

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

**Doc Type:** Working Group Document  
**Title:** Proposal to encode Western Cham in the UCS  
**Source:** Martin Hosken  
**Status:** Individual contribution  
**Action:** For consideration by UTC and ISO  
**Date:** 2019-06-12

**Executive Summary.** This proposal is to add 99 characters to a block: 1E200-1E26F in the supplementary plane, named Western Cham. The proposal is a modification of L2/16-198 (N4734) and while largely in harmony with it does remove some proposed characters

**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. 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.

**Repertoire.** There are various extra characters for addition to Western Cham and some proposed for removal from the existing proposals. They consist of a few different categories. In this discussion, codes presented in italics are for characters not present in the chart (even if the code is used for a different character). They are used for discussion of deleted characters based on their alignment with Eastern Cham.

1. *U+1E241* (FINAL G) is a borrowed sound into Eastern Cham from the surrounding Vietnamese. There is no such sound in Khmer and so Western Cham has no need for this sound. Thus we remove the character and *U+1E240* (FINAL K) may be used instead.
2. *U+1E243* (FINAL NG) and *U+1E22E* (VOWEL OE) are identical glyphs in Western Cham and are considered the same character by Western Cham users. To avoid confusion we remove *U+1E243* (FINAL NG) and keep *U+1E22E* (VOWEL OE OR FINAL NG) to act for both functions. This means that *U+1E22E* (VOWEL OE OR FINAL NG) may occur twice as in *ঃ* *U+1E221* (LETTER BBA) *U+1E22E* (VOWEL OE OR FINAL NG) *U+1E22E* (VOWEL OE OR FINAL NG). Also, in order for there to be a consistent location for the character in a diacritic sequence it is stored after any final vowels. See the following section on combining order.
3. The 5 characters from N4734, *U+1E23B* (FINAL PH) ... *U+1E23F* (FINAL S) were proposed to represent final sounds only occurring in Khmer words that have been borrowed into Western Cham. Of these *U+1E23C* (FINAL B) and *U+1E23D* (FINAL M) have been found to be redundant by the Western Cham community in Cambodia. *U+1E23C* (FINAL B) may be represented by *U+1E247* (FINAL P) and *U+1E23D* (FINAL M) by *U+1E24C* (SIGN FINAL M). We close the gap, thus *U+1E23B* (FINAL PH) from N4734 becomes *U+1E23D* (FINAL PH).

4. The 6 textual symbols U+1E260 (SIGN TANA TAMAT PHON) ... U+1E265 (SIGN ARABIC END) are textual symbols with discourse functions. Specifically:

Codepoint	Character name	Glyph	Description	Script
1E260	SIGN TANA TAMA PHON	﴿	Closing Koranic quotation	Arabic
1E261	SIGN TANA CHAPEDAEM PHON	﴿	Section beginning	Cham
1E262	SIGN TANA CHAPEDAEM IU	﴿	Alternate section beginning	Cham
1E263	SIGN TANA TAMAT AYET	﴿	Closing affirmation	Arabic
1E264	SIGN TANA TAMAT TAKUE	﴿	Closing chants	Arabic
1E265	SIGN ARABIC END	﴾	Closing text	Arabic

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.

5. 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
1E23A	CONSONANT SIGN BRAY	ڧ	1E21D 1E234 1E22C	ڧ
1E23B	LETTER NAN	ڻڻ	1E217 1E246	ڻڻ
1E23C	CONSONANT SIGN SHALOM	ڦڙ	1E227 1E224 1E24C	ڦڙ

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

6. U+1E25B (SIGN TANA PATOK SAP) reduplication mark and U+1E25C (PUNCTUATION SPIRAL) ‘one’ or ‘first’ are the same as in N4374. Note that while U+1E25C (PUNCTUATION SPIRAL) has lexical value, it takes no diacritics.
7. U+1E237 (CONSONANT SIGN TKAJ KA) is an abstract character that only ever ligates with the previous character. It does not occur following a vowel and so acts like a vowel, and is not followed by a final letter either. Ligated sequences involving U+1E237 have a visual similarity to those using U+1E24D (CONSONANT SIGN FINAL H). For example:

Sequence	Glyphs	Meaning	Sequence	Glyphs	Meaning
1E206 1E237	ڦڻ	tie	1E206 1E24D	ڦڻ	push away
1E208 1E237	ڦڻ	stutter	1E208 1E24D	ڦڻ	side

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

8. The character U+1E25A (SIGN 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. ڦ 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 ڦ ڦ 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 occurring on day 15. For example 10 days after a full moon is represented by ڦ ڦ U+1E25A (SIGN MOON TEN) U+1E22E (SIGN OE).

