

JTC1/SC2/WG2 N4718
2016-04-21

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 to encode two ideographic combining pluta marks
Source: Andrew West
Status: Individual Contribution
Action: For consideration by JTC1/SC2/WG2 and UTC
Date: 2016-04-21

1. Introduction

This is a proposal to encode two combining pluta marks for ideographic scripts.

16FFE	IDEOGRAPHIC MARK PLUTA	◦
16FFF	IDEOGRAPHIC MARK TRIPLE PLUTA	⦿

The pluta sign is used to indicate overlong vowels in Sanskrit, and is used in number of different scripts that can be used to write Sanskrit, including Tibetan (U+0F85 TIBETAN MARK PALUTA) and Mongolian (U+1885 MONGOLIAN LETTER ALI GALI BALUDA and U+1886 MONGOLIAN LETTER ALI GALI THREE BALUDA). The use of a single pluta or a triple pluta was also borrowed into the Han script when used for transcribing Sanskrit. In this case, in traditional vertically oriented text, the pluta is represented as a small circle placed to the right of a Han character, and the triple pluta is represented as three small circles placed to the right of a Han character (see Fig. 1).

As the pluta is placed to the right of a character in vertical orientation, this character needs to be encoded as a combining mark, as has been proposed for the Mongolian Baluda and Three Baluda characters. We have no evidence for how the pluta should be placed in modern horizontally oriented text, but it is possible that it would be placed above the Han character, with the triple pluta rotated 90 degrees. However, if the single and triple ideographic pluta signs are encoded as combining marks, the placement of the pluta can be determined at the font level.

We currently have no evidence for the pluta in other ideographic scripts such as Tangut, Large Khitan and Jurchen, but it is quite possible that the circular pluta signs could occur with such scripts, so we suggest making these characters script-common. We suggest encoding these two characters in the Ideographic Symbols and Punctuation block.

2. Unicode Properties

16FFE;IDEOGRAPHIC MARK PLUTA; Mn;0;NSM;;;;;N;;;;;

16FFF;IDEOGRAPHIC MARK TRIPLE PLUTA; Mn;0;NSM;;;;;N;;;;;

Linebreak: CM

Script: Common

3. Figures

Fig. 1: *Tóngwén yùntǒng* 同文韻統 Table 6 (第六譜) p. 131

卷二第統韻文同

			市𠄎𠄎𠄎	市𠄎	第七譜舉鍋字爲例加巴魯達記號成二字
			𠄎𠄎𠄎𠄎	𠄎𠄎	
			𠄎𠄎𠄎𠄎	𠄎𠄎	
			鍋𠄎	鍋𠄎	

131

Single and triple pluta signs in Tibetan, Mongolian, Manchu and Chinese

4. Bibliography

Yunlu 允祿 (1695–1767). *Qīndìng tóngwén yùntǒng* 欽定同文韻統. 1750 woodblock edition.

5. Proposal Summary Form

ISO/IEC JTC 1/SC 2/WG 2 PROPOSAL SUMMARY FORM TO ACCOMPANY SUBMISSIONS FOR ADDITIONS TO THE REPERTOIRE OF ISO/IEC 10646¹ Please fill all the sections A, B and C below. Please read Principles and Procedures Document (P & P) from http://www.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://www.dkuug.dk/JTC1/SC2/WG2/docs/summaryform.html . See also http://www.dkuug.dk/JTC1/SC2/WG2/docs/roadmaps.html for latest Roadmaps.
--

A. Administrative

1. Title:	Proposal to encode two ideographic combining pluta marks
2. Requester's name:	Andrew West
3. Requester type (Member body/Liaison/Individual contribution):	Individual contribution
4. Submission date:	2016-04-21
5. Requester's reference (if applicable):	
6. Choose one of the following:	
This is a complete proposal:	YES
(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): NO Proposed name of script: _____ b. The proposal is for addition of character(s) to an existing block: YES Name of the existing block: Ideographic Symbols and Punctuation
2. Number of characters in proposal:	2
3. Proposed category (select one from below - see section 2.2 of P&P document):	A-Contemporary <input type="checkbox"/> B.1-Specialized (small collection) <input checked="" type="checkbox"/> B.2-Specialized (large collection) <input type="checkbox"/> C-Major extinct <input type="checkbox"/> D-Attested extinct <input type="checkbox"/> E-Minor extinct <input type="checkbox"/> F-Archaic Hieroglyphic or Ideographic <input type="checkbox"/> G-Obscure or questionable usage symbols <input type="checkbox"/>
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?	Andrew West
b. Identify the party granting a license for use of the font by the editors (include address, e-mail, ftp-site, etc.):	Andrew West
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)?	NO
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.	

¹ Form number: N4102-F (Original 1994-10-14; Revised 1995-01, 1995-04, 1996-04, 1996-08, 1999-03, 2001-05, 2001-09, 2003-11, 2005-01, 2005-09, 2005-10, 2007-03, 2008-05, 2009-11, 2011-03, 2012-01)

C. Technical - Justification

1. Has this proposal for addition of character(s) been submitted before?	NO
If YES explain	
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?	
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)	rare
Reference:	
5. Are the proposed characters in current use by the user community?	YES
If YES, where? Reference:	
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?	NO
If YES, is a rationale for its inclusion provided?	
If YES, reference:	
11. Does the proposal include use of combining characters and/or use of composite sequences?	NO
If YES, is a rationale for such use provided?	
If YES, reference:	
Is a list of composite sequences and their corresponding glyph images (graphic symbols) provided?	
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:	