

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

Doc Type: Working Group Document

Title: Proposal for encoding the Chakma script in the UCS

Source: UC Berkeley Script Encoding Initiative (Universal Scripts Project)

Authors: Michael Everson and Martin Hosken

Status: Liaison Contribution

Action: For consideration by JTC1/SC2/WG2 and UTC

Date: 2009-07-28

Replaces: N3428

1. Introduction. The Chakma people were in origin Tibeto-Burman, related to the Burmese. The language which they now speak is Indo-European, part of the Southeastern Bengali branch of Eastern Indo-Aryan. Its better-known closest relatives are Bengali, Assamese, Chittagonian, Bishnupriya, and Sylheti. It is spoken by 312,000 people in southeast Bangladesh near Chittagong City, and another 176,000 in India in Mizoram, Assam, Tripura, and Arunachal Pradesh. Literacy in Chakma script is low. The script itself is also called ᱡᱷᱟ ᱛᱟᱞᱟ *Ajhā pāṭh*, sometimes romanized *Ojhopath*.

There is a certain amount of glyph variation between the script as used in India and Bangladesh. Some fonts are rounder, similar to the style used in Myanmar; compare a similar variation in the Tai Tham script as used in Khün Tai. The glyphs used in this proposal are based on the Chadigang font, with some alterations toward more “generic” shapes for some characters.

The Chakma script is currently being adapted for use in Tanchangya, a language which is closely related to Chakma. An effort to develop the orthography is currently underway, and it appears that there may be additional letters, vowel signs, and tone marks added to cover this script. These extensions are a subject for future standardization, as the orthography for Tanchangya is still under development and testing.

2. Structure. Chakma is of the Brahmic type: the consonant letters contain an inherent vowel. Consonant clusters are written with conjunct characters, and a visible vowel killer shows the deletion of the inherent vowel when there is no conjunct.

3. Independent vowels. Four independent vowels exist: ᱠ *a*, ᱡ *i*, ᱣ *u*, and ᱤ *e*. Other vowels in initial position are formed by adding the vowel sign to ᱠ *a*, as in ᱠᱢ *ī*, ᱠᱣ *ū*, ᱠᱤ *ai*, ᱠᱥ *oi*. Some modern writers are generalizing this spelling in ᱠᱢ *i*, ᱠᱣ *u*, and ᱠᱤ *e*.

4. Dependent vowels. Independent vowel signs have been encoded according to their phonetic value, as in Balinese or Telugu. However, because some of the independent vowels appear to be made out of smaller units, decomposition of the vowels has been taken into consideration. In order to avoid any ambiguous encoding, canonical equivalences have been given in order to resolve the issue of multiple representations. Two glyph fragments (CHAKMA O MARK and CHAKMA AU MARK) have been encoded in order to account for this, and to allow for any users who desire to create texts using the lower glyph fragments on their own. The Chakma vowel signs are given with the letter ᱠ *ka* below.

ᳵ	kā	=	ᳵ	kā
ᳶ	ka	=	ᳵ	kā + ᳶ -a (11127)
᳷	ki	=	ᳵ	kā + ᳷ i (11128)
᳸	kī	=	ᳵ	kā + ᳸ -ī (11129)
᳹	ku	=	ᳵ	kā + ᳹ -u (1112A)
ᳺ	kū	=	ᳵ	kā + ᳺ -ū (1112B)
᳻	ke	=	ᳵ	kā + ᳻ -e (1112C)
᳼	kāi	=	ᳵ	kā + ᳼ -āi (1112D)
᳽	ko	=	ᳵ	kā + ᳽ -o (1112E – or – 11131 + 11127)
᳾	kau	=	ᳵ	kā + ᳾ -au (1112F – or – 11132 + 11127)
᳿	koi	=	ᳵ	kā + ᳿ -oi (11130)
᳠	kaṁ	=	ᳵ	kā + ᳠ -ṁ (11100)
᳡	kaṃ	=	ᳵ	kā + ᳡ -ṃ (11101)
᳢	kaḥ	=	ᳵ	kā + ᳢ -ḥ (11102)
᳣	k	=	ᳵ	kā + ᳣ MAAYYAA

One of the interesting features of Chakma writing is that CANDRABINDU (*cānaphupudā*) can be used together with ANUSVARA (*ekaphudā*) and VISARGA (*dviphudā*):

᳤ᳶ᳷	aḥṁ	=	᳤	ā + ᳶ ḥ + ᳷ ṁ
᳤ᳶ᳸	aṃṁ	=	᳤	ā + ᳶ ṃ + ᳸ ṁ
᳤᳷᳸	uṃṁ	=	᳤	u + ᳷ ṃ + ᳸ ṁ
᳤ᳶ᳹	muṁ	=	᳤	mā + ᳹ u + ᳶ ṁ

Glyph rendering of CANDRABINDU with other vowel signs (like ᳶ *a* or ᳷ *i*) is as yet unattested; nevertheless the encoding is clear.

5. Consonants with killed vowels and conjunct consonants. Like other Brahmic scripts, Chakma makes use of the MAAYYAA (*killer*) to invoke conjoined consonants. In the past, practice was much more common than it is today. Like the Myanmar script, Chakma is encoded with two vowel-killing characters in order to conform to modern user expectations. As shown above, most letters have their vowels killed with the use of the explicit MAAYYAA character:

᳣	k	=	ᳵ	kā + ᳣ MAAYYAA
---	---	---	---	----------------

In 2001 an orthographic reform was recommended in the book *Cānmā pattham pāt* which would limit the standard repertoire of conjuncts to those composed with the five letters ᳶ *yā*, ᳷ *rā*, ᳸ *lā*, ᳹ *wā*, and ᳺ *nā*. The four here are the most widely-accepted repertoire of conjuncts.

ya: X + ◻ VIRAMA + ဝ yā:

က ရ ဝ ယ င - ပ ခ ဓ ဟ ဇ - ဇ ဇ ဉ ဟ ဟ
က ဝ ဉ ဝ ဇ - ပ ဓ ဟ ဟ ဝ - ဝ ယ ဟ ဟ ဝ ယ

ra: X + ◻ VIRAMA + ဝ rā:

က ဝ ဝ ယ င - ပ ခ ဓ ဟ ဇ - ဇ ဇ ဉ ဟ ဟ
က ဝ ဉ ဝ ဇ - ပ ဓ ဟ ဟ ဝ - ဝ ယ ဟ ဟ ဝ ယ

la: X + ◻ VIRAMA + ဝ lā:

က လ ဝ ယ င - ပ ခ ဓ ဟ ဇ - ဇ ဇ ဉ ဟ ဟ
က လ ဝ ဉ ဝ ဇ - ပ ဓ ဟ ဟ ဝ - ဝ ယ ဟ ဟ ဝ ယ

wa: X + ◻ VIRAMA + ဝ wā:

က ဝ ဝ ယ င - ပ ခ ဓ ဟ ဇ - ဇ ဇ ဉ ဟ ဟ
က ဝ ဉ ဝ ဇ - ပ ဓ ဟ ဟ ဝ - ဝ ယ ဟ ဟ ဝ ယ

No separate conjunct forms of subjoined full-form *-yā* or *-rā* appear to exist. The fifth of these conjuncts, the *-na* conjunct, is exemplary of the orthographic shift which has taken place in Chakma.

na: X + ◻ VIRAMA + ဝ nā:

က ဝ ဝ ယ င - ပ ခ ဓ ဟ ဇ - ဇ ဇ ဉ ဟ ဟ
က ဝ ဉ ဝ ဇ - ပ ဓ ဟ ဟ ဝ - ဝ ယ ဟ ဟ ဝ ယ

While some writers would indeed write *kakna* (in ligating style) as ကက or (in subjoining style) as ကက, most now would probably expect it to be written as ကက. The ligating style of glyphs is now considered old-fashioned. Thus, taking the letter ဝ *mā* as the second element, while the glyph shapes ကက *kmā*, ကက *tmā*, ကက *nmā*, ကက *bbā*, ကက *mmā*, ကက *llā*, ကက *smā*, and ကက *hmā* are attested, most users now prefer the glyph shapes ကက *kmā*, ကက *tmā*, ကက *nmā*, ကက *bbā*, ကက *mmā*, ကက *llā*, ကက *smā*, and ကက *hmā*. Again, this distinction is stylistic and not orthographic.

As with Myanmar and Meetei Mayek, encoding a visible killer for modern users alongside an explicit conjoin-former permits the user to make specific choices about spelling more easily. Both the Myanmar encoding model and the Devanagari encoding model have been explained to the user community and feedback is that the Myanmar model fits the script better. (This is little surprise considering the close relationship between the Myanmar and Chakma scripts.)

In principle, nothing prevents the visible killer from appearing together with a stack. Thus while က *kā* + ◻ VIRAMA + ဝ *sā* gives ကက *ksā*, the sequence က *kā* + ◻ VIRAMA + ဝ *sā* + ◻ MAAYYAA could give ကက *ks*. Neither the string ◻ VIRAMA + ◻ MAAYYAA nor the string ◻ MAAYYAA + ◻ VIRAMA following a consonant would be meaningful, however, as both kill the inherent vowel, and “double-killing” a vowel makes no sense; either sequence should generate an error presentation.

The 2004 book *Phadagan* shows examples of the five conjuncts above together alongside conjuncts formed with က *bā*, က *mā*, and က *hā*. These are all formed by simple subjoining.

ba: X + VIRAMA + ூ bā:

ூ ௃ ௄ ௅ ெ - ே ை ௉ ொ ோ - ௌ ் ௎ ௐ ௑
௒ ௓ ௔ ௕ ௖ - ே ை ௉ ொ ோ ௌ - ் ௎ ௐ ௑ ௒ ௓

ma: X + VIRAMA + ் mā:

ூ ௃ ௄ ௅ ெ - ே ை ௉ ொ ோ - ௌ ் ௎ ௐ ௑
௒ ௓ ௔ ௕ ௖ - ே ை ௉ ொ ோ ௌ - ் ௎ ௐ ௑ ௒ ௓

ha: X + VIRAMA + ே hā:

ூ ௃ ௄ ௅ ெ - ே ை ௉ ொ ோ - ௌ ் ௎ ௐ ௑
௒ ௓ ௔ ௕ ௖ - ே ை ௉ ொ ோ ௌ - ் ௎ ௐ ௑ ௒ ௓

In the 1982 book *Cāṅmār āg pudhi* a much wider range of conjunct pairs is shown, some of them with fairly complicated glyphs.

- ூூ kkā = ூ kā + VIRAMA + ூ kā
- ூ௃ kṭā = ூ kā + VIRAMA + ூ௃ ṭā
- ூ௄ ktā = ூ kā + VIRAMA + ூ௄ tā
- ூ௅ kmā = ூ kā + VIRAMA + ூ௅ mā
- ூெ kcā = ூ kā + VIRAMA + ூெ cā (*conjunct shows old-style glyph*)
- ூே ṅkā = ூே ṅā + VIRAMA + ூ kā
- ூை ṅgā = ூை ṅā + VIRAMA + ூை gā
- ூ௉ ccā = ூ௉ cā + VIRAMA + ூ௉ cā (*conjunct shows old-style glyph*)
- ூொ cchā = ூொ cā + VIRAMA + ூொ chā (*conjunct shows old-style glyph*)
- ூோ ṅcā = ூோ ṅā + VIRAMA + ூ௉ cā (*conjunct shows old-style glyph*)
- ூௌ ṅjā = ூௌ ṅā + VIRAMA + ூௌ jā
- ூ் ṅjhā = ூ் ṅā + VIRAMA + ூ் jhā
- ூ௎ ṭṭā = ூ௎ ṭā + VIRAMA + ூ௎ ṭā
- ூௐ ttā = ூௐ tā + VIRAMA + ூௐ tā
- ூ௑ tmā = ூ௑ tā + VIRAMA + ூ௑ mā
- ூ௒ tthā = ூ௒ tā + VIRAMA + ூ௒ thā
- ூ௓ ddā = ூ௓ dā + VIRAMA + ூ௓ dā
- ூ௔ ddhā = ூ௔ dā + VIRAMA + ூ௔ dhā

𑄎𑄏	ntā	=	𑄎	nā	+	𑄏	VIRAMA	+	𑄏	tā
𑄎𑄏𑄏	nthā	=	𑄎	nā	+	𑄏	VIRAMA	+	𑄏𑄏	thā
𑄎𑄏𑄏	nmā	=	𑄎	nā	+	𑄏	VIRAMA	+	𑄏𑄏	mā
𑄏𑄏	ppā	=	𑄏	pā	+	𑄏	VIRAMA	+	𑄏	pā
𑄏𑄏	bbā	=	𑄏	bā	+	𑄏	VIRAMA	+	𑄏	bā
𑄏𑄏	mmā	=	𑄏	mā	+	𑄏	VIRAMA	+	𑄏	mā
𑄏𑄏	jjā	=	𑄏	jā	+	𑄏	VIRAMA	+	𑄏	jā
𑄏𑄏	lkā	=	𑄏	lā	+	𑄏	VIRAMA	+	𑄏	kā
𑄏𑄏	lgā	=	𑄏	lā	+	𑄏	VIRAMA	+	𑄏	gā
𑄏𑄏	llā	=	𑄏	lā	+	𑄏	VIRAMA	+	𑄏	lā
𑄏𑄏	ltā	=	𑄏	lā	+	𑄏	VIRAMA	+	𑄏	tā
𑄏𑄏	lpā	=	𑄏	lā	+	𑄏	VIRAMA	+	𑄏	pā
𑄏𑄏	schā	=	𑄏	sā	+	𑄏	VIRAMA	+	𑄏	chā (<i>conjunct shows old-style glyph</i>)
𑄏𑄏	stā	=	𑄏	sā	+	𑄏	VIRAMA	+	𑄏	tā
𑄏𑄏	skā	=	𑄏	sā	+	𑄏	VIRAMA	+	𑄏	kā
𑄏𑄏	spā	=	𑄏	sā	+	𑄏	VIRAMA	+	𑄏	pā
𑄏𑄏	smā	=	𑄏	sā	+	𑄏	VIRAMA	+	𑄏	mā
𑄏𑄏	hmā	=	𑄏	hā	+	𑄏	VIRAMA	+	𑄏	mā

