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:	**2020-11-18**

Executive Summary. This proposal is to add 87 characters to a block: 1E200-1E26F in the supplementary plane, named Western Cham. In addition, 1 character is proposed for inclusion in the Arabic block and 8 characters are proposed for inclusion in the Arabic Supplemental Symbols block: 1EF00-1EF07. 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 4 textual symbols U+1E260 (SIGN TANA TAMAT PHON) … U+1E262 (SIGN TANA CHAPEDAEM IU) and U+061D (ARABIC END OF TEXT) are textual symbols with discourse functions. Specifically:

Codepoint	Character name	Glyph	Description	Script	Figure
1E260	SIGN TANA TAMAT 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
061D	ARABIC END OF TEXT	⸱⸪⸳	Closing text	Arabic	8, 23

In styling, U+1E260 (✑ SIGN TANA TAMAT 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+2055 (⁕ FLOWER PUNCTUATION MARK) 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 SALAM	✑	1E227 1E224 1E24C	✑	3, 13

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

3. U+1E25B (𞉛 SIGN TANA PATOK SAP) reduplication mark (Figure 45) and U+1E25C (◉ PUNCTUATION SPIRAL) 'one' or 'first' (Figure 46) are the same as in N4374. Note that while U+1E25C (◉ PUNCTUATION SPIRAL) has lexical value, it takes no diacritics.

4. U+1E24E (◌ 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 (ॗ 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 𞈗𞈚.

5. The character U+1E25A (◉ DIGIT MOON TEN) is used when describing days in the lunar calendar. U+1E24C (◌ FINAL M) is used together with a digit to mark the days after a full moon, e.g. 𞉑 U+1E251 (𞉑 DIGIT 1) U+1E24C (◌ FINAL M) for the first day after the full moon. U+1E25A (◉ SIGN MOON TEN) is used for the 10s digit in such contexts: thus ◉𞉑 U+1E25A (◉ SIGN MOON TEN) U+1E251 (𞉑 DIGIT 1) U+1E24C (◌ 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 (◌ SIGN OE) instead of U+1E24C (◌ FINAL M), with the full moon ocurring on day 15. For example 10 days after a full moon is represented by ◉ U+1E25A (◉ DIGIT MOON TEN) U+1E22E (◌ SIGN OE). See Figure 48.

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

6. 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 Supplemental Symbols block. The order of the months is U+1EF07 (ﺝ SEVEN DOTS LUNAR MONTH), U+1EF01 (ﺏ TWO DOTS LUNAR MONTH), U+1EF02 (ﺝ THREE DOTS LUNAR MONTH), U+1EF04 (ﺝ FIVE DOTS LUNAR MONTH), U+1EF06 (ﺝ SIX DOTS LUNAR MONTH), U+1EF00 (ﺍ ONE DOT LUNAR MONTH), U+1EF01 (ﺏ TWO DOTS LUNAR MONTH), U+1EF03 (ﺝ FOUR DOTS LUNAR MONTH), U+1EF04 (ﺝ FIVE DOTS LUNAR MONTH), U+1EF07 (ﺝ SEVEN DOTS LUNAR MONTH), U+1EF00 (ﺍ ONE DOT LUNAR MONTH), U+1EF03 (ﺝ FOUR DOTS LUNAR MONTH). 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+1EF05 (ﺝ FIVE DOTS LUNAR MONTH 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.

7. 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.

8. 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
1E203 (AA03)	1E201 1E22F	⁊ᣀᣥ
1E204 (AA04)	1E201 1E230	ᣀᣥ

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

9. Western Cham has two dandas. The double danda U+1E25E (ᣁ̈ 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 (ᣁᣁᣁᣁᣁ" END OF TEXT) 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 end of text. 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 an end of text. But it is assumed that the single danda does exist, given its existence in the end of text, 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.

    Some texts show a contrasting triple single danda character, but this is always as part of an end of text and always occurs in conjunction with another danda, just as other end of text examples. Therefore there is no need to encode a specific triple danda for this character sub component. Doing so would only cause confusion over how to encode an end of text.

10. U+1E21B (ᣛ PPA) is not found in the standard list of consonants in Cham[1]. It is a special spelling of U+1E21A (ᣚ PA) used in some special words. For example ⁊ᣛᣁ U+1E21B (ᣛ PPA) U+1E22F (⁊ VOWEL E) U+1E22E (ᣁ VOWEL OE) /po/ 'lord'. See also Figure 52.

Ligation. Western Cham is visually highly ligating. For example, ᣚᣁ U+1E206 (ᣚ KA) U+1E22A (ᣁ VOWEL I) and then ᣚᣁ U+1E206 (ᣚ KA) U+1E236 (ᣁ SIGN WA) U+1E22A (ᣁ 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 E	ᣀᣥ	1E224 1E24E	ᣀᣥ	*does not occur*
1E256	DIGIT SIX	ᣜ	**1E223 1E236** 1E24D	ᣜᣁ	choose
1E257	DIGIT SEVEN	ᣝ	**1E223 1E24E** 1E240	ᣝᣚ	vein
1E259	DIGIT NINE	ᣟ	1E21D 1E24E	ᣟ	carry (on shoulder)
1EF02	LUNAR 3 DOTS	ċ̇			
1EF03	LUNAR 4 DOTS	ᣐ̈	0690	ᣐ̈	

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

---

1 Much as æ is not listed in the English alphabet.

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 (ᵰ RA).

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

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
ᵱ	1E206	ᵱ	1E240	ᵱ	1E220	◌	1E24C
ᵱ	1E20A	ᵱ	1E242	ᵱ	1E222	ᵱ	1E248
		◌	1E22E	ᵱ	1E223	ᵱ	1E249
ᵱ	1E20C	ᵱ	1E244	ᵱ	1E224	ᵱ	1E24A
ᵱ	1E213	ᵱ	1E245	ᵱ	1E225	ᵱ	1E23E
ᵱ	1E217	ᵱ	1E246	ᵱ	1E226	ᵱ	1E24B
ᵱ	1E21C	ᵱ	1E23D	ᵱ	1E227	ᵱ	1E23F
ᵱ	1E21A	ᵱ	1E247	ᵱ	1E228	ᵱ	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: ᵱ ᵱ `bank (of river)` U+1E21D (ᵱ LETTER BA) U+1E217 (ᵱ 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 (ᵱ 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
ᵱ ᵱ ᵱ	1E21A 1E214 1E233 1E230	Pre vowels go before the last initial.
ᵱ	1E206 1E22E 1E232	No visual difference based on order.
ᵱ ᵱ	1E215 1E223 1E229 1E22E	Interaction between upper and final vowels.
ᵱ	1E200 1E233 1E229	Interaction between following medial and following vowel.
ᵱ	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: । U+0964 (DEVANAGARI DANDA), ॥ U+0965 (DEVANAGARI DOUBLE DANDA).

◌ͣ U+0353 (COMBINING X BELOW) is used, rarely, in transcribing handwritten texts where the mark is used to indicate a spelling error.

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 SALAM;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;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 END OF TEXT;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;;;;;

1EF00;ARABIC SYMBOL ONE DOT LUNAR MONTH;So;0;ON;;;;;N;;;;;
1EF01;ARABIC SYMBOL TWO DOTS LUNAR MONTH;So;0;ON;;;;;N;;;;;
1EF02;ARABIC SYMBOL THREE DOTS LUNAR MONTH;So;0;ON;;;;;N;;;;;
1EF03;ARABIC SYMBOL FOUR DOTS LUNAR MONTH;So;0;ON;;;;;N;;;;;
1EF04;ARABIC SYMBOL FIVE DOTS LUNAR MONTH;So;0;ON;;;;;N;;;;;
1EF05;ARABIC SYMBOL FIVE DOTS LUNAR MONTH OVER HEH;So;0;ON;;;;;N;;;;;
1EF06;ARABIC SYMBOL SIX DOTS LUNAR MONTH;So;0;ON;;;;;N;;;;;
1EF07;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+1EF00 .. U+1EF07.

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

1E200 (ꈀ) < 1E201 (ꈁ) < 1E202 (ꈂ) < 1E201 (ꈁ) 1E22F (ꈯ) < 1E201 (ꈁ) 1E230 (ꈰ) < 1E205 (ꈅ)

1E206 (ꈆ) < 1E207 (ꈇ) < 1E208 (ꈈ) < 1E209 (ꈉ) < 1E20A (ꈊ) < 1E20C (ꈌ) < 1E20D (ꈍ) < 1E20E (ꈎ) < 1E20F (ꈏ) < 1E210 (ꈐ) < 1E212 (ꈒ) < 1E213 (ꈓ) < 1E214 (ꈔ) < 1E215 (ꈕ) < 1E216 (ꈖ) < 1E217 (ꈗ) < 1E218 (ꈘ) < 1E219 (ꈙ) < 1E21A (ꈚ) < 1E21B (ꈛ) < 1E21C (ꈜ) < 1E21D (ꈝ) < 1E21E (ꈞ) < 1E21F (ꈟ) < 1E220 (ꈠ) < 1E221 (ꈡ) < 1E222 (ꈢ) < 1E223 (ꈣ) < 1E224 (ꈤ) < 1E225 (ꈥ) < 1E226 (ꈦ) < 1E227 (ꈧ) < 1E228 (ꈨ)

1E233 (ꈳ) < 1E234 (ꈴ) < 1E235 (ꈵ) < 1E236 (ꈶ)

1E229 (ꈩ) < 1E22A (ꈪ) < 1E22B (ꈫ) < 1E22C (ꈬ) < 1E22F (ꈯ) < 1E230 (ꈰ) < 1E231 (ꈱ) < 1E232 (ꈲ)

1E240 (ꉀ) < 1E242 (ꉂ) < 1E22E (ꈮ) < 1E244 (ꉄ) < 1E245 (ꉅ) < 1E246 (ꉆ) < 1E247 (ꉇ) < 1E23D (ꈽ) < 1E24C (ꉌ) < 1E248 (ꉈ) < 1E249 (ꉉ) < 1E24A (ꉊ) < 1E23E (ꈾ) < 1E24B (ꉋ) < 1E23F (ꈿ) < 1E24D (ꉍ) < 1E24E (ꉎ)

1E21D 1E234 1E22C << 1E23A (ꈺ)

1E217 1E246 << 1E23B (ꈻ)

1E227 1E224 1E24C << 1E23C (ꈼ)

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

Cluster Patterns. The traditional approach to handling cluster patterns has been through the Canonical Combing Class. 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.[2] This does not mean that relative ordering of marks is not important. Here is a list of mark classes in their relative order. Those occuring first are listed first:

Class	Characters
Medial ra	?◌ 1E234 (CONSONANT SIGN RA)
Medial la	◌ 1E235 (CONSONANT SIGN LA)
Medial ya	◌ 1E233 (CONSONANT SIGN YA)
Medial wa	◌ 1E236 (CONSONANT SIGN WA)
Pre vowels	◌ 1E22F (VOWEL SIGN E), ◌ U+1E230 (VOWEL SIGN AAI)
Final vowels	◌ 1E229 (VOWEL SIGN AA), ◌ 1E22A (VOWEL SIGN I), ◌ 1E22B (VOWEL SIGN II), ◌ 1E22C (VOWEL SIGN AI), ◌ 1E231 (VOWEL SIGN AU)
Upper vowels	◌ 1E22E (VOWEL SIGN OE)
Lower vowels	◌ 1E232 (VOWEL SIGN U)
Finals	◌ 1E24C (CONSONANT SIGN FINAL M), ◌ 1E24D (CONSONANT SIGN FINAL H), ◌ 1E24E (CONSONANT SIGN TKAJ KA)

No more than one character in each class may occur following the same base character U+1E200 (LETTER A) … 1E228 (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 (FINAL PH) … U+1E24B (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]`

---

2 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.

Encoding Model. The encoding model follows that of Eastern Cham and Khmer, in being Indic with vowels following base characters, as described in the cluster patterns. An alternative would be to use a visual encoding order where the medial ra would be encoded before the base consonant. This makes no logical sense in that none of the confusion arguments used for visual ordering apply here. Could a visual ordering be used? Yes. But it would confuse everyone and have to answer the question why differ from Eastern Cham and Khmer when there is no need?

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
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
```

In review, it is noted that 1E233 and 1E236 are both right medials. The Universal Shaping Engine has the unfortunate constraint of only allowing one medial in each position. The author is open to suggestions as to how to resolve this problem. Either the USE has to change to support multiple medials in a position, or one of the right medials has to be arbitrarily marked as above just to conform to a technological constraint. It should be noted that allowing multiple right medials would not cause any ambiguity in the erroneous situation that a user enters multiple right medials where only one is expected, since right medials are assumed to be spacing and not attach to the base.

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

10

Western Cham

Independent Vowels

1E200	𞈀	WESTERN CHAM LETTER A
1E201	𞈁	WESTERN CHAM LETTER E
1E202	𞈂	WESTERN CHAM LETTER O
1E205	𞈅	WESTERN CHAM LETTER OH

Consonants

1E206	𞈆	WESTERN CHAM LETTER KA
1E207	𞈇	WESTERN CHAM LETTER KHA
1E208	𞈈	WESTERN CHAM LETTER GA
1E209	𞈉	WESTERN CHAM LETTER GHAA
1E20A	𞈊	WESTERN CHAM LETTER NGUE
1E20C	𞈌	WESTERN CHAM LETTER CHA
1E20D	𞈍	WESTERN CHAM LETTER CHHA
1E20E	𞈎	WESTERN CHAM LETTER JA
1E20F	𞈏	WESTERN CHAM LETTER JHA
1E210	𞈐	WESTERN CHAM LETTER NHUE
1E212	𞈒	WESTERN CHAM LETTER NHJA
1E213	𞈓	WESTERN CHAM LETTER TA
1E214	𞈔	WESTERN CHAM LETTER THA
1E215	𞈕	WESTERN CHAM LETTER DA
1E216	𞈖	WESTERN CHAM LETTER DHAA
1E217	𞈗	WESTERN CHAM LETTER NUE
1E218	𞈘	WESTERN CHAM LETTER NA
1E219	𞈙	WESTERN CHAM LETTER DDA
1E21A	𞈚	WESTERN CHAM LETTER PA
1E21B	𞈛	WESTERN CHAM LETTER PPA
1E21C	𞈜	WESTERN CHAM LETTER PHA
1E21D	𞈝	WESTERN CHAM LETTER BA
1E21E	𞈞	WESTERN CHAM LETTER BHAA
1E21F	𞈟	WESTERN CHAM LETTER MUE
1E220	𞈠	WESTERN CHAM LETTER MA
1E221	𞈡	WESTERN CHAM LETTER BBA
1E222	𞈢	WESTERN CHAM LETTER YA
1E223	𞈣	WESTERN CHAM LETTER RA
1E224	𞈤	WESTERN CHAM LETTER LA
1E225	𞈥	WESTERN CHAM LETTER VA
1E226	𞈦	WESTERN CHAM LETTER SSA
1E227	𞈧	WESTERN CHAM LETTER SA
1E228	𞈨	WESTERN CHAM LETTER HA

Dependent Vowel Signs

1E229	𞈩	WESTERN CHAM VOWEL SIGN AA
1E22A	𞈪	WESTERN CHAM VOWEL SIGN I
1E22B	𞈫	WESTERN CHAM VOWEL SIGN II
1E22C	𞈬	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 SALAM

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	𞉊	WESTERN CHAM LETTER FINAL L
1E24B	𞉋	WESTERN CHAM LETTER FINAL SS
1E24C	𞉌	WESTERN CHAM CONSONANT SIGN FINAL M
1E24D	𞉍	WESTERN CHAM LETTER FINAL H
1E24E	𞉎	WESTERN CHAM CONSONANT SIGN TKAJ KA

Digits

1E250	𞉐	WESTERN CHAM DIGIT ZERO
1E251	𞉑	WESTERN CHAM DIGIT ONE
1E252	𞉒	WESTERN CHAM DIGIT TWO
1E253	𞉓	WESTERN CHAM DIGIT THREE
1E254	𞉔	WESTERN CHAM DIGIT FOUR
1E255	𞉕	WESTERN CHAM DIGIT FIVE
1E256	𞉖	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	𞉛	WESTERN CHAM SIGN TANA PATOK SAP
1E25C	𞉜	WESTERN CHAM PUNCTUATION SPIRAL
1E25E	𞉞	WESTERN CHAM PUNCTUATION DOUBLE DANDA
