
空白				C		
.	.	.	12-3-8	BA8 21	Halt	
¤	¤	)	12-4-8	CBA84	Clear Word Mark	*Lozenge*
[			12-5-8	BA84 1		
<			12-6-8	BA842		*Less Than*
⫲			12-7-8	CBA8421		*Group Mark*

Figure 17. The Japanese Wikipedia page on IBM 1401 uses a bitmap image for the group mark. [20]

GM	⫲	(Stop)	(Stop)	0 8421	E (Stop)	0-7-8	1-2-4-8		8421		*Group Mark*

Figure 18. The Wikipedia page on IBM 1620 also uses a bitmap image for the group mark. [21]

19 IBM ISG, *Type Catalog*, 1983.
20 Wikipedia (Japanese), *IBM 1401*, 2015.
21 Wikipedia (English), *IBM 1620* , 2015.

Properties

The proposed properties for the group mark are:

```
23FA;GROUP MARK;So;0;ON;;;;;N;;;;;
```

Property	Suggested Value
Code point	23FB
Name	GROUP MARK
General Category	So
Canonical Combining Class	0
Bidirectional Class	ON
Decomposition Type / Decomposition Mapping	
Numeric Type	
Numeric Value	
Bidi Mirrored	N
Unicode 1 Name	
ISO Comment	
Simple Uppercase Mapping	
Simple Lowercase Mapping	
Simple Titlecase Mapping	

Table 1: Suggested character properties for group mark.

Name: The name GROUP MARK is suggested since this is the traditional name for the character. This is in accordance with ISO/IEC 10646:2014 Annex L Guideline 10.

Range: The character seems suited for Miscellaneous Technical. The proposed code point is 23FB. Another possibility would be Miscellaneous Symbols and Arrows, with proposed code point 2BD2.

Collation: The group mark symbol is part of a well-defined collating sequence [22, page N-1] for Standard BCD Interchange Code. The group mark is near the beginning of the collating sequence, after "(" and "<", before "&" and "$". It would make most sense for the group mark symbol to have default Unicode collation in that area, but since Standard BCD Interchange Code has a very different collation order, the collation doesn't matter too much.

The group mark is uncased and has no special line-breaking behavior. The character is not meant for use in identifiers. It is a standalone symbol. It is not a white-space character and has no numeric value. It is not a combining character or punctuation.

22 *System Operation Reference Manual. IBM 1401 Data Processing System. IBM 1460 Data Processing System*, 1966.

Potential issues

Can an existing Unicode symbol represent the group mark? The CJK character U+4E30 丰 looks kind of like the group mark. It's not really satisfactory since the lines are different length and have hooks. The character NOT IDENTICAL TO U+2262 ≢also somewhat resembles the group mark, except the vertical line is slanted, which makes it unsatisfactory. One possibility from an earlier email thread [23] is an overstrike: IDENTICAL TO, COMBINING LONG VERTICAL LINE OVERLAY. This isn't a good representation since the group mark is not a combination of two underlying characters. (See earlier email discussion [24].) As an aside, the group mark with word mark, shown in Figure 12 is clearly a combination of the group mark with an underline, and would not make sense as a separate character. (In the 1401, any character can have a word mark on its storage location and this is represented in text by underlining the character.[25])

Unicode has graphic Pictures for Control Codes including: U+241D “SYMBOL FOR GROUP SEPARATOR”. [26] with glyph: ␝. This raises the question of whether the group mark could be represented as the SYMBOL FOR GROUP SEPARATOR. However, a new group mark symbol is desired for both semantic and visual reasons.

The Pictures for Control Codes are intended for displaying control codes within a data stream, while the proposed group mark is intended as a symbol for running text, which is semantically very different. The group mark was used in the IBM 1401 and other computers as a delimiter for I/O operations. For instance, a tape inter-record gap was represented in the computer by a group mark, and a group mark indicated the end of an I/O buffer. However, this proposal does not deal with the group mark as a character in storage or a data stream, but as a symbol that appears in text. For storing legacy data in a file using Unicode, the SYMBOL FOR GROUP SEPARATOR could be used. But in text, the proposed group mark symbol fills a role that is different from SYMBOL FOR GROUP SEPARATOR.

Visually, the SYMBOL FOR GROUP SEPARATOR glyph looks totally different from the group mark, so it doesn't make sense to consider them glyph variants. The SYMBOL FOR GROUP SEPARATOR glyph ␝can't be substituted in text for the group mark without totally changing the meaning to the reader.

One question is if the group mark symbol is sufficient, or additional symbols will be necessary to support discussions of the 1401 and related computers. The Standard BCD Interchange Code has several special characters in addition to the group mark:

---

23 026 keypunch, Unicode mail archive, 2007.

24 Productive Glyph Design vs. Productive Character Representation (was: Re: Quick survey of Apple symbol fonts ...), Unicode mail archive, 2011.

25 When discussing some other computers, such as the IBM 1410 and 7010, an inverted circumflex (caron) over the character indicates a word mark. See *IBM 1410 Principles of Operation*, p7. Either way, a Unicode combining character can be used.

