UC Berkeley
Proposals from the Script Encoding Initiative

Title

Proposal for encoding the Chakma script in the UCS

Permalink

https://escholarship.org/uc/item/55b6007h

Authors

Everson, Michael
Hosken, Martin

Publication Date

2009-07-28

Peer reviewed

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 𑄃𑄎𑄳𑄣 𑄛𑄖𑄴 *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: 𑄃 *a*, 𑄄 *i*, 𑄅 *u*, and 𑄆 *e*. Other vowels in initial position are formed by adding the vowel sign to 𑄃 *a*, as in 𑄃𑄩 *ī*, 𑄃𑄫 *ū*, 𑄃𑄱 *ai*, 𑄃𑄭 *oi*. Some modern writers are generalizing this spelling in 𑄃𑄨 *i*, 𑄃𑄪 *u*, and 𑄄 *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 𑄇 *ka* below.

ꠇ	kā	= ꠇ kā	
ꠇ	ka	= ꠇ kā + ◌́ -a	(11127)
ꠇ	ki	= ꠇ kā + ◌̊ i	(11128)
ꠇ	kī	= ꠇ kā + ◌̊ -ī	(11129)
ꠇ	ku	= ꠇ kā + ◌̞ -u	(1112A)
ꠇ	kū	= ꠇ kā + ◌̞ -ū	(1112B)
ꠄꠇ	ke	= ꠇ kā + ꠆◌ -e	(1112C)
ꠇ	kāi	= ꠇ kā + ◌� -āi	(1112D)
ꠇ	ko	= ꠇ kā + ◌́ -o	(1112E – *or* – 11131 + 11127)
ꠇ	kau	= ꠇ kā + ◌́ -au	(1112F – *or* – 11132 + 11127)
ꠇ	koi	= ꠇ kā + ◌̷̷ -oi	(11130)
ꠇ̆	kaṁ	= ꠇ kā + ◌̆ -ṁ	(11100)
ꠇ̇	kaṃ	= ꠇ kā + ◌̇ -ṃ	(11101)
ꠇ̈	kaḥ	= ꠇ kā + ◌̈ -ḥ	(11102)
ꠇ̄	k	= ꠇ kā + ◌̄ MAAYYAA	

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

ꠀ̈̆	aḥṁ	= ꠀ ā + ◌̈ ḥ + ◌̆ ṁ	
ꠀ̇̆	aṃṁ	= ꠀ ā + ◌̇ ṃ + ◌̆ ṁ	
ꠃ̇̆	uṃṁ	= ꠃ u + ◌̇ ṃ + ◌̆ ṁ	
ꠝ̆	muṁ	= ꠝ mā + ◌̞ u + ◌̆ ṁ	

Glyph rendering of CANDRABINDU with other vowel signs (like ◌́ *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:

ꠇ̄	k	= ꠇ kā + ◌̄ MAAYYAA	

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

ya: X + ⬚ VIRAMA + ယ *yā*:

ကျ ချ ဂျ ဃျ ၚ - စျ ဆျ ၛ ၜ ၝ - ၐ ၑ ၒ ၓ ၔ

 တျ ထျ ဒ ဓ နျ - ပျ ဖျ ဗျ ဘျ မျ - ယျ ရျ လျ ဝျ သျ

ra: X + ⬚ VIRAMA + ရ *rā*:

ကြ ခြ ဂြ ဃြ ၚ - စြ ဆြ ၛ ၜ ၝ - ၐ ၑ ၒ ၓ ၔ

တြ ထြ ၃ ၄ နြ - ပြ ဖြ ဗြ ဘြ မြ - ယြ ရြ လြ ဝြ သြ

la: X + ⬚ VIRAMA + လ *lā*:

ကၠ ခၠ ဂၠ ဃၠ ၚ - စၠ ဆၠ ၛ ၜ ၝ - ၐ ၑ ၒ ၓ ၔ

တၠ ထၠ ၃ ၄ နၠ - ပၠ ဖၠ ဗၠ ဘၠ မၠ - ယၠ ရၠ လၠ ဝၠ သၠ

wa: X + ⬚ VIRAMA + ဝ *wā*:

ကွ ခွ ဂွ ဃွ ၚ - စွ ဆွ ၛ ၜ ၝ - ၐ ၑ ၒ ၓ ၔ

တွ ထွ ၃ ၄ နွ - ပွ ဖွ ဗွ ဘွ မွ - ယွ ရွ လွ ဝွ သွ

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 + န *nā*:

ကန ခန ဂန ဃန ၚန - စန ဆန ၛန ၜန ၝ - ၐ ၑ ၒ ၓ ၔ

တ ထ ၃ ၄ နန - ပန ဖန ဗန ဘန - ယန ရန လ ဝ သ

While some writers would indeed write *kakna* (in ligating style) as ကကန or (in subjoining style) as ကက, most now would probably expect it to be written as ကၵန. The ligating style of glyphs is now considered old-fashioned. Thus, taking the letter ဝ *mā* as the second element, while the glyph shapes ကဝ *kmā*, တဝ *tmā*, နဝ *nmā*, ဗဝ *bbā*, မဝ *mmā*, လ *llā*, သဝ *smā*, and ဟဝ *hmā* are attested, most users now prefer the glyph shapes ကွ *kmā*, တွ *tmā*, နွ *nmā*, ဗွ *bbā*, မွ *mmā*, လွ *llā*, သွ *smā*, and ဟွ *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 + သ *sā* gives ကွ *ksā*, the sequence က *kā* + ⬚ VIRAMA + သ *sā* + ◌̄ MAAYYAA could give ကွ̄ *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 *Phadagaṅ* shows examples of the five conjuncts above together alongside conjuncts formed with ဗ *bā*, ဝ *mā*, and ဟ *hā*. These are all formed by simple subjoining.

ba: X + ⬚ VIRAMA + ♡ bā:

[script glyphs row 1]

[script glyphs row 2]

ma: X + ⬚ VIRAMA + ⍵ mā:

[script glyphs row 1]

[script glyphs row 2]

ha: X + ⬚ VIRAMA + ∽ hā:

[script glyphs row 1]

[script glyphs row 2]

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

[glyph]	kkā	=	[glyph] kā + ⬚ VIRAMA + [glyph] kā
[glyph]	kṭā	=	[glyph] kā + ⬚ VIRAMA + [glyph] ṭā
[glyph]	ktā	=	[glyph] kā + ⬚ VIRAMA + [glyph] tā
[glyph]	kmā	=	[glyph] kā + ⬚ VIRAMA + [glyph] mā
[glyph]	kcā	=	[glyph] kā + ⬚ VIRAMA + [glyph] cā (*conjunct shows old-style glyph*)
[glyph]	ṅkā	=	[glyph] ṅā + ⬚ VIRAMA + [glyph] kā
[glyph]	ṅgā	=	[glyph] ṅā + ⬚ VIRAMA + [glyph] gā
[glyph]	ccā	=	[glyph] cā + ⬚ VIRAMA + [glyph] cā (*conjunct shows old-style glyph*)
[glyph]	cchā	=	[glyph] cā + ⬚ VIRAMA + [glyph] chā (*conjunct shows old-style glyph*)
[glyph]	ñcā	=	[glyph] ñā + ⬚ VIRAMA + [glyph] cā (*conjunct shows old-style glyph*)
[glyph]	ñjā	=	[glyph] ñā + ⬚ VIRAMA + [glyph] jā
[glyph]	ñjhā	=	[glyph] ñā + ⬚ VIRAMA + [glyph] jhā
[glyph]	ṭṭā	=	[glyph] ṭā + ⬚ VIRAMA + [glyph] ṭā
[glyph]	ttā	=	[glyph] tā + ⬚ VIRAMA + [glyph] tā
[glyph]	tmā	=	[glyph] tā + ⬚ VIRAMA + [glyph] mā
[glyph]	tthā	=	[glyph] tā + ⬚ VIRAMA + [glyph] thā
[glyph]	ddā	=	[glyph] dā + ⬚ VIRAMA + [glyph] dā
[glyph]	ddhā	=	[glyph] dā + ⬚ VIRAMA + [glyph] dhā

৩৬৩	ntā	=	৩৬ nā	+	⸬ VIRAMA	+	৩ tā
৩৬৩৩	nthā	=	৩৬ nā	+	⸬ VIRAMA	+	৩৩ thā
৩৬৩	nmā	=	৩৬ nā	+	⸬ VIRAMA	+	৩ mā
৪	ppā	=	৪ pā	+	⸬ VIRAMA	+	৪ pā
৪৩	bbā	=	৪ bā	+	⸬ VIRAMA	+	৪ bā
৩৩	mmā	=	৩ mā	+	⸬ VIRAMA	+	৩ mā
৩৩	jjā	=	৩ jā	+	⸬ VIRAMA	+	৩ jā
৩৩	lkā	=	৩ lā	+	⸬ VIRAMA	+	৩ kā
৩৩	lgā	=	৩ lā	+	⸬ VIRAMA	+	৩ gā
৩৩	llā	=	৩ lā	+	⸬ VIRAMA	+	৩ lā
৩৩	lṭā	=	৩ lā	+	⸬ VIRAMA	+	৩ ṭā
৩৩	lpā	=	৩ lā	+	⸬ VIRAMA	+	৪ pā
৩৩	schā	=	৩ sā	+	⸬ VIRAMA	+	৪ chā *(conjunct shows old-style glyph)*
৩৩	sṭā	=	৩ sā	+	⸬ VIRAMA	+	৩ ṭā
৩৩	skā	=	৩ sā	+	⸬ VIRAMA	+	৩ kā
৩৩	spā	=	৩ sā	+	⸬ VIRAMA	+	৪ pā
৩৩	smā	=	৩ sā	+	⸬ VIRAMA	+	৩ mā
৩৩	hmā	=	৩ hā	+	⸬ VIRAMA	+	৩ 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 ꘓ QUESTION MARK, and a ⸰ 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.

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.
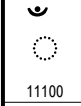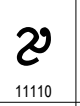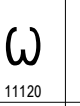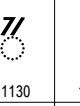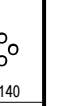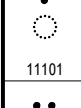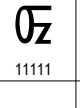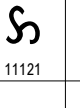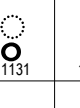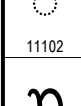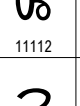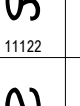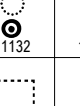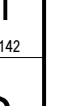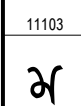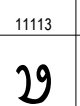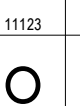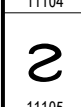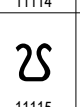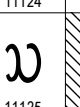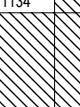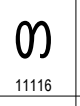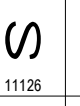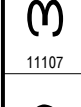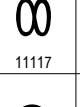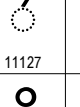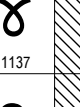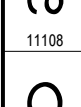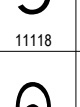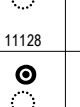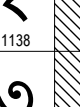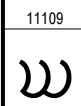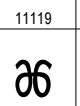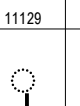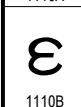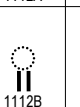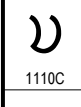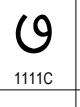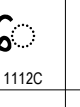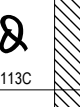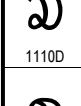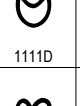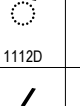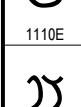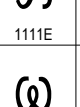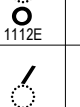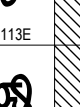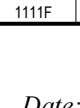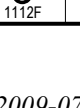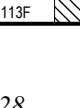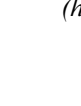
```
11100;CHAKMA SIGN CANDRABINDU;Mn;230;NSM;;;;;N;;;;;
11101;CHAKMA SIGN ANUSVARA;Mn;230;NSM;;;;;N;;;;;
11102;CHAKMA SIGN VISARGA;Mc;230;L;;;;;N;;;;;
11103;CHAKMA LETTER AA;Lo;0;L;;;;;N;;;;;
11104;CHAKMA LETTER I;Lo;0;L;;;;;N;;;;;
11105;CHAKMA LETTER U;Lo;0;L;;;;;N;;;;;
11106;CHAKMA LETTER E;Lo;0;L;;;;;N;;;;;
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.

Cāṅmā, Cirajyoti and Maṅgal Cāṅgmā. 1982. *Cāṅmār āg pudhi* (*Chakma primer*). Rāṅamāṭi: Cāṅmābhāṣā Prakāśanā Pariṣad.

Khisa, Bhagadatta. 2001. *Cāṅmā pattham pāt = Chakma primer*. Rāṅamāṭi: Tribal Cultural Institute (TCI).

Singā. 2004. *Phagadāṅ*.

	1110	1111	1112	1113	1114
0	◌̆ 11100	২ 11110	ധ 11120	◌̷̷ 11130	◦°◦ 11140
1	◌̇ 11101	৳ 11111	৯ 11121	◌̥ 11131	‖ 11141
2	◌̈ 11102	৳ 11112	৪ 11122	◌̥ 11132	‖ 11142
3	൩ 11103	੨ 11113	৶ 11123	⊡ 11133	⸱ 11143
4	൴ 11104	൬ 11114	০ 11124	◌̄ 11134	
5	ᦧ 11105	൮ 11115	൵ 11125		
6	৯ 11106	ൻ 11116	൰ 11126	০ 11136	
7	ൾ 11107	ൽ 11117	◌́ 11127	৬ 11137	
8	৫ 11108	ൗ 11118	◦̊ 11128	২ 11138	
9	↻ 11109	൨ 11119	◉̊ 11129	৬ 11139	
A	൝ 1110A	൫ 1111A	◌̇ 1112A	൳ 1113A	
B	ε 1110B	∪ 1111B	◌̈ 1112B	ஜ 1113B	
C	ധ 1110C	৩ 1111C	৬◦ 1112C	ষ 1113C	
D	ৰ 1110D	৩ 1111D	◌̇⁷ 1112D	੧ 1113D	
E	ঞ 1110E	൫ 1111E	◌̥́ 1112E	৮ 1113E	
F	৩ 1110F	ൽ 1111F	◌̥́ 1112F	ষ 1113F	

Various signs

11100 ◌̆ CHAKMA SIGN CANDRABINDU
= caanaphupudaa
11101 ◌̇ CHAKMA SIGN ANUSVARA
= ekaphudaa
11102 ◌̈ CHAKMA SIGN VISARGA
= dviphudaa

Independent vowels

11103 ꄃ CHAKMA LETTER AA
= pichapujhaa aa
11104 ꄄ CHAKMA LETTER I
= delabhaangagaa i
11105 ꄅ CHAKMA LETTER U
= bacacu u
11106 ꄆ CHAKMA LETTER E
= lejaubaa e

Consonants

11107 ꄇ CHAKMA LETTER KAA
= cucyaangyaa kaa
11108 ꄈ CHAKMA LETTER KHAA
= grajaangyaa khaa
11109 ꄉ CHAKMA LETTER GAA
= caandyaa gaa
1110A ꄊ CHAKMA LETTER GHAA
= tinaddaalyaa ghaa
1110B ꄋ CHAKMA LETTER NGAA
= cilaama ngaa
1110C ꄌ CHAKMA LETTER CAA
= dvibhalyaa caa
1110D ꄍ CHAKMA LETTER CHAA
= majaraa chaa
1110E ꄎ CHAKMA LETTER JAA
= dvipadalaa haa
1110F ꄏ CHAKMA LETTER JHAA
= uraauraa jhaa
11110 ꄐ CHAKMA LETTER NYAA
= silaacyaa nyaa
11111 ꄑ CHAKMA LETTER TTAA
= dviyaadaat ttaa
11112 ꄒ CHAKMA LETTER TTHAA
= phudaadviyaat tthaa
11113 ꄓ CHAKMA LETTER DDAA
= aadudaangaat ddaa
11114 ꄔ CHAKMA LETTER DDHAA
= lejabharaat ddhaa
11115 ꄕ CHAKMA LETTER NNAA
= pettttuyaa nnaa
11116 ꄖ CHAKMA LETTER TAA
= ghangadaat taa
11117 ꄗ CHAKMA LETTER THAA
= jagadaat thaa
11118 ꄘ CHAKMA LETTER DAA
= dolaniit daa
11119 ꄙ CHAKMA LETTER DHAA
= talamuyaat dhaa
1111A ꄚ CHAKMA LETTER NAA
= phaarabaanyaa naa
1111B ꄛ CHAKMA LETTER PAA
= paalyaa paa
1111C ꄜ CHAKMA LETTER PHAA
= ubaraphudaa phaa

11111D ꄝ CHAKMA LETTER BAA
= ubaramuyaa baa
1111E ꄞ CHAKMA LETTER BHAA
= ciraddaalyaa bhaa
1111F ꄟ CHAKMA LETTER MAA
= bugatpadalaa maa
11120 ꄠ CHAKMA LETTER YYAA
= cimayyaa yyaa
11121 ꄡ CHAKMA LETTER YAA
= jilyaa yaa
11122 ꄢ CHAKMA LETTER RAA
= dvidaayyaa raa
11123 ꄣ CHAKMA LETTER LAA
= talamuyaa laa
11124 ꄤ CHAKMA LETTER WAA
= bajhonyaa waa
11125 ꄥ CHAKMA LETTER SAA
= bhudibukyaa saa
11126 ꄦ CHAKMA LETTER HAA
= ubaramuyaa haa

Dependent vowel signs

11127 ◌̖ CHAKMA VOWEL SIGN A
= ubaratulyaa a
11128 ◌̊ CHAKMA VOWEL SIGN I
= bahryaa i
11129 ◌̊ CHAKMA VOWEL SIGN II
= baaniiphadaa ii
1112A ◌̩ CHAKMA VOWEL SIGN U
= ekattaana u
1112B ◌̤ CHAKMA VOWEL SIGN UU
= dvittaana uu
1112C ꄬ◌ CHAKMA VOWEL SIGN E
= ekaara e
1112D ◌̓ CHAKMA VOWEL SIGN AI
= delabhaanga ai
1112E ◌̥ CHAKMA VOWEL SIGN O
= okaara o
1112F ◌̥ CHAKMA VOWEL SIGN AU
= aukaara au
11130 ◌̌ CHAKMA VOWEL SIGN OI
= oikaara oi
11131 ◌̥ CHAKMA O MARK
11132 ◌̥ CHAKMA AU MARK

Various signs

11133 ◌ CHAKMA VIRAMA
• used to form conjuncts
→ 1039 ◌ myanmar sign virama
11134 ◌̄ CHAKMA MAAYYAA
• killer
→ 103A ◌ myanmar sign asat

Digits

11136 ꄶ CHAKMA DIGIT ZERO
11137 ꄷ CHAKMA DIGIT ONE
11138 ꄸ CHAKMA DIGIT TWO
11139 ꄹ CHAKMA DIGIT THREE
1113A ꄺ CHAKMA DIGIT FOUR
1113B ꄻ CHAKMA DIGIT FIVE
1113C ꄼ CHAKMA DIGIT SIX
1113D ꄽ CHAKMA DIGIT SEVEN
1113E ꄾ CHAKMA DIGIT EIGHT

1113F   ꄏ   CHAKMA DIGIT NINE

Punctuation

11140   ❖   CHAKMA SECTION MARK
      = phulacihna
11141   ∣   CHAKMA DANDA
      = ekacilyaa
11142   ∥   CHAKMA DOUBLE DANDA
      = dvicilyaa
11143   ꜿ   CHAKMA QUESTION MARK
      = pujhaar

DR. G. A. Grierson সংগৃহীত চাকমা বর্ণ(১৯০৩ খৃ:)এবং এগুলি সম্বর্ধে
তাঁর মতামত

The following account of the Chākmā alphabet is based on information provided
by Dewan Kristo Chandra, a gentleman of Chākmā nationality, and forwarded to me
by Mr. J. A. Cave-Browne. Assistant Commissioner, Chittagong Hill Tracts.

The Chākmā alphabet is as follows :—

ka	*khā*	*gā*	*ghā*	*ṅā*
cha (sā)	*chhā*	*jā*	*jhā*	*ñā*
ṭa	*ṭhā*	*ḍā*	*ḍhā*	*ṇā*
tā	*thā*	*dā*	*dhā*	*nī*
pā	*phā*	*bā*	*bhā*	*mā*
yī	*rā*	*lā*	*wā*	*shā*
hā	*hlā*	*ā.*		

Bengali.

323   BENGALI.

The most important point to notice in this alphabet is that the vowel inherent in
each consonant is, not *a* as in other Indian languages, but *ā*. Note also that ꠥ the
initial form (there is, of course, no non-initial form) of *ā* is treated as a consonant,
much as the letter *alif* is treated as a consonant in Arabic.

For purposes of comparison, I here give the usual Burmese forms of the conso-
nants :—

ka, *kha,* *ga,* *gha,* *ṅ,*
cha, *chha,* *ja,* *jha,* *ña,*
ṭa, *ṭha,* *ḍa,* *ḍha,* *na,*
ta, *tha,* *da.* *dha.* *na,*
pa, *pha,* *ba,* *bha,* *ma,*
ya, *ro,* *la,* *wa*
sa, *ha*

Figure 1. Chakma chart from Grierson's *Linguistic Survey of India*, 1903.

চাক্মা বর্ণমালা

হুচ্চ্যাঙ্যা - কা	ঙল্যাঙ্যা - খা	চান্দ্যা - গা	তিনধাল্যা - ঘা	চিলাম - ঙা
ধিধাল্যা - চা	মোঙ্ঝল্যা - হ্যা	দ্বিপদল্যা - জা	উরাউবি - ঝা	চিলোস্যা - এ্রা
ধিহদাত্ৰ - টা	ফুদাদিয়াত্ৰ - ঠা	হ্যাদুভাঙাত্ৰ - ডা	লেজভরাত - ঢা	পেণ্ডুয়া - ণা
ঘড়দাত্ৰ - তা	ঙগদাত্ৰ - থা	দোলনিয়ত - দা	তলমুয়াত্ৰ - ধা	ফারবান্যা - না
পাল্যা - পা	উবরপদল্যা - ফা	উবরমুয়া - বা	চেইরধাল্যা - ভা	বুগতপদল্যা - মা
ঝিল্যা - য়া	দ্বিদার্জ্যা - রা	তলমুয়া - লা	বায়োন্যা - ওয়া	উবরমুয়া - হ্যা
ভুদিবুকা - সা	চিমোল্যা - খা			

অ	আ	ই	উ	এ
ও	ঐ			

চিহ্নের ব্যবহার

বর্ণ		চিহ্ন			যেমন
ম	উবর তুল্যে	´		ক	যেমন
ম	মাজ্যা	¯		ক্	যেমন
ম	বান্যে	o		কি	যেমন
ম	একটান দিলে	ι		কূ	যেমন
ম	এ-কার দিলে	ে		কে	যেমন
ম	ও-কার দিলে	ৗ		কো	যেমন
ম	দেল ভাঙিলে	৺		কাই	যেমন
ম	য়া - দিলে	⌐		ক্যা	যেমন
ম	রা - দিলে	‿		ক্রা	যেমন
ম	লা - দিলে			ক্লা	যেমন
ম	ওয়া - দিলে	o		কোয়া	যেমন
ম	হা - দিলে			কাহ্	যেমন
ম	না - দিলে			ন্না	যেমন
ম	একফুদা দিলে	·		কাং	যেমন
ম	দ্বিফুদা দিলে	··		কাঃ	যেমন
ম	চান ফুদা দিলে	˘		কঁ	যেমন

৬৬

৬৭

Figure 2. Charts taken from a paper written by Mr Sugata Chakma of the Tribal Cultural Institute, on "the Primary classification of languages".

Figure 3. Example of poetry from the book *Phagadāṅ*, 2004.

Figure 4. Alphabet chart from Khisa 2001.

চিহ্নের ব্যবহার

বর্ণ		চিহ্ন				
(Chakma)	উবর তুল্যে	´	(sign)	ক	যেমন	(example)
(Chakma)	মাজ্যা	−	(sign)	ক্	যেমন	(example)
(Chakma)	বান্যে	০	(sign)	কি	যেমন	(example)
(Chakma)	একটান দিলে	৷	(sign)	কু	যেমন	(example)
(Chakma)	এ-কার দিলে	৬	(sign)	কে	যেমন	(example)
(Chakma)	ও-কার দিলে	৹	(sign)	কো	যেমন	(example)
(Chakma)	দেল ভাঙিলে	(sign)	(sign)	কাই	যেমন	(example)
(Chakma)	য়া - দিলে	(sign)	(sign)	ক্যা	যেমন	(example)
(Chakma)	রা - দিলে	(sign)	(sign)	ক্রা	যেমন	(example)