1E25F	𞉟	WESTERN CHAM END OF TEXT
1E260	𞉠	WESTERN CHAM SIGN TANA TAMAT PHON
1E261	𞉡	WESTERN CHAM SIGN TANA CHAPEDAEM PHON
1E262	𞉢	WESTERN CHAM SIGN TANA CHAPEDAEM IU

	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

	1EF0
0	أ 1EF00
1	ت 1EF01
2	ج 1EF02
3	ر 1EF03
4	خ 1EF04
5	ڤ 1EF05
6	و 1EF06
7	ر 1EF07
8	
9	
A	
B	
C	
D	
E	
F	

Arabic Supplemental Symbols

Lunar Symbols

1EF00	ا	ARABIC SYMBOL ONE DOT LUNAR MONTH
1EF01	ت	ARABIC SYMBOL TWO DOTS LUNAR MONTH
1EF02	ث	ARABIC SYMBOL THREE DOTS LUNAR MONTH
1EF03	د	ARABIC SYMBOL FOUR DOTS LUNAR MONTH
1EF04	ح	ARABIC SYMBOL FIVE DOTS LUNAR MONTH
1EF05	ة	ARABIC SYMBOL FIVE DOTS LUNAR MONTH OVER HEH
1EF06	ق	ARABIC SYMBOL SIX DOTS LUNAR MONTH
1EF07	ل	ARABIC SYMBOL SEVEN DOTS LUNAR MONTH

Bibliography.

សហគមន៍ឥស្លាមកនអ៊ុំម៉ុំសាន់កម្ពុជា *SahakumIslamKromKanImamSanKampucia* (Islamic Community Kan Imam-San of Cambodia), 2011
ꩰꨬꨛ ꩪꨬ ꨆꨮꩴꨢ ꨔꨆꨮ ꩰꨬ ꩴꨮ *Procum Neug TakajThanaw* (Books and Ethics)

សហគមន៍ឥស្លាមកនអ៊ុំម៉ុំសាន់កម្ពុជា *SahakumIslamKromKanImamSanKampucia* (Islamic Community Kan Imam-San of Cambodia), 2011
ꨮꩴꨮ ꨆꨮ ꨆ ꨮ ꩪ ꨆꨮꩴ ꩪꨬꨬ ꨮꩴ *SiophouAksosaSgamCam* (Book of literature learning

Cham)

Revision 1. The first revision differs from the original document:

- Moved one symbol into the Arabic Extended A block and closed up the gap. Added extra block chart, etc.
- Adding sections on disunification, ligation and finals
- Expanding section text in introduction, combining orders, confusability
- renaming 1E26B from WESTERN CHAM SIGN LUNAR 5 DOTS OVER HAH to WESTERN CHAM SIGN LUNAR 5 DOTS OVER HEH.
- Adding two sequence independent vowels to collation order. Fixing 1E22E collation specials.
- Adding sample glyphs to all character name references
- Added subheadings in names list after the chart

Revision 2. These are the changes from revision 1 (June 2019) to revision 2 (Sept 2019)

- Added Indic syllabic and positional categories
- Added extra examples in response to comments.

There is no change to the character repertoire.

Revision 3. These are the changes from revision 2 (Sept 2019) to revision 3 (November 2019)

- Renamed U+1E22E from WESTERN CHAM VOWEL SIGN OE OR FINAL NG to WESTERN CHAM VOWEL SIGN OE.
- Added many more examples.
- Moved lunar symbols from U+1E266 ..U+1E26D to U+1EEF8..U+1EEFF and rename from WESTERN CHAM SIGN LUNAR to ARABIC SYMBOL LUNAR.
- Added a bibliography.

L2/20-061. These are the changes from revision 3 (November 2019) to L2/20-061 (January 2020)

- Add figure references to the discussion of proposed characters.
- Remove references to unencoded characters removed from Eastern Cham, with some remaining.
- Remove discussion of alternative final character mark.
- Rename lunar month characters.
- Rename ligatures from consonant signs to LIGATURE.
- Change all punctuation characters to have bidi property ON.
- Remove U+1E25F (PUNCTUATION TRIPLE DANDA) in favour of a sequence of double and single dandas.

L2/20-000. These are the changes from L2/20-061 (January 2020) to L2/20-000 (March 2020)

- Reintroduce U+1E25F (PUNCTUATION TRIPLE DANDA) in response to direct request from community.

16

- 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.

L2/20-061R3. These are the changes from L2/20-061R2 (May 2020) to L2/20-000 (November 2020)

- Add examples of U+1E21B, in Fig 52 and the main text (which is also found in Eastern Cham (U+AA1B), for which no evidence was provided in its proposal document). Explain why it does not occur in the alphabet lists from the Chief Imam.
- Yet again, move the lunar symbols to another block. This time U+1EF00 … U+1EF07.
- Remove U+1E263 SIGN TANA TAMAT AYET to use U+2055 FLOWER PUNCTUATION MARK instead.
- Remove U+1E264 SIGN TANA TAMAT TAKUE for presumed lack of evidence, despite other characters having been encoded on far less evidence. The example figure remains to save renumbering all the figures yet again and for future evidence should a request be made for encoding later.
- Rename U+1E25F PUNCTUATION TRIPLE DANDA to END OF TEXT. Justify why there is no need for an extra triple danda since such a thing never occurs alone and is always a component of an END OF TEXT, for example the figure on L2/20-185 p3 shows an example of END OF TEXT and not of a separate triple danda.
- Explain the implications of the Chief Immam's imprimatur in Fig 24, 25.
- Add comment on U+0353 COMBINING X BELOW to the section on punctuation.

Samples
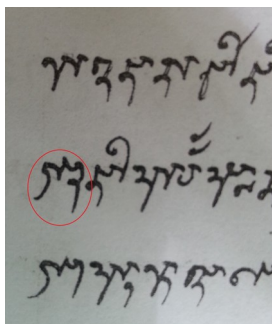
Fig 1. Showing U+1E206 (�ᵖ LETTER KA) U+1E24E (◌ᵤ CONSONANT SIGN TKAJ KA)
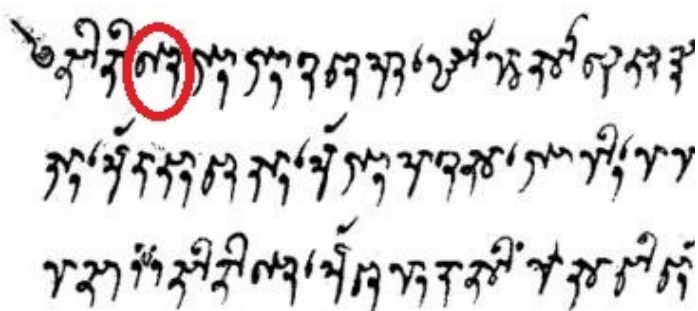
Fig 2. Showing U+1E213 (6ᵖ LETTER TA) U+1E24E (◌ᵤ CONSONANT SIGN TKAJ KA). The sign to the far left is U+1E262 (◌ SIGN TANA CHAPEDAEM IU)
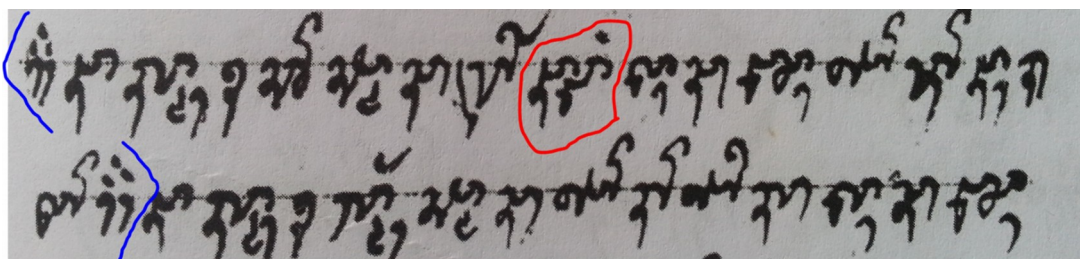
Fig 3. Showing U+1E23C (◌ CONSONANT SIGN SALAM) and the use of U+1E25E (◌ DOUBLE DANDA) around a quotation
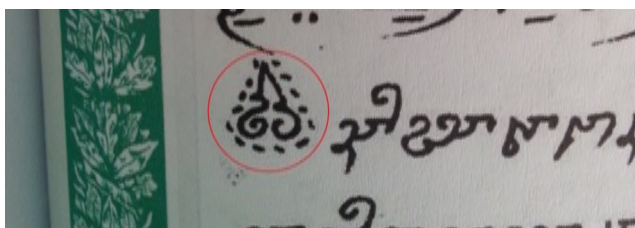
Fig 4. Showing U+1E261 (◌ SIGN TANA CHAPEDAEM PHON)
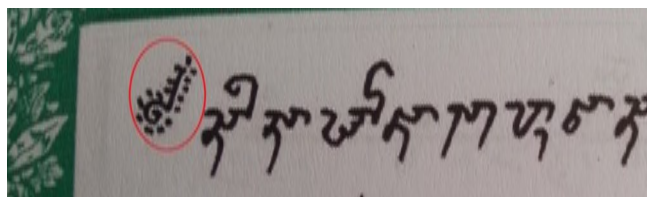
Fig 5. Showing U+1E262 (◌ SIGN TANA CHAPEDAEM IU)
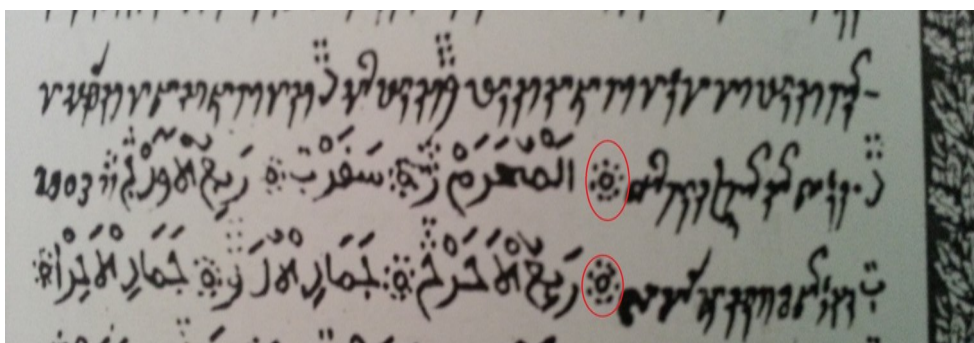
Fig 6. Showing U+2055 (✺ FLOWER PUNCTUATION MARK). Notice the mix of Arabic and Western Cham and also the lunar characters.
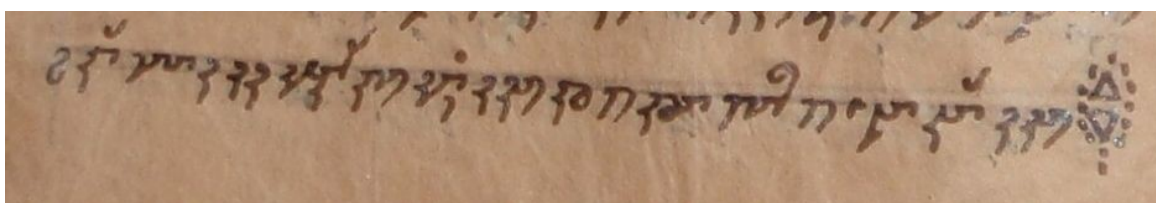
Fig 7. Showing unencoded symbol (꣸ SIGN TANA TAMA TAKUE).
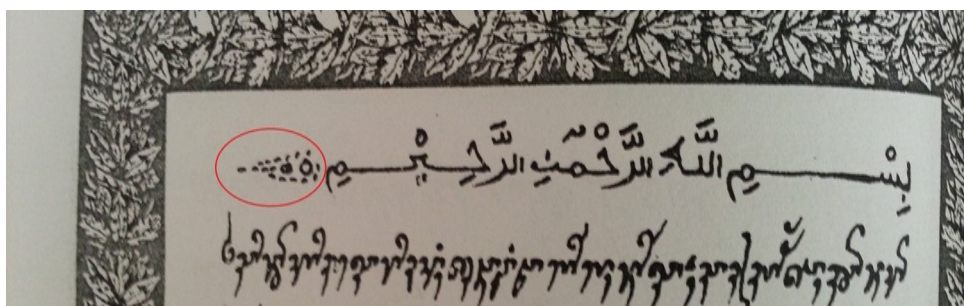
Fig 8. Showing U+061D (ⵗ؞ ARABIC END OF TEXT) at the end of a line of Arabic text before Western Cham text.The first character of the Western Cham text is U+1E262 (ꉢ SIGN TAHA CHAPEDAEM IU)

Fig 9. Showing all the digits and imprimatur from the Chief Imam. U+1E251 … U+1E25A.
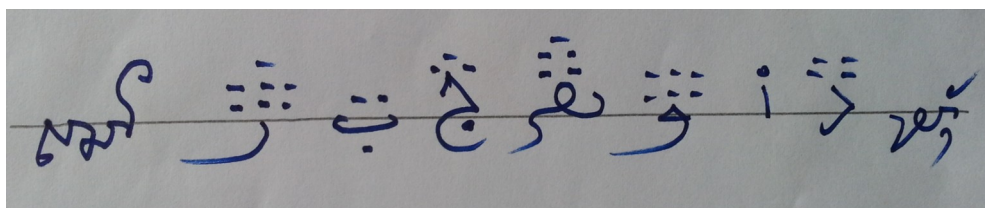
Fig 10. Showing the relative positions of Western Cham letters and lunar month symbols relative to a conceptual baseline.
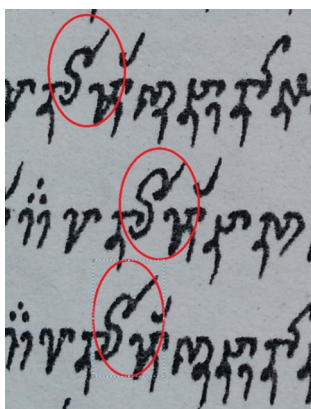
Fig 11. Showing academic variants of Western Cham vowel sequences.
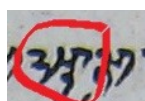
Fig 12. Showing the sequence U+1E20A (ꞩ NGUE) U+1E235 (◌ꞵ SIGN LA)
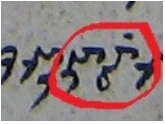
Fig 13. Showing U+1E23C (ᤞ LIGATURE SALAM)
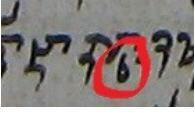
Fig 14. Showing an alternate of U+1E235 (◌ SIGN LA).
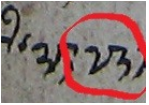
Fig 15. Showing the sequence U+1E21D (𞈝 BA) U+1E24E (◌ SIGN TKAJ KA)
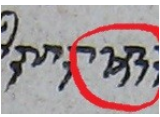
Fig 16. Showing the sequence U+1E220 (𞈠 MA) U+1E24E (◌ SIGN TKAJ KA)
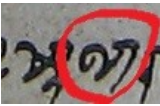
Fig 17. Showing a scribal contraction of U+1E246 (𞉆 FINAL N)
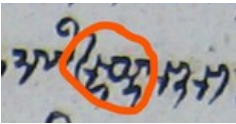
Fig 18. Showing the sequence U+1E207 (𞈇 KHA) U+1E234 (◌ SIGN RA) U+1E235 (◌ SIGN LA)

Fig 19. Showing error marks U+0353 (◌ COMBINING X BELOW)
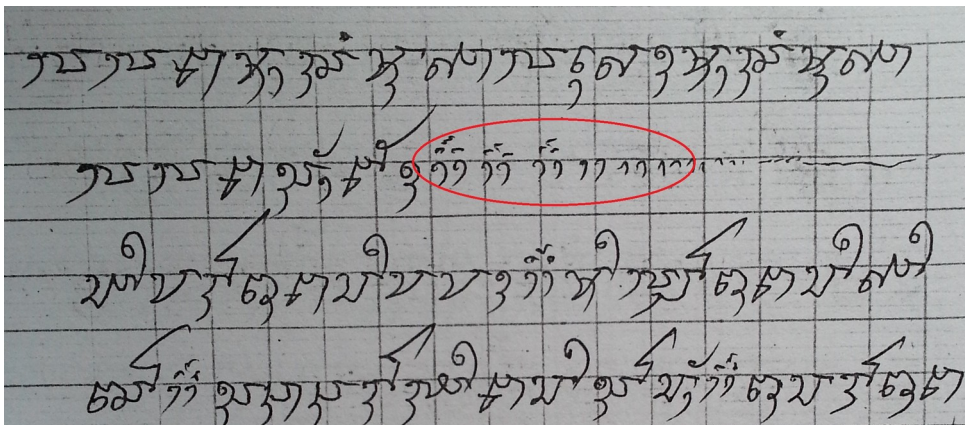
Fig 20. Triple danda.

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

Code	West	East	Name
1E200			A
1E201			I
1E202			U
1E203			E
1E204			AI
1E205			O
1E206			KA
1E207			KHA
1E208			GA
1E209			GHA
1E20A			NGUE
1E20B			NGA
1E20C			CHA
1E20D			CHHA
1E20E			JA
1E20F			JHA
1E210			NHUE
1E211			NHA
1E212			NHJA
1E213			TA
1E214			THA
1E215			DA
1E216			DHA
1E217			NUE
1E218			NA
1E219			DDA
1E21A			PA
1E21B			PPA
1E21C			PHA
1E21D			BA
1E21E			BHA
1E21F			MUE
1E220			MA
1E221			BBA
1E222			YA
1E223			RA
1E224			LA
1E225			VA
1E226			SSA
1E227			SA
1E228			HA
1E229			-AA
1E22A			-I
1E22B			-II
1E22C			-EI

Code	West	East	Name
1E22D			-U
1E22E			-OE
1E22F			-O
1E230			-AI
1E231			-AU
1E232			-UE
1E233			-YA
1E234			-RA
1E235			-LA
1E236			-WA
1E23B		( )	-PH
1E23C		( )	-B
1E23D		( )	-M
1E23E		( )	-V
1E23F		( )	-S
1E240			-K
1E241			-G
1E242			-NG
1E243			-NG (cons. sign)
1E244			-CH
1E245			-T
1E246			-N
1E247			-P
1E248			-Y
1E249			-R
1E24A			-L
1E24B			-SS
1E24C			-M (cons. sign)
1E24D			-H (cons. sign)
1E24F		( )	NAAN
1E250			0
1E251			1
1E252			2
1E253			3
1E254			4
1E255			5
1E256			6
1E257			7
1E258			8
1E259			9
1E25B		( )	TANA PATOK SAP
1E25C			SPIRAL
1E25D			DANDA
1E25E			DOUBLE DANDA
1E25F			TRIPLE DANDA

Fig 21. Glyph Comparison chart from L2/16-198 (n4734)
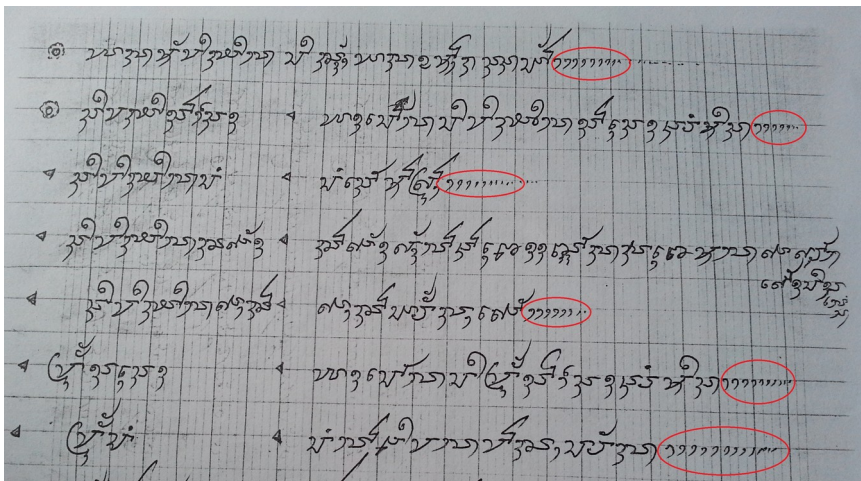
Fig 22. Use of triple danda
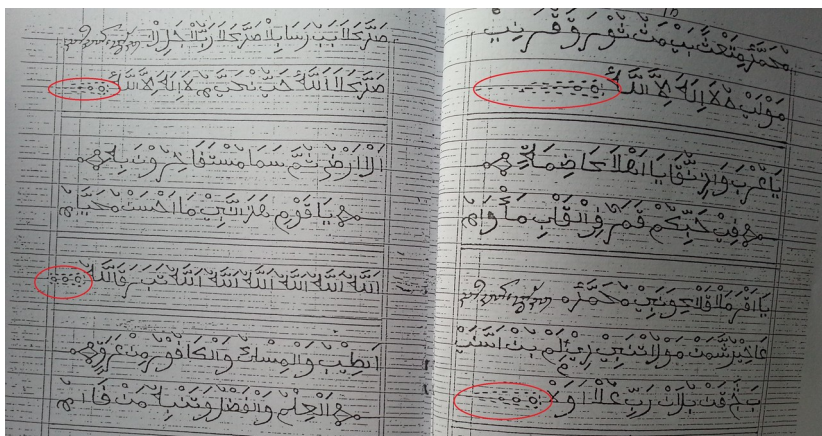
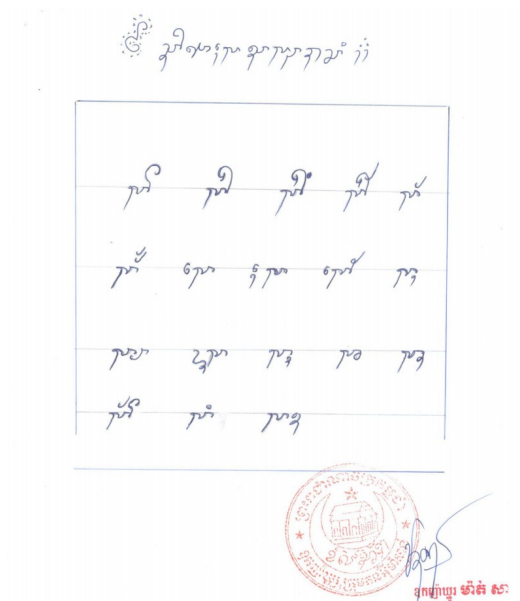
Fig 23. Examples of U+061D (arabic end of text)
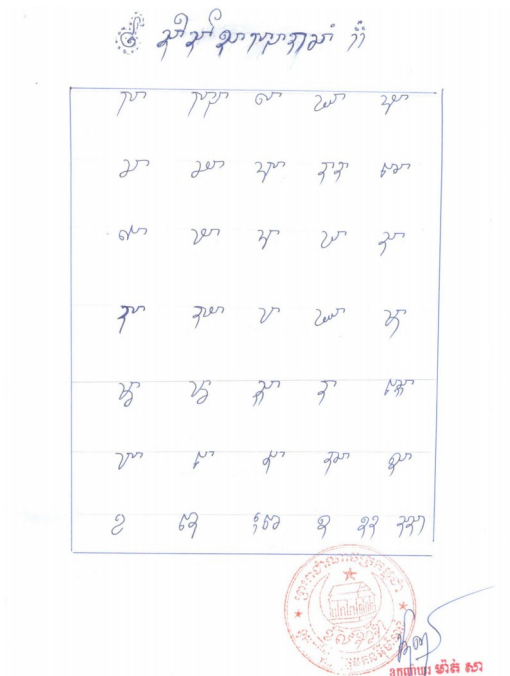
Fig 24. Consonants with imprimatur from the Chief Imam as the authoritative alphabet listing coming from the language community. Consonants in row order: 1E206, 1E207, 1E208, 1E209, 1E20A; 1E20C, 1E20D, 1E20E, 1E20F[3], 1E210; 1E213, 1E214, 1E215, 1E216, 1E217; 1E21A, 1E21C, 1E21D, 1E21E, 1E21F; 1E225, 1E221, 1E219,

---

3   Archaic glyph form used.

1E223, 1E212; 1E222, 1E224, 1E227, 1E228, 1E200; 1E202, 1E201, 1E201+1E230, 1E205, 1E23B.

Fig 25. Vowels with imprimatur from the Chief Imam as the authoritative alphabet listing coming from the language community. 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.
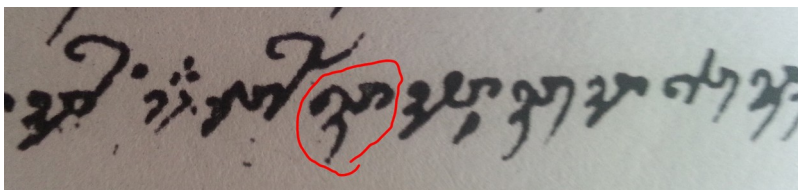
Fig 26. U+1E200 (ꈀ LETTER A)
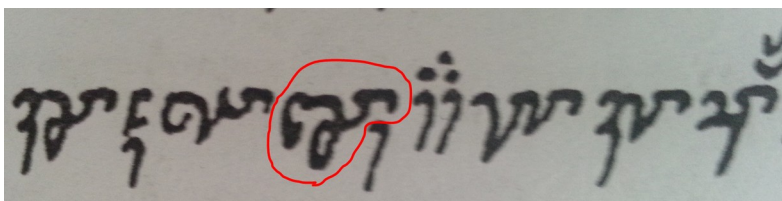
Fig 27. U+1E210 (ꈐ LETTER NHUE)
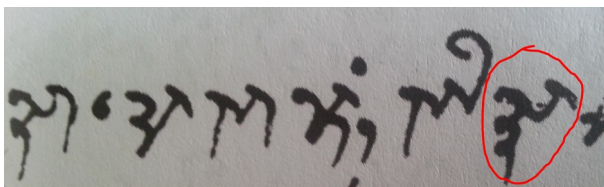
Fig 28. U+1E218 (ꈘ LETTER NA)
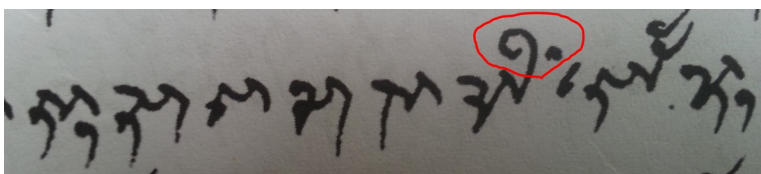
Fig 29. U+1E22B (ဉ် SIGN II)
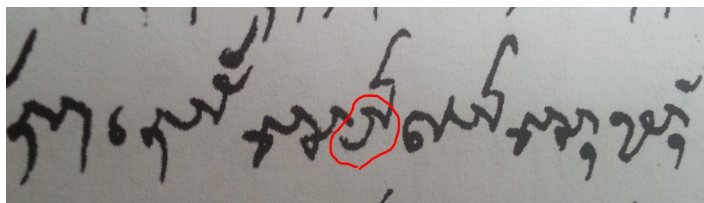
Fig 30. U+1E233 (ဴ SIGN YA)
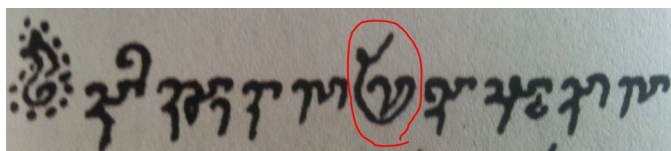
Fig 31. U+1E23A (ဪ LIGATURE BRAY)
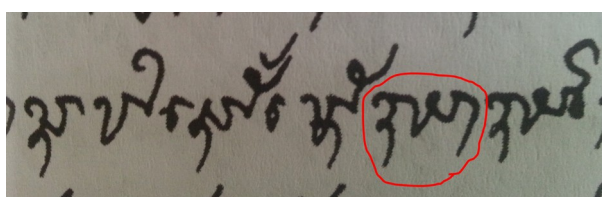
Fig 32. U+1E23D (ဿ LETTER FINAL PH)
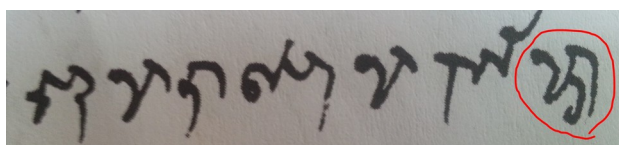
Fig 33. U+1E23E (ဿ LETTER FINAL V)
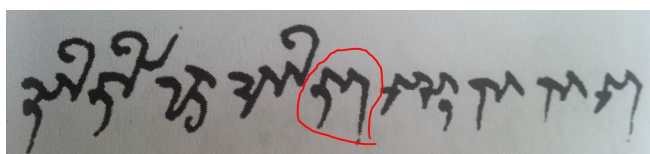
Fig 32a. U+1E23F (ဿ LETTER FINAL S)
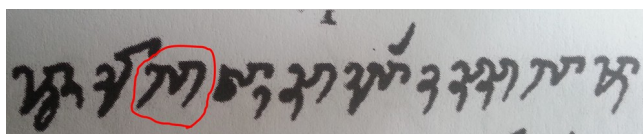
Fig 33a. U+1E240 (ဿ LETTER FINAL K)
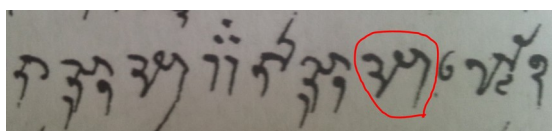
Fig 34. U+1E242 (ဿ LETTER FINAL NG)
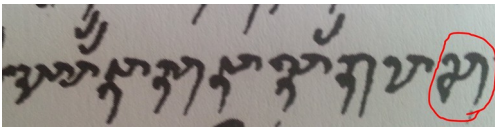
Fig 35. U+1E244 (�C LETTER FINAL CH)
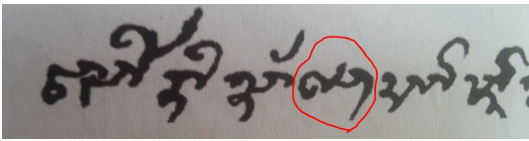
Fig 36. U+1E245 (ꓔ LETTER FINAL T)
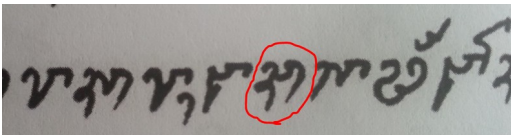
Fig 37. U+1E246 (ꓔ LETTER FINAL N)
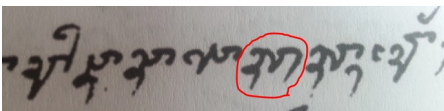
Fig 38. U+1E247 (ꓔ LETTER FINAL P)
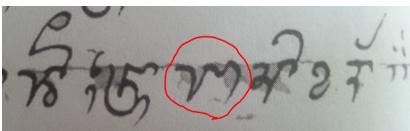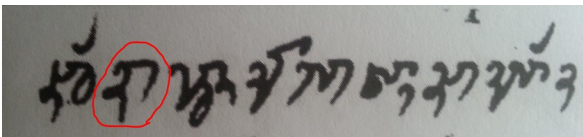

Fig 39. U+1E248 (ꓔ LETTER FINAL Y)
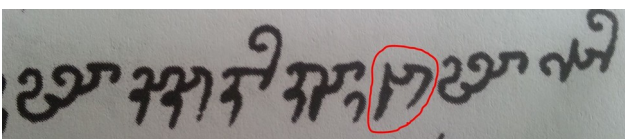

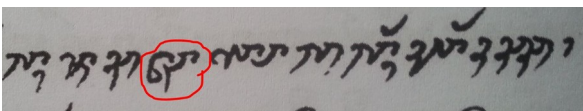*Fig 40. U+1E249 (ꓔ LETTER FINAL R)*

Fig 41. U+1E24A (ꓔ LETTER FINAL L)

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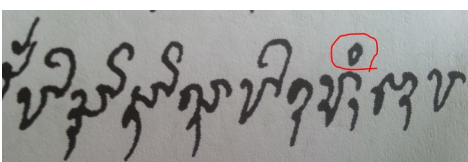

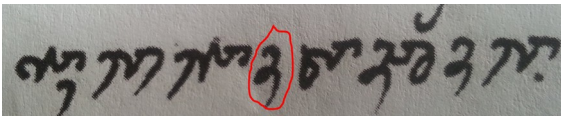*Fig 43. U+1E24C (◌ CONSONANT SIGN FINAL M)*

Fig 44. U+1E24D (ꠍ CONSONANT LETTER FINAL H)
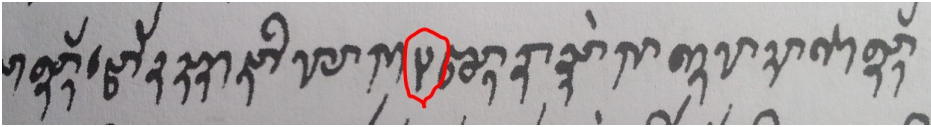

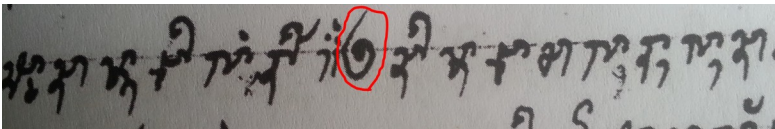*Fig 45. U+1E25B (ꠛ SIGN TANA PATOK SAP)*

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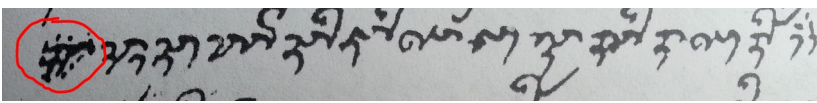

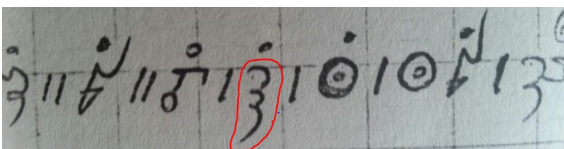*Fig 47. Variant form of U+1E260 (ꠠ SIGN TANA TAMAT PHON)*

Fig 48. Examples of moon numbers

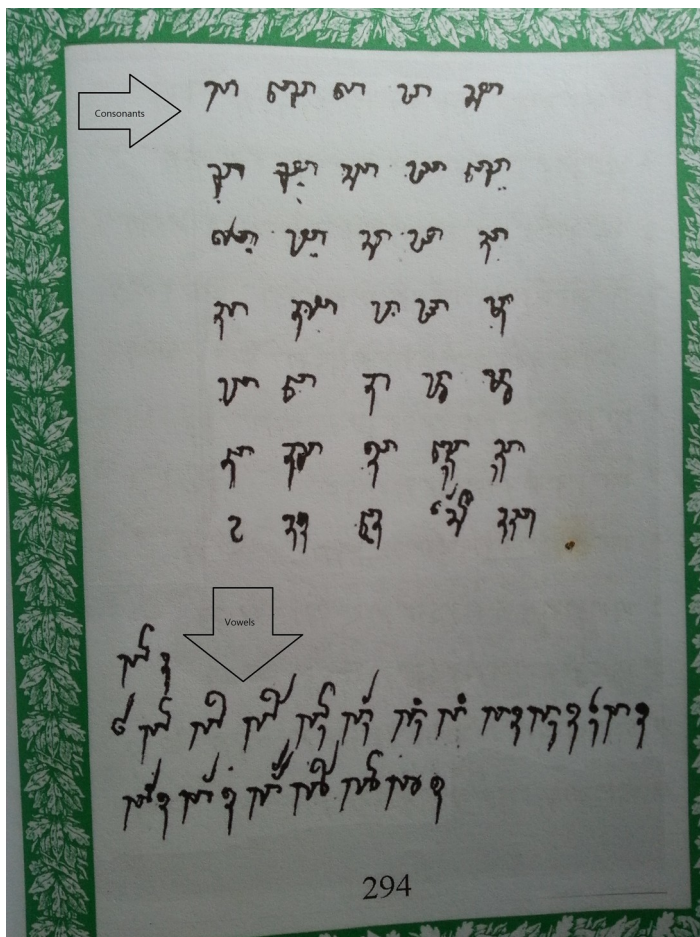*Fig 49. A handwritten consonant a digits chart*

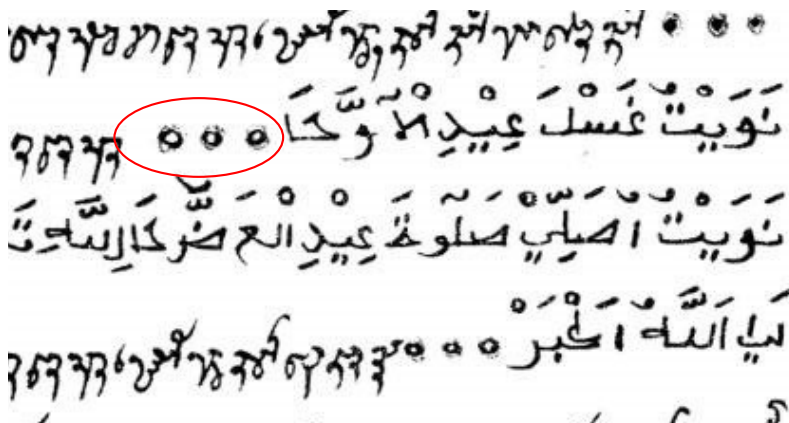*Fig 50. Another handwritten consonant and vowels chart*

Fig 51. Examples of sequences of U+2055 (❀ FLOWER PUNCTUATION MARK). Notice the space after the left most one on line 2.
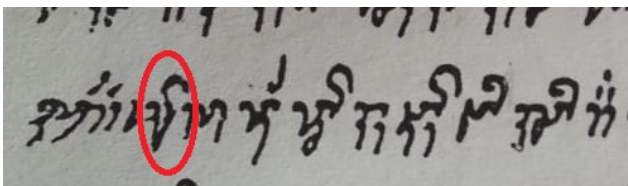

Fig 52. Example of U+1E21B (ℬ LETTER PPA).

នូវការសុំច្បាប់ការឆ្លងការបារស្ត្រីក្រុងសោ

A letter of endorsement

ឆ្នាំ ០៩ ឧសភា ឆ្នាំ ២០១៩
Date : May 04, 2019

ការស្នើមកស៊ីវីន្តការសុំមកឆ្នាំក្រុងសោនត្រខ្ញុំសោការល្បីសោ ៕
To Unicode international committee:

ក្រុមជីវីមកក្រុននផ្គ ក្រំចការត្រានក្នុំនការកុំឧស្សោ ក្រុឌនចក្រុមត្រនក្រការសការសន្តការសោ ត្រនការស្រនសម៉ាុំការក្រុងសោការល្បី ៕
