Title: Proposal to Encode Section Marks for Siddham in ISO/IEC 10646
Source: Scripts Encoding Initiative (SEI)
Shingon Buddhist International Institute, Fresno, California
Author: Anshuman Pandey (pandey@umich.edu)
Status: Liaison Contribution
Action: For consideration by UTC and WG2
Date: 2012-09-30

1 Introduction

This is a proposal to encode fourteen new characters in the Siddham block of the Universal Character Set (ISO/IEC 10646). The proposed characters and their properties are:

GLYPH	CODE	CHARACTER NAME	GC	CCC	BIDI	MIRRORED	LINEBREAK
	115CA	SIDDHAM SECTION MARK-1	So	0	L	N	BA
	115CB	SIDDHAM SECTION MARK-2	So	0	L	N	BA
	115CC	SIDDHAM SECTION MARK-3	So	0	L	N	BA
	115CD	SIDDHAM SECTION MARK-4	So	0	L	N	BA
	115CE	SIDDHAM SECTION MARK-5	So	0	L	N	BA
	115D0	SIDDHAM SECTION MARK-6	So	0	L	N	BA
	115D1	SIDDHAM SECTION MARK-7	So	0	L	N	BA
	115D2	SIDDHAM SECTION MARK-8	So	0	L	N	BA
	115D3	SIDDHAM SECTION MARK-9	So	0	L	N	BA
	115D4	SIDDHAM SECTION MARK-10	So	0	L	N	BA
	115D5	SIDDHAM SECTION MARK-11	So	0	L	N	BA
	115CF	SIDDHAM SECTION MARK-12	So	0	L	N	BA
	115D6	SIDDHAM SECTION MARK-13	So	0	L	N	BA
	115D7	SIDDHAM SECTION MARK-14	So	0	L	N	BA

The Siddham script was proposed for encoding by the present author (see N4294 L2/12-234R) and accepted for inclusion in the UCS at the meeting of the Unicode Technical Committee in August 2012. These charac-

ters were identified in original proposal for Siddham, but were not proposed for inclusion because additional information regarding their usage was required.

2   Description

The section marks are used for structuring text, such as marking the ends of sentences, phrases, verses (*śloka*), and sections. They are attested in manuscripts (figures 1, 2) and script manuals (figures 3, 4, 5). Tokuzan describes them specifically as "symbols written at the end of a Sanskrit text" (1974: 227). Kodama writes that these marks are "attached at the end of sentence". In Shingon philosophy, as expressed by Jōgon and Kodama, the section marks represent an abbreviation of the 'aṃ'-character (1991: 222).

The section marks can be grouped into five categories based upon their graphical structure. Group 1 consists of marks based upon tridents radiating from a circle: ⁂ SECTION MARK-1, ⁂ SECTION MARK-2. Group 2 consists of marks based upon rays exuding from a circle: ⁂ SECTION MARK-3, ⁂ SECTION MARK-4, ⁂ SECTION MARK-5. Group 3 contains marks based upon concentric circles: ◎ SECTION MARK-6, ◒ SECTION MARK-7. Group 4 consists of marks formed from arcs above and below a circle: ⁑ SECTION MARK-8, ⁑ SECTION MARK-9, ⁑ SECTION MARK-10, ⁑ SECTION MARK-11. Group 5 contains marks based upon circles arranged in a lattice: ⁂ SECTION MARK-12, ⁂ SECTION MARK-13 ⁂ SECTION MARK-14.

According to users of the script from the Shingon community, the proposed characters are part of the Siddham canon. Beyond their function as punctuation used at the ends of sections, in Shingon philosophy the characters possess esoteric qualities that convey information regarding the interpretation of the text, such as a *mantra* or *dhāranī*, that precedes a section mark. They are required for the complete representation of Siddham documents.

3   References

淨嚴 [Jōgon]. 1682. 三密鈔 [*Shittan Sanmitsushō*]. In *Taishō Shinshū Daizōkyō* (大正新脩大藏經), vol. 84, no. 2710, 715–810. Accessed electronically at the SAT Daizōkyō Text Database `http://21dzk.l.u-tokyo.ac.jp/SAT/ddb-sat3.php?s=&mode=detail&useid=2710_,84`

児玉義隆 [Kodama, Giryū]. 1991. 梵字必携 ： 書写と解読 [*Bonji hikkei : shosha to kaidoku*]. 初版 [Shohan]. 大阪：朱鷺書房 [Ōsaka : Toki Shobō].

———. 2005 [平成 Heisei 17]. 梵字の書法：真言密教・ほとけの文字 [*Bonji no shohō : Shingon Mikkyō hotoke no moji*. 東京：大法輪閣 [Tōkyō : Daihōrinkaku].

Lokesh Chandra; Jiun Sonja. 1972. *Sanskrit Manuscripts from Japan*. Vols. 1 & 2. Śata-Piṭaka Series. v. 93 & 94. New Delhi.

Pandey, Anshuman. 2012. "Proposal to Encode the Siddham Script in ISO/IEC 10646". ISO/IEC JTC1/SC2/WG2 N4294 L2/12-234R. August 1, 2012. `http://std.dkuug.dk/jtc1/sc2/wg2/docs/n4294.pdf`

徳山暉純 [Tokuzan, Kijun]. 1974. 図說梵字 – 密敎の壁：悉曇參究 [*Zusetsu Bonji – mikkyō no kabe : shittan sankyū*]. 東京：木耳社 [Tōkyō : Mokujisha].

———. 1976 [昭和 Shōwa 51]. 梵字手帖 [*Bonji techō*]. 東京：木耳社 [Tōkyō : Mokujisha].

4   Acknowledgments

FACSIMILE OF THE TWO PALM-LEAVES OF HORIUZI, FROM THE ASHARAGIO.

SEE CATAL. OF JAPANESE BOOKS & MSS. IN THE BODLEIAN LIBRARY; No. 46.

ANECDOTA OXONIENSIA, ARYAN SERIES, VOL. I, PART III.

BUDDHIST TEXTS FROM JAPAN,

EDITED BY F. MAX MÜLLER.

AUTOTYPE, LONDON.

Figure 1: The use of ◌ SIDDHAM SECTION MARK-2 and ◌ SIDDHAM SECTION MARK-12 in the Hōryū-ji (法隆寺) *Prajñāpāramitā Hṛdaya Sūtra* manuscript (from Müller and Nanjio 1884: Plate II).

4

Figure 2: The use of ⁂ SIDDHAM SECTION MARK-14 and ⁂ SIDDHAM SECTION MARK-13 for indicating end of paragraph (from Lokesh Chandra 1972: 410).
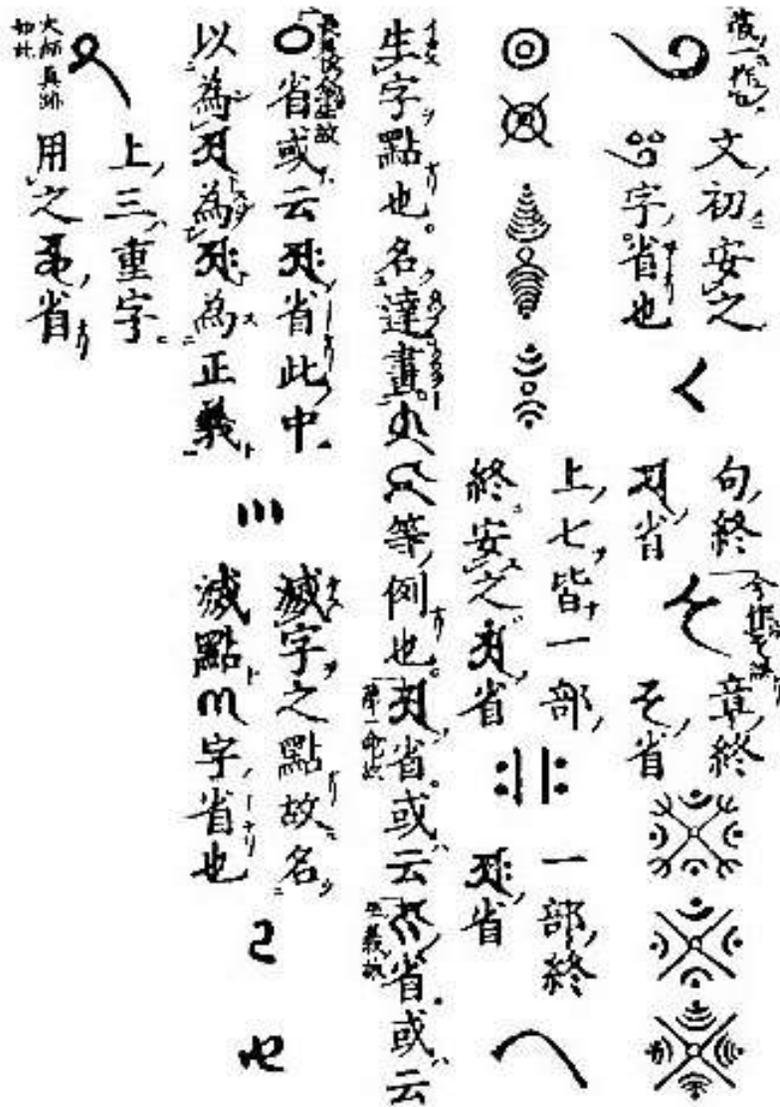
Figure 3: Description of various Siddham signs in the *Shittan Sanmitsushō* (from Jōgon 1682: 724). The section marks shown here are ※ SIDDHAM SECTION MARK-2, ※ SIDDHAM SECTION MARK-3, ※ SIDDHAM SECTION MARK-5, ◎ SIDDHAM SECTION MARK-6, a variant of ⊗ SIDDHAM SECTION MARK-7 with intersecting rays, ⚇ SIDDHAM SECTION MARK-11, ⚇ SIDDHAM SECTION MARK-8. Jōgon writes that "the seven characters above are all placed at the end of a section" and that they represent the abbreviation of the syllable "aṃ".

Figure 4: Punctuation and ornaments used in Siddham (from Tokuzan 1974: 227). Ornaments shown are ⚜ SIDDHAM SECTION MARK-1, ⚜ SIDDHAM SECTION MARK-2, ⚜ SIDDHAM SECTION MARK-3, ⚜ SIDDHAM SECTION MARK-5, ◎ SIDDHAM SECTION MARK-6, ⚷ SIDDHAM SECTION MARK-7, ⸙ SIDDHAM SECTION MARK-10, ⚜ SIDDHAM SECTION MARK-12. The character ꞏ||ꞏ has been accepted for encoding as U+1159C SIDDHAM END OF TEXT MARK.

記号	記号の説明
①	悉達(たたっ)（生の画）。母音符（a）を除いた子音字（半体文）を作るときに体文の下部につける。
②	文頭に書く。ｉ字の省略。
③	句末。読点にあたる。ma字の省略。
④	文末。句点にあたる。da字の省略。
⑤	畳句(じょうく)（同じ言葉をくり返すときに用いる）。cha字の省略。
⑥	減滅点（書き損じたときに、文字の上あるいは下にこの記号を符す）。na・字の省略。
⑦	文章の終りに符す。am・字の省略。
⑧	文章の終りに符す。ah・字の省略。

Figure 5: Punctuation and other signs used in Siddham (from Kodama 1991: 222). Column (7) shows the signs used at the end of a text section, specifically ※ SIDDHAM SECTION MARK-1, ※ SIDDHAM SECTION MARK-4, ≣ SIDDHAM SECTION MARK-9.

	1158	1159	115A	115B	115C	115D	115E	115F
0	11580	11590	115A0	115B0	115C0	115D0		
1	11581	11591	115A1	115B1	115C1	115D1		
2	11582	11592	115A2	115B2	115C2	115D2		
3	11583	11593	115A3	115B3	115C3	115D3		
4	11584	11594	115A4	115B4	115C4	115D4		
5	11585	11595	115A5	115B5	115C5	115D5		
6	11586	11596	115A6		115C6	115D6		
7	11587	11597	115A7		115C7	115D7		
8	11588	11598	115A8	115B8	115C8			
9	11589	11599	115A9	115B9	115C9			
A	1158A	1159A	115AA	115BA	115CA			
B	1158B	1159B	115AB	115BB	115CB			
C	1158C	1159C	115AC	115BC	115CC			
D	1158D	1159D	115AD	115BD	115CD			
E	1158E	1159E	115AE	115BE	115CE			
F	1158F	1159F	115AF	115BF	115CF			

The script is also known as 'Siddhamatrika'.

Independent vowels

11580	𑖀	SIDDHAM LETTER A
11581	𑖁	SIDDHAM LETTER AA
11582	𑖂	SIDDHAM LETTER I
11583	𑖃	SIDDHAM LETTER II
11584	𑖄	SIDDHAM LETTER U
11585	𑖅	SIDDHAM LETTER UU
11586	𑖆	SIDDHAM LETTER VOCALIC R
11587	𑖇	SIDDHAM LETTER VOCALIC RR
11588	𑖈	SIDDHAM LETTER VOCALIC L
11589	𑖉	SIDDHAM LETTER VOCALIC LL
1158A	𑖊	SIDDHAM LETTER E
1158B	𑖋	SIDDHAM LETTER AI
1158C	𑖌	SIDDHAM LETTER O
1158D	𑖍	SIDDHAM LETTER AU

Consonants

1158E	𑖎	SIDDHAM LETTER KA
1158F	𑖏	SIDDHAM LETTER KHA
11590	𑖐	SIDDHAM LETTER GA
11591	𑖑	SIDDHAM LETTER GHA
11592	𑖒	SIDDHAM LETTER NGA
11593	𑖓	SIDDHAM LETTER CA
11594	𑖔	SIDDHAM LETTER CHA
11595	𑖕	SIDDHAM LETTER JA
11596	𑖖	SIDDHAM LETTER JHA
11597	𑖗	SIDDHAM LETTER NYA
11598	𑖘	SIDDHAM LETTER TTA
11599	𑖙	SIDDHAM LETTER TTHA
1159A	𑖚	SIDDHAM LETTER DDA
1159B	𑖛	SIDDHAM LETTER DDHA
1159C	𑖜	SIDDHAM LETTER NNA
1159D	𑖝	SIDDHAM LETTER TA
1159E	𑖞	SIDDHAM LETTER THA
1159F	𑖟	SIDDHAM LETTER DA
115A0	𑖠	SIDDHAM LETTER DHA
115A1	𑖡	SIDDHAM LETTER NA
115A2	𑖢	SIDDHAM LETTER PA
115A3	𑖣	SIDDHAM LETTER PHA
115A4	𑖤	SIDDHAM LETTER BA
115A5	𑖥	SIDDHAM LETTER BHA
115A6	𑖦	SIDDHAM LETTER MA
115A7	𑖧	SIDDHAM LETTER YA
115A8	𑖨	SIDDHAM LETTER RA
115A9	𑖩	SIDDHAM LETTER LA
115AA	𑖪	SIDDHAM LETTER VA
115AB	𑖫	SIDDHAM LETTER SHA
115AC	𑖬	SIDDHAM LETTER SSA
115AD	𑖭	SIDDHAM LETTER SA
115AE	𑖮	SIDDHAM LETTER HA

Dependent vowel signs

115AF	◌𑖯	SIDDHAM VOWEL SIGN AA
115B0	𑖰◌	SIDDHAM VOWEL SIGN I
115B1	◌𑖱	SIDDHAM VOWEL SIGN II
115B2	◌𑖲	SIDDHAM VOWEL SIGN U
115B3	◌𑖳	SIDDHAM VOWEL SIGN UU
115B4	◌𑖴	SIDDHAM VOWEL SIGN VOCALIC R
115B5	◌𑖵	SIDDHAM VOWEL SIGN VOCALIC RR
115B6	▨	<reserved>
115B7	▨	<reserved>
115B8	◌𑖸	SIDDHAM VOWEL SIGN E
115B9	◌𑖹	SIDDHAM VOWEL SIGN AI
115BA	◌𑖺	SIDDHAM VOWEL SIGN O
		≡ 115B8 ◌𑖸 115AF ◌𑖯

