

Universal Coded Character Set  
International Organization for Standardization  
Organisation Internationale de Normalisation  
Международная организация по стандартизации

**Doc Type:** Working Group Document  
**Title:** Proposal to encode two Tangut ideographs  
**Source:** Eiso Chan (陈永聪/𐺑𐺑𐺑),  
Jerry You (游程宇/𐺑𐺑𐺑),  
Viacheslav Zaytsev (IOM RAS)  
**Status:** Individual Contribution  
**Action:** For consideration by JTC1/SC2/WG2 and UTC  
**Date:** 2024-10-14

## 1. Introduction

While indexing the Tangut head characters (= ideographs) presented in Hán Xiǎománg's recently published nine-volume dictionary of Tangut words (*Xìxiàwén Cídiǎn*, 2021), we identified two characters that are not currently represented among the Tangut characters included in the UCS and Unicode. Having checked this issue by original Tangut texts from the Western Xia period and consulted with other Tangut scholars (see Acknowledgements), we conclude that these two characters are used in the original texts and modern printed sources, are not variants of existing encoded characters, and need to be encoded. This document is a proposal to encode them in the UCS and Unicode as part of the Tangut Supplement block (U+18D00–U+18D7F) (see Table 1).

**Table 1. Proposed Tangut ideographs**

Code Point	Glyph	IDS	Source Reference	RS
*U+18D1D	𐺑𐺑𐺑	𐺑𐺑𐺑	H2021-309801	106.13
*U+18D1E	𐺑𐺑𐺑	𐺑𐺑𐺑	H2021-834001	579.14

## 2. Code Point Allocation and Character Properties

### 2.1. Code Point Allocation

To encode new (recently identified) Tangut ideographs, there are free code points only in the Tangut Supplement block (U+18D00–U+18D7F). The last occupied code point in this block is U+18D1C. This code point was occupied for further version of standard by the Tangut ideograph, which among others was proposed to be encoded in Proposal *WG2 N5217R = L2/23-246* (see References). In this regard, we propose to encode two new characters with the next two free code points \*U+18D1D and \*U+18D1E.

### 2.2. Unicode Character Properties

The entries for Tangut ideographs in the UnicodeData.txt file should be amended to:

```
18D00;<Tangut Ideograph Supplement, First>;Lo;0;L;;;;;N;;;;;
18D1E;<Tangut Ideograph Supplement, Last>;Lo;0;L;;;;;N;;;;;
```

### 2.3. Tangut Sources

The TangutSources.txt file should be modified to add a new source reference for the two proposed additional Tangut ideographs after the existing entry H2004:

```
# H2021 =Hán Xiǎománg (韓小忙), 西夏文词典: 世俗文献部分 (Xīxiàwén Cídiǎn: Shísú
Wénxiàn Bùfēn)
# [Tangut Word Dictionary: Secular Literature Part, 9 vols.]. 2021.
```

*Note:* We write the author’s name in traditional Chinese characters and the title of the work in simplified ones, as done for source L2006, although in fact both the author’s name and the title are written in simplified characters in the original book (see *Xīxiàwén Cídiǎn*, 2021 in References for full bibliographic information).

We propose the following format for coding new reference: H2021–dddddd, where the new source code H2021 comes first, followed by the ‘-’ delimiter used for other sources as well, followed by the six-digit character code in the source. The digits in the character code read as follows:

- 1 — volume number (1..9);
- 2–4 — page number (001..999) (page number is never four digits);
- 5–6 — sequence number of the character on the page (1..99).

The following entries should be added to the TangutSources.txt file:

```
U+18D1D kTGT_MergedSrc H2021-309801
U+18D1D kRSTUnicode 106.13
U+18D1E kTGT_MergedSrc H2021-834001
U+18D1E kRSTUnicode 579.14
```

If this is within the scope of consideration, we also propose to update Table 6 “Format details of the tags used in the source reference file for Tangut ideographs” of the standard text (*ISO/IEC 10646:2020 (E)*, p. 41) with a new source H2021 as follows:

Tag Value	Tag description	Third field format
kTGT_MergedSrc	Tangut source	(H2004-A-dddd), (H2004-B-dddd), (H2021-dddddd), (L1986-dddd), <...>

and paragraph “25.1 List of source references” (*ISO/IEC 10646:2020 (E)*, p. 40) as follows:

H2021 Hán Xiǎománg 2021 (Tangut Word Dictionary: Secular Literature Part, 9 vols.)

### 3. Evidence and Notes

#### 3.1. \*U+18D1D 𪗇

The ideograph was included as a head character entry in Hán Xiǎománg's dictionary and is not found in other modern Tangut dictionaries (see Fig. 1).

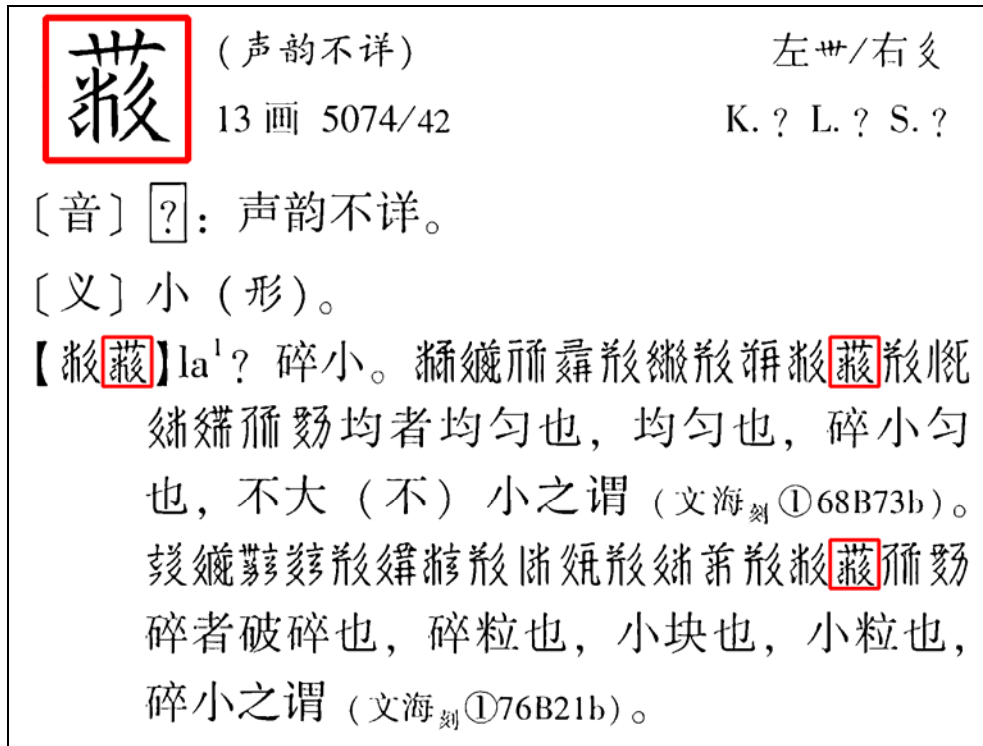


Fig. 1. *Xìxiàwén Cídiǎn*, 2021, vol. 3, p. 98

According to Hán Xiǎománg, it occurs in the explanatory text of two entries in the block-printed edition of the original Tangut dictionary called “Sea of Writing” (the title is known as *Wénhǎi* 文海 in modern Chinese translation from Tangut)

The first entry is for the head character 68B73 𪗇 [U+18099] (see Fig. 2, 3). The text of the entry says in Tangut:

𪗇: 𪗇𪗇𪗇𪗇 𪗇𪗇𪗇𪗇 𪗇𪗇𪗇𪗇 𪗇𪗇𪗇𪗇 𪗇𪗇𪗇𪗇 𪗇𪗇𪗇𪗇 𪗇𪗇𪗇𪗇 𪗇𪗇𪗇𪗇.

The second entry is for the head character 76B21 𪗇 [U+17B6D] = 𪗇 [U+17B67] (see Fig. 2, 3). The text of the entry says in Tangut:

𪗇: 𪗇𪗇𪗇𪗇 𪗇𪗇𪗇𪗇 𪗇𪗇𪗇𪗇 𪗇𪗇𪗇𪗇 𪗇𪗇𪗇𪗇 𪗇𪗇𪗇𪗇 𪗇𪗇𪗇𪗇 𪗇𪗇𪗇𪗇.

Hán Xiǎománg deciphered the meaning of proposed ideograph 𪗇 as “small” (Chinese *xiǎo* 小), and the meaning of binomial *la¹ ? 𪗇𪗇* (a compound word consisting of two characters), which includes the ideograph 𪗇, as “younger family members (sons, daughters)” (Chinese *suìxiǎo* 碎小). The character appears as part of this word in both dictionary entries of the “Sea of Writing”, which further confirms that it is the same character.

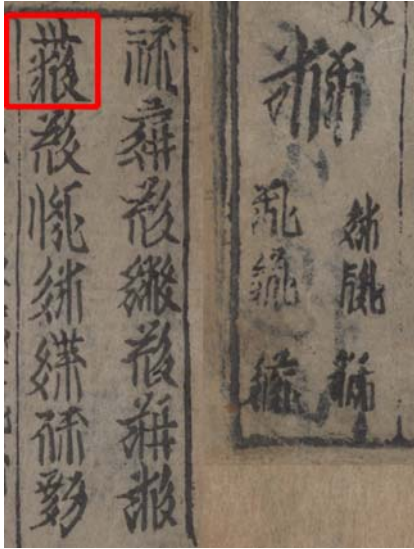
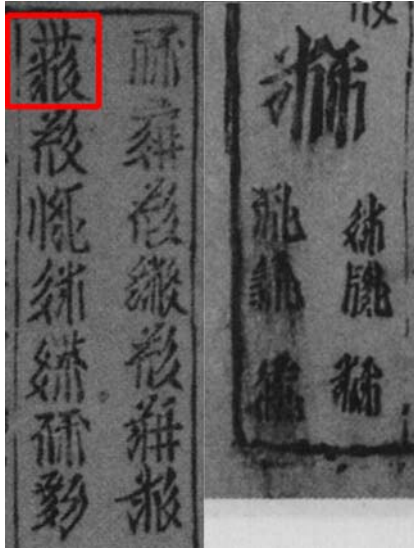
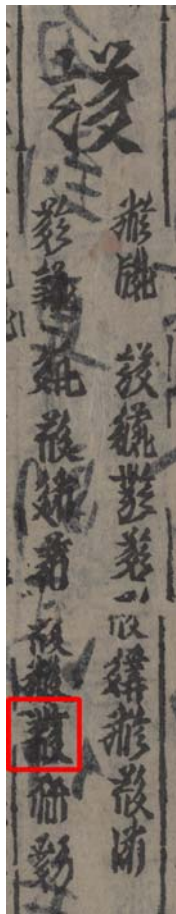
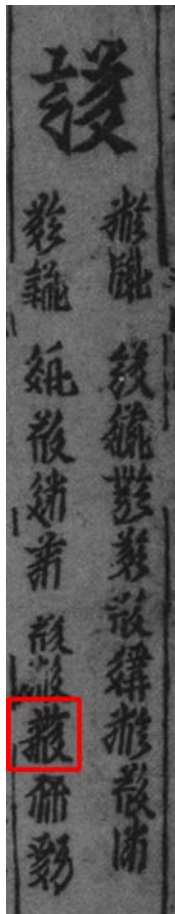

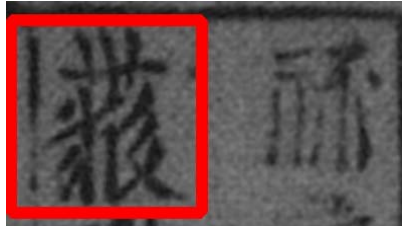

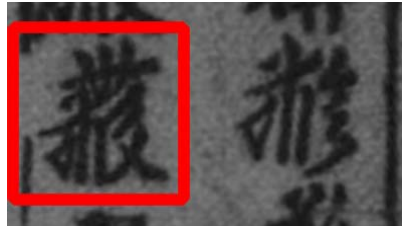
Folio	IOM RAS pressmark	
	Tang 14/1, old. inv. no. 211	Tang 14/2, old. inv. no. 212
folio 68b-69a, entry character 68B73 𦵑	 <p>(Kepping et al., 1969, pt. 1, p. 561-562; Shǐ Jīnbō et al., 1983, p. 622-623, 263-264)</p>	 <p>(ECHW 7, p. 153-154, photos 110-64, 110-65)</p>
folio 76b, entry character 76B21 𦵑	 <p>(Kepping et al., 1969, pt. 1, p. 569; Shǐ Jīnbō et al., 1983, p. 630, 279)</p>	 <p>(ECHW 7, p. 157, photo 110-72)</p>

Fig. 2. IOM RAS. *Sea of Writing I*

Folio	IOM RAS pressmark	
	Tang 14/1, old. inv. no. 211	Tang 14/2, old. inv. no. 212
folio 69a, entry character 68B73 𪛗	 <p>(Kepping et al., 1969, pt. 1, p. 562; Shǐ Jīnbō et al., 1983, p. 623, 264)</p>	 <p>(ECHW 7, p. 154, photo 110-65)</p>
folio 76b, entry character 76B21 𪛗	 <p>(Kepping et al., 1969, pt. 1, p. 569; Shǐ Jīnbō et al., 1983, p. 630, 279)</p>	 <p>(ECHW 7, p. 157, photo 110-72)</p>

**Fig. 3. IOM RAS. *Sea of Writing I* (enlarged fragments)**

Both head characters can be found in other modern dictionaries. In this regard, it is interesting to compare how other authors “read” the ideograph proposed for encoding in these two entries. The well-known Tangut-Chinese dictionary by Lǐ Fànwén for both head characters cites the same entries from “Sea of Writing”, but reads (= identifies) the ideograph and the compound word differently. In the first case as 𪛗 [U+17734] and 𪛗𪛗 (see Fig. 4), in the second as 𪛗 [U+1776A] and 𪛗𪛗 (see Fig. 5). Both graphic forms of the characters (𪛗 and 𪛗) do not fully correspond to the graphic form of the character in the original (see Fig. 2, 3) and, as is now clear, are simply erroneous readings (identifications) of it.

*Note 1:* Graphic form of U+17734 𪛗 is very similar to proposed \*U+18D1D 𪛗, but meaning of these ideographs is different. U+17734 𪛗 means “crooked, hollow”, and \*U+18D1D 𪛗 means “small”, meaning they are non-cognates.

*Note 2:* One of “Sea of Writing” printing copies preserved today (IOM RAS, Tang 14/1, old. inv. no. 211) seems to have an extra stroke (𪛗 instead of 𪛗) in the character 𪛗 that occurred in the entry 68B73 𪛗 (see Fig. 3). We believe that this is a mistake of the board carver, which is confirmed by the second copy of the dictionary (IOM RAS, Tang 14/2, old. inv. no. 212) (see Fig. 3). We should note that because of such inconsistencies (differences between different copies of the same work printed from the same set of boards, often accompanied by their repair and correction of text either at the time of printing or for the next print run), which are very common in Tangut written monuments, researchers’ conclusions (e.g., the reading of a particular ideograph) may depend on which copy of the text they used. We observe this in Lǐ Fànwén’s dictionary, where a copy of the “Sea of Writing” dictionary under pressmark Tang. 14/1, old. inv. no. 211 was clearly used.



3.2. \*U+18D1E 𪛗

The ideograph was included as a head character entry in Hán Xiǎománg's dictionary and is not found in other modern Tangut dictionaries (see Fig. 6).


	Ⅲ 切韵 2. 60 tji [底]	左 𪛗、𪛗/右 支
	15 画 1344/20	K. ? L. ? S. ?
[音] Ⅲ 切韵: 切韵 <sub>丁</sub> R70。		
2. 60 tji: 𪛗𪛗 tju nji (拟测 <sub>𪛗</sub> 切韵 R70)。		
[底]: 来源于同音字 (不同调 1. 67 Ⅲ68) 𪛗的对音, 见掌中珠 <sub>甲</sub> 22A12。		

Fig. 6. *Xixiàwén Cídiǎn*, 2021, vol. 8, p. 340–341

At the moment, the ideograph 𪛗 is documented only in one of the multiple manuscript editions of the Tangut rhyme tables, called “Corresponding Rhymes of the Five Sounds” (the title is known as *Wǔ Shēng Qiè Yùn* 五聲切韻 or *Wǔ Yīn Qiē Yùn* 五音切韻 in modern Chinese translation from Tangut). The character appears in Rhyme 70 on folio 73b of the Edition D of the Tables (see Fig. 7).

Hán Xiǎománg in his dictionary proposed *fǎnqiè* (反切; lit. ‘reverse cut’) of this ideograph, indicating pronunciation, as 𪛗𪛗 tju + nji = tji (see Fig. 6). The parts of these two characters compose the graphical form of the ideograph in question as well (𪛗𪛗).



Fig. 7. IOM RAS. *Rhyme Tables, Edition D*, folio 73b  
(*ECHW* 7, p. 354, photo 122–93; *Lǐ Fànwén*, 2006, p. 206)

## 4. References

### 4.1. Technical Documents

**ISO/IEC 10646:2020 (E).** *International Standard ISO/IEC 10646: Information technology — Universal coded character set (UCS)*. Sixth edition. 2020-12.

URL: [https://standards.iso.org/ittf/PubliclyAvailableStandards/c076835\\_ISO\\_IEC\\_10646\\_2020\(E\).zip](https://standards.iso.org/ittf/PubliclyAvailableStandards/c076835_ISO_IEC_10646_2020(E).zip)

**N5217R.** Andrew West. *Proposal to encode 2 Tangut components and 28 Tangut ideographs*. ISO/IEC JTC1/SC2/WG2 N5217R (L2/23-246). 2023-10-02.

WG2 URL: <https://www.unicode.org/wg2/docs/n5217r2-Tangut-components-ideographs.pdf>

L2 URL: <https://www.unicode.org/L2/L2023/23246-wg2-n5217r2.pdf>

### 4.2. Tangut Sources

**Rhyme Tables, Edition D.** *ngwə ·êi we mbju* 儻𑖇𑖇𑖇 [Corresponding Rhymes of the Five Sounds = *Wǔ Shēng Qiè Yùn* 五聲切韻]. Manuscript. IOM RAS, Tangut Collection. Pressmark Tang 22/4, old inventory no. 623. Published as Edition D (丁種本) in: ECHW 7, p. 308–368. Other facsimile publications: Lǐ Fànwén 2006, p. 6–283.

**Sea of Writing I.** *·iwə ngôn* 𑖇𑖇 [Sea of Writing = *Wénhǎi* 文海]. Blockprint. IOM RAS, Tangut Collection. Pressmark Tang 14/1, old inventory no. 211 (folios 04b–45a, 47b–93a). Pressmark Tang 14/2, old inventory no. 212 (folios 19a, 24b, 25a, 57b–78a). Published as Edition A (甲種本) in: ECHW 7, p. 122–176. Other facsimile publications (based on Tang 14/1): Kepping et al. 1969, pt. 1, p. 499–607; Shǐ Jīnbō et al. 1983, p. 559–668, cf. p. 135–355.

### 4.3. Academic Literature

**ECHW 7.** Shǐ Jīnbō 史金波, Wèi Tóngxián 魏同賢, and E. I. Kychanov (chief eds.). *É Cáng Hēishuǐchéng Wénxiàn* 俄藏黑水城文獻 = *Памятники письменности из Хара-Хото хранящиеся в России = Heishuicheng Manuscripts Collected in Russia*. Vol. 7. Shànghǎi 上海: Shànghǎi gǔjí chūbǎnshè 上海古籍出版社, 1997. ISBN 7-5325-2213-X.

**Kepping et al., 1969.** Кепинг, К. Б. [K. B. Kepping], В. С. Колоколов [V. S. Kolokolov], Е. И. Кычанов [E. I. Kychanov], and А. П. Терентьев-Катанский [A. P. Terentiev-Katansky] (trans. from Tangut, introductory papers and appendices). *Море письмен. Факсимиле тангутских ксилографов. Часть 1–2* [Sea of Writing. Facsimile of the Tangut blockprints. Part 1–2]. Москва [Moscow]: Издательство «Наука», Главная редакция восточной литературы [“Nauka” Publishing House, The Chief Editorial Office for Oriental Literature], 1969. (Памятники письменности Востока [Written Monuments of the Orient]; XXV, [1]–[2]).

**Lǐ Fànwén, 2006.** Lǐ Fànwén 李范文 (author), Jiǎ Chángyè 贾常业 (chief ed.), Lǐ Hǎitāo 李海涛 (deputy chief ed.). “*Wǔyīn Qiēyùn*” yǔ “*Wénhǎi Bǎoyùn*” Bǐjiào Yánjiū 《五音切韵》与《文海宝韵》比较研究 [Comparative Study of *Wǔyīn Qiēyùn* and *Wénhǎi Bǎoyùn*]. In *Xīxià Yánjiū* 西夏研究 [Xixia Studies], Part II. Běijīng 北京: Zhōngguó shèhuì kēxué chūbǎnshè 中国社会科学出版社, 2006. ISBN 7-5004-5688-3.

**Shǐ Jīnbō et al., 1983.** Shǐ Jīnbō 史金波, Bái Bīn 白滨, Huáng Zhènghuà 黄振华. *Wénhǎi Yánjiū* 文海研究 [Study of the Sea of Writing]. Běijīng 北京: Zhōngguó shèhuì kēxué chūbǎnshè 中国社会科学出版社, 1983.



- Xià-Hàn Zìdiǎn, 2008.** Lǐ Fànwén 李範文 (comp.) and Jiǎ Chángyè 賈常業 (rev. and exp.). *Xià-Hàn Zìdiǎn* 夏漢字典 [Tangut-Chinese Dictionary (Revised Ed.)]. Běijīng 北京: Zhōngguó shèhuì kēxué chūbǎnshè 中國社會科學出版社, 2008. ISBN 978-7-5004-2113-9.
- Xīxiàwén Cídiǎn, 2021.** Hán Xiǎománg 韓小忙 (comp.). *Xīxiàwén Cídiǎn (Shísú Wénxiàn Bùfēn)* 西夏文词典（世俗文献部分）：全 9 册 [Tangut Word Dictionary (Secular Literature Part). 9 vols.]. Běijīng 北京: Zhōngguó shèhuì kēxué chūbǎnshè 中国社会科学出版社, 2021. (*Xīxià Wénzì yǔ Wénxiàn Yánjiū* 西夏文字与文献研究 [Studies of Tangut characters and documents]). ISBN 978-7-5203-8011-9.

## 5. Acknowledgements

We are grateful to the Institute of Oriental Manuscripts of the Russian Academy of Sciences (IOM RAS, Saint Petersburg, Russia) for encouraging our research and providing color digital images of Tangut manuscripts from their Tangut Collection for illustration purpose in this proposal, to Andrew West (魏安) and Sūn Bójūn (孫伯君), Tangut scholars from the UK and China respectively, for reading the draft and helpful comments.

## 6. Proposal Summary Form

**SO/IEC JTC 1/SC 2/WG 2  
PROPOSAL SUMMARY FORM TO ACCOMPANY SUBMISSIONS  
FOR ADDITIONS TO THE REPERTOIRE OF ISO/IEC 10646<sup>1</sup>**

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:	<b>Proposal to encode two Tangut ideographs</b>
2. Requester's name:	<i>Eiso Chan, Jerry You, Viacheslav Zaytsev</i>
3. Requester type (Member body/Liaison/Individual contribution):	<i>Individual contribution</i>
4. Submission date:	<i>2024-10-14</i>
5. Requester's reference (if applicable):	
6. Choose one of the following:	
This is a complete proposal:	<input checked="" type="checkbox"/> YES
(or) More information will be provided later:	<input type="checkbox"/>

### B. Technical – General

1. Choose one of the following:	
a. This proposal is for a new script (set of characters):	<input type="checkbox"/> NO
Proposed name of script:	
b. The proposal is for addition of character(s) to an existing block:	<input checked="" type="checkbox"/> YES
Name of the existing block:	<i>Tangut Supplement</i>
2. Number of characters in proposal:	<i>2</i>
3. Proposed category (select one from below - see section 2.2 of P&P document):	
A-Contemporary <input type="checkbox"/>	B.1-Specialized (small collection) <input type="checkbox"/>
B.2-Specialized (large collection) <input type="checkbox"/>	C-Major extinct <input type="checkbox"/>
D-Attested extinct <input type="checkbox"/>	E-Minor extinct <input type="checkbox"/>
F-Archaic Hieroglyphic or Ideographic <input checked="" type="checkbox"/>	G-Obscure or questionable usage symbols <input type="checkbox"/>
4. Is a repertoire including character names provided?	
a. If YES, are the names in accordance with the "character naming guidelines" in Annex L of P&P document?	<input checked="" type="checkbox"/> YES
b. Are the character shapes attached in a legible form suitable for review?	<input checked="" type="checkbox"/> YES
5. Fonts related:	
a. Who will provide the appropriate computerized font to the Project Editor of 10646 for publishing the standard?	<i>Andrew West</i>
b. Identify the party granting a license for use of the font by the editors (include address, e-mail, ftp-site, etc.):	<i>Andrew West</i>
6. References:	
a. Are references (to other character sets, dictionaries, descriptive texts etc.) provided?	<input checked="" type="checkbox"/> YES
b. Are published examples of use (such as samples from newspapers, magazines, or other sources) of proposed characters attached?	<input checked="" type="checkbox"/> 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)?	<input type="checkbox"/> NO

### 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.

<sup>1</sup> Form number: N4102-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? If YES explain	NO
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.)? If YES, with whom? If YES, available relevant documents:	YES <i>Tangut experts</i>
3. Information on the user community for the proposed characters (for example: size, demographics, information technology use, or publishing use) is included? Reference:	NO
4. The context of use for the proposed characters (type of use; common or rare) Reference:	Rare
5. Are the proposed characters in current use by the user community? If YES, where? Reference:	YES
6. After giving due considerations to the principles in the P&P document must the proposed characters be entirely in the BMP? If YES, is a rationale provided? If YES, reference:	NO
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? If YES, is a rationale for its inclusion provided? If YES, reference:	NO
9. Can any of the proposed characters be encoded using a composed character sequence of either existing characters or other proposed characters? If YES, is a rationale for its inclusion provided? If YES, reference:	NO
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? If YES, is a rationale for its inclusion provided? If YES, reference:	NO
11. Does the proposal include use of combining characters and/or use of composite sequences? 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:	NO
12. Does the proposal contain characters with any special properties such as control function or similar semantics? If YES, describe in detail (include attachment if necessary)	NO
13. Does the proposal contain any Ideographic compatibility characters? If YES, are the equivalent corresponding unified ideographic characters identified? If YES, reference:	NO