26 Unicode Standard 7.0*, Unicode Control Pictures*, Range: 2400-243F, 1991-2014.

- Mode change: Δ.
- Word separator: ⟙
- Tape segment mark: ⧺
- Blank: ␣
- Tape mark: √
- Record mark ‡

These characters can all be represented reasonably well with existing Unicode symbols. These are displayed above as GREEK CAPITAL LETTER DELTA, CURLY LOGICAL OR [27], TRIPLE PLUS, BLANK SYMBOL, SQUARE ROOT, and DOUBLE DAGGER respectively. Adding the group mark is sufficient to complete the character set. (Since the existing characters are in the BMP, it would be preferable for group mark to also be in the BMP.) While new characters could be added to Unicode for MODE CHANGE, etc. the new characters would look nearly identical to the existing characters. There would be some semantic benefit of separating MODE CHANGE from GREEK CAPITAL LETTER DELTA, for instance. But I don't see it as being useful enough to merit introducing multiple new characters.

One question that could be raised is if the group mark symbol is of relevance only to IBM, as the majority of the usage references are IBM documents. However, the symbol has been used in books written by external authors [9, 13, 14, 16]. It has also been used by Univac [4] and Scientific Data Systems [8]. Finally, the symbol is currently in use on Wikipedia pages [20, 21]. Thus, the symbol is in general use, not just IBM.

One user community that could use the group mark symbol is the IBM 1401 discussion group [28]. The author has contacted this community and received moderate interest and has incorporated their feedback into this proposal.

In addition to being useful for written text, the group mark symbol could be used by software for user display. IBM 1401 simulators [29] currently use ASCII, representing the group mark with a somewhat arbitrary character such as " or }. Providing the group mark in Unicode would permit simulators to display characters as intended. Assigning a code point in the BMP to the group mark would simplify its use in software, since all other characters in the character set have representations in the BMP.

---

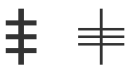
27 There is some variability in the “word separator” glyph used in text. It sometimes has a loop and looks more like ⟟ . In documentation for the IBM 1410, the symbol has longer “wings” and looks like ⩚. (See *IBM 1410 Principles of Operation*, p6.) Unicode could add a WORD SEPARATOR code point to deal with these glyph variants, but it's probably not worth the effort.

28 The 1401_software mailing list is one place to find the relevant user community.

29 The ROPE system is probably the easiest to use if you're looking for a 1401 simulator. It is built on top of the SIMH simulator.

Drawing the glyph

The sample texts above show variant glyphs that have been used for the group mark. The variants can be roughly grouped into two styles:



The first has short horizontals that are widely spaced, with a line weight that matches the surrounding text; Figure 5 is an example. The second has the three horizontal lines thin, close together, and long (about the same length as the vertical); Figure 8 is an example. These are not different symbols, but variant glyphs - I've seen several documents that use both glyphs for the group mark in the same document. I recommend a glyph similar to the first glyph - it harmonizes better with the DOUBLE DAGGER if that is used for the record mark. The first glyph also renders better than the second at small size.

Included with this proposal [30] is a TrueType font called GroupMark that contains both glyphs for the proposed group mark character, the first one at U+E000 and the second at U+E002.

Conclusion

Looking at the Unicode symbol guidelines, the group mark ⋕ is a good fit for addition to Unicode. It occurs in running text. It has a well defined user community. Being able to search for it in text would be useful. It has well-defined semantics that make it appropriate for computer processing. It completes a class of symbols already in the standard (the symbols in Standard BCD Interchange Code). Finally, it is letter-like in the sense that it should match the surrounding font style.

The group mark symbol has been used in running text since the 1950s. Providing the group mark symbol in Unicode would be beneficial.

Acknowledgments: Many people provided helpful comments including Frank da Cruz, Roozbeh Pournader, Ken Whistler, Jean-François Colson, Frédéric Grosshans, Van Snyder, Michael Albaugh, Jay Jaeger, Carl Claunch, Ed Thelen, and Jay Jaeger. The author benefited from the advice in Proposal to add IEC power symbols to Unicode. The font was created with FontForge and Fontastic.

30 The font, along with other documents, are at github.com/shirriff/groupmark

ISO/IEC JTC 1/SC 2/WG 2
PROPOSAL SUMMARY FORM TO ACCOMPANY SUBMISSIONS
FOR ADDITIONS TO THE REPERTOIRE OF ISO/IEC 10646[1]
Please fill all the sections A, B and C below.
Please read Principles and Procedures Document (P & P) from http://std.dkuug.dk/JTC1/SC2/WG2/docs/principles.html **for guidelines and details before filling this form.**
Please ensure you are using the latest Form from http://std.dkuug.dk/JTC1/SC2/WG2/docs/summaryform.html.
See also http://std.dkuug.dk/JTC1/SC2/WG2/docs/roadmaps.html **for latest *Roadmaps*.**

A. Administrative

1. **Title:** *Addition of group mark symbol*
2. Requester's name: *Ken Shirriff*
3. Requester type (Member body/Liaison/Individual contribution): *individual*
4. Submission date: *02/14/15*
5. Requester's reference (if applicable):
6. Choose one of the following:
   This is a complete proposal: *X*
   (or) More information will be provided later:

B. Technical – General

1. Choose one of the following:
   a. This proposal is for a new script (set of characters): *No*
      Proposed name of script:
   b. The proposal is for addition of character(s) to an existing block: *Yes*
      Name of the existing block: *Miscellaneous Technical*
2. Number of characters in proposal: *1*
3. Proposed category (select one from below - see section 2.2 of P&P document):
   A-Contemporary ( ) B.1-Specialized (small collection) (X) B.2-Specialized (large collection) ( )
   C-Major extinct ( ) D-Attested extinct ( ) E-Minor extinct ( )
   F-Archaic Hieroglyphic or Ideographic ( ) G-Obscure or questionable usage symbols ( )
4. Is a repertoire including character names provided? *Yes*
   a. If YES, are the names in accordance with the "character naming guidelines" in Annex L of P&P document? *Yes*
   b. Are the character shapes attached in a legible form suitable for review? *Yes*
5. Fonts related:
   a. Who will provide the appropriate computerized font to the Project Editor of 10646 for publishing the standard?
   Ken Shirriff
   b. Identify the party granting a license for use of the font by the editors (include address, e-mail, ftp-site, etc.):
   Ken Shirriff
6. References:
   a. Are references (to other character sets, dictionaries, descriptive texts etc.) provided? *Yes*
   b. Are published examples of use (such as samples from newspapers, magazines, or other sources) of proposed characters attached? *Yes*
7. Special encoding issues:
   Does the proposal address other aspects of character data processing (if applicable) such as input, presentation, sorting, searching, indexing, transliteration etc. (if yes please enclose information)? *Yes*
   see text

8. Additional Information:

Submitters are invited to provide any additional information about Properties of the proposed Character(s) or Script that will assist in correct understanding of and correct linguistic processing of the proposed character(s) or script. Examples of such properties are: Casing information, Numeric information, Currency information, Display behaviour information such as line breaks, widths etc., Combining behaviour, Spacing behaviour, Directional behaviour, Default Collation behaviour, relevance in Mark Up contexts, Compatibility equivalence and other Unicode normalization related information. See the Unicode standard at http://www.unicode.org for such information on other scripts. Also see Unicode Character Database (http://www.unicode.org/reports/tr44/) and associated Unicode Technical Reports for information needed for consideration by the Unicode Technical Committee for inclusion in the Unicode Standard.

[1] Form number: N4502-F (Original 1994-10-14; Revised 1995-01, 1995-04, 1996-04, 1996-08, 1999-03, 2001-05, 2001-09, 2003-11, 2005-01, 2005-09, 2005-10, 2007-03, 2008-05, 2009-11, 2011-03, 2012-01)

C. Technical - Justification

1. Has this proposal for addition of character(s) been submitted before? *No*
   If YES explain
2. Has contact been made to members of the user community (for example: National Body, user groups of the script or characters, other experts, etc.)? *Yes*
   If YES, with whom? *1401_software mailing list*
   If YES, available relevant documents: *see text*
3. Information on the user community for the proposed characters (for example: size, demographics, information technology use, or publishing use) is included? *Yes*
   Reference: *see text*
4. The context of use for the proposed characters (type of use; common or rare) *rare*
   Reference: *see text*
5. Are the proposed characters in current use by the user community? *Yes*
   If YES, where? Reference: *see text*
6. After giving due considerations to the principles in the P&P document must the proposed characters be entirely in the BMP? *preferable*
   If YES, is a rationale provided? *Yes*
   If YES, reference: *see text*
7. Should the proposed characters be kept together in a contiguous range (rather than being scattered)? *n/a*
8. Can any of the proposed characters be considered a presentation form of an existing character or character sequence? *No*
   If YES, is a rationale for its inclusion provided?
   If YES, reference: *see text*
9. Can any of the proposed characters be encoded using a composed character sequence of either existing characters or other proposed characters? *No*
   If YES, is a rationale for its inclusion provided?
   If YES, reference: *see text*
10. Can any of the proposed character(s) be considered to be similar (in appearance or function) to, or could be confused with, an existing character? *No*
    If YES, is a rationale for its inclusion provided?
    If YES, reference: *see text*
11. Does the proposal include use of combining characters and/or use of composite sequences? *No*
    If YES, is a rationale for such use provided?
    If YES, reference:
    Is a list of composite sequences and their corresponding glyph images (graphic symbols) provided? *No*
    If YES, reference:
12. Does the proposal contain characters with any special properties such as control function or similar semantics? *No*
    If YES, describe in detail (include attachment if necessary)
13. Does the proposal contain any Ideographic compatibility characters? *No*
    If YES, are the equivalent corresponding unified ideographic characters identified?
    If YES, reference: