

To: Unicode Technical Committee
 From: Debbie Anderson, SEI, UC Berkeley
 Subject: Comments on [L2/20-061R](#) Final Proposal to encode Western Cham in the UCS
 Date: April 16, 2020

The following comments are meant to assist in the review of Western Cham proposal [L2/20-061R](#), with additional comments for the author of [L2/20-061R](#).

1. Changes in the Western Cham proposal dated March 25, 2020 ([L2/20-061R](#)) from the January 28, 2020 version ([L2/20-061](#)):

- a. WESTERN CHAM CONSONANT SIGN TKAJ KA was moved in [L2/20-061R](#) from the set of Consonant Signs (U+1E237) to the set of Final Consonants (U+1E24E). Discussion and rationale are included in [L2/20-063](#), the Response to Western Cham Script Adhoc Comments L2/20-046

There has been some confusion on the author's part over this character. But it is a final. This begs the question why it is therefore not listed among the finals. First it is confusable with U+1E24D (SIGN FINAL H), which encouraged moving it away from that spot. The rest of the finals list is encoded to parallel the list in Eastern Cham. While none of these are very strong reasons for not moving it, neither is it a problem where it is. If it has to move, then the recommendation is that U+1E24D be change general category from Mc to Lo with a requisite name change to LETTER FINAL H and Tkaj ka move to U+1E24E

- b. WESTERN CHAM LETTER FINAL H was changed to Lo from Mc. (The earlier name [in [L2/20-061](#)] was WESTERN CHAM CONSONANT SIGN FINAL H.)

- c. Attached is a version of the Western Cham proposal [L2/20-061R](#) with highlighting by Debbie Anderson showing significant new text or changes to the earlier version. Notable is new text was added in the Repertoire section and in the Finals section.

2. Responses to [L2/20-046](#) Script Ad Hoc Recs Jan 2020 (with the comment repeated in bold), responding to the r3 version of the 2019 proposal ([L2/19-217r3](#))

- **General**

- **Encoding model**
[Script Ad Hoc]

Justify the encoding model and justify the phonetic order of left-side combining marks.

[Response by Martin Hosken in [L2/20-063](#)]

Justify the encoding model. There are two parts to this question. The first is that the proposer is expected to justify their encoding model against some arbitrary encoding suggestion, hinted at in the adhoc comments. If the proposer of this alternative model feels strongly that their suggestion is in the best interests of the community, the encoding and the standard, then the author would welcome them making a formal counter proposal and would engage with that. On that basis, all discussion of such a final character marker has been removed from the proposal.

The second question of justifying why a logical order is required seems counter to all other proposals where a visual order has to be justified because a logical order is presumed. It is not the duty of a proposal writer to justify all the presuppositional bases of the Unicode standard and its principles for every new character and script. Again, if the logical order is considered wrong, then a counter proposal would be welcomed. As to justification, the statement that Western Cham follows the same encoding model to Eastern Cham is considered sufficient.

- **Chart**

[Script Ad Hoc]

Provide a chart of the characters, if possible. (One example is found under "ALPHABET ČAM DU CAMBODGE" at

<https://archive.org/details/dictionnaireamf00cabagoog/page/n24>, although the work dates to 1906 and varies from the proposed repertoire.)

[Response by Martin Hosken in [L2/20-063](#)]

The charts provided have been stamped with the authority of the community leader. They are about as authoritative as one can get. They are certainly more authoritative than something published in a book that hasn't been checked by the community. Figures 49 and 50 have been added for those readers who must have a page number on a sample document. No attempt has been made to find the particular book references.

[Comment from Debbie Anderson]

Pages 49 and 50 are handwritten charts with page numbers.

- **Repertoire**

- **SIGN TANA TAMAT AYET**

[Script Ad Hoc]

Move U+1E263 SIGN TANA TAMAT AYET to the Supplemental Punctuation block, since it occurs at the boundary between Arabic and Western Cham.

[Response by Martin Hosken in [L2/20-063](#)]

Moving punctuation characters. One school of thought considers that all punctuation characters that might be used in other scripts should go into a common block of punctuation. It is noticeable that everything in the supplemental punctuation block is there because it shares between at least one script and Latin script. Here there is no sharing with Latin script. The proposal follows the school of thought that one follows the script and keeps as much of the script in one block as one can. This is particularly true for non-Roman scripts. Therefore the proposer asks the UTC to justify its requirement that the character be moved and compare with other scripts that share characters between scripts not including Latin.

- **SIGN TANA TAMAT TAKUE**

[Script Ad Hoc]

Are there examples of U+1E264 SIGN TANA TAMAT TAKUE beside Arabic text? If so, it should also be moved to Supplemental Punctuation block.

[Comment from Debbie Anderson]

No additional example added beyond figure 7, which was in earlier version of the proposal.

○ **ARABIC END OF TEXT**

[Script Ad Hoc]

The ARABIC END OF TEXT appears to be a sequence of U+1E263 SIGN TANA TAMAT AYET characters repeated, and getting smaller, in a manner similar to the way the dandas extend out to fill the end of the line in figure 22. Based on the evidence, we recommend this character be removed.

[Comment from Debbie Anderson]

No new discussion in the revised version of the proposal; below is what the proposal says. The character is still in the proposal and in a figure.

[Wording from the proposal L2/20-061R]

U+061D (ARABIC END OF TEXT) is an Arabic symbol that is used in Arabic text only. As such it is proposed for inclusion in the Arabic block. It is distinct from a sequence of U+1E263 (SIGN TANA TAMAT AYET) as can be seen in Figure 51.

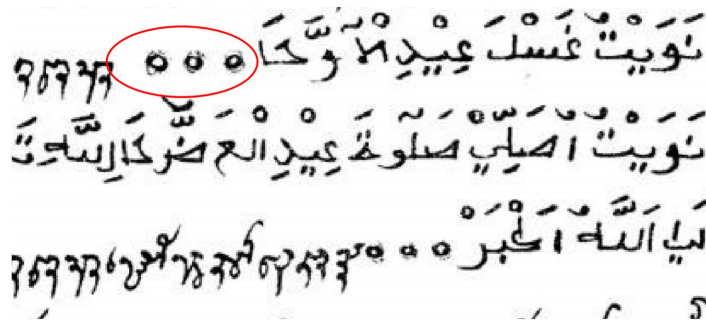


Fig 51. Examples of sequences of U+1E263 (SIGN TANA TAMAT AYET). Notice the space after the left most one on line 2.

○ **TRIPLE DANDA**

[Script Ad Hoc]

In our opinion, the “TRIPLE DANDA” as shown in figure 20 is actually a sequence of double dandas and single dandas. In our view, the TRIPLE DANDA should be removed as a separate character.

[Comment from Debbie Anderson]

No new discussion in the revised version of the proposal; below is what the proposal says. The character is still in the proposal.

[Wording from the proposal L2/20-061R]

Western Cham has two dandas. The double danda U+1E25E (PUNCTUATION DOUBLE DANDA) occurs on its own (Figures 3, 26, 27, 34, 46) as a sentence final symbol and also as part of a triple danda. The U+1E25F (PUNCTUATION TRIPLE DANDA) occurs in different styles (see Figures 20 and 22), but they all represent the same semantic

paragraph/verse end. This is a single concept to users rather than some arbitrary sequence of double and single dandas that a user might want to type, making it effectively impossible to search for and select the triple danda. In handwritten texts the variability occurs within the same document, but this is not expected in print.

The only known occurrences of a single danda are as part of a triple danda. But it is assumed that the single danda does exist, given its existence in the triple danda, and has simply fallen out of use. To this end, a space is left should the community decide to encode it later. Devanagari dandas are also used, for example to break lists into items.

[Comments from Debbie Anderson]

The caption for figure 20 says, “A sequence of double and single dandas often referred to as a triple danda.” The wording in the caption seems to support the opinion by the Script Ad Hoc: the triple danda is a sequence of single and double dandas. Indeed, figure 22 contains a single danda, though the character was removed in the 2020 versions of the proposal.

Figure 20:

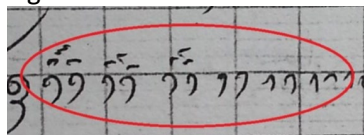
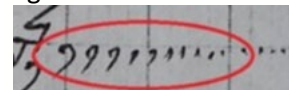


Figure 22 with “U+1E25Dm” the single danda. One circled portion of figure 22 shows:



Among the list of changes on page 16 of the proposal ([L2/20-061R](#)) is: “Reintroduce U+1E25F (PUNCTUATION TRIPLE DANDA) in response to direct request from community.” However, the TRIPLD DANDA was in the various versions of [L2/19-217](#) and [L2/20-061](#). Presumably it was removed at some point, then re-inserted.

- **Properties / Other**

- **[Script Ad Hoc]**

- Note that “Final vowels” in the “Combining Orders” chart don’t work with USE’s cluster pattern when U+1E232 VOWEL SIGN U precedes. “Medial ya” + “Medial wa” are also not allowed by USE. These cases may not currently be used actively, but they should be discussed if the author is trying to specify the “combining order”.

- **[Comment from Norbert Lindenberg]**

- It’s true – the USE currently doesn’t allow multiple medial consonants on the same side of the base. If these two can occur together, then the USE needs to be updated to allow this.)

- **[Script Ad Hoc]**

- What does “vowel” or “consonant” refer to in the proposal: its phonological analysis, Unicode Indic properties, or the OpenType USE shaping model?

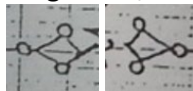
- [Script Ad Hoc]
Revise the section on “Combining Orders” and write about “cluster pattern” generally.

- **Figures**

- **Figure 23**

[Script Ad Hoc]

In figure 23, what are the following? Which scripts do they occur beside?



[Response by Martin Hosken in [L2/20-063](#)]

The debated shape in Figure 23 has been analysed as a non-textual decoration that they do not want to be encoded.

- **Figures 24-5**

[Script Ad Hoc]

What are figs. 24-25? Where do they come from and what is the context?

[Comment from Debbie Anderson]

No response to this question, but presumably they come from the source as the charts, which have been stamped with the authority of the community leader.

3. Additional comments on L2/20-061R by Debbie Anderson:

- Fix the names in Unicode properties on page 7 for Lunar Symbols.
- Correct the character name for U+1E262 in figures 2, 5, and 8 to SIGN TANA CHAPEDAEM IU.
- Figure 17 shows FINAL N, not T.
- Figure 19 shows an error mark represented by U+0353, but this is not discussed in the prose section of the proposal.
- Figure 31 says “SIGN BRAY”, but the character is now LIGATURE BRAY.
- The caption for Figure 44 needs to be update to the name LETTER FINAL H.
- Figure 24 lists 1E22D, but no character has been proposed at that code point. It appears to be 1E205.
- Figure 24 has “1E201”, presumably a typo for 1E210.
- Is there an example of 1E21B PPA? (Note: It was difficult to confirm whether each character appeared in a figure. Many vowels and consonants are attested only in the handwritten lists in figures 24 and 25. A more organized listing of all the characters and figures would be helpful.)
- Remove figure 22 if the author does not mean to propose the single danda.

4. Other comments (by Debbie Anderson)

a. Printed materials in the script

The proposal ([L2/20-063](#)):states (p. 2): “It should be noted that there is very little print material in Western Cham. All the core samples are taken from handwritten books and documents.”

However, the response document [L2/20-018](#) provides print examples from textbooks and teacher training materials, which have been used with the permission of the Cambodian Ministry of Education.

Other examples in the response doc come from the “seasonal magazine” Mukva (figs. 5, 6, 8, 9, 10). One of the co-authors of some publications is himself a prominent member of the Imam San community, and his way of writing Western Cham reflects how he was taught by his father and grandfather.

The authors of the response document [L2/20-018](#) also state that the current Western Cham proposal reflects one of the several pre-existing spelling systems, namely, the one followed by the current leadership of the Krom Kan Imam San. The authors of [L2/20-018](#) state, however, it does not reflect materials currently deposited in the National Library of Cambodia.

b. Disagreement

Response in [L2/20-063](#):

It is strongly hoped that this proposal is the final version. The language community is very small and has been worn down by the incessant request for yet more examples and justification. There is a reason Socrates was executed.

Looking forward, there are signs that the two schools of thought on how Western Cham should be written using the script may resolve their differences. On this basis, there may well be a proposal to add what the discussions decide is missing. But without an initial encoding in Unicode, such discussions may not happen. Communities need to be able work with what they have if they are to be able develop. On this basis, this proposal is proffered as sufficient for the needs of the Western Cham as they understand their script at the moment.