115BB	◌𑖻	SIDDHAM VOWEL SIGN AU
		≡ 115B9 ◌𑖹 115AF ◌𑖯

Various signs

115BC	◌𑖼	SIDDHAM SIGN CANDRABINDU
115BD	◌𑖽	SIDDHAM SIGN ANUSVARA
115BE	◌𑖾	SIDDHAM SIGN VISARGA
115BF	◌𑖿	SIDDHAM SIGN VIRAMA
115C0	◌𑗀	SIDDHAM SIGN NUKTA
		• used for writing foreign sounds

Head mark

115C1	𑗁	SIDDHAM SIGN SIDDHAM
		= siddhirastu
		• used at the beginning of texts

Punctuation

115C2	𑗂	SIDDHAM DANDA
115C3	𑗃	SIDDHAM DOUBLE DANDA
115C4	𑗄	SIDDHAM SEPARATOR-1
		• used for separating syllables and words
115C5	𑗅	SIDDHAM SEPARATOR-2
		• used for separating syllables and words

Repetition marks

115C6	𑗆	SIDDHAM REPETITION MARK-1
115C7	𑗇	SIDDHAM REPETITION MARK-2
115C8	𑗈	SIDDHAM REPETITION MARK-3

Terminal mark

115C9	𑗉	SIDDHAM END OF TEXT MARK
		• indicates the completion of a text

Ornaments

These characters are used at the end of a section or text

115CA	𑗊	SIDDHAM SECTION MARK-1
115CB	𑗋	SIDDHAM SECTION MARK-2
115CC	𑗌	SIDDHAM SECTION MARK-3
115CD	𑗍	SIDDHAM SECTION MARK-4
115CE	𑗎	SIDDHAM SECTION MARK-5
115CF	𑗏	SIDDHAM SECTION MARK-6
115D0	𑗐	SIDDHAM SECTION MARK-7
115D1	𑗑	SIDDHAM SECTION MARK-8
115D2	𑗒	SIDDHAM SECTION MARK-9
115D3	𑗓	SIDDHAM SECTION MARK-10
115D4	𑗔	SIDDHAM SECTION MARK-11
115D5	𑗕	SIDDHAM SECTION MARK-12
115D6	𑗖	SIDDHAM SECTION MARK-13
115D7	𑗗	SIDDHAM SECTION MARK-14

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://www.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://www.dkuug.dk/JTC1/SC2/WG2/docs/summaryform.html**.**
See also http://www.dkuug.dk/JTC1/SC2/WG2/docs/roadmaps.html **for latest *Roadmaps*.**

A. Administrative

1. **Title:** *Proposal to Encode Section Marks for Siddham in ISO/IEC 10646*
2. Requester's name: *Script Encoding Initiative (SEI) / Anshuman Pandey <pandey@umich.edu>*
   Shingon Buddhist International Institute, Fresno
3. Requester type (Member body/Liaison/Individual contribution): *Liaison contribution*
4. Submission date: *2012-09-30*
5. Requester's reference (if applicable):
6. Choose one of the following:
   This is a complete proposal: *Yes*
   (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):
      Proposed name of script:
   b. The proposal is for addition of character(s) to an existing block:
      Name of the existing block: *Siddham*
2. Number of characters in proposal: *14*
3. Proposed category (select one from below - see section 2.2 of P&P document):
   A-Contemporary **X**   B.1-Specialized (small collection)   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?
   Anshuman Pandey and Lee Collins (Apple)
   b. Identify the party granting a license for use of the font by the editors (include address, e-mail, ftp-site, etc.):
   Shingon Buddhist International Institute and Lee Collins (Apple)
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 of proposal

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.

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? — *Shingon Buddhist community, Indic script experts*
   If YES, available relevant documents:

3. Information on the user community for the proposed characters (for example: size, demographics, information technology use, or publishing use) is included? — *Yes*
   Reference:

4. The context of use for the proposed characters (type of use; common or rare) — *Common*
   Reference:

5. Are the proposed characters in current use by the user community? — *Yes*
   If YES, where? Reference:

6. After giving due considerations to the principles in the P&P document must the proposed characters be entirely in the BMP? — *N/A*
   If YES, is a rationale provided?
   If YES, reference:

7. Should the proposed characters be kept together in a contiguous range (rather than being scattered)? — *Yes*

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:

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:

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:

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?
    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: