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 (Revised)**
Source:	**Martin Hosken**
Status:	**Individual contribution**
Action:	**For consideration by UTC and ISO**
Date:	**2019-09-01**

Executive Summary. This proposal is to add 98 characters to a block: 1E200-1E26F in the supplementary plane, named Western Cham. In addition, 1 character is proposed for inclusion in the Arabic Extended A block: 08A0-08FF. The proposal is a modification of L2/16-198 (N4734) and while largely in harmony with it does remove some proposed characters.

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.

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, and therefore having no representative glyph (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 (ꬲ 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 ꬦ U+1E221 (ꬡ 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 (ꬦ 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 (ꬲ 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 (ꬲ FINAL PH).

4. The 6 textual symbols U+1E260 (SIGN TANA TAMAT PHON) … U+1E264 (SIGN TANA TAMAT TAKUE) are textual symbols with discourse functions. Specifically:

Codepoint	Character name	Glyph	Description	Script
1E260	SIGN TANA TAMA PHON	ꭱ	Closing Korannic quotation	Arabic
1E261	SIGN TANA CHAPEDAEM PHON	ꭱ	Section beginning	Cham
1E262	SIGN TANA CHAPEDAEM IU	ꭱ	Alternate section beginning	Cham
1E263	SIGN TANA TAMAT AYET	◌	Closing affirmation	Arabic
1E264	SIGN TANA TAMAT TAKUE	◌	Closing chants	Arabic
08D2	ARABIC END OF TEXT	·∴◌◌	Closing text	Arabic

In styling, U+1E260 (ꭱ SIGN TANA TAMA PHON) and U+1E262 (ꭱ SIGN TANA CAPEDAEM IU) may lose their dotted outline and U+1E261 (ꭱ SIGN TANA CHAPEDAEM PHON) may gain one. The script indiciation 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+08D2 (⸱⸱⸳⸴ 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.

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	౬	1E21D 1E234 1E22C	౷
1E23B	LETTER NAN	౫౫	1E217 1E246	౫ ౫
1E23C	CONSONANT SIGN SHALOM	౫ఀ	1E227 1E224 1E24C	౫ ఀ

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

6. U+1E25B (ⵏ 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 (⦿ PUCTUATION SPIRAL) has lexical value, it takes no diacritics.

7. U+1E237 (◌ఌ 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 (◌ఌ CONSONANT SIGN FINAL H). For example:

Sequence	Glyphs	Meaning	Sequence	Glyphs	Meaning
1E206 1E237	౫౩	tie	1E206 1E24D	౫౩	push away
1E208 1E237	౫౩	stutter	1E208 1E24D	౫౩	side

U+1E24D (CONSONANT SIGN FINAL H) may occur following a vowel as in ౫౫౫.

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. ౫ U+1E251 (౫ 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 ⦿ ౫ U+1E25A (⦿ SIGN MOON TEN) U+1E251 (౫ 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 (⦿ 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, 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. The order of the months is U+1E26C (ﺝ LUNAR 7 DOTS), U+1E266 (ﺏ LUNAR 2 DOTS), U+1E267 (ﺥ LUNAR 3 DOTS), U+1E269 (ﺥ LUNAR 5 DOTS), U+1E26B (ﺝ LUNAR 6 DOTS), U+1E265 (ﺍ LUNAR 1 DOT), U+1E266 (ﺏ LUNAR 2 DOTS), U+1E268 (ﺥ LUNAR 4 DOTS), U+1E269 (ﺥ LUNAR 5 DOTS), U+1E26C (ﺝ LUNAR 7 DOTS), U+1E265 (ﺍ LUNAR 1 DOT), U+1E268 (ﺥ LUNAR 4 DOTS). Notice the repetitions in the order. These symbols are used both in Western Cham script and also in Arabic script, when writing Western Cham. They may occur next to each other and with Western Cham digits but never with Arabic digits.

In addition, there are two forms of the lunar 5 dots, that occur in free variation in a text, hence U+1E26A (ﻩ LUNAR 5 DOTS OVER HEH).

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

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
1E203 (AA03)	1E201 1E22F	ᷛᷫᷩ
1E204 (AA04)	1E201 1E230	ᷛᷫᷩ

Note that if such a letter were followed by a medial as in *1E203 1E234*, the corresponding sequence would be 1E201 (ᷫᷩ A) 1E234 (ᷛᷩ SIGN RA) 1E22F (ᷛ◌ SIGN E), which also appropriately resolves the shaping issue.

Ligation. Western Cham is visually highly ligating. So for example, ᴾᵛᵖ U+1E206 (ᴾᵛᵖ KA) U+1E22A (◌ᷩ VOWEL I) and then ᴾᵛᵖ U+1E206 (ᴾᵛᵖ KA) U+1E236 (◌ᷩ SIGN WA) U+1E22A (◌ᷩ 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	Sample Word Sequence	Glyphs	Meaning
1E201	LETTER I	ᷫᷩ	1E224 1E237	ᷫᷩ	*does not occur*
1E256	DIGIT SIX	ᷩᵖ	1E223 1E236 1E24D	ᷩᵖᷩ	choose
1E257	DIGIT SEVEN	ᷩᷩ	1E223 1E237 1E240	ᷩᷩᴾᵛ	veil
1E259	DIGIT NINE	ᷩᷩ	1E21D 1E237	ᷩᷩ	carry (on shoulder)
1E267	LUNAR 3 DOTS	◌̇	08B5		
1E268	LUNAR 4 DOTS	◌̈	0690	◌̈	

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	ᷩᷩᷩ	ᷩᷩ
1E223 1E233 1E229	ᷩᷩᷩ	ᷩᷩ

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

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

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

Most finals are productive modifications of their corresponding initials. There is a limited set of finals and feelings run strongly over whether other finals may or may not occur. Use of a separate 'finalising' character would not be appropriate given cultural sensibilities. In addition it breaks with the Unicode principle of not using modifiers as productive of other characters. This would make Western Cham unique if it were introduced (cf. New Tai Lue and Cham).

U+1E22E (◌́ SIGN VOWEL OE OR FINAL NG) and U+1E24C (◌̊ SIGN FINAL M) are both considered marks because they are non-spacing. U+1E24D (◌ꯣ SIGN FINAL H) inherits its mark status from the corresponding character in Eastern Cham. 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 (◌ꯣ SIGN 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
꯫ ꯦ ꯢ	1E21A 1E214 1E233 1E230	Pre vowels go before the last initial.
ꯞ̊	1E206 1E232 1E22E	No visual difference based on order.
ꯤ ꯣ	1E215 1E223 1E229 1E22E	Interaction between upper and final vowels.
ꯨꯥ	1E200 1E233 1E229	Interaction between following medial and following vowel.
꯰	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.

```
08D2;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 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;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;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 LUNAR 1 DOTS;Po;0;ON;;;;;N;;;;;
1E266;WESTERN CHAM SIGN LUNAR 2 DOTS;Po;0;ON;;;;;N;;;;;
1E267;WESTERN CHAM SIGN LUNAR 3 DOTS;Po;0;ON;;;;;N;;;;;
1E268;WESTERN CHAM SIGN LUNAR 4 DOTS;Po;0;ON;;;;;N;;;;;
1E269;WESTERN CHAM SIGN LUNAR 5 DOTS;Po;0;ON;;;;;N;;;;;
1E26A;WESTERN CHAM SIGN LUNAR 5 DOTS OVER HEH;Po;0;ON;;;;;N;;;;;
1E26B;WESTERN CHAM SIGN LUNAR 6 DOTS;Po;0;ON;;;;;N;;;;;
1E26C;WESTERN CHAM SIGN LUNAR 7 DOTS;Po;0;ON;;;;;N;;;;;
```

U+08D2 (ﬣﬢﬤﬡ ARABIC END OF TEXT) has a joining behaviour of non-joining.

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

1E200 (ꢀ) < 1E201 (ꢁ) < 1E202 (ꢂ) < 1E201 (ꢁ) 1E22F (ꢯ◌) < 1E201 (ꢁ) 1E230 (ꢰ◌) < 1E205 (ꢅ)

1E206 (ꢆ) < 1E207 (ꢇ) < 1E208 (ꢈ) < 1E209 (ꢉ) < 1E20A (ꢊ) < 1E20C (ꢌ) < 1E20D (ꢍ) < 1E20E (ꢎ) < 1E20F (ꢏ) < 1E210 (ꢐ) < 1E212 (ꢒ) < 1E213 (ꢓ) < 1E214 (ꢔ) < 1E215 (ꢕ) < 1E216 (ꢖ) < 1E217 (ꢗ) < 1E218 (ꢘ) < 1E219 (ꢙ) < 1E21A (ꢚ) < 1E21B (ꢛ) < 1E21C (ꢜ) < 1E21D (ꢝ) < 1E21E (ꢞ) < 1E21F (ꢟ) < 1E220 (ꢠ) < 1E221 (ꢡ) < 1E222 (ꢢ) < 1E223 (ꢣ) < 1E224 (ꢤ) < 1E225 (ꢥ) < 1E226 (ꢦ) < 1E227 (ꢧ) < 1E228 (ꢨ)

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

1E229 (◌ꢩ) < 1E22A (◌ꢪ) < 1E22B (◌ꢫ) < 1E22C (◌ꢬ) < 1E22F (ꢯ◌) < 1E230 (ꢰ◌) < 1E231 (◌ꢱ) < 1E232 (◌ꢲ)

1E240 (ꢀꢁ) < 1E242 (ꢊꢁ) < 1E22E (◌ꢮ) < 1E244 (ꢌ) < 1E245 (ꢓ) < 1E246 (ꢊꢁ) < 1E247 (ꢧꢁ) < 1E23D (ꢜꢁ) < 1E24C (◌) < 1E248 (ꢢꢁ) < 1E249 (ꢣꢁ) < 1E24A (ꢤꢁ) < 1E23E (ꢚꢁ) < 1E24B (ꢦꢁ) < 1E23F (ꢧꢁ) < 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 vowel:

```
U+1E22C U+1E23D < U+1E22E U+1E23D
U+1E22C U+1E23E < U+1E22E U+1E23E
U+1E22C U+1E23F < U+1E22E U+1E23F
U+1E22C U+1E240 < U+1E22E U+1E240
U+1E22C U+1E241 < U+1E22E U+1E241
U+1E22C U+1E242 < U+1E22E U+1E242
U+1E22C U+1E22E < U+1E22E U+1E22E
U+1E22C U+1E244 < U+1E22E U+1E244
U+1E22C U+1E245 < U+1E22E U+1E245
U+1E22C U+1E246 < U+1E22E U+1E246
U+1E22C U+1E247 < U+1E22E U+1E247
U+1E22C U+1E248 < U+1E22E U+1E248
U+1E22C U+1E249 < U+1E22E U+1E249
U+1E22C U+1E24A < U+1E22E U+1E24A
U+1E22C U+1E24B < U+1E22E U+1E24B
```

U+1E22C U+1E24C < U+1E22E U+1E24C
U+1E22C U+1E24D < U+1E22E U+1E24D

Combining Orders. Due to the immutability of Canonical Combining Class (CCC) values and the confusion between whether the CCC value is an order or a class, it is required that the CCC values for all characters in this block SHALL BE 0.[1] This does not mean that relative ordering of marks is not important. Here is a list of mark classes in their relative order. Those 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)
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), (CONSONANT SIGN FINAL H)

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). The Finals class also includes all the spacing final letters U+1E23D (FINAL PH) … U+1E24B (FINAL SS).

Indic Properties. Based on the ordering rules above, we can set the following Indic properties for Western Cham. Ranges may include unassigned characters for convenience.

```
1E200..1E205   ; Vowel_Independent
1E206..1E228   ; Consonant
1E229..1E232   ; Vowel_Dependent
1E233..1E236   ; Consonant_Medial
1E237          ; Vowel_Dependent
1E23A..1E23C   ; Consonant_Placeholder
1E23D..1E24D   ; Consonant_Final
1E250..1E25A   ; Digits

1E229..1E22E   ; Top
1E22F..1E230   ; Left
1E231          ; Top
1E232          ; Bottom
1E233          ; Right
1E234          ; Left
1E235          ; Bottom
1E236..1E237   ; Right
1E24C          ; Top
1E24D          ; Right
```

---

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

Revisions. 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.
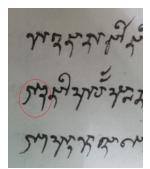
Samples

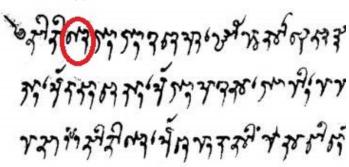*Fig 1. Showing U+1E206 (ꈆ LETTER KA) U+1E237 (◌ CONSONANT SIGN TKAJ KA)*

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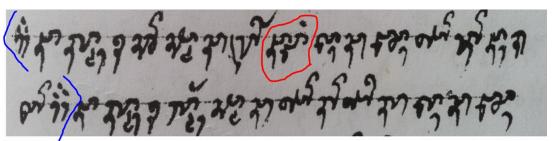

Fig 3. Showing U+1E23C (◌ CONSONANT SIGN SHALOM) and the use of U+1E25E (◌ DOUBLE DANDA) around a quotation
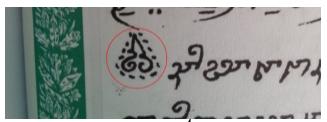

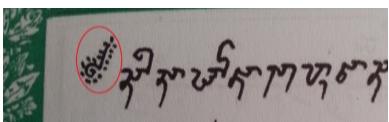*Fig 4. Showing U+1E261 (◌ SIGN TANA CHAPEDAEM PHON)*
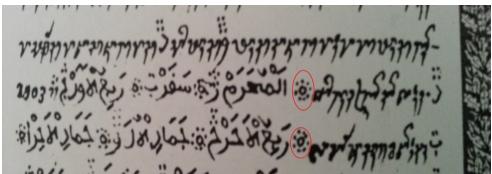

Fig 5. Showing U+1E262 (◌ SIGN CHAPEDAEM IU)

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

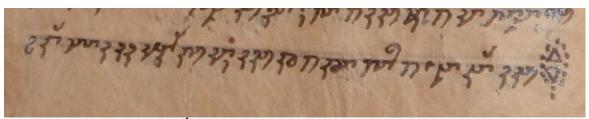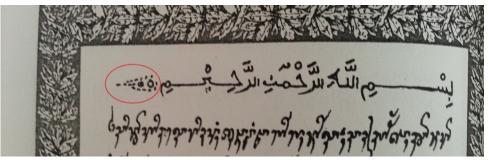*Fig 7. Showing U+1E264 (⸱ SIGN TANA TAMA TAKUE)*

Fig 8. Showing U+08D2 (⸱⸱⸱⸱ ARABIC END OF TEXT) at the end of a line of Arabic text before Western Cham text.The first character of the Western Cham text is U+1E262 (⸱ SIGN TAHA CHAPEDAEM PHON)
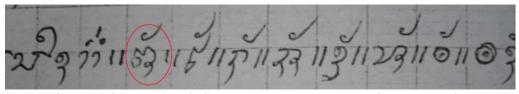

Fig 9. Showing digits 3 (circled), 4, 6, 7, 8, 9, moon 10, taking a vowel, separated by U+0965 (॥ DEVANAGARI DOUBLE DANDA)
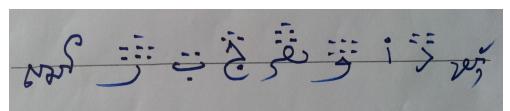

Fig 10. Showing the relative positions of Western Cham letters and lunar month symbols relative to a conceptual baseline.

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

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

Fig 13. Showing a font variant of U+1E236 (◌ꞩ SIGN WA)

Fig 14. Showing two alternates of U+1E235 (◌ꞩ SIGN LA). The second two samples show font variants of the same alternate glyph.

Fig 15. Showing the sequence U+1E21D (Ꝟ BA) U+1E237 (◌ꞩ SIGN TKAJ KA)

Fig 16. Showing the sequence U+1E220 (ꞨꞨ MA) U+1E237 (◌ꞩ SIGN TKAJ KA)

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

Fig 18. Showing the sequence U+1E207 (ꈀꐚ KHA) U+1E234 (ꐚ SIGN RA) U+1E235 (ꐚ SIGN LA)

Fig 19. Showing error marks U+0353 (ꐚ COMBINING X BELOW)
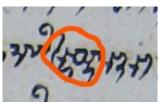
A letter of endorsement

Date : May 04, 2019

To Unicode international committee:

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.

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

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

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

3. Head Imam of Phum Keoh village (Math Soum)

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

Sincerely

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[2]
Please fill all the sections A, B and C below.
Please read Principles and Procedures Document (P & P) from http://std.dkuug.dk/JTC1/SC2/WG2/docs/principles.html **for guidelines and details before filling this form.**
Please ensure you are using the latest Form from http://std.dkuug.dk/JTC1/SC2/WG2/docs/summaryform.html.
See also http://std.dkuug.dk/JTC1/SC2/WG2/docs/roadmaps.html **for latest *Roadmaps*.**

A. Administrative

1. **Title:** Western Cham
2. Requester's name: *Martin Hosken*
3. Requester type (Member body/Liaison/Individual contribution): *Individual Contribution*
4. Submission date: *20/May/19*
5. Requester's reference (if applicable):
6. Choose one of the following:
   This is a complete proposal: *X*
   (or) More information will be provided later:

B. Technical – General

1. Choose one of the following:
   a. This proposal is for a new script (set of characters): *X*
   Proposed name of script: *Western Cham*
   b. The proposal is for addition of character(s) to an existing block:
   Name of the existing block:
2. Number of characters in proposal: *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 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 — *This finalises N4734 L2/16-198*
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:

	08A	08B	08C	08D	08E	08F
0	پ 08A0	گ 08B0			ؙ 08E0	ؠ 08F0
1	ﭑ 08A1	ﭐ 08B1			ؚ 08E1	ﭡ 08F1
2	ﺥ 08A2	ﺭ 08B2		⣀ 08D2	◯ 08E2	ﭢ 08F2
3	ﻅ 08A3	ﻉ 08B3		◌ 08D3	◌ 08E3	ﭣ 08F3
4	ﻑ 08A4	ﻙ 08B4		الربع 08D4	◌ 08E4	◌ 08F4
5	ﻕ 08A5			ص◌ 08D5	◌ 08E5	◌ 08F5
6	ﻝ 08A6	ﺏ 08B6		ع◌ 08D6	◌ 08E6	◌ 08F6
7	ﺙ 08A7	ﭖ 08B7		ق◌ 08D7	◌ 08E7	◌ 08F7
8	ﺉ 08A8	ﺕ 08B8		ن◌ 08D8	◌ 08E8	◌ 08F8
9	ﻱ 08A9	ﻥ 08B9		◌ 08D9	◌ 08E9	◌ 08F9
A	ﻁ 08AA	ﻱ 08BA		الثلثة◌ 08DA	◌ 08EA	◌ 08FA
B	ﻭ 08AB	ﻑ 08BB		السجدة◌ 08DB	◌ 08EB	◌ 08FB
C	ﻊ 08AC	ﻭ 08BC		النصف◌ 08DC	◌ 08EC	◌ 08FC
D	ﺍ 08AD	ﻥ 08BD		سكتة◌ 08DD	◌ 08ED	◌ 08FD
E	ﺩ 08AE			قف◌ 08DE	◌ 08EE	◌ 08FE
F	ﺹ 08AF			وقفة◌ 08DF	◌ 08EF	◌ 08FF

Arabic Extended A

08D2   ٜ̈   ARABIC END OF TEXT

	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	1E263
4		1E214	1E224	1E234	1E244	1E254	1E264
5	1E205	1E215	1E225	1E235	1E245	1E255	1E265
6	1E206	1E216	1E226	1E236	1E246	1E256	1E266
7	1E207	1E217	1E227	1E237	1E247	1E257	1E267
8	1E208	1E218	1E228		1E248	1E258	1E268
9	1E209	1E219	1E229		1E249	1E259	1E269
A	1E20A	1E21A	1E22A	1E23A	1E24A	1E25A	1E26A
B		1E21B	1E22B	1E23B	1E24B	1E25B	1E26B
C	1E20C	1E21C	1E22C	1E23C	1E24C	1E25C	1E26C
D	1E20D	1E21D		1E23D	1E24D	1E25D	
E	1E20E	1E21E	1E22E	1E23E		1E25E	
F	1E20F	1E21F	1E22F	1E23F		1E25F	

Western Cham

Independent Vowels

1E200	ꯀ	WESTERN CHAM LETTER A
1E201	ꯁ	WESTERN CHAM LETTER E
1E202	ꯂ	WESTERN CHAM LETTER O
1E205	ꯃ	WESTERN CHAM LETTER OS

Consonants

1E206		WESTERN CHAM LETTER KA
1E207		WESTERN CHAM LETTER KHA
1E208		WESTERN CHAM LETTER GA
1E209		WESTERN CHAM LETTER KHAA
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 PHAA
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 THAA
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 SIGN VOWEL OE OR 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
1E237		WESTERN CHAM CONSONANT SIGN TKAJ KA
1E23A		WESTERN CHAM CONSONANT SIGN BRAY
1E23B		WESTERN CHAM LETTER NAN
1E23C		WESTERN CHAM CONSONANT SIGN SHALOM

Final Consonants

1E23D		WESTERN CHAM LETTER FINAL PH
1E23E		WESTERN CHAM LETTER FINAL V
1E23F		WESTERN CHAM LETTER FINAL S
1E240		WESTERN CHAM LETTER FINAL K
1E242		WESTERN CHAM LETTER FINAL NG
1E244		WESTERN CHAM LETTER FINAL CH
1E245		WESTERN CHAM LETTER FINAL T
1E246		WESTERN CHAM LETTER FINAL N
1E247		WESTERN CHAM LETTER FINAL P
1E248		WESTERN CHAM LETTER FINAL Y
1E249		WESTERN CHAM LETTER FINAL R
1E24A		WESTERN CHAM LETTER FINAL L
1E24B		WESTERN CHAM LETTER FINAL SS
1E24C		WESTERN CHAM CONSONANT SIGN FINAL M
1E24D		WESTERN CHAM CONSONANT SIGN FINAL H

Digits

1E250	O	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
1E25D		WESTERN CHAM PUNCTUATION DANDA
1E25E		WESTERN CHAM PUNCTUATION DOUBLE DANDA
1E25F		WESTERN CHAM PUNCTUATION TRIPLE DANDA
1E260		WESTERN CHAM SIGN TANA TAMAT PHON
1E261		WESTERN CHAM SIGN TANA CHAPEDAEM PHON
1E262		WESTERN CHAM SIGN TANA CHAPEDAEM IU
1E263		WESTERN CHAM SIGN TANA TAMAT AYET
1E264		WESTERN CHAM SIGN TANA TAMAT TAKUE

Lunar Symbols

1E265		WESTERN CHAM SIGN LUNAR 1 DOTS
1E266		WESTERN CHAM SIGN LUNAR 2 DOTS
1E267		WESTERN CHAM SIGN LUNAR 3 DOTS
1E268		WESTERN CHAM SIGN LUNAR 4 DOTS
1E269		WESTERN CHAM SIGN LUNAR 5 DOTS
1E26A		WESTERN CHAM SIGN LUNAR 5 DOTS OVER HEH
1E26B		WESTERN CHAM SIGN LUNAR 6 DOTS
1E26C		WESTERN CHAM SIGN LUNAR 7 DOTS