Personal comment from Jorge Lopez Cortina (personal communication to Debbie Anderson):

I believe it would be better for everyone to have a Western Cham block that includes at least both spelling systems currently used. That way, even if in several decades one becomes universal, old books could still be encoded and read. Leaving the system used by the current teachers, publishers, and scholars, and ignoring all the existing printed Western Cham materials (crucially, the textbooks) can only lead to divisions and difficulties in implementation.

Universal Multiple-Octet Coded Character Set
International Organization for Standardization
Organisation Internationale de Normalisation

Doc Type: Working Group Document
Title: Final Proposal to encode Western Cham in the UCS
Source: Martin Hosken
Status: Individual contribution
Action: For consideration by UTC and ISO
Date: 2020-03-25

Executive Summary. This proposal is to add 89 characters to a block: 1E200-1E26F in the supplementary plane, named Western Cham. In addition, 1 character is proposed for inclusion in the Arabic Extended A block: 08A0-08FF and 8 characters are proposed for inclusion in the Arabic Mathematical Alphabetic Symbols block: 1EEF0-1EEFF. The proposal is a revision of L2/20-061.

A revision history is given at the end of the main text of this document.

Introduction. Eastern Cham is already encoded in the range AA00-AA5F. As stated in N4734, Western Cham while closely related to Eastern Cham, is sufficiently different in style for Eastern Cham characters in plain text to be unintelligible to Western Cham readers and therefore a separate block is required. As such, the proposal to encode Western Cham in a separate block from (Eastern) Cham can be considered a disunification. But since Western Cham was never supported in the (Eastern) Cham block, it can be argued that this is a proposal for a new script. This is further discussed in the section on disunification.

The Western Cham encoding follows the same encoding model as for (Eastern) Cham. There is no halant or virama that calls for a Brahmic model. The basic structure is a base character followed by a sequence of marks as described in the section on combining orders. The basic orthographic syllable structure is: (C | I) M* V* F? with relative orders as shown in the section on combining orders. A final may also be a spacing base character.

The proposal in N4734 has been closely reviewed by script experts and community leaders among the Western Cham in Cambodia and this proposal arises from their considerations. Where a direct correspondance exists between a Western Cham character and an Eastern Cham character, the character is encoded in a corresponding position in the Western Cham block. Extra final characters are used in Western Cham for words borrowed from Khmer. Special ligatures and symbols are also added. These result in the Western Cham block needing an extra column over Eastern Cham.

The Western Cham language is written in 3 scripts: Arabic (Jawi), Western Cham and some informal Latin Romanizations. Speakers of the Western Cham language are found almost entirely in Cambodia. There are small pockets of speakers in Vietnam, but they do not use the Western Cham script, only the Jawi script.

Disunification. While there are some similarities between many characters of both scripts and some consistent features between both scripts, the scripts have diverged for a long period of time in which there was little contact between the two communities. Eastern Cham plain text characters are not identifiable for Western Cham readers, unless people have been trained by someone knowledgeable in both scripts and vice versa. Contact between the two groups today is very limited and exchange nearly unknown, which means that exposure to the other group's script is virtually zero for more than more than 99% of the population. Therefore the probability of them learning from each other is small. There are probably only a handful of Western Cham people who could even identify Eastern Cham letters as being Eastern Cham, let alone being able to transpose the letters in their minds. The Western Cham community would feel very strongly, not to say violently, about one of their texts being displayed in Eastern Cham characters by default and the same is almost certainly true for the Eastern Cham people as well.







The languages themselves also have diverged to the point of no longer being mutually intelligible, with even basic words like ‘house’ being different. Apart from being geographically and politically isolated from each other, the two groups of Cham people are also separated in culture and religion, the Western Cham being Muslims and the Eastern Cham predominately Buddhist.




Reports of Eastern Cham people from Vietnam visiting a Western Cham village in Cambodia show that the Eastern Cham could not be understood by young or old. The attempt to introduce Eastern Cham letters into that Western Cham community was rejected strongly.



Some fonts have been made to use Eastern Cham codepoints for Western Cham, and while they are somewhat successful, it is necessary for them to use unassigned codepoints from the block to complete basic Western Cham support. One thing such fonts do indicate is a desire to keep the same basic encoding model between the two scripts. It should also be noted that there is very little print material in Western Cham. All the core samples are taken from handwritten books and documents.

Repertoire. The basic structure of Western Cham text consists of initial consonants: U+1E206 (LETTER KA) ... U+1E228 (LETTER HA); medial consonants: U+1E233 (CONSONANT SIGN YA) ... U+1E236 (CONSONANT SIGN WA); dependent vowels: U+1E229 (VOWEL SIGN AA) ... U+1E232 (VOWEL SIGN U) and final consonants: U+1E23D (FINAL PH) ... U+1E24E (CONSONANT SIGN TKAJ KA). There are also a few independent vowels: U+1E200 (LETTER A) ... U+1E205 (LETTER OH). The overall structure of the Western Cham block follows that of Eastern Cham closely, removing some correspondances and introducing many new characters.

1. The 6 textual symbols U+1E260 (SIGN TANA TAMAT PHON) ... U+1E264 (SIGN TANA TAMAT TAKUE) and U+061D (ARABIC END OF TEXT) are textual symbols with discourse functions. Specifically:

Codepoint	Character name	Glyph	Description	Script	Figure
1E260	SIGN TANA TAMA PHON		Closing Korannic quotation	Arabic	47
1E261	SIGN TANA CHAPEDAEM PHON		Section beginning	Cham	4
1E262	SIGN TANA CHAPEDAEM IU		Alternate section beginning	Cham	2, 5
1E263	SIGN TANA TAMAT AYET		Closing affirmation	Arabic	6
1E264	SIGN TANA TAMAT TAKUE		Closing chants	Arabic	7
061D	ARABIC END OF TEXT		Closing text	Arabic	8, 23

In styling, U+1E260 ( SIGN TANA TAMA PHON) and U+1E262 ( SIGN TANA CAPEDAEM IU) may lose their dotted outline and U+1E261 ( SIGN TANA CHAPEDAEM PHON) may gain one. The script indication for each symbol gives its script context. Arabic script symbols occur in Western Cham script, but often at the boundary of some Arabic text either in Arabic language or Western Cham language. But as can be seen in the examples, some symbols claiming to be used for Arabic script also get used in Western Cham script. It is felt, therefore, best to encode these as part of a Western Cham block and to allow their use in Arabic script.

U+061D ( ARABIC END OF TEXT) is an Arabic symbol that is used in Arabic text only. As such it is proposed for inclusion in the Arabic block. It is distinct from a sequence of U+1E263 ( SIGN TANA TAMAT AYET) as can be seen in Figure 51.

2. There are three special ligatures, after Arabic style ligatures, which have contrasting rendering with the normal sequence. Both the ligature and the normal sequence may occur in text. The sequence is the compatibility normal form for the ligature.

Codepoint	Character name	Glyph	Sequence	Glyphs	Figure
1E23A	LIGATURE BRAY		1E21D 1E234 1E22C		31
1E23B	LIGATURE NAN		1E217 1E246		24
1E23C	LIGATURE SHALOM		1E227 1E224 1E24C		3, 13

All other ligatures are rendered the same as their character sequence.

- U+1E25B (𑄛 SIGN TANA PATOK SAP) reduplication mark (Figure 45) and U+1E25C (𑄜 PUNCTUATION SPIRAL) ‘one’ or ‘first’ (Figure 46) are the same as in N4374. Note that while U+1E25C (𑄜 PUNCTUATION SPIRAL) has lexical value, it takes no diacritics.
- U+1E24E (𑄞 CONSONANT SIGN TKAJ KA) is an abstract character that only ever ligates with the previous character (figures 1, 2, 15, 16). It does not occur following a vowel and is also not followed by a final letter either. Its occurrence after a consonant usually indicates a vowel or it may indicate that the consonant is acting as a final. Ligated sequences involving U+1E24E have a visual similarity to those using U+1E24D (𑄝 CONSONANT SIGN FINAL H) (Figures 25, 44). For example:

Sequence	Glyphs	Meaning	Sequence	Glyphs	Meaning
1E206 1E24E		tie	1E206 1E24D		push away
1E208 1E24E		stutter	1E208 1E24D		side

U+1E24D (CONSONANT SIGN FINAL H) may occur following a vowel as in

- The character U+1E25A (𑄟 SIGN MOON TEN) is used when describing days in the lunar calendar. U+1E24C (𑄚 FINAL M) is used together with a digit to mark the days after a full moon, e.g. 𑄟 U+1E251 (𑄠 DIGIT 1) U+1E24C (𑄚 FINAL M) for the first day after the full moon. U+1E25A (𑄟 SIGN MOON TEN) is used for the 10s digit in such contexts: thus 𑄟 𑄟 U+1E25A (𑄟 SIGN MOON TEN) U+1E251 (𑄠 DIGIT 1) U+1E24C (𑄚 FINAL M) for the eleventh day after a full moon. The new moon occurs on day 14 or 15. The days after the new moon work in the same way but are marked with U+1E22E (𑄞 SIGN OE) instead of U+1E24C (𑄚 FINAL M), with the full moon occurring on day 15. For example 10 days after a full moon is represented by 𑄟 U+1E25A (𑄟 SIGN MOON TEN) U+1E22E (𑄞 SIGN OE). See Figure 48.

U+1E25A (𑄟 SIGN MOON TEN) is not a digit, in that it has no numeric value.

- Western Cham has special characters, unique to itself and unknown to other languages, that can be used for indicating month names. They are based on the transliterated month name into Arabic followed by 1 to 7 dots. Since these are Arabic symbols, they are proposed for inclusion in the Arabic Mathematical Alphabetic Symbols block. The order of the months is U+1EEFF (𑄛 LUNAR 7 DOTS), U+1EEF9 (𑄛 LUNAR 2 DOTS), U+1EEFA (𑄛 LUNAR 3 DOTS), U+1EEFC (𑄛 LUNAR 5 DOTS), U+1EEFE (𑄛 LUNAR 6 DOTS), U+1EEF8 (𑄛 LUNAR 1 DOT), U+1EEF9 (𑄛 LUNAR 2 DOTS), U+1EEFB (𑄛 LUNAR 4 DOTS), U+1EEFC (𑄛 LUNAR 5 DOTS), U+1EEFF (𑄛 LUNAR 7 DOTS), U+1EEF8 (𑄛 LUNAR 1 DOT), U+1EEFB (𑄛 LUNAR 4 DOTS). Notice the repetitions in the order. These symbols are used both in Western Cham script and also in Arabic script, when writing Western Cham. They may occur next to each other and with Western Cham digits but never with Arabic digits.

In addition, there are two forms of the lunar 5 dots, that occur in free variation in a text, hence U+1EEFD (𑄛 LUNAR 5 DOTS OVER HEH).

In font design, these characters contrast to Arabic characters in their position. Western Cham conceptually hangs from a baseline and the Lunar characters are positioned relative to that upper baseline. See Figure 10.

- The reserved codepoints at *1E20B*, *1E211* and *1E22D* correspond to characters in Eastern Cham that do not occur in Western Cham. The slots are kept to keep character alignment with the Eastern Cham encoding.
- The reserved codepoints at *1E203* and *1E204*, correspond to characters in Eastern Cham that while they occur in Western Cham have no visual distinction from other characters in the block. Thus

Potential Codepoint and Eastern Cham Correspondance	Sequence	Glyphs
<i>1E203</i> (AA03)	1E201 1E22F	ᵇᵇ
<i>1E204</i> (AA04)	1E201 1E230	ᶇᵇ

Note that the Western Cham analogue of U+AA03 (letter e) U+AA34 (sign ra) is ᵇᶇᶇᵇ U+1E201 (ᵇᵇ A) U+1E234 (ᶇᵇ SIGN RA) U+1E22F (ᵇᵇ SIGN E), which also appropriately resolves the shaping issue.

- Western Cham has two dandas. The double danda U+1E25E (ᶇᶇ PUNCTUATION DOUBLE DANDA) occurs on its own (Figures 3, 26, 27, 34, 46) as a sentence final symbol and also as part of a triple danda. The U+1E25F (ᶇᶇᶇᶇᶇᶇ PUNCTUATION TRIPLE DANDA) occurs in different styles (see Figures 20 and 22), but they all represent the same semantic paragraph/verse end. This is a single concept to users rather than some arbitrary sequence of double and single dandas that a user might want to type, making it effectively impossible to search for and select the triple danda. In handwritten texts the variability occurs within the same document, but this is not expected in print. The only known occurrences of a single danda are as part of a triple danda. But it is assumed that the single danda does exist, given its existence in the triple danda, and has simply fallen out of use. To this end, a space is left should the community decide to encode it later. Devanagari dandas are also used, for example to break lists into items.