(Chakma)	লা - দিলে	(sign)	(sign)	ক্লা	যেমন	(example)
(Chakma)	ওয়া - দিলে	(sign)	(sign)	কোয়া	যেমন	(example)
(Chakma)	হা - দিলে	(sign)	(sign)	কাহ্	যেমন	(example)
(Chakma)	না - দিলে	(sign)	(sign)	ক্না	যেমন	(example)
(Chakma)	একফুদা দিলে	·	(sign)	কাং	যেমন	(example)
(Chakma)	দ্বিফুদা দিলে	··	(sign)	কাঃ	যেমন	(example)
(Chakma)	চান ফুদা দিলে	(sign)	(sign)	কাঁ	যেমন	(example)

১। চাকমা বর্ণমালায় কোন যুক্তাক্ষর থাকবে না। তবে য, র, ল, ব, হ ও ন ফলা থাকবে। যেমনঃ (Chakma examples) ইত্যাদি।

২। দীর্ঘ ই -কার, দীর্ঘ উ-কার থাকবে না। উ ( ) বর্ণের পরিবর্তে (Chakma) - উ, ই ( ) বর্ণের পরিবর্তে (Chakma) - ই এবং এ ( ) বর্ণের পরিবর্তে (Chakma) - এ লিখতে হবে। হ্ম ( ) বর্ণটির ব্যবহার হবে না।

৩। অংকের সংখ্যাগুলি বাংলা সংখ্যার ন্যায় লিখতে হবে- ১, ২, ৩, ৪, ৫, ৬, ৭, ৮, ৯, ১০

Figure 5. Chart of vowel signs and conjuncts from Khisa 2001.

Figure 6. Chart with old-style conjuncts from Cāṅmā 1982.

Figure 7. Chart with punctuation and digits from Cāṅmā 1982. Shown are: — EM DASH, , COMMA, ∣ CHAKMA DANDA, ∥ CHAKMA DOUBLE DANDA, ꜱ CHAKMA QUESTION MARK, = EQUALS SIGN; (LEFT PARENTHESIS,) RIGHT PARENTHESIS, ⟪ LEFT DOUBLE PARENTHESIS, ⟫ RIGHT DOUBLE PARENTHESIS, < LESS-THAN SIGN, > GREATER-THAN SIGN (unless these are single guillemets ‹ › or angle brackets ⟨ ⟩), SOLIDUS repeated (the name given is *udkār* 'quotation' so these could likely be quotation marks ""); ⸖ CHAKMA SECTION MARK, + PLUS SIGN, − MINUS SIGN, × MULTIPLICATION SIGN, and ÷ DIVISION SIGN.

Figure 8. Handwritten text with DANDA and DOUBLE DANDA.

k	kh	g	gh	n
c	ch	j	jh	n
ṭ	ṭh	ḍ	ḍh	ṇ
t	th	d	dh	n
p	ph	b	bh	m
y	r	l	s	h
y				
a	i	u	é	o
kā	ka	ki	ku	ké
kai	kua	ko	kāṅ	kã
kyā	krā	klā	knā	k
1	2	3	4	5
6	7	8	9	0

Figure 9. Chakma alphabet chart from Bernot 1972.

A. Administrative

1. Title
Proposal for encoding the Chakma script in the UCS
2. Requester's name
UC Berkeley Script Encoding Initiative (Universal Scripts Project)
3. Requester type (Member body/Liaison/Individual contribution)
Liaison contribution.
4. Submission date
2009-07-28
5. Requester's reference (if applicable)
6. Choose one of the following:
6a. This is a complete proposal
No.
6b. More information will be provided later
Yes.

B. Technical – General

1. Choose one of the following:
1a. This proposal is for a new script (set of characters)
Yes.
1b. Proposed name of script
Chakma.
1c. The proposal is for addition of character(s) to an existing block
No.
1d. Name of the existing block
2. Number of characters in proposal
67.
3. Proposed category (A-Contemporary; 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)
Category A.
4a. Is a repertoire including character names provided?
Yes.
4b. If YES, are the names in accordance with the "character naming guidelines" in Annex L of P&P document?
Yes.
4c. Are the character shapes attached in a legible form suitable for review?
Yes.
5a. Who will provide the appropriate computerized font (ordered preference: True Type, or PostScript format) for publishing the standard?
Michael Everson and Hangendra Chakma.
5b. If available now, identify source(s) for the font (include address, e-mail, ftp-site, etc.) and indicate the tools used:
Michael Everson, Fontographer.
6a. Are references (to other character sets, dictionaries, descriptive texts etc.) provided?
Yes.
6b. Are published examples of use (such as samples from newspapers, magazines, or other sources) of proposed characters attached?
Yes.
7. 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. 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/Public/UNIDATA/ UnicodeCharacterDatabase.html and associated Unicode Technical Reports for information needed for consideration by the Unicode Technical Committee for inclusion in the Unicode Standard.
See above.

C. Technical – Justification

1. Has this proposal for addition of character(s) been submitted before? If YES, explain.
No.
2a. 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.
2b. If YES, with whom?
Hagendra Chakma, Provungshu Chakma, John Clifton, Keisuke Huziwara, Pragya Joyoti, Saikat Khisa, Helen Leake, Chandra Roy
2c. 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?

People living in Bangladesh and in India.

4a. The context of use for the proposed characters (type of use; common or rare)

Common.

4b. Reference

5a. Are the proposed characters in current use by the user community?

Yes.

5b. If YES, where?

In Bangladesh and in India.

6a. After giving due considerations to the principles in the P&P document must the proposed characters be entirely in the BMP?

Yes.

6b. If YES, is a rationale provided?

Yes.

6c. If YES, reference

Contemporary use and accordance with the Roadmap.

7. Should the proposed characters be kept together in a contiguous range (rather than being scattered)?

Yes.

8a. Can any of the proposed characters be considered a presentation form of an existing character or character sequence?

No.

8b. If YES, is a rationale for its inclusion provided?

8c. If YES, reference

9a. Can any of the proposed characters be encoded using a composed character sequence of either existing characters or other proposed characters?

No.

9b. If YES, is a rationale for its inclusion provided?

9c. If YES, reference

10a. Can any of the proposed character(s) be considered to be similar (in appearance or function) to an existing character?

No.

10b. If YES, is a rationale for its inclusion provided?

10c. If YES, reference

11a. Does the proposal include use of combining characters and/or use of composite sequences (see clauses 4.12 and 4.14 in ISO/IEC 10646-1: 2000)?

No.

11b. If YES, is a rationale for such use provided?

11c. If YES, reference

11d. Is a list of composite sequences and their corresponding glyph images (graphic symbols) provided?

No.

11e. If YES, reference

12a. Does the proposal contain characters with any special properties such as control function or similar semantics?

No.

12b. If YES, describe in detail (include attachment if necessary)

13a. Does the proposal contain any Ideographic compatibility character(s)?

No.

13b. If YES, is the equivalent corresponding unified ideographic character(s) identified?