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

9. Western Cham has special characters that can be used for indicating month names. They are based on the transliterated month name into Arabic followed by 1 to 7 dots. The order of the months is U+1E26D (LUNAR 7 DOTS), U+1E267 (LUNAR 2 DOTS), U+1E268 (LUNAR 3 DOTS), U+1E26A (LUNAR 5 DOTS), U+1E26C (LUNAR 6 DOTS), U+1E266 (LUNAR 1 DOT), U+1E267 (LUNAR 2 DOTS), U+1E269 (LUNAR 4 DOTS), U+1E26A (LUNAR 5 DOTS), U+1E26D (LUNAR 7 DOTS), U+1E266 (LUNAR 1 DOT), U+1E269 (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.

In addition, there are two forms of the lunar 5 dots, that occur in free variation in a text, hence U+1E26B (LUNAR 5 DOTS OVER HAH).

10. The characters at *U+1E20B*, *U+1E211* and *U+1E22D* do not occur in Western Cham. The slots are kept to keep character alignment with the Eastern Cham encoding.
11. The characters that would be at *U+1E203* and *U+1E204*, if directly corresponding to Eastern Cham, have no contrastive shape from their corresponding sequences, unlike in Eastern Cham where a visual contrast is evident. Thus

Potential Codepoint and Eastern Cham Correspondance	Sequence	Glyphs
<i>IE203 (AA03)</i>	1E201 1E22F	ڦڻڻ
<i>IE204 (AA04)</i>	1E201 1E230	ڦڻڻ

Note that if such a letter were followed by a medial as in *IE203 IE234*, the corresponding sequence would be 1E201 1E234 1E22F, which also appropriately resolves the shaping issue.

**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	Sample Word Sequence	Glyphs	Meaning
1E201	LETTER I	ڦ	1E224 1E237	ڦڻ	<i>does not occur</i>
1E256	DIGIT SIX	ڻ	1E223 1E236 1E24D	ڻڻڻ	choose
1E257	DIGIT SEVEN	ڻ	1E223 1E237 1E240	ڻڻڻ	veil
1E259	DIGIT NINE	ڻ	1E21D 1E237	ڻڻ	carry (on shoulder)

In addition, U+1E251 (DIGIT ONE) can be confusable with U+1E24D (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 (RA).

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

**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
ଫ୍ରେଣ୍ଡ୍ସ୍	1E21A 1E214 1E233 1E230	Pre vowels go before the last initial.
ପ୍ରେଟ୍ସ୍	1E206 1E232 1E22E	No visual difference based on order.
ପ୍ରେଟ୍ୟୁସ୍ନ୍	1E215 1E223 1E229 1E22E	Interaction between upper and final vowels.
ପ୍ରେଟ୍ୟୁସ୍ନ୍	1E200 1E233 1E229	Interaction between following medial and following vowel.
ଲେନ୍ୟୁସ୍ନ୍	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 are a discourse marker 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).

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

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 OS;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 KHAA;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 PHAA;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 THAA;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 SIGN VOWEL OE OR FINAL NG;Mn;  
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;Mn;0;NSM;;;;N;;;;  
 1E237;WESTERN CHAM CONSONANT SIGN TKAJ KA;Mc;0;NSM;;;;N;;;;  
 1E23A;WESTERN CHAM CONSONANT SIGN BRAY;Lo;0;L;;;;N;;;;  
 1E23B;WESTERN CHAM LETTER NAN;Lo;0;L;;;;N;;;;  
 1E23C;WESTERN CHAM CONSONANT SIGN 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 CONSONANT SIGN FINAL H;Mc;0;L;;;;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;L;;;;N;;;;  
 1E25D;WESTERN CHAM PUNCTUATION DANDA;Po;0;L;;;;N;;;;  
 1E25E;WESTERN CHAM PUNCTUATION DOUBLE DANDA;Po;0;L;;;;N;;;;  
 1E25F;WESTERN CHAM PUNCTUATION TRIPLE DANDA;Po;0;L;;;;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;;;;  
 1E265;WESTERN CHAM SIGN ARABIC END;Po;0;ON;;;;N;;;;  
 1E266;WESTERN CHAM SIGN LUNAR 1 DOTS;Po;0;ON;;;;N;;;;  
 1E267;WESTERN CHAM SIGN LUNAR 2 DOTS;Po;0;ON;;;;N;;;;  
 1E268;WESTERN CHAM SIGN LUNAR 3 DOTS;Po;0;ON;;;;N;;;;  
 1E269;WESTERN CHAM SIGN LUNAR 4 DOTS;Po;0;ON;;;;N;;;;  
 1E26A;WESTERN CHAM SIGN LUNAR 5 DOTS;Po;0;ON;;;;N;;;;  
 1E26B;WESTERN CHAM SIGN LUNAR 5 DOTS OVER HAH;Po;0;ON;;;;N;;;;  
 1E26C;WESTERN CHAM SIGN LUNAR 6 DOTS;Po;0;ON;;;;N;;;;  
 1E26D;WESTERN CHAM SIGN LUNAR 7 DOTS;Po;0;ON;;;;N;;;;

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

1E200 (ঃ) < 1E201 (ঁ) < 1E202 (ং) < 1E205 (ঃঃ)

1E206 (ঃঃ) < 1E207 (ঃঃঃ) < 1E208 (ঃঃঃ) < 1E209 (ঃঃঃ) < 1E20A (ঃঃঃ) < 1E20D (ঃঃঃ) <  
 1E20E (ঃঃঃ) < 1E20F (ঃঃঃ) < 1E210 (ঃঃঃ) < 1E212 (ঃঃঃ) < 1E213 (ঃঃঃ) < 1E214 (ঃঃঃ) < 1E215 (ঃঃঃ) <  
 1E216 (ঃঃঃ) < 1E217 (ঃঃঃ) < 1E218 (ঃঃঃ) < 1E219 (ঃঃঃ) < 1E21A (ঃঃঃ) < 1E21B (ঃঃঃ) < 1E21C (ঃঃঃ) <  
 1E21D (ঃঃঃ) < 1E21E (ঃঃঃ) < 1E21F (ঃঃঃ) < 1E220 (ঃঃঃ) < 1E221 (ঃঃঃ) < 1E222 (ঃঃঃ) < 1E223 (ঃঃঃ) < 1E224  
 (ঃঃঃ) < 1E225 (ঃঃঃ) < 1E226 (ঃঃঃ) < 1E227 (ঃঃঃ) < 1E228 (ঃঃঃ)

1E233 (ঃঃঃ) < 1E234 (ঃঃঃ) < 1E235 (ঃঃঃ) < 1E236 (ঃঃঃ) < 1E237 (ঃঃঃ)

1E229 (ঃঃঃ) < 1E22A (ঃঃঃ) < 1E22B (ঃঃঃ) < 1E22C (ঃঃঃ) < 1E22E (ঃঃঃ) < 1E22F (ঃঃঃ) < 1E230 (ঃঃঃ) < 1E231  
 (ঃঃঃ) < 1E232 (ঃঃঃ)

1E240 (ঃঃঃ) < 1E242 (ঃঃঃ) < 1E244 (ঃঃঃ) < 1E245 (ঃঃঃ) < 1E246 (ঃঃঃ) < 1E247 (ঃঃঃ) < 1E23D (ঃঃঃ)

1E24C (ঃঃঃ) < 1E248 (ঃঃঃ) < 1E249 (ঃঃঃ) < 1E24A (ঃঃঃ) < 1E23E (ঃঃঃ) < 1E24B (ঃঃঃ) < 1E23F (ঃঃঃ) < 1E24D  
 (ঃঃঃ)

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 final:

U+1E229 U+1E242 < U+1E229 U+1E22E  
U+1E22A U+1E242 < U+1E22A U+1E22E  
U+1E22B U+1E242 < U+1E22B U+1E22E  
U+1E22C U+1E242 < U+1E22C U+1E22E  
U+1E22E U+1E242 < U+1E22E U+1E22E  
U+1E22F U+1E242 < U+1E22F U+1E22E  
U+1E230 U+1E242 < U+1E230 U+1E22E  
U+1E231 U+1E242 < U+1E231 U+1E22E  
U+1E232 U+1E242 < U+1E232 U+1E22E

**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.<sup>1</sup> 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	ঁ 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), ঁ 1E230 (VOWEL SIGN AAI)
Lower vowels	ঁ 1E232 (VOWEL SIGN U)
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)
Finals	ঁ 1E237 (CONSONANT SIGN TKAJ KA), ঁ 1E24C (CONSONANT SIGN FINAL M), ঁ 1E238 (CONSONANT SIGN FINAL H)

No more than one character in each class may occur following the same base character (1E200 ... 1E228). The only exception to this U+1E22E (VOWEL SIGN OE).

**Acknowledgements.** Thanks go to Payap University Linguistics Institute, Chiang Mai, Thailand, under whose auspices this work is done. Special thanks go to Diethelm Kanjahn who created the font.

<sup>1</sup> 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.

## Samples

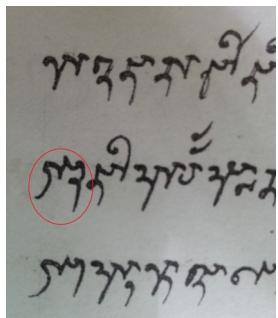


Fig 1. Showing U+1E206 (LETTER KA) U+1E237 (CONSONANT SIGN TKAJ KA)

Three lines of Korean handwriting. The first line shows the letter TA. The second line shows the consonant sign TKAJ KA circled in red. The third line shows another character.

Fig 2. Showing U+1E213 (LETTER TA) U+1E237 (CONSONANT SIGN TKAJ KA). The sign to the far left is U+1E262 (SIGN TANA CHAPEDAEM PHON)

Two lines of Korean handwriting. The first line starts with a blue bracket and ends with a blue double danda. The second line starts with a blue double danda and ends with a blue bracket. A red circle highlights a specific character in the middle of the text.

Fig 3. Showing salam and the use of double dandas around a quotation

Two lines of Korean handwriting. A red circle highlights a specific character in the middle of the text.

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

Two lines of Korean handwriting. A red circle highlights a specific character in the middle of the text.

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

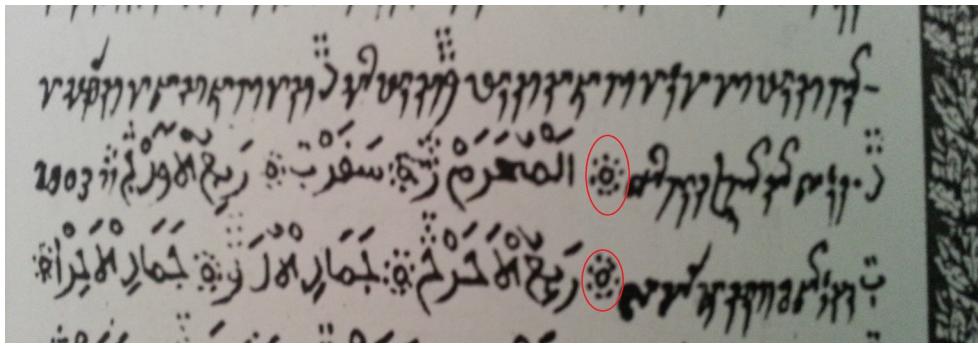


Fig 6. Showing U+1E263 (SIGN TANA TAMAT AYET). Notice the mix of Arabic and Western Cham and also the lunar characters.

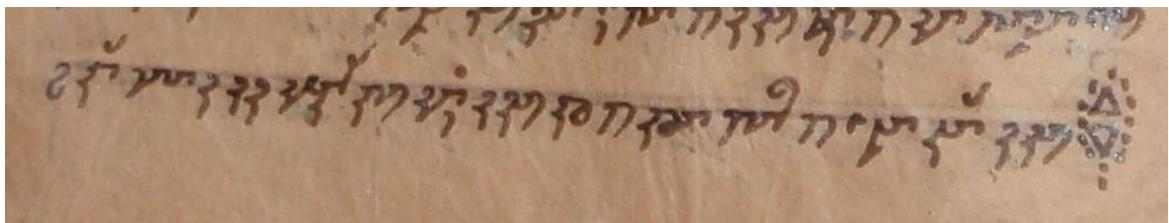


Fig 7. Showing U+1E264 (SIGN TANA TAMA TAKUE)

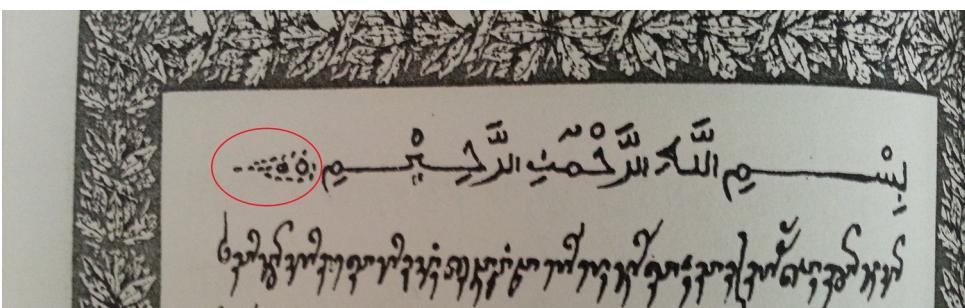


Fig 8. Showing U+1E265 (SIGN ARABIC END) at the end of a line of Arabic text before Western Cham text. The first character of the Western Cham text is U+1E262 (SIGN TAHA CHAPEDAEM PHON)

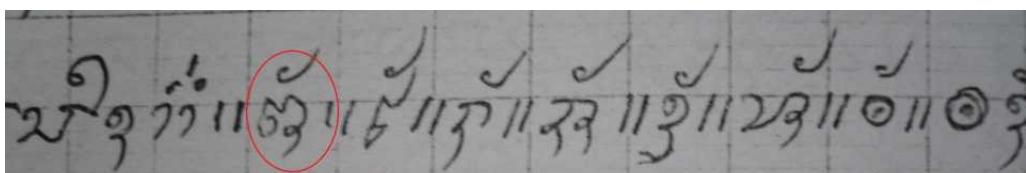


Fig 9. Showing digits 3 (circled), 4, 6, 7, 8, 9, moon 10, taking a vowel, separated by U+0965 (DEVANAGARI DOUBLE DANDA)

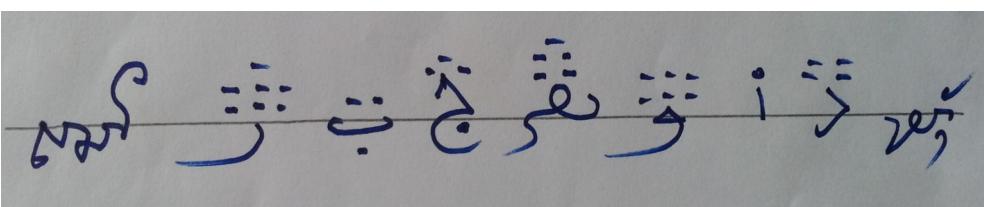


Fig 10. Showing the relative positions of Western Cham letters and lunar month symbols relative to a conceptual baseline.

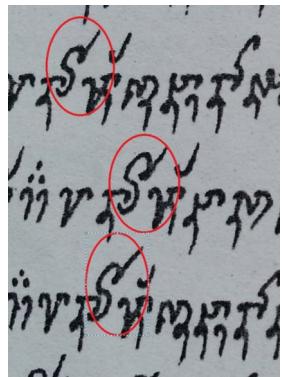


Fig 11. Showing academic variants of Western Cham vowel sequences.

## ଆମ୍ବାଦାର ପାତ୍ରକାରୀ ଲେଖନ

፳፻፲፭ ዓ.ም ከፃዬ ምት በግብር ተከራክሯል ቤትና ቅዱን  
Date : May 04, 2019

ଓ রাজ্যীয় মন্ত্রণালয়ের সমূহ বিভিন্ন বিষয়ে উকিল এবং প্রতিক্রিয়া দেওয়া গুরুত্ব এবং  
To Unicode international committee:

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

*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.*

ପ୍ରକାଶ ମହିନେ ଏକଟି ପରିବର୍ତ୍ତନ ଆବଶ୍ୟକ ହେଲା ଏବଂ ଏହା କିମ୍ବା ଏହାରେ ଏକାକିଳୀଁ ଏହା କିମ୍ବା ଏହାରେ ଏକାକିଳୀଁ

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

၆ (မြတ် ၁၅။) ခလောက်သာဆုံး၊ ရှာနိုင်ရန်သာဆုံးများ

1. Chief Imam of entire Cham Iman San Group (OngKhnuo Math Sa)

២. Imam right hand of Srae Prey village (Kokk Math)

ឯ (ផ្លូវ ភ៉ា) នាមីមិន់រាជការបានស្វែងរកពីរាជការណ៍របស់រាជការ

၄. Head Imam of Phum Caghoh village (Ol Lee)

*Sincerely*

မွေး၏ နှစ်ခုပြုလောင်းများ  
ကျေးဇူးတွေများ၊ ကျေးဇူးတွေများ၊ နှစ်ခုပြုလောင်းများ

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



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

Please fill all the sections A, B and C below.

Please read Principles and Procedures Document (P & P) from <http://std.dkug.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.dkug.dk/JTC1/SC2/WG2/docs/summaryform.html>.  
See also <http://std.dkug.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:	99	
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 <a href="http://www.unicode.org">http://www.unicode.org</a> for such information on other scripts. Also see Unicode Character Database ( <a href="http://www.unicode.org/reports/tr44/">http://www.unicode.org/reports/tr44/</a> ) and associated Unicode Technical Reports for information needed for consideration by the Unicode Technical Committee for inclusion in the Unicode Standard.		

2 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>	yes
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?	<i>supporting letter enclosed</i>	yes
3. Information on the user community for the proposed characters (for example: size, demographics, information technology use, or publishing use) is included? Reference:		no
4. The context of use for the proposed characters (type of use; common or rare) Reference:		common
5. Are the proposed characters in current use by the user community? If YES, where? Reference:	<i>This document</i>	yes
6. After giving due considerations to the principles in the P&P document must the proposed characters be entirely in the BMP? If YES, is a rationale provided? If YES, reference:		no
7. Should the proposed characters be kept together in a contiguous range (rather than being scattered)?		yes
8. Can any of the proposed characters be considered a presentation form of an existing character or character sequence? If YES, is a rationale for its inclusion provided? If YES, reference:		no
9. Can any of the proposed characters be encoded using a composed character sequence of either existing characters or other proposed characters? If YES, is a rationale for its inclusion provided? If YES, reference:		no
10. Can any of the proposed character(s) be considered to be similar (in appearance or function) to, or could be confused with, an existing character? If YES, is a rationale for its inclusion provided? If YES, reference:	<i>This document</i>	yes yes
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>	yes yes
12. Does the proposal contain characters with any special properties such as control function or similar semantics? If YES, describe in detail (include attachment if necessary)		no
13. Does the proposal contain any Ideographic compatibility characters? If YES, are the equivalent corresponding unified ideographic characters identified? If YES, reference:		no

1E200

# Western Cham

**1E26F**

# Western Cham

Western Cham		1E23C	ቁ	WESTERN CHAM CONSONANT SIGN SHALOM
1E200	፩	WESTERN CHAM LETTER A	1E23D	ቁ
1E201	፪	WESTERN CHAM LETTER E	1E23E	ቃ
1E202	፫	WESTERN CHAM LETTER O	1E23F	ቄ
1E205	፬	WESTERN CHAM LETTER OS	1E240	ቅ
1E206	፭	WESTERN CHAM LETTER KA	1E242	ቆ
1E207	፮	WESTERN CHAM LETTER KHA	1E244	ቈ
1E208	፯	WESTERN CHAM LETTER GA	1E245	቉
1E209	፱	WESTERN CHAM LETTER KHAA	1E246	ቊ
1E20A	፲	WESTERN CHAM LETTER NGUE	1E247	ቋ
1E20C	፳	WESTERN CHAM LETTER CHA	1E248	ቌ
1E20D	፴	WESTERN CHAM LETTER CHHA	1E249	ቍ
1E20E	፵	WESTERN CHAM LETTER JA	1E24A	቏
1E20F	፶	WESTERN CHAM LETTER JHA	1E24B	ቊ
1E210	፷	WESTERN CHAM LETTER NHUE	1E24C	ሸ
1E212	፹	WESTERN CHAM LETTER NHJA	1E24D	ሹ
1E213	፻	WESTERN CHAM LETTER TA		WESTERN CHAM CONSONANT SIGN FINAL H
1E214	፼	WESTERN CHAM LETTER THA	1E250	፦
1E215	፽	WESTERN CHAM LETTER DA	1E251	፾
1E216	፿	WESTERN CHAM LETTER PHAA	1E252	፿
1E217	፿	WESTERN CHAM LETTER NUE	1E253	፿
1E218	፿	WESTERN CHAM LETTER NA	1E254	፿
1E219	፿	WESTERN CHAM LETTER DDA	1E255	፿
1E21A	፿	WESTERN CHAM LETTER PA	1E256	፿
1E21B	፿	WESTERN CHAM LETTER PPA	1E257	፿
1E21C	፿	WESTERN CHAM LETTER PHA	1E258	፿
1E21D	፿	WESTERN CHAM LETTER BA	1E259	፿
1E21E	፿	WESTERN CHAM LETTER THAA	1E25A	፦
1E21F	፿	WESTERN CHAM LETTER MUE	1E25B	፯
1E220	፿	WESTERN CHAM LETTER MA	1E25C	፿
1E221	፿	WESTERN CHAM LETTER BBA	1E25D	፿
1E222	፿	WESTERN CHAM LETTER YA	1E25E	፿
1E223	፿	WESTERN CHAM LETTER RA	1E25F	፿፿፿፿፿
1E224	፿	WESTERN CHAM LETTER LA	1E260	፿
1E225	፿	WESTERN CHAM LETTER VA	1E261	፿
1E226	፿	WESTERN CHAM LETTER SSA	1E262	፿
1E227	፿	WESTERN CHAM LETTER SA	1E263	፿
1E228	፿	WESTERN CHAM LETTER HA	1E264	፿
1E229	፿	WESTERN CHAM VOWEL SIGN AA	1E265	፿፿፿፿፿
1E22A	፿	WESTERN CHAM VOWEL SIGN I	1E266	፿
1E22B	፿	WESTERN CHAM VOWEL SIGN II	1E267	፿
1E22C	፿	WESTERN CHAM VOWEL SIGN AI	1E268	፿፿፿፿፿
1E22E	፿	WESTERN CHAM SIGN VOWEL OE OR FINAL NG	1E269	፿፿፿፿፿
1E22F	፿	WESTERN CHAM VOWEL SIGN E	1E26A	፿፿፿፿፿
1E230	፿	WESTERN CHAM VOWEL SIGN AAI	1E26B	፿፿፿፿፿
1E231	፿	WESTERN CHAM VOWEL SIGN AU		WESTERN CHAM SIGN ARABIC END
1E232	፿	WESTERN CHAM VOWEL SIGN U		WESTERN CHAM SIGN LUNAR 1 DOTS
1E233	፿	WESTERN CHAM CONSONANT SIGN YA		WESTERN CHAM SIGN LUNAR 2 DOTS
1E234	፿	WESTERN CHAM CONSONANT SIGN RA		WESTERN CHAM SIGN LUNAR 3 DOTS
1E235	፿	WESTERN CHAM CONSONANT SIGN LA		WESTERN CHAM SIGN LUNAR 4 DOTS
1E236	፿	WESTERN CHAM CONSONANT SIGN WA		WESTERN CHAM SIGN LUNAR 5 DOTS
1E237	፿	WESTERN CHAM CONSONANT SIGN TKAJ KA		WESTERN CHAM SIGN LUNAR 5 DOTS OVER HAH
1E23A	፿	WESTERN CHAM CONSONANT SIGN BRAY	1E26C	፿
1E23B	፿	WESTERN CHAM LETTER NAN	1E26D	፿