Ligation. Western Cham is visually highly ligating. For example, ᵇᶇ U+1E206 (ᵇᶇ KA) U+1E22A (ᵇᵇ VOWEL I) and then ᵇᶇ U+1E206 (ᵇᶇ KA) U+1E236 (ᵇᵇ SIGN WA) U+1E22A (ᵇᵇ VOWEL I). This calls for most non-initial characters to be marks. Spacing characters are reserved for those characters that do not ligate with anything previous to them, which consists only of initials and most finals.

Confusability. The Western Cham alphabet has a number of characters that are visually equivalent to letter sequences. These sequences can occur in text and font designers are advised to ensure a visual contrast to avoid confusion.

Codepoint	Character name	Glyph	Confusable with	Glyphs	Meaning
1E201	LETTER I	ᵇᵇ	1E224 1E24E	ᵇᵇ	<i>does not occur</i>
1E256	DIGIT SIX	ᶇᶇ	1E223 1E236 1E24D	ᶇᶇ	choose
1E257	DIGIT SEVEN	ᶇᶇ	1E223 1E24E 1E240	ᶇᶇ	vein
1E259	DIGIT NINE	ᶇᶇ	1E21D 1E24E	ᶇᶇ	carry (on shoulder)
1EEFA	LUNAR 3 DOTS	ᵇᵇᵇ	08B5		
1EEFB	LUNAR 4 DOTS	ᵇᵇᵇᵇ	0690	ᵇᵇᵇᵇ	

In addition, U+1E251 (ᶇᶇ DIGIT ONE) can be confusable with U+1E24D (ᶇᶇ FINAL H) in some fonts, although a good font will endeavour to contrast them.

Variants. There are two ‘literary’ variants for a pair of vowels. Such variants are used in religious and ethical texts. The examples are shown following an arbitrary base character of U+1E223 (ᶇᶇ RA).

Sequence	Normal	Literary variant
1E223 1E22E 1E232		
1E223 1E233 1E229		

There is also a free stylistic font variant of U+1E235 (ൊ CONSONANT SIGN LA) ോ (figure 14).

Finals. For most finals there is a visual link between the initial and the final:

Initial	Encoding	Final	Encoding	Initial	Encoding	Final	Encoding
	1E206		1E240		1E220		1E24C
	1E20A		1E242		1E222		1E248
			1E22E		1E223		1E249
	1E20C		1E244		1E224		1E24A
	1E213		1E245		1E225		1E23E
	1E217		1E246		1E226		1E24B
	1E21C		1E23D		1E227		1E23F
	1E21A		1E247		1E228		1E24D

Most finals are productive modifications of their corresponding initials. See figures 32-41, 43, 44.

U+1E22E (ൈ VOWEL SIGN OE) and U+1E24C (൉ SIGN FINAL M) are both considered marks because they are non-spacing (figure 25). U+1E22E (ൈ VOWEL SIGN OE) acts as a vowel or a final and there may be a sequence involving both of them. For example: ൎ ൈ ൎ `bank (of river)` U+1E21D (ൎ LETTER BA) U+1E217 (ൎ LETTER NUE) U+1E22E (ൈ VOWEL SIGN OE) U+1E22E (ൈ VOWEL SIGN OE). Those viewing this script through a Sanskrit lens may want to refer to U+1E24C (൉ SIGN FINAL M) as an anusvara, but this is an alien term and concept to Western Cham speakers and is not used in a number of scripts in the region. Likewise for U+1E24D (൏ FINAL H) as visarga, which also loses its connotation in a number of scripts in the region.

Spelling. The details of the relative ordering of marks in a sequence are listed in the section on character properties. Here we list some exemplar spellings.

Word	Spelling	Notes
	1E21A 1E214 1E233 1E230	Pre vowels go before the last initial.
	1E206 1E22E 1E232	No visual difference based on order.
	1E215 1E223 1E229 1E22E	Interaction between upper and final vowels.
	1E200 1E233 1E229	Interaction between following medial and following vowel.
	1E253 1E22E	For lunar days, also 1E24C

Layout. Text is laid out left to right, although due to mixing with Arabic script, regularly occurs in a bidirectional context. There are no word spaces in Western Cham. Spaces act like or with a comma as per Khmer. Western Cham follows other Indic fonts in considering base characters as hanging from a baseline rather than building up from one. This even applies to the use of Arabic shapes in Western Cham symbols.

Punctuation. In addition to specific Western Cham punctuation marks, the script also borrows punctuation from other scripts. Regular use is made of Western European punctuation: « U+00AB (LEFT-POINTING DOUBLE ANGLE QUOTATION MARK), » U+00BB (RIGHT-POINTING DOUBLE ANGLE QUOTATION MARK), “ U+201C (LEFT DOUBLE QUOTATION MARK), ” U+201D (RIGHT DOUBLE QUOTATION MARK), ‘ U+2018 (LEFT SINGLE QUOTATION MARK), ’ U+2019 (RIGHT QUOTATION MARK), ? U+003F (QUESTION MARK), ! U+0021 (EXCLAMATION MARK), (U+0028 (LEFT PARENTHESIS),) U+0029 (RIGHT PARENTHESIS), [U+005B (LEFT SQUARE BRACKET),] U+005D (RIGHT SQUARE BRACKET).

In addition: I U+0964 (DEVANAGARI DANDA), II U+0965 (DEVANAGARI DOUBLE DANDA).

Unicode Properties. These are the Unicode database properties for the characters proposed.

```
061D;ARABIC END OF TEXT;Po;0;ON;;;;;N;;;;;
1E200;WESTERN CHAM LETTER A;Lo;0;L;;;;;N;;;;;
1E201;WESTERN CHAM LETTER E;Lo;0;L;;;;;N;;;;;
1E202;WESTERN CHAM LETTER O;Lo;0;L;;;;;N;;;;;
1E205;WESTERN CHAM LETTER OH;Lo;0;L;;;;;N;;;;;
1E206;WESTERN CHAM LETTER KA;Lo;0;L;;;;;N;;;;;
1E207;WESTERN CHAM LETTER KHA;Lo;0;L;;;;;N;;;;;
1E208;WESTERN CHAM LETTER GA;Lo;0;L;;;;;N;;;;;
1E209;WESTERN CHAM LETTER GHAA;Lo;0;L;;;;;N;;;;;
1E20A;WESTERN CHAM LETTER NGUE;Lo;0;L;;;;;N;;;;;
1E20C;WESTERN CHAM LETTER CHA;Lo;0;L;;;;;N;;;;;
1E20D;WESTERN CHAM LETTER CHHA;Lo;0;L;;;;;N;;;;;
1E20E;WESTERN CHAM LETTER JA;Lo;0;L;;;;;N;;;;;
1E20F;WESTERN CHAM LETTER JHA;Lo;0;L;;;;;N;;;;;
1E210;WESTERN CHAM LETTER NHUE;Lo;0;L;;;;;N;;;;;
1E212;WESTERN CHAM LETTER NHJA;Lo;0;L;;;;;N;;;;;
1E213;WESTERN CHAM LETTER TA;Lo;0;L;;;;;N;;;;;
1E214;WESTERN CHAM LETTER THA;Lo;0;L;;;;;N;;;;;
1E215;WESTERN CHAM LETTER DA;Lo;0;L;;;;;N;;;;;
1E216;WESTERN CHAM LETTER DHAA;Lo;0;L;;;;;N;;;;;
1E217;WESTERN CHAM LETTER NUE;Lo;0;L;;;;;N;;;;;
1E218;WESTERN CHAM LETTER NA;Lo;0;L;;;;;N;;;;;
1E219;WESTERN CHAM LETTER DDA;Lo;0;L;;;;;N;;;;;
1E21A;WESTERN CHAM LETTER PA;Lo;0;L;;;;;N;;;;;
1E21B;WESTERN CHAM LETTER PPA;Lo;0;L;;;;;N;;;;;
1E21C;WESTERN CHAM LETTER PHA;Lo;0;L;;;;;N;;;;;
1E21D;WESTERN CHAM LETTER BA;Lo;0;L;;;;;N;;;;;
1E21E;WESTERN CHAM LETTER BHAA;Lo;0;L;;;;;N;;;;;
1E21F;WESTERN CHAM LETTER MUE;Lo;0;L;;;;;N;;;;;
1E220;WESTERN CHAM LETTER MA;Lo;0;L;;;;;N;;;;;
1E221;WESTERN CHAM LETTER BBA;Lo;0;L;;;;;N;;;;;
1E222;WESTERN CHAM LETTER YA;Lo;0;L;;;;;N;;;;;
1E223;WESTERN CHAM LETTER RA;Lo;0;L;;;;;N;;;;;
1E224;WESTERN CHAM LETTER LA;Lo;0;L;;;;;N;;;;;
1E225;WESTERN CHAM LETTER VA;Lo;0;L;;;;;N;;;;;
1E226;WESTERN CHAM LETTER SSA;Lo;0;L;;;;;N;;;;;
1E227;WESTERN CHAM LETTER SA;Lo;0;L;;;;;N;;;;;
1E228;WESTERN CHAM LETTER HA;Lo;0;L;;;;;N;;;;;
1E229;WESTERN CHAM VOWEL SIGN AA;Mn;0;NSM;;;;;N;;;;;
1E22A;WESTERN CHAM VOWEL SIGN I;Mn;0;NSM;;;;;N;;;;;
1E22B;WESTERN CHAM VOWEL SIGN II;Mn;0;NSM;;;;;N;;;;;
1E22C;WESTERN CHAM VOWEL SIGN AI;Mn;0;NSM;;;;;N;;;;;
1E22E;WESTERN CHAM VOWEL SIGN OE;Mn;0;NSM;;;;;N;;;;;
1E22F;WESTERN CHAM VOWEL SIGN E;Mc;0;L;;;;;N;;;;;
1E230;WESTERN CHAM VOWEL SIGN AAI;Mc;0;L;;;;;N;;;;;
1E231;WESTERN CHAM VOWEL SIGN AU;Mn;0;NSM;;;;;N;;;;;
1E232;WESTERN CHAM VOWEL SIGN U;Mn;0;NSM;;;;;N;;;;;
1E233;WESTERN CHAM CONSONANT SIGN YA;Mc;0;L;;;;;N;;;;;
1E234;WESTERN CHAM CONSONANT SIGN RA;Mc;0;L;;;;;N;;;;;
1E235;WESTERN CHAM CONSONANT SIGN LA;Mn;0;NSM;;;;;N;;;;;
1E236;WESTERN CHAM CONSONANT SIGN WA;Mc;0;NSM;;;;;N;;;;;
1E23A;WESTERN CHAM LIGATURE BRAY;Lo;0;L;;;;;N;;;;;
1E23B;WESTERN CHAM LIGATURE NAN;Lo;0;L;;;;;N;;;;;
1E23C;WESTERN CHAM LIGATURE SHALOM;Lo;0;L;;;;;N;;;;;
1E23D;WESTERN CHAM LETTER FINAL PH;Lo;0;L;;;;;N;;;;;
1E23E;WESTERN CHAM LETTER FINAL V;Lo;0;L;;;;;N;;;;;
1E23F;WESTERN CHAM LETTER FINAL S;Lo;0;L;;;;;N;;;;;
1E240;WESTERN CHAM LETTER FINAL K;Lo;0;L;;;;;N;;;;;
1E242;WESTERN CHAM LETTER FINAL NG;Lo;0;L;;;;;N;;;;;
1E244;WESTERN CHAM LETTER FINAL CH;Lo;0;L;;;;;N;;;;;
1E245;WESTERN CHAM LETTER FINAL T;Lo;0;L;;;;;N;;;;;
1E246;WESTERN CHAM LETTER FINAL N;Lo;0;L;;;;;N;;;;;
1E247;WESTERN CHAM LETTER FINAL P;Lo;0;L;;;;;N;;;;;
1E248;WESTERN CHAM LETTER FINAL Y;Lo;0;L;;;;;N;;;;;
1E249;WESTERN CHAM LETTER FINAL R;Lo;0;L;;;;;N;;;;;
1E24A;WESTERN CHAM LETTER FINAL L;Lo;0;L;;;;;N;;;;;
1E24B;WESTERN CHAM LETTER FINAL SS;Lo;0;L;;;;;N;;;;;
1E24C;WESTERN CHAM CONSONANT SIGN FINAL M;Mn;0;NSM;;;;;N;;;;;
1E24D;WESTERN CHAM LETTER FINAL H;Lo;0;L;;;;;N;;;;;
1E24E;WESTERN CHAM CONSONANT SIGN TKAJ KA;Mc;0;NSM;;;;;N;;;;;
1E250;WESTERN CHAM DIGIT ZERO;Nd;0;L;;0;0;N;;;;;
```

1E251;WESTERN CHAM DIGIT ONE;Nd;0;L;;1;1;1;N;;;;;
 1E252;WESTERN CHAM DIGIT TWO;Nd;0;L;;2;2;2;N;;;;;
 1E253;WESTERN CHAM DIGIT THREE;Nd;0;L;;3;3;3;N;;;;;
 1E254;WESTERN CHAM DIGIT FOUR;Nd;0;L;;4;4;4;N;;;;;
 1E255;WESTERN CHAM DIGIT FIVE;Nd;0;L;;5;5;5;N;;;;;
 1E256;WESTERN CHAM DIGIT SIX;Nd;0;L;;6;6;6;N;;;;;
 1E257;WESTERN CHAM DIGIT SEVEN;Nd;0;L;;7;7;7;N;;;;;
 1E258;WESTERN CHAM DIGIT EIGHT;Nd;0;L;;8;8;8;N;;;;;
 1E259;WESTERN CHAM DIGIT NINE;Nd;0;L;;9;9;9;N;;;;;
 1E25A;WESTERN CHAM DIGIT MOON TEN;No;0;L;;;N;;;;;
 1E25B;WESTERN CHAM SIGN TANA PATOK SAP;Lm;0;L;;;N;;;;;
 1E25C;WESTERN CHAM PUNCTUATION SPIRAL;Po;0;ON;;;N;;;;;
 1E25E;WESTERN CHAM PUNCTUATION DOUBLE DANDA;Po;0;ON;;;N;;;;;
 1E25F;WESTERN CHAM PUNCTUATION TRIPLE DANDA;Po;0;ON;;;N;;;;;
 1E260;WESTERN CHAM SIGN TANA TAMAT PHON;Po;0;ON;;;N;;;;;
 1E261;WESTERN CHAM SIGN TANA CHAPEDAEM PHON;Po;0;ON;;;N;;;;;
 1E262;WESTERN CHAM SIGN TANA CHAPEDAEM IU;Po;0;ON;;;N;;;;;
 1E263;WESTERN CHAM SIGN TANA TAMAT AYET;Po;0;ON;;;N;;;;;
 1E264;WESTERN CHAM SIGN TANA TAMAT TAKUE;Po;0;ON;;;N;;;;;

 1EEF8;ARABIC SYMBOL ONE DOT LUNAR MONTH;So;0;ON;;;N;;;;;
 1EEF9;ARABIC SYMBOL TWO DOTS LUNAR MONTH;So;0;ON;;;N;;;;;
 1EEFA;ARABIC SYMBOL THREE DOTS LUNAR MONTH;So;0;ON;;;N;;;;;
 1EEFB;ARABIC SYMBOL FOUR DOTS LUNAR MONTH;So;0;ON;;;N;;;;;
 1EEFC;ARABIC SYMBOL FIVE DOTS LUNAR MONTH;So;0;ON;;;N;;;;;
 1EEFD;ARABIC SYMBOL FIVE DOTS LUNAR MONTH OVER HEH;So;0;ON;;;N;;;;;
 1EEFE;ARABIC SYMBOL SIX DOTS LUNAR MONTH;So;0;ON;;;N;;;;;
 1EEFF;ARABIC SYMBOL SEVEN DOTS LUNAR MONTH;So;0;ON;;;N;;;;;

U+061D (﷌ ARABIC END OF TEXT) has a joining behaviour of non-joining as do all the lunar symbols in the range U+1EEF8 .. U+1EEFF.

Collation. For the most part Western Cham follows Eastern Cham with regard to default collation:

1E200 (𑜀) < 1E201 (𑜁) < 1E202 (𑜂) < 1E201 (𑜁) 1E22F (𑜀) < 1E201 (𑜁) 1E230 (𑜀) < 1E205 (𑜅)
 1E206 (𑜆) < 1E207 (𑜇) < 1E208 (𑜈) < 1E209 (𑜉) < 1E20A (𑜊) < 1E20C (𑜌) < 1E20D (𑜍) <
 1E20E (𑜎) < 1E20F (𑜏) < 1E210 (𑜐) < 1E212 (𑜒) < 1E213 (𑜓) < 1E214 (𑜔) < 1E215 (𑜕) <
 1E216 (𑜖) < 1E217 (𑜗) < 1E218 (𑜘) < 1E219 (𑜙) < 1E21A (𑜚) < 1E21B (𑜛) < 1E21C (𑜜) <
 1E21D (𑜝) < 1E21E (𑜞) < 1E21F (𑜟) < 1E220 (𑜠) < 1E221 (𑜡) < 1E222 (𑜢) < 1E223 (𑜣) < 1E224
 (𑜤) < 1E225 (𑜥) < 1E226 (𑜦) < 1E227 (𑜧) < 1E228 (𑜨)
 1E233 (𑜩) < 1E234 (𑜪) < 1E235 (𑜫) < 1E236 (𑜬)
 1E229 (𑜭) < 1E22A (𑜮) < 1E22B (𑜯) < 1E22C (𑜰) < 1E22F (𑜀) < 1E230 (𑜀) < 1E231 (𑜱) < 1E232
 (𑜲)
 1E240 (𑜳) < 1E242 (𑜵) < 1E22E (𑜱) < 1E244 (𑜶) < 1E245 (𑜷) < 1E246 (𑜸) < 1E247 (𑜹) <
 1E23D (𑜺) < 1E24C (𑜻) < 1E248 (𑜼) < 1E249 (𑜽) < 1E24A (𑜾) < 1E23E (𑜿) < 1E24B (𺀀) < 1E23F
 (𺀁) < 1E24D (𺀂) < 1E24E (𺀃)
 1E21D 1E234 1E22C << 1E23A (𺀄)
 1E217 1E246 << 1E23B (𺀅)
 1E227 1E224 1E24C << 1E23C (𺀆)

The punctuation, digits and other signs sort appropriately with similar characters from other scripts. Special handling is required for handling U+1E22E as a vowel:

U+1E22C U+1E23D < U+1E22E U+1E23D
 U+1E22C U+1E23E < U+1E22E U+1E23E
 U+1E22C U+1E23F < U+1E22E U+1E23F
 U+1E22C U+1E240 < U+1E22E U+1E240
 U+1E22C U+1E241 < U+1E22E U+1E241

U+1E22C U+1E242 < U+1E22E U+1E242
 U+1E22C U+1E22E < U+1E22E U+1E22E
 U+1E22C U+1E244 < U+1E22E U+1E244
 U+1E22C U+1E245 < U+1E22E U+1E245
 U+1E22C U+1E246 < U+1E22E U+1E246
 U+1E22C U+1E247 < U+1E22E U+1E247
 U+1E22C U+1E248 < U+1E22E U+1E248
 U+1E22C U+1E249 < U+1E22E U+1E249
 U+1E22C U+1E24A < U+1E22E U+1E24A
 U+1E22C U+1E24B < U+1E22E U+1E24B
 U+1E22C U+1E24C < U+1E22E U+1E24C
 U+1E22C U+1E24D < U+1E22E U+1E24D

Combining Orders. Due to the immutability of Canonical Combining Class (CCC) values and the confusion between whether the CCC value is an order or a class, it is required that the CCC values for all characters in this block SHALL BE 0.¹ This does not mean that relative ordering of marks is not important. Here is a list of mark classes in their relative order. Those occurring first are listed first:

Class	Characters
Medial ra	Ꞩ 1E234 (CONSONANT SIGN RA)
Medial la	ꞩ 1E235 (CONSONANT SIGN LA)
Medial ya	Ɦ 1E233 (CONSONANT SIGN YA)
Medial wa	Ɜ 1E236 (CONSONANT SIGN WA)
Pre vowels	Ɡ 1E22F (VOWEL SIGN E), Ɬ U+1E230 (VOWEL SIGN AAI)
Final vowels	Ɪ 1E229 (VOWEL SIGN AA), ꞯ 1E22A (VOWEL SIGN I), Ʞ 1E22B (VOWEL SIGN II), Ʇ 1E22C (VOWEL SIGN AI), Ʝ 1E231 (VOWEL SIGN AU)
Upper vowels	Ꭓ 1E22E (VOWEL SIGN OE)
Lower vowels	Ꞵ 1E232 (VOWEL SIGN U)
Finals	ꞵ 1E24C (CONSONANT SIGN FINAL M), Ꞷ 1E24D (CONSONANT SIGN FINAL H), ꞷ 1E24E (CONSONANT SIGN TKAJ KA)

No more than one character in each class may occur following the same base character U+1E200 (Ʞ LETTER A) ... 1E228 (ꞰꞰ LETTER HA). The only exception to this U+1E22E (VOWEL SIGN OE), which can be a vowel or a final and may occur twice in sequence. The Finals class also includes all the spacing final letters U+1E23D (ꞰꞰꞰ FINAL PH) ... U+1E24B (ꞰꞰꞰ FINAL SS). From this we can establish a set of basic script level regular expressions. This description allows sequences that would not occur in Western Cham but is sufficient for rendering and keyboarding.

- [1E200-1E228] 1E234? 1E235? 1E233? 1E236? [1E22F-1E230]? [1E229-1E22C 1E231]? 1E22E? 1E22E? 1E232? [1E23D-1E24D 1E24E]?
- [1E23A-1E23C]

Indic Properties. Based on the ordering rules above, we can set the following Indic properties for Western Cham.

1E200..1E205 ; Vowel_Independent
 1E206..1E228 ; Consonant
 1E229..1E232 ; Vowel_Dependent
 1E233..1E236 ; Consonant_Medial

¹ It is unfortunate that the effect of making CCC values immutable means they have to be set to 0, resulting in the exact opposite effect that making CCC values immutable intends to achieve of enforcing a consistent order.

1E23A..1E23C ; Consonant_Placeholder
1E23D..1E24E ; Consonant_Final
1E250..1E25A ; Digits

1E229..1E22E ; Top
1E22F..1E230 ; Left
1E231 ; Top
1E232 ; Bottom
1E233 ; Right
1E234 ; Left
1E235 ; Bottom
1E236 ; Right
1E24C ; Top
1E24E ; Right

	1E20	1E21	1E22	1E23	1E24	1E25	1E26
0	ᩈᩢ᩠ᨦ 1E200	ᩈᩢ᩠ᨦ 1E210	ᩈᩢ᩠ᨦ 1E220	ᩈᩢ᩠ᨦ 1E230	ᩈᩢ᩠ᨦ 1E240	ᩈᩢ᩠ᨦ 1E250	ᩈᩢ᩠ᨦ 1E260
1	ᩈᩢ᩠ᨦ 1E201		ᩈᩢ᩠ᨦ 1E221	ᩈᩢ᩠ᨦ 1E231		ᩈᩢ᩠ᨦ 1E251	ᩈᩢ᩠ᨦ 1E261
2	ᩈᩢ᩠ᨦ 1E202	ᩈᩢ᩠ᨦ 1E212	ᩈᩢ᩠ᨦ 1E222	ᩈᩢ᩠ᨦ 1E232	ᩈᩢ᩠ᨦ 1E242	ᩈᩢ᩠ᨦ 1E252	ᩈᩢ᩠ᨦ 1E262
3		ᩈᩢ᩠ᨦ 1E213	ᩈᩢ᩠ᨦ 1E223	ᩈᩢ᩠ᨦ 1E233		ᩈᩢ᩠ᨦ 1E253	ᩈᩢ᩠ᨦ 1E263
4		ᩈᩢ᩠ᨦ 1E214	ᩈᩢ᩠ᨦ 1E224	ᩈᩢ᩠ᨦ 1E234	ᩈᩢ᩠ᨦ 1E244	ᩈᩢ᩠ᨦ 1E254	ᩈᩢ᩠ᨦ 1E264
5	ᩈᩢ᩠ᨦ 1E205	ᩈᩢ᩠ᨦ 1E215	ᩈᩢ᩠ᨦ 1E225	ᩈᩢ᩠ᨦ 1E235	ᩈᩢ᩠ᨦ 1E245	ᩈᩢ᩠ᨦ 1E255	
6	ᩈᩢ᩠ᨦ 1E206	ᩈᩢ᩠ᨦ 1E216	ᩈᩢ᩠ᨦ 1E226	ᩈᩢ᩠ᨦ 1E236	ᩈᩢ᩠ᨦ 1E246	ᩈᩢ᩠ᨦ 1E256	
7	ᩈᩢ᩠ᨦ 1E207	ᩈᩢ᩠ᨦ 1E217	ᩈᩢ᩠ᨦ 1E227		ᩈᩢ᩠ᨦ 1E247	ᩈᩢ᩠ᨦ 1E257	
8	ᩈᩢ᩠ᨦ 1E208	ᩈᩢ᩠ᨦ 1E218	ᩈᩢ᩠ᨦ 1E228		ᩈᩢ᩠ᨦ 1E248	ᩈᩢ᩠ᨦ 1E258	
9	ᩈᩢ᩠ᨦ 1E209	ᩈᩢ᩠ᨦ 1E219	ᩈᩢ᩠ᨦ 1E229		ᩈᩢ᩠ᨦ 1E249	ᩈᩢ᩠ᨦ 1E259	
A	ᩈᩢ᩠ᨦ 1E20A	ᩈᩢ᩠ᨦ 1E21A	ᩈᩢ᩠ᨦ 1E22A	ᩈᩢ᩠ᨦ 1E23A	ᩈᩢ᩠ᨦ 1E24A	ᩈᩢ᩠ᨦ 1E25A	
B		ᩈᩢ᩠ᨦ 1E21B	ᩈᩢ᩠ᨦ 1E22B	ᩈᩢ᩠ᨦ 1E23B	ᩈᩢ᩠ᨦ 1E24B	ᩈᩢ᩠ᨦ 1E25B	
C	ᩈᩢ᩠ᨦ 1E20C	ᩈᩢ᩠ᨦ 1E21C	ᩈᩢ᩠ᨦ 1E22C	ᩈᩢ᩠ᨦ 1E23C	ᩈᩢ᩠ᨦ 1E24C	ᩈᩢ᩠ᨦ 1E25C	
D	ᩈᩢ᩠ᨦ 1E20D	ᩈᩢ᩠ᨦ 1E21D		ᩈᩢ᩠ᨦ 1E23D	ᩈᩢ᩠ᨦ 1E24D		
E	ᩈᩢ᩠ᨦ 1E20E	ᩈᩢ᩠ᨦ 1E21E	ᩈᩢ᩠ᨦ 1E22E	ᩈᩢ᩠ᨦ 1E23E	ᩈᩢ᩠ᨦ 1E24E	ᩈᩢ᩠ᨦ 1E25E	
F	ᩈᩢ᩠ᨦ 1E20F	ᩈᩢ᩠ᨦ 1E21F	ᩈᩢ᩠ᨦ 1E22F	ᩈᩢ᩠ᨦ 1E23F		ᩈᩢ᩠ᨦ 1E25F	

Western Cham

Independent Vowels

1E200	ꨀ	WESTERN	CHAM	LETTER	A
1E201	ꨁ	WESTERN	CHAM	LETTER	E
1E202	ꨂ	WESTERN	CHAM	LETTER	O
1E205	ꨃ	WESTERN	CHAM	LETTER	OH

Consonants

1E206	ꨄ	WESTERN	CHAM	LETTER	KA
1E207	ꨅ	WESTERN	CHAM	LETTER	KHA
1E208	ꨆ	WESTERN	CHAM	LETTER	GA
1E209	ꨇ	WESTERN	CHAM	LETTER	GHAA
1E20A	ꨈ	WESTERN	CHAM	LETTER	NGUE
1E20C	ꨉ	WESTERN	CHAM	LETTER	CHA
1E20D	ꨊ	WESTERN	CHAM	LETTER	CHHA
1E20E	ꨋ	WESTERN	CHAM	LETTER	JA
1E20F	ꨌ	WESTERN	CHAM	LETTER	JHA
1E210	ꨍ	WESTERN	CHAM	LETTER	NHUE
1E212	ꨎ	WESTERN	CHAM	LETTER	NHJA
1E213	ꨏ	WESTERN	CHAM	LETTER	TA
1E214	ꨐ	WESTERN	CHAM	LETTER	THA
1E215	ꨑ	WESTERN	CHAM	LETTER	DA
1E216	ꨒ	WESTERN	CHAM	LETTER	DHAA
1E217	ꨓ	WESTERN	CHAM	LETTER	NUE
1E218	ꨔ	WESTERN	CHAM	LETTER	NA
1E219	ꨕ	WESTERN	CHAM	LETTER	DDA
1E21A	ꨖ	WESTERN	CHAM	LETTER	PA
1E21B	ꨗ	WESTERN	CHAM	LETTER	PPA
1E21C	ꨘ	WESTERN	CHAM	LETTER	PHA
1E21D	ꨙ	WESTERN	CHAM	LETTER	BA
1E21E	ꨚ	WESTERN	CHAM	LETTER	BHAA
1E21F	ꨛ	WESTERN	CHAM	LETTER	MUE
1E220	ꨜ	WESTERN	CHAM	LETTER	MA
1E221	ꨝ	WESTERN	CHAM	LETTER	BBA
1E222	ꨞ	WESTERN	CHAM	LETTER	YA
1E223	ꨟ	WESTERN	CHAM	LETTER	RA
1E224	ꨠ	WESTERN	CHAM	LETTER	LA
1E225	ꨡ	WESTERN	CHAM	LETTER	VA
1E226	ꨢ	WESTERN	CHAM	LETTER	SSA
1E227	ꨣ	WESTERN	CHAM	LETTER	SA
1E228	ꨤ	WESTERN	CHAM	LETTER	HA

Dependent Vowel Signs

1E229	ꨥ	WESTERN	CHAM	VOWEL SIGN	AA
1E22A	ꨦ	WESTERN	CHAM	VOWEL SIGN	I
1E22B	ꨧ	WESTERN	CHAM	VOWEL SIGN	II
1E22C	ꨨ	WESTERN	CHAM	VOWEL SIGN	AI
1E22E	ꨩ	WESTERN	CHAM	VOWEL SIGN	OE
					• also used for final ng
1E22F	ꨪ	WESTERN	CHAM	VOWEL SIGN	E
1E230	ꨫ	WESTERN	CHAM	VOWEL SIGN	AAI
1E231	ꨬ	WESTERN	CHAM	VOWEL SIGN	AU
1E232	ꨭ	WESTERN	CHAM	VOWEL SIGN	U

Consonant Signs

1E233	ꨮ	WESTERN	CHAM	CONSONANT SIGN	YA
1E234	ꨯ	WESTERN	CHAM	CONSONANT SIGN	RA
1E235	ꨰ	WESTERN	CHAM	CONSONANT SIGN	LA
1E236	ꨱ	WESTERN	CHAM	CONSONANT SIGN	WA
1E23A	ꨲ	WESTERN	CHAM	LIGATURE	BRAY
1E23B	ꨳ	WESTERN	CHAM	LIGATURE	NAN
1E23C	ꨴ	WESTERN	CHAM	LIGATURE	SHALOM

Final Consonants

1E23D	ꨵ	WESTERN	CHAM	LETTER FINAL	PH
1E23E	ꨶ	WESTERN	CHAM	LETTER FINAL	V
1E23F	꨷	WESTERN	CHAM	LETTER FINAL	S
1E240	꨸	WESTERN	CHAM	LETTER FINAL	K
1E242	꨹	WESTERN	CHAM	LETTER FINAL	NG
1E244	꨺	WESTERN	CHAM	LETTER FINAL	CH
1E245	꨻	WESTERN	CHAM	LETTER FINAL	T
1E246	꨼	WESTERN	CHAM	LETTER FINAL	N
1E247	꨽	WESTERN	CHAM	LETTER FINAL	P
1E248	꨾	WESTERN	CHAM	LETTER FINAL	Y
1E249	꨿	WESTERN	CHAM	LETTER FINAL	R
1E24A	ꩀ	WESTERN	CHAM	LETTER FINAL	L
1E24B	ꩁ	WESTERN	CHAM	LETTER FINAL	SS
1E24C	ꩂ	WESTERN	CHAM	CONSONANT SIGN	FINAL M
1E24D	ꩃ	WESTERN	CHAM	LETTER FINAL	H
1E24E	ꩄ	WESTERN	CHAM	CONSONANT SIGN	TKAJ KA

Digits

1E250	ꩅ	WESTERN	CHAM	DIGIT	ZERO
1E251	ꩆ	WESTERN	CHAM	DIGIT	ONE
1E252	ꩇ	WESTERN	CHAM	DIGIT	TWO
1E253	ꩈ	WESTERN	CHAM	DIGIT	THREE
1E254	ꩉ	WESTERN	CHAM	DIGIT	FOUR
1E255	ꩊ	WESTERN	CHAM	DIGIT	FIVE
1E256	ꩋ	WESTERN	CHAM	DIGIT	SIX
1E257	ꩌ	WESTERN	CHAM	DIGIT	SEVEN
1E258	ꩍ	WESTERN	CHAM	DIGIT	EIGHT
1E259	꩎	WESTERN	CHAM	DIGIT	NINE
1E25A	꩏	WESTERN	CHAM	DIGIT	MOON TEN
















Punctuation

1E25B	꩐	WESTERN	CHAM	SIGN	TANA PATOK SAP
1E25C	꩑	WESTERN	CHAM	PUNCTUATION	SPIRAL
1E25E	꩒	WESTERN	CHAM	PUNCTUATION	DOUBLE DANDA
1E25F	꩓	WESTERN	CHAM	PUNCTUATION	TRIPLE DANDA
1E260	꩔	WESTERN	CHAM	SIGN	TANA TAMAT PHON
1E261	꩕	WESTERN	CHAM	SIGN	TANA CHAPEDAEM PHON
1E262	꩖	WESTERN	CHAM	SIGN	TANA CHAPEDAEM IU
1E263	꩗	WESTERN	CHAM	SIGN	TANA TAMAT AYET
1E264	꩘	WESTERN	CHAM	SIGN	TANA TAMAT TAKUE

0610











Arabic

061F

	061
0	 0610
1	 0611
2	 0612
3	 0613
4	 0614
5	 0615
6	 0616
7	 0617
8	 0618
9	 0619
A	 061A
B	 061B
C	 061C
D	 061D
E	 061E
F	 061F

Arabic

061D ۞ ARABIC END OF TEXT

	1EEF
0	 1EEF0
1	 1EEF1
2	
3	
4	
5	
6	
7	
8	 1EEF8
9	 1EEF9
A	 1EEFA
B	 1EEFB
C	 1EEFC
D	 1EEFD
E	 1EEFE
F	 1EEFF

Arabic Mathematical Alphabetic Symbols

Lunar Symbols

1EEF8	١	ARABIC SYMBOL LUNAR 1 DOTS
1EEF9	٢	ARABIC SYMBOL LUNAR 2 DOTS
1EEFA	٣	ARABIC SYMBOL LUNAR 3 DOTS
1EEFB	٤	ARABIC SYMBOL LUNAR 4 DOTS
1EEFC	٥	ARABIC SYMBOL LUNAR 5 DOTS
1EEFD	٥	ARABIC SYMBOL LUNAR 5 DOTS OVER HEH
1EEFE	٦	ARABIC SYMBOL LUNAR 6 DOTS
1EEFF	٧	ARABIC SYMBOL LUNAR 7 DOTS

- Move U+1E237 to U+1E24E (CONSONANT SIGN TKAJ KA).
- Reorder lower vowels to come after upper vowels to fit with Universal Shaping Engine expectations.
- New introductory paragraph to the Repertoire section.
- Fix chart names for ligatures U+1E23A ... U+1A23C.
- Change the letter name U+1E205 WESTERN CHAM LETTER OS to a more conventional WESTERN CHAM LETTER OH.

Samples

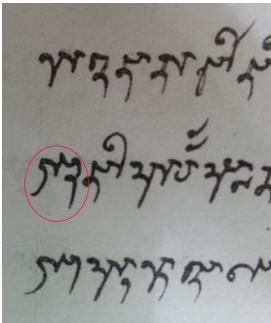


Fig 1. Showing U+1E206 (Რ LETTER KA) U+1E24E (Ე CONSONANT SIGN TKAJ KA)

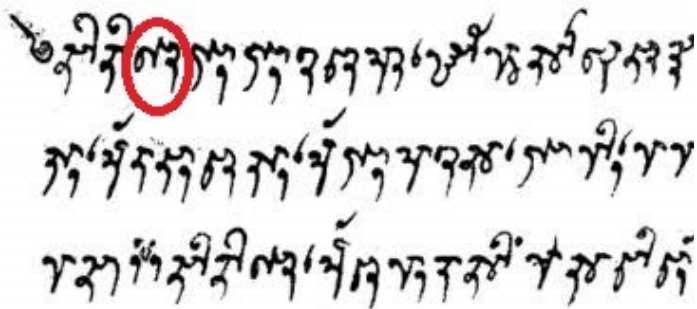


Fig 2. Showing U+1E213 (Ს LETTER TA) U+1E24E (Ე CONSONANT SIGN TKAJ KA). The sign to the far left is U+1E262 (Გ SIGN TANA CHAPEDAEM PHON)

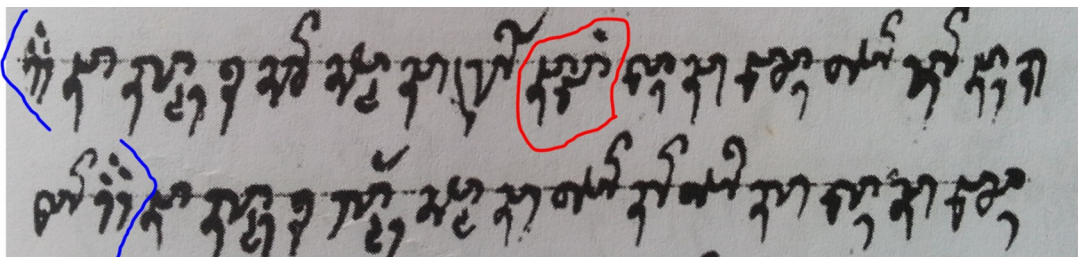


Fig 3. Showing U+1E23C (Მ CONSONANT SIGN SHALOM) and the use of U+1E25E (Პ DOUBLE DANDA) around a quotation

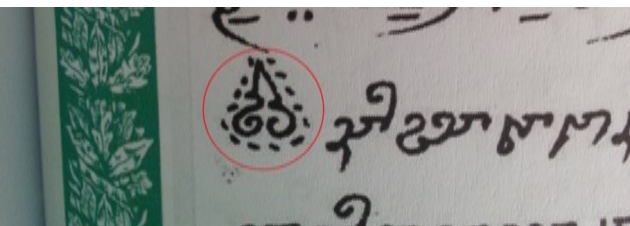


Fig 4. Showing U+1E261 (Გ SIGN TANA CHAPEDAEM PHON)

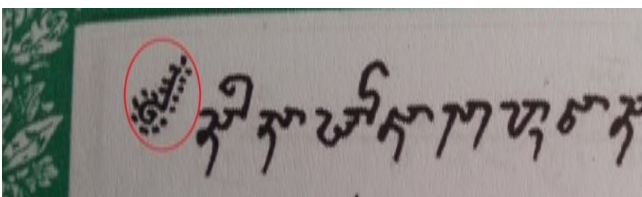


Fig 5. Showing U+1E262 (Გ SIGN CHAPEDAEM IU)

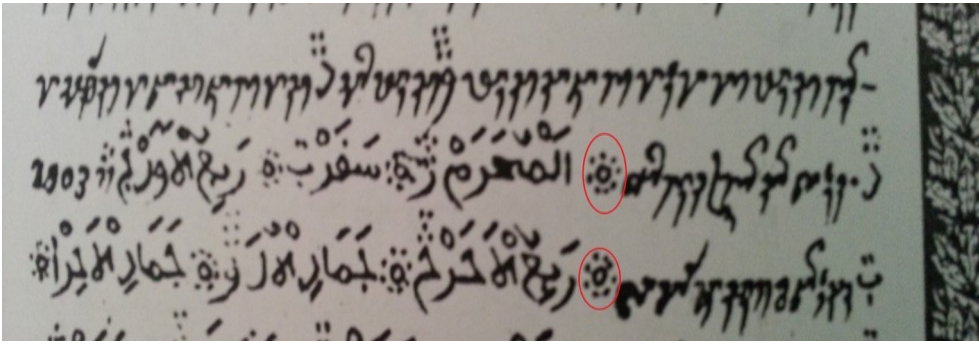


Fig 6. Showing U+1E263 (☉ SIGN TANA TAMAT AYET). Notice the mix of Arabic and Western Cham and also the lunar characters.

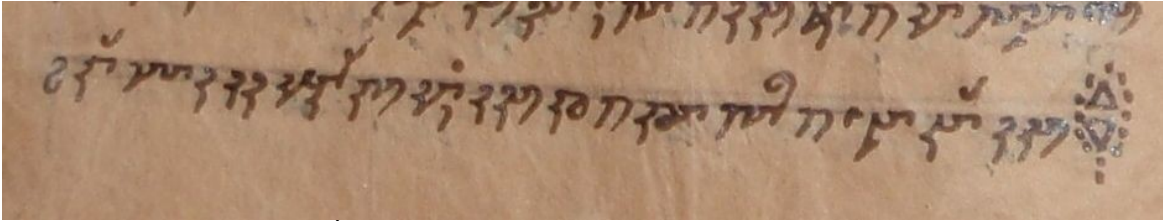


Fig 7. Showing U+1E264 (☼ SIGN TANA TAMA TAKUE)

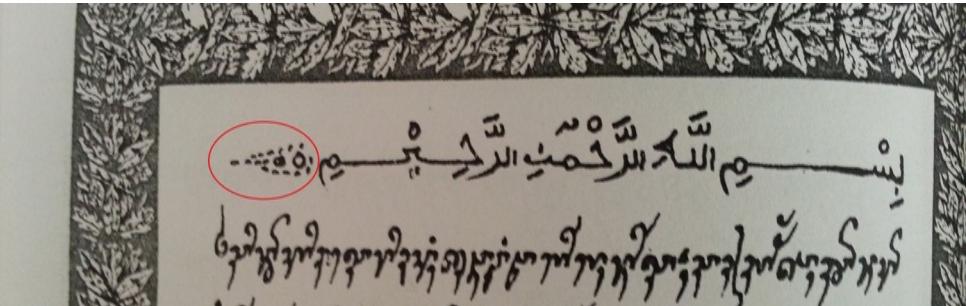


Fig 8. Showing U+061D (·:·:·: ARABIC END OF TEXT) at the end of a line of Arabic text before Western Cham text. The first character of the Western Cham text is U+1E262 (☉ SIGN TAHA CHAPEDAEM PHON)



Fig 9. Showing all the digits and imprimatur from the Chief Imam. U+1E251 ... U+1E25A.

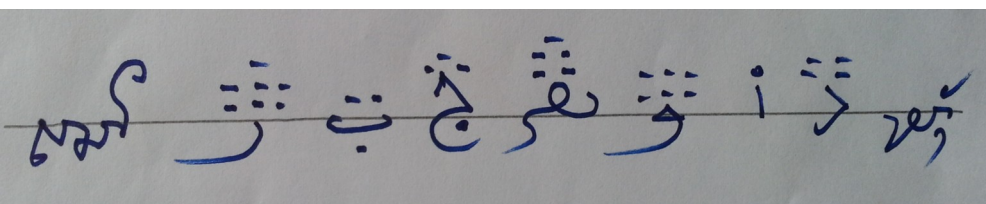


Fig 10. Showing the relative positions of Western Cham letters and lunar month symbols relative to a conceptual baseline.

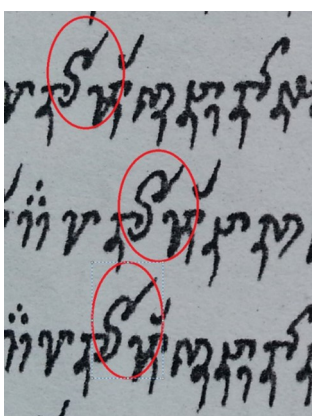


Fig 11. Showing academic variants of Western Cham vowel sequences.

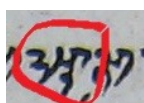


Fig 12. Showing the sequence U+1E20A (ꠘꠞ NGUE) U+1E235 (ꠘꠞ SIGN LA)

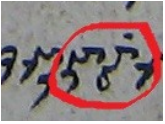


Fig 13. Showing U+1E23C (שָׁלוֹם LIGATURE SHALOM)

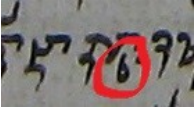


Fig 14. Showing an alternate of U+1E235 (לָא SIGN LA).

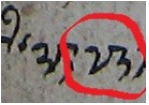


Fig 15. Showing the sequence U+1E21D (בָּא BA) U+1E24E (תְּכַאֵם SIGN TKAJ KA)

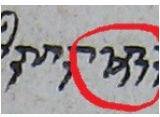


Fig 16. Showing the sequence U+1E220 (מָא MA) U+1E24E (תְּכַאֵם SIGN TKAJ KA)

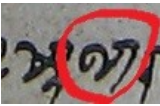


Fig 17. Showing a scribal contraction of U+1E246 (תְּכַאֵם FINAL T)

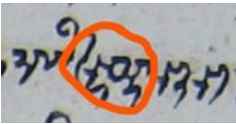


Fig 18. Showing the sequence U+1E207 (כַּחַה KHA) U+1E234 (רָאֵם SIGN RA) U+1E235 (לָא SIGN LA)



Fig 19. Showing error marks U+0353 (כַּחַח COMBINING X BELOW)

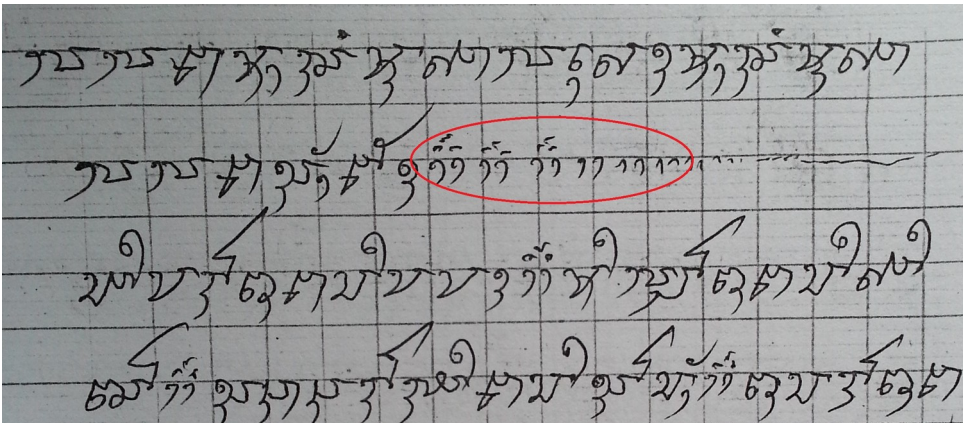


Fig 20. A sequence of double and single dandas often referred to as a triple danda.

4. Glyph comparison. Western Cham characters are given here next to the Eastern Cham characters already encoded to facilitate glyph comparison.