5.1 Specific recommendation for Chakma fonts. In Chakma, the encoding model supports conjunct behaviour and Chakma fonts *by default* should display the subjoined form of letters when following virama, to ensure legibility. Whether a conjunct is required or not is part of the spelling of a word; it is not a stylistic issue. (We have seen no examples of conjuncts with more than one consonant, and while the encoding handles (in principle) any length of stacking examples of such would probably be spelling errors.)

6. Collating order. As an Indo-European language, the standard Brahmic sorting order applies to Chakma.

7. Character names. Consonant letter names use the typical Brahmic transliteration used in the UCS. Chakma letters have a descriptive name followed by a traditional Brahmic consonant. These latter are given in annotations to the character names.

8. Punctuation and digits. Alongside a | DANDA and || DOUBLE DANDA punctuation, Chakma has a unique 𑄏 QUESTION MARK, and a 𑄏 SECTION MARK. There is some variation in the glyphs for the SECTION MARK, some looking like flowers or leaves. A set of digits exists and is encoded, although Bengali digits are also used. The Tanchangya use Myanmar digits.

9. Linebreaking. Letters and digits behave as in Bengali. Both CHAKMA DANDA and CHAKMA DOUBLE DANDA behave as in Devanagari. The CHAKMA QUESTION MARK behaves like U+003F QUESTION MARK. The CHAKMA SECTION MARK behaves like U+2055 FLOWER PUNCTUATION MARK.

10. Unicode Character Properties.

11100;CHAKMA SIGN CANDRABINDU;Mn;230;NSM;;;;N;;;;;
11101;CHAKMA SIGN ANUSVARA;Mn;230;NSM;;;;N;;;;;
11102;CHAKMA SIGN VISARGA;Mc;230;L;;;;N;;;;;
11103;CHAKMA LETTER AA;Lo;0;L;;;;N;;;;;
11104;CHAKMA LETTER I;Lo;0;L;;;;N;;;;;
11105;CHAKMA LETTER U;Lo;0;L;;;;N;;;;;
11106;CHAKMA LETTER E;Lo;0;L;;;;N;;;;;
11107;CHAKMA LETTER KAA;Lo;0;L;;;;N;;;;;
11108;CHAKMA LETTER KHAA;Lo;0;L;;;;N;;;;;
11109;CHAKMA LETTER GAA;Lo;0;L;;;;N;;;;;
1110A;CHAKMA LETTER GHAA;Lo;0;L;;;;N;;;;;
1110B;CHAKMA LETTER NGAA;Lo;0;L;;;;N;;;;;
1110C;CHAKMA LETTER CAA;Lo;0;L;;;;N;;;;;
1110D;CHAKMA LETTER CHAA;Lo;0;L;;;;N;;;;;
1110E;CHAKMA LETTER JAA;Lo;0;L;;;;N;;;;;
1110F;CHAKMA LETTER JHAA;Lo;0;L;;;;N;;;;;
11110;CHAKMA LETTER NYAA;Lo;0;L;;;;N;;;;;
11111;CHAKMA LETTER TTAA;Lo;0;L;;;;N;;;;;
11112;CHAKMA LETTER TTHAA;Lo;0;L;;;;N;;;;;
11113;CHAKMA LETTER DDAA;Lo;0;L;;;;N;;;;;
11114;CHAKMA LETTER DDHAA;Lo;0;L;;;;N;;;;;
11115;CHAKMA LETTER NNAA;Lo;0;L;;;;N;;;;;
11116;CHAKMA LETTER TAA;Lo;0;L;;;;N;;;;;
11117;CHAKMA LETTER THAA;Lo;0;L;;;;N;;;;;
11118;CHAKMA LETTER DAA;Lo;0;L;;;;N;;;;;
11119;CHAKMA LETTER DHAA;Lo;0;L;;;;N;;;;;
1111A;CHAKMA LETTER NAA;Lo;0;L;;;;N;;;;;
1111B;CHAKMA LETTER PAA;Lo;0;L;;;;N;;;;;
1111C;CHAKMA LETTER PHAA;Lo;0;L;;;;N;;;;;
1111D;CHAKMA LETTER BAA;Lo;0;L;;;;N;;;;;
1111E;CHAKMA LETTER BHAA;Lo;0;L;;;;N;;;;;
1111F;CHAKMA LETTER MAA;Lo;0;L;;;;N;;;;;
11120;CHAKMA LETTER YYAA;Lo;0;L;;;;N;;;;;
11121;CHAKMA LETTER YAA;Lo;0;L;;;;N;;;;;
11122;CHAKMA LETTER RAA;Lo;0;L;;;;N;;;;;
11123;CHAKMA LETTER LAA;Lo;0;L;;;;N;;;;;
11124;CHAKMA LETTER WAA;Lo;0;L;;;;N;;;;;
11125;CHAKMA LETTER SAA;Lo;0;L;;;;N;;;;;
11126;CHAKMA LETTER HAA;Lo;0;L;;;;N;;;;;
11127;CHAKMA VOWEL SIGN A;Mn;230;NSM;;;;N;;;;;
11128;CHAKMA VOWEL SIGN I;Mn;230;NSM;;;;N;;;;;
11129;CHAKMA VOWEL SIGN II;Mn;230;NSM;;;;N;;;;;
1112A;CHAKMA VOWEL SIGN U;Mn;220;NSM;;;;N;;;;;
1112B;CHAKMA VOWEL SIGN UU;Mn;220;NSM;;;;N;;;;;
1112C;CHAKMA VOWEL SIGN E;Mc;224;L;;;;N;;;;;
1112D;CHAKMA VOWEL SIGN AI;Mn;230;NSM;;;;N;;;;;
1112E;CHAKMA VOWEL SIGN O;Mn;0;NSM;11131 11127;;;;N;;;;;
1112F;CHAKMA VOWEL SIGN AU;Mn;0;NSM;11132 11127;;;;N;;;;;
11130;CHAKMA VOWEL SIGN OI;Mn;230;NSM;;;;N;;;;;
11131;CHAKMA O MARK;Mn;220;NSM;;;;N;;;;;
11132;CHAKMA AU MARK;Mn;220;NSM;;;;N;;;;;
11133;CHAKMA VIRAMA;Mn;9;NSM;;;;N;;;;;
11134;CHAKMA MAAYYAA;Mn;9;NSM;;;;N;;;;;
11136;CHAKMA DIGIT ZERO;Nd;0;L;0;0;0;N;;;;;
11137;CHAKMA DIGIT ONE;Nd;0;L;1;1;1;N;;;;;
11138;CHAKMA DIGIT TWO;Nd;0;L;2;2;2;N;;;;;
11139;CHAKMA DIGIT THREE;Nd;0;L;3;3;3;N;;;;;
1113A;CHAKMA DIGIT FOUR;Nd;0;L;4;4;4;N;;;;;
1113B;CHAKMA DIGIT FIVE;Nd;0;L;5;5;5;N;;;;;
1113C;CHAKMA DIGIT SIX;Nd;0;L;6;6;6;N;;;;;
1113D;CHAKMA DIGIT SEVEN;Nd;0;L;7;7;7;N;;;;;
1113E;CHAKMA DIGIT EIGHT;Nd;0;L;8;8;8;N;;;;;
1113F;CHAKMA DIGIT NINE;Nd;0;L;9;9;9;N;;;;;
11140;CHAKMA SECTION MARK;Po;0;L;;;;N;;;;;
11141;CHAKMA DANDA;Po;0;L;;;;N;;;;;
11142;CHAKMA DOUBLE DANDA;Po;0;L;;;;N;;;;;
11143;CHAKMA QUESTION MARK;Po;0;L;;;;N;;;;;

11. Bibliography.

Cānmā, Cirajyoti and Maṅgal Cāngmā. 1982. *Cānmār āg pudhi (Chakma primer)*. Rānamāṭi: Cānmābhāṣā Prakāśanā Pariṣad.

Khisa, Bhagadatta. 2001. *Cāṅmā pattham pāt = Chakma primer*. Rāṅmāṭi: Tribal Cultural Institute (TCI).
Singā. 2004. *Phagadāṅ*.

12. Acknowledgements. This project was made possible in part by a grant from the U.S. National Endowment for the Humanities, which funded the Universal Scripts Project (part of the Script Encoding Initiative at UC Berkeley) in respect of the Chakma encoding. Any views, findings, conclusions or recommendations expressed in this publication do not necessarily reflect those of the National Endowment of the Humanities.