# Proposal to encode 0D5F MALAYALAM LETTER ARCHAIC II

Shriramana Sharma, jamadagni-at-gmail-dot-com, India

2012-May-22

## §1. Introduction

In the Malayalam Unicode encoding, the independent letter form for the long vowel Ī is:

ഈ

where the length mark ◌ൗ is appended to the short vowel ഇ to parallel the symbols for U/UU i.e. ഉ/ഊ. However, Ī was originally written as:

ഀ൧ഀ

As these are entirely different representations of Ī, although their sound value may be the same, it is proposed to encode the archaic form as a separate character.
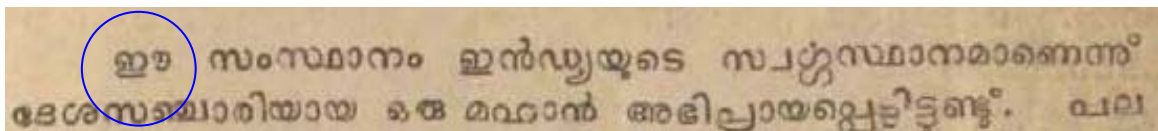
## §2. Background

As the core Unicode encoding for the major Indic scripts is based on ISCII, and the ISCII code chart for Malayalam only contained the modern form for the independent Ī [1, p 24]:
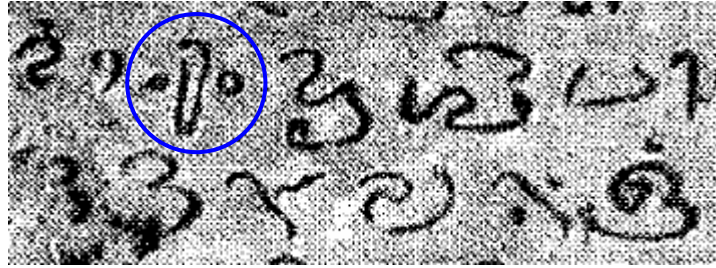


... thus it is this written form that came to be encoded as 0D08 MALAYALAM LETTER II.

While this "new" written form is seen in print as early as 1936 CE [2]:



... there is no doubt that the much earlier form was ഀ�൦ഀ parallel to the modern Grantha ◌ௐ◌, both of which are derived from the old Grantha Ī as seen below:

*ma - īdṛgvidhā | tatarāya īṇ*

Old Grantha from the Iḷaiyānputtūr copper plates [3, p 13]. Also seen is the Vatteḷuttu Ī of the same time (line 2, 2nd char from right) which also exhibits the two dots. Of course, both are derived from old South Indian Brahmi Ī ·|· which again has the two dots.

It is said [entire paragraph: Radhakrishna Warrier, personal communication] that it was the poet Vaḷḷattōḷ Nārāyaṇa Mēnōn (1878–1958) who introduced the new form of Ī ഈ. The old form �൦഻഻ seems to be found only in books printed at least 75 years ago. People of the senior generation however continued to use it in handwriting, and handwritten horoscopes around 60 years old are seen where the old Ī is used. As such, the Malayali people are clearly aware of the distinction between the old and new forms of Ī.
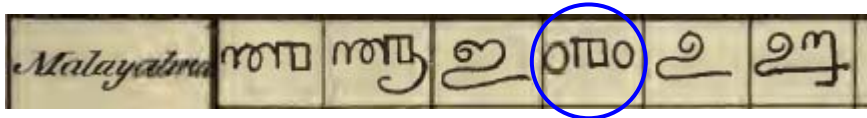
## §3. Attestation

Some of the earliest attestations of the archaic Malayalam Ī can be found in the 18th century work *Alphabetum Grandonico-Malabaricum Sive Samscrudonicum* at a period when the script was termed "Malayalam Grantha" or "Malabar Grantha" [4, pp 55 & 41]:

Even about sixty years later (in the first half of the 19th century) one sees the same [5]:

This later became standardized into the all-curve form [6, pp 7 and 8]:
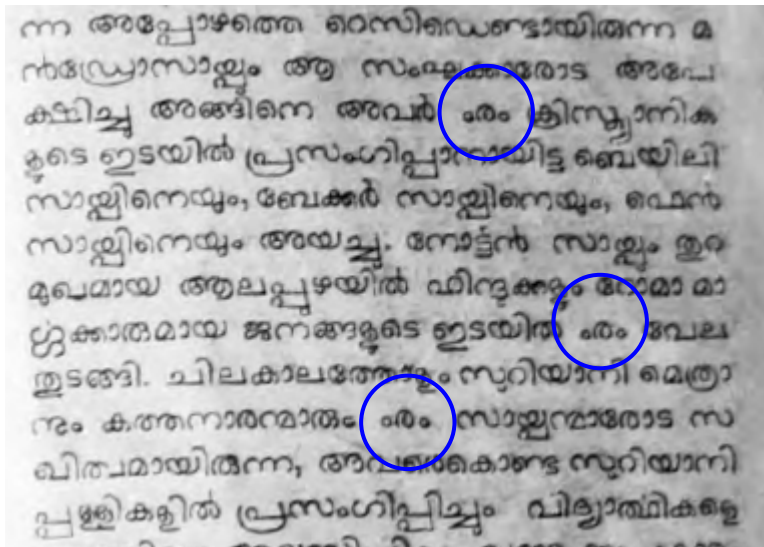
Hermann Gundert's grammatical publications in Malayalam from the 19th century also show this form exclusively: [8 p 2] [9 p XII]

5. Malayalam vowels. മലയാള സ്വരങ്ങൾ (ഉയിരുകൾ) ൧൨ ആകുന്നു.

(ആയ്യ) അ ആ ഇ ൦ൟ൦ ഉ ഊ എ ഏ ഐ ഒ ഓ ഔ

| അ | a | as a in about |
| ആ | ā | „ a „ far |
| ഇ | i | „ i „ pin |
| ൦ൟ൦ | ī | „ i „ police |

3

... whereas other works of the early 20th century show both (always indicating the older form first): [10 p 1] [11 p 352]

## Vowels.

| | | | | | | |
|---|---|---|---|---|---|---|
| 1. | [Malayalam] | = a | as | a | in | Aft. |
| 2. | [Malayalam] | = ā | „ | a | „ | Ah! |
| 3. | [Malayalam] | = i | „ | i | „ | It. |
| 4. | [Malayalam] or [Malayalam] | = ī | „ | ie | „ | Field. |

### VOWELS.

[Malayalam] a; [Malayalam] ā; [Malayalam] i; [Malayalam] or [Malayalam] ī; [Malayalam] u; [Malayalam] ū; [Malayalam] ṛu; [Malayalam] ṛū; [Malayalam] e;
[Malayalam] ē; [Malayalam] o; [Malayalam] ō; [Malayalam] ei; [Malayalam] au.

This chronology indicates that the new form came into being about the latter part of the 19th century, which corroborates the aforementioned theory that it was the poet Vaḷḷattōḷ (1878–1958) who introduced the new form.

Bhāratīya Prācīna Lipi Mālā, an authoritative text on Indian palaeography from the second half of the 20th century, still lists both forms [7 plate 81]:

मनयाळम् लिपि.

| अ | आ | इ | ई | ई | उ | ऋ | ऋ | ऋ | ए | ए | ओ | ओ | ऐ | औ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| [Malayalam script characters] | | | | | | | | | | | | | | |

| क | ख | ग | घ | ङ | च | छ | ज | झ | ञ | ट | ठ | ड | ढ | ण | त | थ | द | ध |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| [Malayalam script characters] | | | | | | | | | | | | | | | | | | |

| न | प | फ | ब | भ | म | य | र | ल | व | श | ष | स | ह | ळ | ऴ | ऱ | का | कि | की |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| [Malayalam script characters] | | | | | | | | | | | | | | | | | | | |

| कु | कू | के | कै | कै | को | कौ | कौ | १ | २ | ३ | ४ | ५ | ६ | ७ | ८ | ९ | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| [Malayalam script characters] | | | | | | | | | | | | | | | | | |

... but perhaps only because it is about "prācīna lipi"-s i.e. "old scripts", as we have seen that later standards such as ISCII did not even list the older form as it fell out of use.

# §4. Justification for encoding

It is thus very clear that the old form of Malayalam Ī i.e. ൦ഌ൦ indeed existed and was in popular use at one point of time. Thus, in textually digitizing such old writings, it would be appropriate to store this old written form ൦ഌ൦ distinctly from the new form ഈ.

Even though the two different forms of Ī ഈ and ൦ഌ൦ signify the same sound, they should be distinctly encoded, as **Unicode encodes written forms and not the sounds thereof.** An excellent precedent exists in the modern Bengali/Assamese script encoding, where the phoneme /ra/ is denoted by 09B0 র in Bengali language orthography and 09F0 ৰ in Assamese, and these are distinctly encoded as part of the same script. The Devanagari script encoding also treats distinctly the two different ways of marking the dependent vowel E: 0947  े and 094E ॎ. The reason for these distinct encodings is that while the sound value may be the same, the written form is not the same and they are not stylistic variants as there is no common glyphic skeleton to the two written forms. For the same reason, the two different ways of writing Ī in Malayalam must be distinctly encoded.

# §5. Character to be encoded

<p style="text-align:center; font-size:2em">൦ഌ൦</p>

<p style="text-align:center">0D5F Mᴀʟᴀʏᴀʟᴀᴍ Lᴇᴛᴛᴇʀ Aʀᴄʜᴀɪᴄ II</p>

# §6. Unicode Character Properties

`0D5F;MALAYALAM LETTER ARCHAIC II;Lo;0;L;;;;;;N;;;;;`

The codepoint chosen for the character is intended to isolate it from the other characters in common use in the contemporary Malayalam orthography. It also effectively places it along with the rare use vowel characters 0D60-0D63.

The name proposed for the character is based on the pattern of <script> <category> <adjective> <identifier> seen in the names of 094E Dᴇᴠᴀɴᴀɢᴀʀɪ Vᴏᴡᴇʟ Sɪɢɴ Pʀɪsʜᴛʜᴀᴍᴀᴛʀᴀ E and other such characters.

It should be noted that while the new form ഇൗ is confusable with 0D07 ഇ + 0D57 ◌ൗ, the old form ൦ൠ൦ is confusable with 0D02 ◌ം + 0D30 ര + 0D02 ◌ം. If necessary, this character may be prohibited from use in IDNs after seeking feedback from the user community.

## §7. References

1. Indian Script Code for Information Interchange, Bureau of Indian Standards, New Delhi, 1991, read from http://varamozhi.sourceforge.net/iscii91.pdf 2012-May-22

2. https://ml.wikisource.org/w/index.php?title=%E0%B4%A4%E0%B4%BE%E0%B5%BE: Geography_textbook_4th_std_tranvancore_1936.djvu/5&oldid=33784

3. Āvaṇam, Tamil Nadu Archaeological Society, Thanjavur, 2007, ISSN 0975-1831

4. Alphabetum Grandonico-Malabaricum Sive Samscrudonicum, Rome, 1772; http://books.google.com/books?id=qmETAAAAQAAJ&dq=Alphabetum%20Grandonico-Malabaricum%20Sive%20Samscrudonicum

5. Ancient and Modern Alphabets of the Popular Hindu Alphabets of the Southern Peninsula of India, Capt Henry Harkness, Royal Asiatic Society, 1837, http://www.archive.org/details/ancientmodernalp00harkrich

6. http://nid.edu/typography11/download.php?f=Hashim-P-M_Birth-of-A-Typeface.pdf, P M Hashim, retrieved 2012-May-22

7. Bhāratīya Prācīna Lipi Mālā, Gaurishankar Ojha, reprinted by Munshiram Manoharlal Publishers, New Delhi, 1971

Courtesy of Elmar Kniprath, Germany:

8. Malayāḷa Bhāṣā Vyākaraṇam, Hermann Gundert, Basel Mission Press, Mangalore, 1868; second reprint: Asian Educational Services, New Delhi, 1991

9. A Malayalam and English Dictionary, Hermann Gundert, Basel Mission Press, Mangalore, 1872; second reprint: Asian Educational Services, New Delhi, 1989

10. A Progressive Grammar of the Malayalam Language, L Johannes Frohnmeyer, Basel Mission Press, Mangalore, 1889, 2$^{nd}$ edn 1913; reprint: Asian Educational Services New Delhi, 1989

11. Linguistic Survey of India – Vol IV, George A Grierson, 1st edn, Office of the Supdt. of Government Printing, Calcutta, 1906: reprint by Motilal Banarsidass, Delhi, 1967

# §8. Official Proposal Summary Form

(Based on N3902-F)

## A. Administrative

1. Title
**Proposal to encode 0D5F MALAYALAM LETTER ARCHAIC II**
2. Requester's name
**Shriramana Sharma**
3. Requester type (Member body/Liaison/Individual contribution)
**Individual contribution**
4. Submission date
**2012-May-22**
5. Requester's reference (if applicable)
6. Choose one of the following: This is a complete proposal (or) More information will be provided later
**This is a complete proposal.**

## B. Technical – General

1. Choose one of the following:
1a. This proposal is for a new script (set of characters), Proposed name of script
**No.**
1b. The proposal is for addition of character(s) to an existing block, Name of the existing block
**Yes. Malayalam.**
2. Number of characters in proposal
**1 (one)**
3. Proposed category
**Category B1, specialized small (for this character, though Malayalam itself is "A, contemporary")**
4. Is a repertoire including character names provided?
**Yes.**
4a. If YES, are the names in accordance with the "character naming guidelines" in Annex L of P&P document?
**Yes.**
4b. 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?
**Shriramana Sharma. (This glyph may be composed easily using Malayalam glyphs already used in the code chart to maintain stylistic consistency. I am however ready to provide one if necessary.)**
b. Identify the party granting a license for use of the font by the editors (include address, e-mail etc.)
**Shriramana Sharma and other contributors to the Lohit Malayalam font, under derivative rights granted by the OFL. See https://fedorahosted.org/lohit/.**
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)?
**Not applicable.**
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.
**See detailed proposal.**

## 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?

**Radhakrishna Warrier, Rajeev Sebastian, Cibu Johny, Shiju Alex. Contact details available on request.**

2c. If YES, available relevant documents

**None specifically. The matter was discussed via email.**

3. Information on the user community for the proposed characters (for example: size, demographics, information technology use, or publishing use) is included?

**People desiring to employ the archaic form of II from the old orthography.**

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

**Not very common in the popular context.**

4b. Reference

**See detailed proposal.**

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

**Yes.**

5b. If YES, where?

**In texts (such as horoscopes) written using the old (i.e. 60+ years) Malayalam orthography.**

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

**It belongs in the Malayalam block which is in the BMP.**

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

**Only one character is proposed. It is proposed to be placed near other rare use characters.**

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?

**It is similar in appearance to a sequence of existing characters i.e. 0D02 0D30 0D02.**

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

**Yes.**

10c. If YES, reference

**The character properties i.e. GC=Lo etc would be different.**

11a. Does the proposal include use of combining characters and/or use of composite sequences?

**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?

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?

13c. If YES, reference:


–o–o–o–