1E200	ꨀ	A	1E22D	ꨀꨀ	-U
1E201	ꨁ	I	1E22E	ꨁꨁ	-OE
1E202	ꨂ	U	1E22F	ꨂꨂ	-O
1E203	ꨃ	E	1E230	ꨃꨃ	-AI
1E204	ꨄ	AI	1E231	ꨄꨄ	-AU
1E205	ꨅ	O	1E232	ꨅꨅ	-UE
1E206	ꨆ	KA	1E233	ꨆꨆ	-YA
1E207	ꨇ	KHA	1E234	ꨇꨇ	-RA
1E208	ꨈ	GA	1E235	ꨈꨈ	-LA
1E209	ꨉ	GHA	1E236	ꨉꨉ	-WA
1E20A	ꨊ	NGUE	1E23B	ꨊꨊ (ꨊꨊ)	-PH
1E20B	ꨋ	NGA	1E23C	ꨋꨋ (ꨋꨋ)	-B
1E20C	ꨌ	CHA	1E23D	ꨌꨌ (ꨌꨌ)	-M
1E20D	ꨍ	CHHA	1E23E	ꨍꨍ (ꨍꨍ)	-V
1E20E	ꨎ	JA	1E23F	ꨎꨎ (ꨎꨎ)	-S
1E20F	ꨏ	JHA	1E240	ꨐꨐ	-K
1E210	ꨑ	NHUE	1E241	ꨑꨑ	-G
1E211	ꨒ	NHA	1E242	ꨒꨒ	-NG
1E212	ꨓ	NHJA	1E243	ꨓꨓ	-NG (cons. sign)
1E213	ꨔ	TA	1E244	ꨔꨔ	-CH
1E214	ꨕ	THA	1E245	ꨕꨕ	-T
1E215	ꨖ	DA	1E246	ꨖꨖ	-N
1E216	ꨗ	DHA	1E247	ꨗꨗ	-P
1E217	ꨘ	NUE	1E248	ꨘꨘ	-Y
1E218	ꨙ	NA	1E249	ꨙꨙ	-R
1E219	ꨚ	DDA	1E24A	ꨚꨚ	-L
1E21A	ꨛ	PA	1E24B	ꨛꨛ	-SS
1E21B	ꨜ	PPA	1E24C	ꨜꨜ	-M (cons. sign)
1E21C	ꨝ	PHA	1E24D	ꨝꨝ	-H (cons. sign)
1E21D	ꨞ	BA	1E24F	ꨞꨞ (ꨞꨞ)	NAAN
1E21E	ꨟ	BHA	1E250	ꨟꨟ	0
1E21F	ꨠ	MUE	1E251	ꨠꨠ	1
1E220	ꨡ	MA	1E252	ꨡꨡ	2
1E221	ꨢ	BBA	1E253	ꨢꨢ	3
1E222	ꨣ	YA	1E254	ꨣꨣ	4
1E223	ꨤ	RA	1E255	ꨤꨤ	5
1E224	ꨥ	LA	1E256	ꨥꨥ	6
1E225	ꨦ	VA	1E257	ꨦꨦ	7
1E226	ꨧ	SSA	1E258	ꨧꨧ	8
1E227	ꨨ	SA	1E259	ꨨꨨ	9
1E228	ꨩ	HA	1E25B	ꨩꨩ (ꨩꨩ)	TANA PATOK SAP
1E229	ꨪ	-AA	1E25C	ꨪꨪ	SPIRAL
1E22A	ꨫ	-I	1E25D	ꨫꨫ	DANDA
1E22B	ꨬ	-II	1E25E	ꨬꨬꨬ	DOUBLE DANDA
1E22C	ꨭ	-EI	1E25F	ꨭꨭꨭꨭ	TRIPLE DANDA

Fig 21. Glyph Comparison chart from L2/16-198 (n4734)

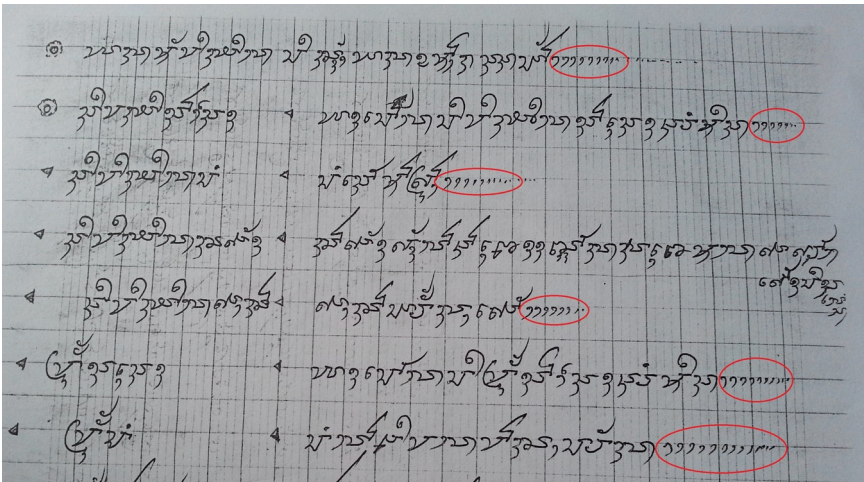


Fig 22. Use of U+1E25D (□ DANDA)

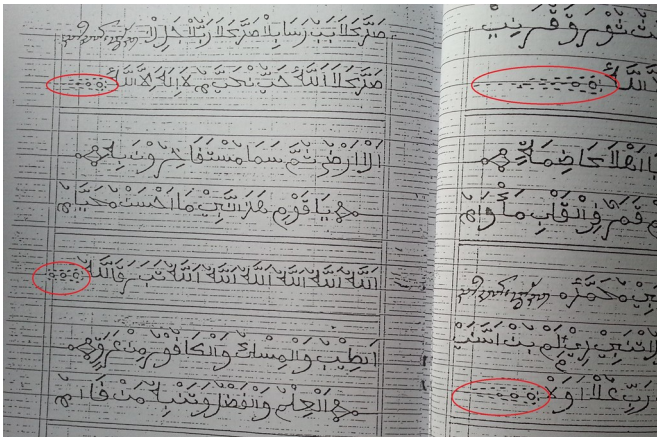


Fig 23. Examples of U+061D (arabic end of text)

ကံ ကံ ကံ ကံ ကံ ကံ

က	က	က	က	က	က
က	က	က	က	က	က
က	က	က	က	က	က
က	က	က	က	က	က
က	က	က	က	က	က
က	က	က	က	က	က
က	က	က	က	က	က



Fig 24. Consonants with imprimatur from the Chief Imam. Consonants in row order: 1E206, 1E207, 1E208, 1E209, 1E20A; 1E20C, 1E20D, 1E20E, 1E20F², 1E201; 1E213, 1E214, 1E215, 1E216, 1E217; 1E21A, 1E21C, 1E21D, 1E21E, 1E21F; 1E225, 1E221, 1E219, 1E223, 1E212; 1E222, 1E224, 1E227, 1E228, 1E200; 1E202, 1E201, 1E201+1E230, 1E22D, 1E23B.

2 Archaic glyph form used.

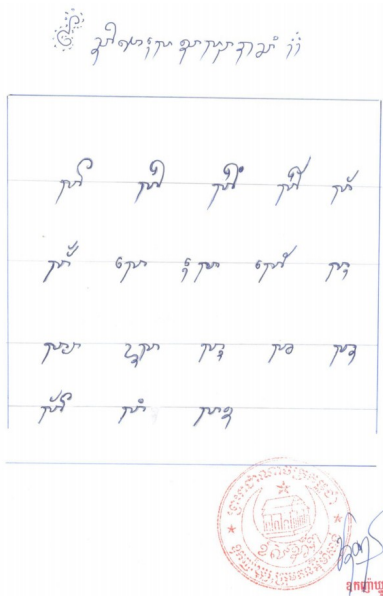


Fig 25. Vowels with imprimatur from the Chief Imam. Vowels in row order in conjunction with U+1E206: 1E229, 1E22A, 1E22B, 1E22C, 1E22E; 1E22E+1E22E, 1E22F, 1E230, 1E22F+1E231, 1E232; 1E233, 1E234, 1E235, 1E236, 1E24E; 1E22E+1E229, 1E24C, 1E24D.

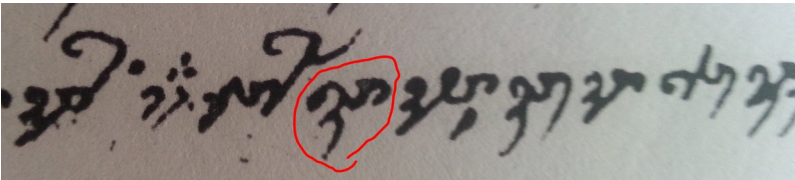


Fig 26. U+1E200 (၅ LETTER A)

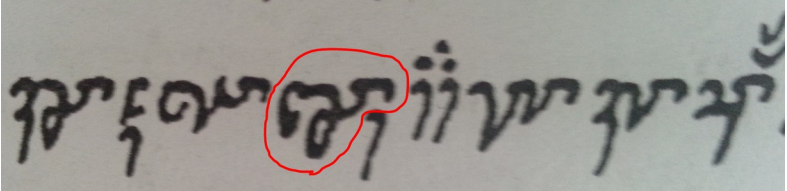


Fig 27. U+1E210 (၈ LETTER NHUE)

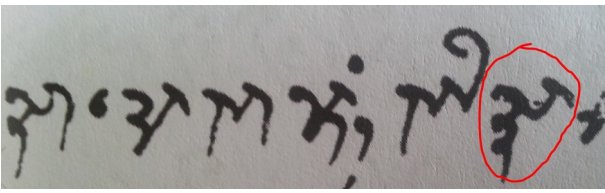


Fig 28. U+1E218 (၉ LETTER NA)

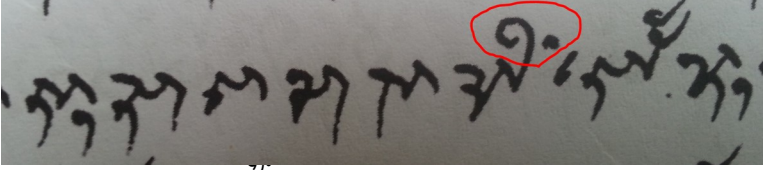


Fig 29. U+1E22B (၀ SIGNI)

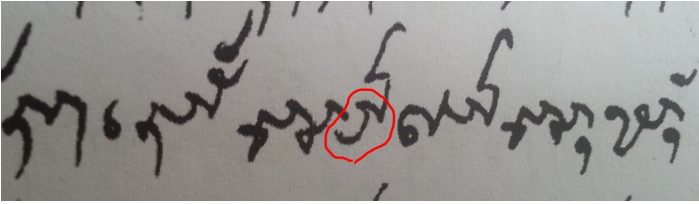


Fig 30. U+1E233 (၀ SIGN YA)

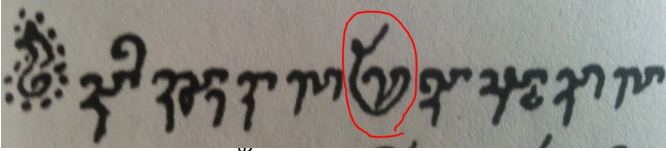


Fig 31. U+1E23A (၀ SIGN BRAY)

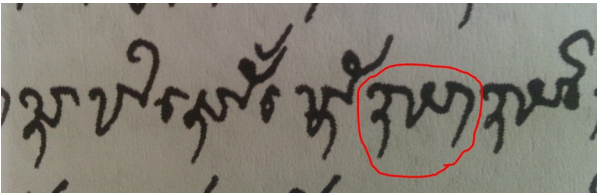


Fig 32. U+1E23D (၀ LETTER FINAL PH)

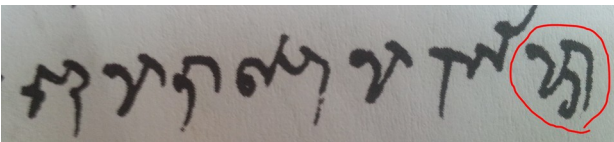


Fig 33. U+1E23E (၀ LETTER FINAL V)

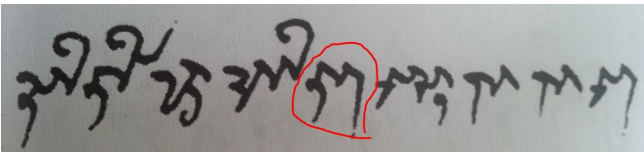


Fig 32. U+1E23F (၀ LETTER FINAL S)

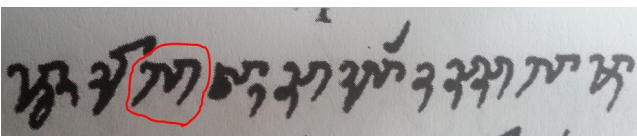


Fig 33. U+1E240 (၀ LETTER FINAL K)

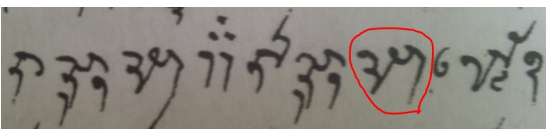


Fig 34. U+1E242 (၀ LETTER FINAL NG)

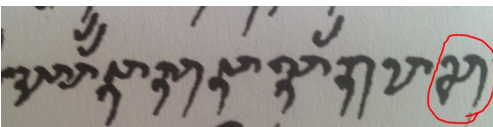


Fig 35. U+1E244 (၀ LETTER FINAL CH)

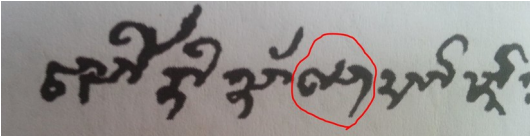


Fig 36. U+1E245 (၆၇) LETTER FINAL T)

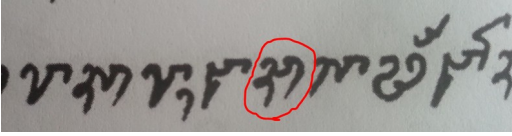


Fig 37. U+1E246 (ဒွဲ) LETTER FINAL N)

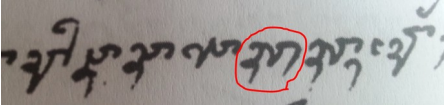


Fig 38. U+1E247 (ဒွဲ) LETTER FINAL P)

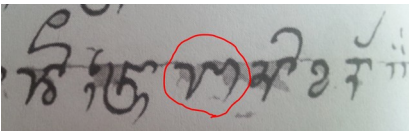


Fig 39. U+1E248 (၇၇) LETTER FINAL Y)

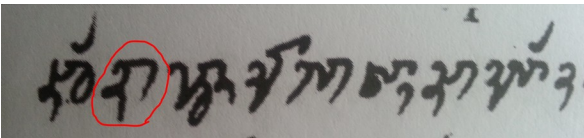


Fig 40. U+1E249 (ဒွဲ) LETTER FINAL R)

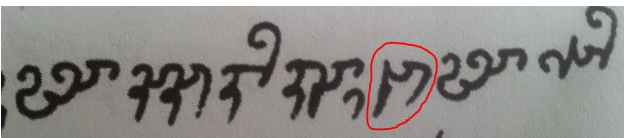


Fig 41. U+1E24A (၆၇) LETTER FINAL L)

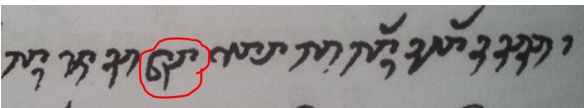


Fig 42. U+1E226 (၉၇) LETTER SSA), the initial for U+1E24B (၉၇) LETTER FINAL SS). No samples of U+1E24B are on hand, but the character is proposed since some teachers consider it valid.

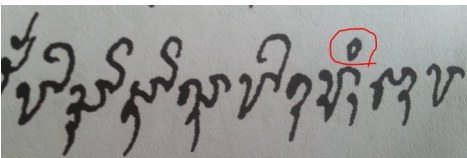


Fig 43. U+1E24C (၆၇) CONSONANT SIGN FINAL M)

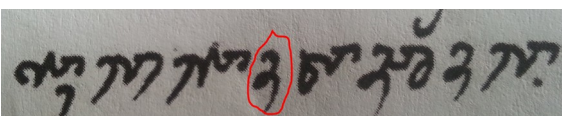


Fig 44. U+1E24D (ဒွဲ) CONSONANT SIGN FINAL H)

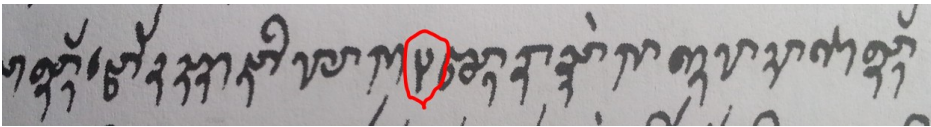


Fig 45. U+1E25B (ʘ SIGN TANA PATOK SAP)

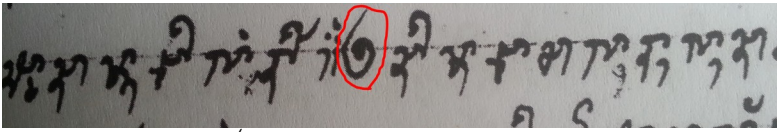


Fig 46. U+1E25C (⊙ PUNCTUATION SPIRAL)

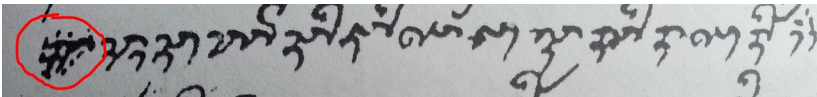


Fig 47. Variant form of U+1E260 (⦿ SIGN TANA TAMAT PHON)

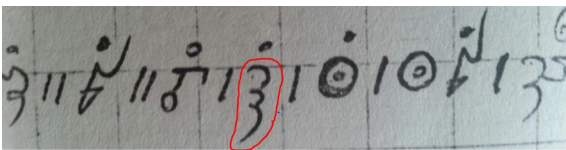


Fig 48. Examples of moon numbers

Consonants				
ᳵ	ᳶ	᳷	᳸	᳹
ᳺ	᳻	᳼	᳽	᳾
᳿	᳠	᳡	᳢	᳣
᳤	᳥	᳦	᳧	᳨
ᳩ	ᳪ	ᳫ	ᳬ	᳭
ᳮ	ᳯ	ᳰ	ᳱ	ᳲ

363

Fig 49. A handwritten consonant a digits chart

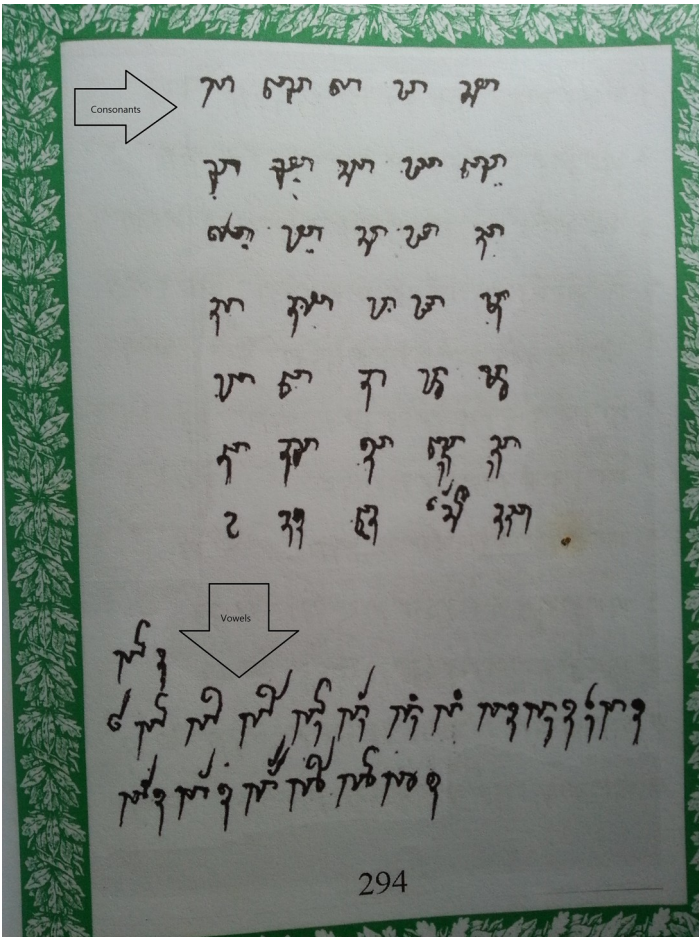


Fig 50. Another handwritten consonant and vowels chart

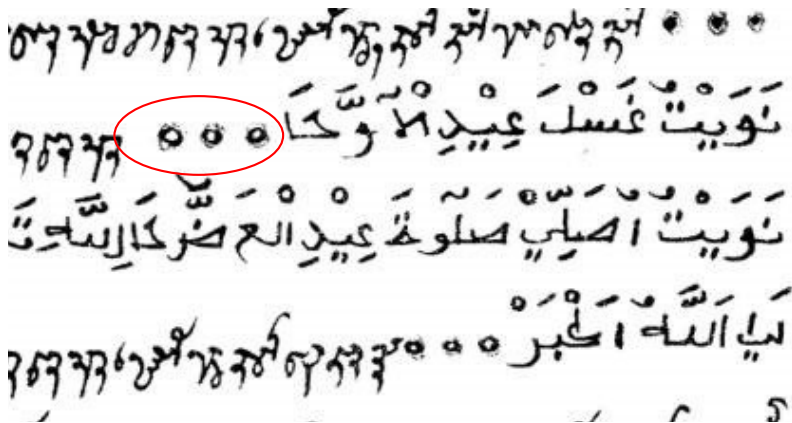


Fig 51. Examples of sequences of U+1E263 (∗ SIGN TANA TAMAT AYET). Notice the space after the left most one on line 2.

ប្រតិបត្តិការណ៍ក្រុមប្រឹក្សាឥស្លាមកម្ពុជា
A letter of endorsement

ចុះថ្ងៃទី ០៤ ខែ ឧសភា ឆ្នាំ ២០១៩
Date : May 04, 2019

គ្រឹះស្ថានប្រឹក្សាឥស្លាមកម្ពុជា
To Unicode international committee:

យើងគាំទ្រគម្រោងបញ្ជូនអក្សរឃ្លាឥស្លាមកម្ពុជា ទៅក្នុងប្រព័ន្ធគ្រប់លំដាប់អក្សរឃ្លាឥស្លាមកម្ពុជា
We, the Islamic Community Kan Imam San of Cambodia, support the Proposal to encode Western Cham in the UCS.

យើងគាំទ្រគម្រោងនេះ ព្រោះវាបានរួមបញ្ចូលទាំងសំឡេង និងសញ្ញាណប្រតិបត្តិការណ៍ ដែលយើងត្រូវការ ដើម្បីសរសេរ និងអាន ក្នុងអក្សរឃ្លាឥស្លាមកម្ពុជា
We support this proposal because it has included all the Cham sounds and symbols that we need to write and read in Cham KakKhak script.

នេះគឺជាឈ្មោះនៃអ្នកគាំទ្រគម្រោងនេះ
These are the names of the people who endorse this proposal with its signatures.

១ (ម្ចាស់ មុនី ខ័ណ្ឌធានី) ប្រតិបត្តិការណ៍ក្រុមប្រឹក្សាឥស្លាមកម្ពុជា
1. Chief Imam of entire Cham Iman San Group (OngKhnuo Math Sa)

២ (ម្ចាស់ មុនី ខាងស្តាំ) ភូមិស្រែចម្ការ (កក ម៉ាត)
2. Imam right hand of Srae Prey village (Kokk Math)

៣ (ម្ចាស់ មុនី ខាងឆ្វេង) ភូមិស្រែចម្ការ (ម៉ាត សូម)
3. Head Imam of Phum Keoh village (Math Soum)

៤ (ម្ចាស់ មុនី ខាងឆ្វេង) ភូមិស្រែចម្ការ (លី ឡេ)
4. Head Imam of Phum Caghoah village (Ol Lee)

Sincerely

ម្ចាស់ មុនី ខ័ណ្ឌធានី
ប្រតិបត្តិការណ៍ក្រុមប្រឹក្សាឥស្លាមកម្ពុជា

The Chief Imam of the Islamic Community Kan Imam San of Cambodia
OngKhnuo Math Sa



អង្គការឥស្លាមកម្ពុជា

**ISO/IEC JTC 1/SC 2/WG 2
PROPOSAL SUMMARY FORM TO ACCOMPANY SUBMISSIONS
FOR ADDITIONS TO THE REPERTOIRE OF ISO/IEC 10646³.**

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:	Western Cham
2. Requester's name:	Martin Hosken
3. Requester type (Member body/Liaison/Individual contribution):	Individual Contribution
4. Submission date:	20/May/19
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):	X
Proposed name of script:	Western Cham
b. The proposal is for addition of character(s) to an existing block:	
Name of the existing block:	
2. Number of characters in proposal:	98
3. Proposed category (select one from below - see section 2.2 of P&P document):	
A-Contemporary <input checked="" type="checkbox"/>	B.1-Specialized (small 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 type="checkbox"/>
G-Obscure or questionable usage symbols <input type="checkbox"/>	
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?	SIL International
b. Identify the party granting a license for use of the font by the editors (include address, e-mail, ftp-site, etc.):	martin_hosken@sil.org
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

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.

³ 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? If YES explain	<i>This finalises N4734 L2/16-198</i>	<i>yes</i>
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:	<i>supporting letter enclosed</i>	<i>yes</i>
3. Information on the user community for the proposed characters (for example: size, demographics, information technology use, or publishing use) is included? Reference:		<i>no</i>
4. The context of use for the proposed characters (type of use; common or rare) Reference:		<i>common</i>
5. Are the proposed characters in current use by the user community? If YES, where? Reference:	<i>This document</i>	<i>yes</i>
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:		<i>no</i>
7. Should the proposed characters be kept together in a contiguous range (rather than being scattered)?		<i>yes</i>
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:		<i>no</i>
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:		<i>no</i>
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:	<i>This document</i>	<i>yes</i> <i>yes</i>
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:	<i>This document</i>	<i>yes</i> <i>yes</i>
Is a list of composite sequences and their corresponding glyph images (graphic symbols) provided? If YES, reference:		<i>no</i>
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)		<i>no</i>
13. Does the proposal contain any Ideographic compatibility characters? If YES, are the equivalent corresponding unified ideographic characters identified? If YES, reference:		<i>no</i>