

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

**Doc Type: Working Group Document**

**Title: Preliminary proposal for encoding the Chakma script in the UCS**

**Source: Michael Everson**

**Status: Individual Contribution**

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

**Date: 2008-04-08**

This is a preliminary proposal to encode the Chakma script in the BMP of the UCS.

**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 310,000 people in southeast Bangladesh near Chittagong City, and another 300,000 in India in Mizoram, Assam, and Tripura. Literacy in Chakma script is low. The script itself is also called  $\text{ᱠᱟ ᱵᱟᱨ}$  *Ajhā pāṭh*, sometimes romanized *Ojhopath*.

**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. The inherent vowel in Chakma is *-ā*, and the VOWEL SIGN UBARATULYAA A is used to shorten it. The other vowel signs function as expected in Brahmic scripts:

$\text{ᱠ}$	kā	=	$\text{ᱠ}$	kā
$\text{ᱡ}$	ka	=	$\text{ᱠ}$	kā + $\text{ᱠ}$ -a
$\text{ᱢ}$	ki	=	$\text{ᱠ}$	kā + $\text{ᱡ}$ i
$\text{ᱣ}$	kī	=	$\text{ᱠ}$	kā + $\text{ᱢ}$ -ī
$\text{ᱤ}$	ku	=	$\text{ᱠ}$	kā + $\text{ᱣ}$ -u
$\text{ᱥ}$	kū	=	$\text{ᱠ}$	kā + $\text{ᱤ}$ -ū
$\text{ᱦ}$	ke	=	$\text{ᱠ}$	kā + $\text{ᱦ}$ -e
$\text{ᱧ}$	ko	=	$\text{ᱠ}$	kā + $\text{ᱧ}$ -o
$\text{ᱨ}$	kāi	=	$\text{ᱠ}$	kā + $\text{ᱨ}$ -āi
$\text{ᱩ}$	kau	=	$\text{ᱠ}$	kā + $\text{ᱩ}$ -au
$\text{ᱪ}$	kaṁ	=	$\text{ᱠ}$	kā + $\text{ᱪ}$
$\text{ᱫ}$	kaṁ	=	$\text{ᱠ}$	kā + $\text{ᱫ}$
$\text{ᱬ}$	kaḥ	=	$\text{ᱠ}$	kā + $\text{ᱬ}$
$\text{ᱭ}$	ka	=	$\text{ᱠ}$	kā + $\text{ᱭ}$ virama + ZWNJ

One of the interesting features of Chakma writing is that CAANAPHUPUDAA (*candrabindu*) can be used together with EKAPHUDAA (*anusvara*) and DVIPUDAA (*visarga*):

$\text{ᳵ}$  aḥm̐ = ᳵ ā + ᳵ ḥ + ᳵ ṁ  
 $\text{ᳶ}$  aṁm̐ = ᳵ ā + ᳶ ṁ + ᳵ ṁ  
 $\text{᳷}$  uṁm̐ = ᳷ u + ᳶ ṁ + ᳵ ṁ  
 $\text{᳸}$  muṁ = ᳸ mā + ᳶ u + ᳵ ṁ

**3. Conjunct consonants.** The most frequently-occurring ligatures are subjoined wa, ya, ra, la, and na. These may occur with any consonant, though there is some :

wa: X + ᳵ virama + ᳶ wā:

ᳵ ᳶ ᳶ ᳶ ᳶ ᳶ - ᳶ ᳶ ᳶ ᳶ ᳶ ᳶ - ᳶ ᳶ ᳶ ᳶ ᳶ ᳶ  
 ᳵ ᳶ ᳶ ᳶ ᳶ ᳶ - ᳶ ᳶ ᳶ ᳶ ᳶ ᳶ - ᳶ ᳶ ᳶ ᳶ ᳶ ᳶ

ya: X + ᳵ virama + ᳸ yā:

ᳵ ᳸ ᳸ ᳸ ᳸ ᳸ - ᳸ ᳸ ᳸ ᳸ ᳸ ᳸ - ᳸ ᳸ ᳸ ᳸ ᳸ ᳸  
 ᳵ ᳸ ᳸ ᳸ ᳸ ᳸ - ᳸ ᳸ ᳸ ᳸ ᳸ ᳸ - ᳸ ᳸ ᳸ ᳸ ᳸ ᳸

ra: X + ᳵ virama + ᳶ rā:

ᳵ ᳶ ᳶ ᳶ ᳶ ᳶ - ᳶ ᳶ ᳶ ᳶ ᳶ ᳶ - ᳶ ᳶ ᳶ ᳶ ᳶ ᳶ  
 ᳵ ᳶ ᳶ ᳶ ᳶ ᳶ - ᳶ ᳶ ᳶ ᳶ ᳶ ᳶ - ᳶ ᳶ ᳶ ᳶ ᳶ ᳶ

la: X + ᳵ virama + ᳶ lā:

ᳵ ᳶ ᳶ ᳶ ᳶ ᳶ - ᳶ ᳶ ᳶ ᳶ ᳶ ᳶ - ᳶ ᳶ ᳶ ᳶ ᳶ ᳶ  
 ᳵ ᳶ ᳶ ᳶ ᳶ ᳶ - ᳶ ᳶ ᳶ ᳶ ᳶ ᳶ - ᳶ ᳶ ᳶ ᳶ ᳶ ᳶ

na: X + ᳵ virama + ᳶ nā:

ᳵ ᳶ ᳶ ᳶ ᳶ ᳶ - ᳶ ᳶ ᳶ ᳶ ᳶ ᳶ - ᳶ ᳶ ᳶ ᳶ ᳶ ᳶ  
 ᳵ ᳶ ᳶ ᳶ ᳶ ᳶ - ᳶ ᳶ ᳶ ᳶ ᳶ ᳶ - ᳶ ᳶ ᳶ ᳶ ᳶ ᳶ

A fairly large number of other conjuncts is formed.

$\text{ᳵᳶ}$  kkā = ᳵ kā + ᳵ virama + ᳶ kā  
 $\text{ᳵ᳸}$  kṭā = ᳵ kā + ᳵ virama + ᳸ ṭā  
 $\text{ᳵᳶ}$  ktā = ᳵ kā + ᳵ virama + ᳶ tā  
 $\text{ᳵ᳸}$  kmā = ᳵ kā + ᳵ virama + ᳸ mā

ක	kcā	=	ඌ	kā + ◌̄	virama +	ඞ	cā
ආ	nkā	=	ඈ	nā + ◌̄	virama +	ඌ	kā
ඈ	ngā	=	ඈ	nā + ◌̄	virama +	ආ	gā
ඞ	ccā	=	ඞ	cā + ◌̄	virama +	ඞ	cā
ච	cchā	=	ඞ	cā + ◌̄	virama +	ඡ	chā
ඣ	ñcā	=	ඡ	ñā + ◌̄	virama +	ඞ	cā
ඤ	ñjā	=	ඡ	ñā + ◌̄	virama +	ඣ	jā
ඦ	ñjhā	=	ඡ	ñā + ◌̄	virama +	ට	jhā
ඨ	ṭṭā	=	ඨ	ṭā + ◌̄	virama +	ඨ	ṭā
ඪ	ttā	=	ඌ	tā + ◌̄	virama +	ඌ	tā
ඬ	tmā	=	ඌ	tā + ◌̄	virama +	ඹ	mā
ඬ	tthā	=	ඌ	tā + ◌̄	virama +	ඹ	thā
ඳ	ddā	=	ඳ	dā + ◌̄	virama +	ඳ	dā
ඳ	ddhā	=	ඳ	dā + ◌̄	virama +	ඳ	dhā
ඳ	ntā	=	ඳ	nā + ◌̄	virama +	ඌ	tā
ඳ	nthā	=	ඳ	nā + ◌̄	virama +	ඹ	thā
ඳ	nmā	=	ඳ	nā + ◌̄	virama +	ඹ	mā
ඳ	ppā	=	ඳ	pā + ◌̄	virama +	ඳ	pā
ඳ	bbā	=	ඳ	bā + ◌̄	virama +	ඳ	bā
ඳ	mmā	=	ඹ	mā + ◌̄	virama +	ඹ	mā
ඳ	jjā	=	ඣ	jā + ◌̄	virama +	ඣ	jā
ඳ	lkā	=	ඬ	lā + ◌̄	virama +	ඌ	kā
ඳ	lgā	=	ඬ	lā + ◌̄	virama +	ආ	gā
ඳ	llā	=	ඬ	lā + ◌̄	virama +	ඬ	lā
ඳ	lṭā	=	ඬ	lā + ◌̄	virama +	ඨ	ṭā
ඳ	lpā	=	ඬ	lā + ◌̄	virama +	ඳ	pā
ඳ	schā	=	ඳ	sā + ◌̄	virama +	ඡ	chā
ඳ	sṭā	=	ඳ	sā + ◌̄	virama +	ඨ	ṭā
ඳ	skā	=	ඳ	sā + ◌̄	virama +	ඌ	kā

$$\begin{aligned} \text{ꠘ} \text{ spā} &= \text{ꠘ} \text{ sā} + \text{ꠘ} \text{ virama} + \text{ꠘ} \text{ pā} \\ \text{ꠘꠘ} \text{ smā} &= \text{ꠘ} \text{ sā} + \text{ꠘ} \text{ virama} + \text{ꠘ} \text{ mā} \\ \text{ꠘꠘ} \text{ hmā} &= \text{ꠘ} \text{ hā} + \text{ꠘ} \text{ virama} + \text{ꠘ} \text{ mā} \end{aligned}$$

And there are some Special conjuncts with ha:

[to be supplied]

**3. Collating order.** As an Indo-European language, it is likely that the standard Brahmic sorting order should apply to Chakma.

**4. 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. Both are used here in the character names

**5. Punctuation.** Alongside a single and double danda punctuation, Chakma has a unique question mark, and a section sign, PHULACIHR.

**6. Linebreaking.** Letters and digits behave as in Bengali.

































































## 7. Unicode Character Properties.

```

1AC0 CHAKMA SIGN CAANAPHUPUDAA;Mn;0;NSM;;;;;N;;;;;
1AC1 CHAKMA SIGN EKAPHUDAA;Mn;0;NSM;;;;;N;;;;;
1AC2 CHAKMA SIGN DVIPHUDAA;Mc;0;L;;;;;N;;;;;
1AC3 CHAKMA LETTER PICHAPUJHAA AA;Lo;0;L;;;;;N;;;;;
1AC4 CHAKMA LETTER DELABHAANGAGAA I;Lo;0;L;;;;;N;;;;;
1AC5 CHAKMA LETTER BACACI U;Lo;0;L;;;;;N;;;;;
1AC6 CHAKMA LETTER LEJAUBAA E;Lo;0;L;;;;;N;;;;;
1AC7 CHAKMA LETTER BAJONYAA WAA;Lo;0;L;;;;;N;;;;;
1AC8 CHAKMA LETTER CUCYAANGYAA KAA;Lo;0;L;;;;;N;;;;;
1AC9 CHAKMA LETTER GRAJAANGYAA KHAA;Lo;0;L;;;;;N;;;;;
1ACA CHAKMA LETTER CAANDYAA GAA;Lo;0;L;;;;;N;;;;;
1ACB CHAKMA LETTER TINADDAALYAA GHAA;Lo;0;L;;;;;N;;;;;
1ACC CHAKMA LETTER CILAAMA NGAA;Lo;0;L;;;;;N;;;;;
1ACD CHAKMA LETTER DVIBHALYAA CAA;Lo;0;L;;;;;N;;;;;
1ACE CHAKMA LETTER MAJARAA CHAA;Lo;0;L;;;;;N;;;;;
1ACF CHAKMA LETTER DVIPADALAA JAA;Lo;0;L;;;;;N;;;;;
1AD0 CHAKMA LETTER URAAURAA JHAA;Lo;0;L;;;;;N;;;;;
1AD1 CHAKMA LETTER SILAACYAA NYAA;Lo;0;L;;;;;N;;;;;
1AD2 CHAKMA LETTER DVIYAADAAT TTAA;Lo;0;L;;;;;N;;;;;
1AD3 CHAKMA LETTER PHUDAADVIYAAT TTHAA;Lo;0;L;;;;;N;;;;;
1AD4 CHAKMA LETTER AADUDAANGAAT DDAA;Lo;0;L;;;;;N;;;;;
1AD5 CHAKMA LETTER LEJABHARAAT DDHAA;Lo;0;L;;;;;N;;;;;
1AD6 CHAKMA LETTER PETTTUYAA NNAA;Lo;0;L;;;;;N;;;;;
1AD7 CHAKMA LETTER GHANGADAAT TAA;Lo;0;L;;;;;N;;;;;
1AD8 CHAKMA LETTER JAGADAAT THAA;Lo;0;L;;;;;N;;;;;
1AD9 CHAKMA LETTER DOLANIIT DAA;Lo;0;L;;;;;N;;;;;
1ADA CHAKMA LETTER TALAMUYAAT DHAA;Lo;0;L;;;;;N;;;;;
1ADB CHAKMA LETTER PHAARABAANYAA NAA;Lo;0;L;;;;;N;;;;;
1ADC CHAKMA LETTER PAALYAA PAA;Lo;0;L;;;;;N;;;;;
1ADD CHAKMA LETTER UBARAPHUDAA PHAA;Lo;0;L;;;;;N;;;;;
1ADE CHAKMA LETTER UBARAMUYAA BAA;Lo;0;L;;;;;N;;;;;
1ADF CHAKMA LETTER CIRADDAALYAA BHAA;Lo;0;L;;;;;N;;;;;
1AE0 CHAKMA LETTER BUGATPADALAA MAA;Lo;0;L;;;;;N;;;;;
1AE1 CHAKMA LETTER JILYAA YAA;Lo;0;L;;;;;N;;;;;
1AE2 CHAKMA LETTER DVIDAAYYAA RAA;Lo;0;L;;;;;N;;;;;
1AE3 CHAKMA LETTER TALAMUYAA LAA;Lo;0;L;;;;;N;;;;;
1AE4 CHAKMA LETTER BHUDIBUKYAA SAA;Lo;0;L;;;;;N;;;;;
1AE5 CHAKMA LETTER UBARAMUYAA HAA;Lo;0;L;;;;;N;;;;;
1AE6 CHAKMA LETTER CIMAYYAA YAA;Lo;0;L;;;;;N;;;;;
1AE7 CHAKMA VOWEL SIGN UBARATULYAA A;Mn;0;NSM;;;;;N;;;;;
1AE8 CHAKMA VOWEL SIGN BHRAYAA I;Mn;0;NSM;;;;;N;;;;;
1AE9 CHAKMA VOWEL SIGN BAANIIPHADAA II;Mn;0;NSM;;;;;N;;;;;
1AEA CHAKMA VOWEL SIGN EKATTAANA U;Mn;0;NSM;;;;;N;;;;;
1AEB CHAKMA VOWEL SIGN DVITTAANA UU;Mn;0;NSM;;;;;N;;;;;
1AEC CHAKMA VOWEL SIGN EKAARA E;Mc;0;L;;;;;N;;;;;

```



	1AC	1AD	1AE	1AF
0	 1AC0	 1AD0	 1AE0	 1AF0
1	 1AC1	 1AD1	 1AE1	 1AF1
2	 1AC2	 1AD2	 1AE2	 1AF2
3	 1AC3	 1AD3	 1AE3	 1AF3
4	 1AC4	 1AD4	 1AE4	 1AF4
5	 1AC5	 1AD5	 1AE5	
6	 1AC6	 1AD6	 1AE6	 1AF6
7	 1AC7	 1AD7	 1AE7	 1AF7
8	 1AC8	 1AD8	 1AE8	 1AF8
9	 1AC9	 1AD9	 1AE9	 1AF9
A	 1ACA	 1ADA	 1AEA	 1AFA
B	 1ACB	 1ADB	 1AEB	 1AFB
C	 1ACC	 1ADC	 1AEC	 1AFC
D	 1ACD	 1ADD	 1AED	 1AFD
E	 1ACE	 1ADE	 1AEE	 1AFE
F	 1ACF	 1ADF	 1AEF	 1AFF



## A. Administrative

1. Title

**Preliminary proposal for encoding the Chakma script in the UCS**

2. Requester's name

**Michael Everson**

3. Requester type (Member body/Liaison/Individual contribution)

**Individual contribution.**

4. Submission date

**2008-04-08**

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

**63.**

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.**

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?

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?