      We, the Islamic Community Kan Imam San of Cambodia, support the Proposal to encode Western Cham in the UCS.

ក្រុឌនចក្រុមត្រនការការនការណ្នក្រ មសជីត្រស្វត្រ វិនន ការមចន្តស្ត្រ ន្តការសោ ននក្រុន ក្រុឌនម្តត្រការសត្រនបន្តនន ការត្រចោម
ជាសម៉ាត្រការការនត្រនការការកការ ៕
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.

នស្តនសន នសក្រននក្រ ចក្រុនស្ត្រត្រ នការនការនញ្ននសន្តនន មសោ ៕
These are the names of the people who endorse this proposal with its signatures.

១ (ក្រុសោ នន វសនននន ក្រុំនស្ត នខ្ញុំ នន
1. Chief Imam of entire Cham Iman San Group (OngKhnuo Math Sa)

២ (ចក្រន ក្រុសោ) ខ្ញុំត្រនសនស ឌននកការនន ក្រ ម៉ាចនភ្តន
2. Imam right hand of Srae Prey village (Kokk Math)

៣ (ក្រុសោ នន) ខ្ញុំត្រនសនស ឌនការ ននននខ្ញុំ
3. Head Imam of Phum Keoh village (Math Soum)

៤ (ខ្ញុំសោ នន) ខ្ញុំត្រនសនស ឌនការ ននននសនក្រ
4. Head Imam of Phum Caghoh village (Ol Lee)

Sincerely

ក្រុសោ នន វសនននន
ចក្រនននន ក្រុំចការននខ្ញុំ នខ្ញុំ នការមនឧស្សោ

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

PROPOSAL SUMMARY FORM TO ACCOMPANY SUBMISSIONS
FOR ADDITIONS TO THE REPERTOIRE OF ISO/IEC 10646.[4]
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: *12/Nov/2020*
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: *X*
      Name of the existing block: *Arabic; Arabic Supplemental Symbols*
2. Number of characters in proposal: *96*
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?
      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.

C. Technical - Justification

1. Has this proposal for addition of character(s) been submitted before?   *yes*
    If YES explain   *L2/20-061*
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?   *supporting letter enclosed*
        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?   *no*
    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: *This document*
6. After giving due considerations to the principles in the P&P document must the proposed characters be entirely
    in the BMP?   *no*
        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?   *yes*
        If YES, is a rationale for its inclusion provided?   *yes*
        If YES, reference: *This document*
11. Does the proposal include use of combining characters and/or use of composite sequences?   *yes*
    If YES, is a rationale for such use provided?   *yes*
        If YES, reference: *This document*
    Is a list of composite sequences and their corresponding glyph images (graphic symbols) provided?   *no*
        If YES, reference:
12. Does the proposal contain characters with any special properties such as
    control function or similar semantics?   *no*
        If YES, describe in detail (include attachment if necessary)

13. Does the proposal contain any Ideographic compatibility characters?   *no*
    If YES, are the equivalent corresponding unified ideographic characters identified?
        If YES, reference: