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 add** HEBREW POINT HOLAM HASER FOR VAV **to the BMP of the UCS**
Source: **Michael Everson & Mark Shoulson**
Status: **Individual Contribution**
Action: **For consideration by JTC1/SC2/WG2 and UTC**
Date: **2004-07-29**

This document requests an additional character to be added to the UCS and contains the proposal summary form.

A. Administrative

1. Title
Proposal to add HEBREW POINT HOLAM HASER FOR VAV to the BMP of the UCS.
2. Requester's name
Michael Everson & Mark Shoulson.
3. Requester type (Member body/Liaison/Individual contribution)
Individual contribution.
4. Submission date
2004-07-29
5. Requester's reference (if applicable)
6. Choose one of the following:
6a. This is a complete proposal
Yes.
6b. More information will be provided later
No.

B. Technical – General

1. Choose one of the following:
1a. This proposal is for a new script (set of characters)
No.
Proposed name of script
1b. The proposal is for addition of character(s) to an existing block
Yes.
1b. Name of the existing block
Hebrew
2. Number of characters in proposal
1
3. Proposed category (see section II, Character Categories)
Category B.1
4a. Proposed Level of Implementation (1, 2 or 3) (see clause 14, ISO/IEC 10646-1: 2000)
Level 3.
4b. Is a rationale provided for the choice?
Yes.
4c. If YES, reference
Combining character.
5a. Is a repertoire including character names provided?
Yes.

5b. If YES, are the names in accordance with the character naming guidelines in Annex L of ISO/IEC 10646-1: 2000?
Yes.

5c. Are the character shapes attached in a legible form suitable for review?
Yes.

6a. Who will provide the appropriate computerized font (ordered preference: True Type, or PostScript format) for publishing the standard?
Michael Everson. TrueType.

6b. If available now, identify source(s) for the font (include address, e-mail, ftp-site, etc.) and indicate the tools used:
Michael Everson. Fontographer.

7a. Are references (to other character sets, dictionaries, descriptive texts etc.) provided?
Yes, see bibliography below.

7b. Are published examples of use (such as samples from newspapers, magazines, or other sources) of proposed characters attached?
Yes.

8. 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, see below.

9. 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.
Yes, see Unicode properties below.

C. Technical – Justification

1. Has this proposal for addition of character(s) been submitted before? If YES, explain.
Yes. In document L2/04-193 this character is described as HEBREW POINT LEFT HOLAM. That document recommends a joiner-based solution for the problem; as noted below, we believe such a solution is not preferable to the new character proposed here.

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?
Discussion on the hebrew@unicode.org list.

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?
Yes.

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

4b. Reference
See examples below.

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

5b. If YES, where?
See examples below.

6a. After giving due considerations to the principles in Principles and Procedures document (a WG 2 standing document) must the proposed characters be entirely in the BMP?
Yes.

6b. If YES, is a rationale provided?
Yes.

6c. If YES, reference
All Hebrew points are in the BMP.

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

8a. Can any of the proposed characters be considered a presentation form of an existing character or character sequence?
No. While it is similar to Hebrew point holam it contrasts in position and sometimes size and height. It it not a positional variant. It is functionally a different character.
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?
Yes.
10b. If YES, is a rationale for its inclusion provided?
Yes.
10c. If YES, reference
It is derived from Hebrew point holam but it has a different placement, sometimes shape, and it has a different interpretation.
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)?
Yes.
11b. If YES, is a rationale for such use provided?
No.
11c. If YES, reference
12a. Is a list of composite sequences and their corresponding glyph images (graphic symbols) provided?
No.
12b. If YES, reference
13a. Does the proposal contain characters with any special properties such as control function or similar semantics?
No.
13b. If YES, describe in detail (include attachment if necessary)
14a. Does the proposal contain any Ideographic compatibility character(s)?
No.
14b. If YES, is the equivalent corresponding unified ideographic character(s) identified?
14c. If YES, reference

D. Proposal

The early Semitic letter vav was originally used consonantally, but by the time the Square Hebrew script (encoded as "Hebrew") came into being, the use of vav as one of the *matres lectionis* was well established. In unpointed Hebrew text, vav can be read as a consonant [w] ([v] in modern dialects), or as an indicator of a vowel [o] or [u]. Various systems of pointing arose to allow for more precise readings. In the Tiberian system most widely used today, one vowel point relevant to the present proposal is U+05B9 Hebrew point holam, which is used to indicate [o].

This point is unusual among the Hebrew vowel points in the Tiberian system, in that it is positioned above the letter which bears it, not below the letter like all the others. Its form is that of a dot, placed either above the left edge of the letter or over the space between it and the next letter. The original, proper position appears to have been between the letters, or at least overhanging a little; placing it on the top left of the letter which bears it may be a concession to printing technology.

When Hebrew point holam is used with vav, the situation is more complex, because vav can be read either with its original consonantal value, or as a *mater lectionis*. It must be understood that there is a terminological difference between *holam* (the vowel [o]) appearing with a consonant and holam (the point indicating that vowel). So it is true to say that in pointed Hebrew, *mem holam* [mo] can be represented as מֹ mem holam or as מוֹ mem vav holam.

The letter vav is troublesome in this context because even when pointed, vav plus point holam can be ambiguous; it can be read [o] or [wo] ([vo]). A vav used as a *mater lectionis* with a point holam is called

holam male 'full holam'; when the POINT HOLAM is used on a consonant without the *mater* VAV, it is called *holam haser* 'deficient holam'. But VAV can also be used as a consonant with *holam haser* – in which case it can look just like the *holam male*.

In order to make a distinction between these two uses of POINT HOLAM and VAV, writers of Hebrew manuscripts as early as the Aleppo Codex (ca. 900 CE) took to placing the dot on the VAV with *holam haser* further to the left of the dot on VAV when *holam male*. Indeed, some of the early manuscripts have the dot on the *holam male* decidedly to the right of the VAV, between it and the letter before, suggesting graphically that it is really a POINT HOLAM on the previous letter with a "silent VAV" *mater* written afterwards. In general, however, the *holam male* is generally drawn with the dot either squarely on top of the VAV or over a bit to the right of its centre (but still atop the VAV): וֹ. The VAV with *holam haser* has its dot over to the left, as POINT HOLAM is generally drawn on other letters: וֹ.

The plain-text distinction made here, from a millennium ago in scribal manuscripts and later in careful typography, cannot be represented in Unicode and ISO/IEC 10646 at present. By far and away the most common use of VAV with POINT HOLAM is that of the *mater lectionis* vowel; the VAV with *holam haser* is the marked case, which is why we propose that a new character be added to represent that case. To quantify the matter, we compared the use of *holam male* vs. VAV with *holam haser* in the online *Biblia Hebraica Stuttgartensia* (*eBHS*): there are 34,699 instances of the former, and 421 instances of the latter – in other words, *holam male* comprises 98.79% of the cases involving HOLAM on VAV and consonantal VAV with *holam haser* comprises only 1.21% of the cases in the same text.

The proposal for a new character here is analogous to the proposal made in N2755R to distinguish QAMATS QATAN from QAMATS. In texts which do not distinguish the two, QAMATS is used as the generic mark; reading rules only distinguish them. In texts which *do* make a distinction, the QAMATS QATAN can be used. Here, we likewise recognize that in texts which do not distinguish the two, HOLAM is used as the generic mark; reading rules only distinguish them. In texts which *do* make a distinction, the HOLAM HASER FOR VAV can be used. The name of the character clearly indicates the intended use of the character.

Discussion on the **hebrew@unicode.org** list has resulted in a different proposal for dealing with this to represent HOLAM HASER FOR VAV by means of a sequence VAV ZWNJ HOLAM. The precedent claimed for this use of ZWNJ after a base character and before a combining character is the sequence defined for Bengali Reph and Ya-phalaa in *The Unicode Standard* version 4.0.1. We contend that the "precedent" claimed in that proposal is not appropriate for the Hebrew usage described here. The sequence BENGALI LETTER RA + BENGALI SIGN VIRAMA + BENGALI LETTER YA is a normal sequence in a quintessentially ligating script where the RA takes a special form when ligated with YA. In the sequence BENGALI LETTER RA + ZWNJ + BENGALI SIGN VIRAMA + BENGALI LETTER YA the ZWNJ does not cause the VIRAMA to change its shape or position – it simply indicates that the script's normal ligation behaviour is changed so that the RA retains its shape and the YA takes on a special shape. We contend that the proposal to use ZWNJ for Hebrew is an inappropriate extension of the use of the joiner because Hebrew is not a ligating script. The letter VAV is a base letter and the POINT HOLAM is a combining diacritical mark; וֹ is no more a ligature than *ö* is.

We believe that the encoding principle accepted for QAMATS QATAN should be applied here as well – a changed shape for the unmarked point should be reflected in plain text with the use of an explicitly-encoded character, for those users who prefer to make that distinction.

| 05B9 | ◌ֹ | HEBREW POINT HOLAM |
| | | • used generically or as holam male with vav in orthography which distinguishes it from holam haser |

| (05C7) | ◌ | HEBREW POINT HOLAM HASER FOR VAV |

Unicode Character Properties

`05C7;HEBREW POINT HOLAM HASER FOR VAV;Mn;19;NSM;;;;;N;;*;;;`

Figures

Here are some examples of texts which distinguish the two points. *Holam male* points are indicated with a featherless arrow, and VAV with *holam haser* points are indicated with a feathered arrow.

Figure 1. From the Aleppo Codex, ca. 900 CE. The word is *baʕăwono* 'by his sin', from Joshua 22:20, with HOLAM HASER FOR VAV on the first (rightmost) VAV and with HOLAM on the second (leftmost) VAV.

Figure 2. From the *Leningrad Codex* 1006. The word is *ʕăwono* 'his sin', from Leviticus 5:1, with HOLAM HASER FOR VAV on the first VAV and with HOLAM on the second VAV.

Figure 3. From the *Lisbon Codex*, 1492. The word is *ʕăwono* 'his sin', Leviticus 5:1, with HOLAM HASER FOR VAV on the first VAV and with HOLAM on the second VAV.

Figure 4. From the *Biblia Rabbinica*, Joseph ben Hayyim, 1524–5. The word is *ʕăwono* 'his sin', from Leviticus 5:1, with HOLAM HASER FOR VAV on the first VAV and with HOLAM on the second VAV.

Figure 5. From Letteris 1866. The word is *ʕăwono* 'his sin', from Leviticus 5:1, with HOLAM HASER FOR VAV on the first VAV and with HOLAM on the second VAV.

עָוֹן ‚ עָוֹן (`ăvŏ'n) *m*, *s*. עָוֹן,
w.s. עֲוֹנִי, *pl.* עֲוֹנִים עֲוֹנוֹת,
perverseness, sin, guilt,
crime; punishment, suf-
fering.

Figure 6. From Feyerabend [1961]. The word is *ʕăwon* 'sin', with HOLAM on the first VAV and with HOLAM HASER FOR VAV on the second VAV.

עֲוֹנוֹ

Figure 7. From the *Biblia Hebraica Stuttgartensia*, Schenker 1977. The word is *ʕăwono* 'his sin', with HOLAM HASER FOR VAV on the first VAV and with HOLAM on the second VAV.

עֲוֹנוֹ

Figure 8. From Snaith [1995]. The word is *ʕăwono* 'his sin', from Leviticus 5:1, with HOLAM on both VAVs. This is an example of the many cases where no distinction is being made.

עֲוֹנוֹ

Figure 9. From Scherman 1996. The word is *ʕăwono* 'his sin', with HOLAM HASER FOR VAV on the first VAV and with HOLAM on the second VAV.

עֲוֹנוֹ

Figure 10. From *The Jersualem Bible*, Fisch 1997. The word is *ʕăwono* 'his sin', with HOLAM HASER FOR VAV on the first VAV and with HOLAM on the second VAV.

בַּעֲוֹנוֹ

Figure 11. From *The Jersualem Bible*, Fisch 1997. The word is *baʕăwono* 'by his sin', from Joshua 22:20, with HOLAM HASER FOR VAV on the first VAV and with HOLAM on the second VAV.

מִצְוֺתַי   חֻקּוֹתַי

Figure 12. From *Humash ha-Menuqad*, 1999. The words are *miṣwotaj* 'my commandments' and *ḥuqqotaj* 'my laws', from Genesis 26:5, with HOLAM HASER FOR VAV on the first VAV and with HOLAM on the second VAV.

וְתַחְתּוֹנִים,   וַעֲוֺנִי,
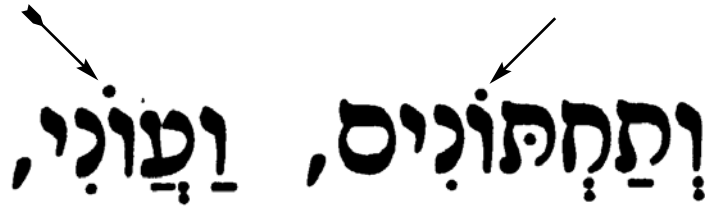
Figure 13. From *Humash ha-Menuqad*, 1999. The Horev Bible is very much intended for children; even all the commentaries are pointed. The examples here are from Rashi's commentary on Genesis 4:13: *wǝtaḥtonīm, waʕăwonī* ('and lower ones, and my sin'), with HOLAM on the first VAV and with HOLAM HASER FOR VAV on the second VAV.

E. PUNCTUATION AND VOWELIZATION MARKS

When designing fonts for printing, one should also design punctuation marks and a system of vowelization marks fitting the style of the fonts.

The Hebrew vowelization system consists of short horizontal lines and dots placed below the letters, as well as dots placed above and inside the letters. The vowelization marks are as follows: Patah – a short horizontal line; Kamatz – a short horizontal line with a dot placed in the centre below it; Hirik – a single dot; Tsere – two dots placed horizontally; Shewa – two dots placed vertically; Kubutz – three dots placed obliquely, descending to the right; Segol – three dots placed in the form of a reversed triangle; Hataf-Patah, Hataf-Kamatz and Hataf-Segol – the respective vowel sign with a Shewa on its right-hand side. Holam – a dot above a Waw or to the left of the upper corner of any letter; Shuruk – Waw with a dot to the left of its centre. In addition to these vowelization marks, there is the Dagesh mark, which is a dot placed inside a letter to indicate that it is doubled or pronounced in a different way. A consonantal He at the end of a word is also marked with a dot, called Mappik. Shin and Sin are marked with diacritical points.

בְּגְזֶמְלִינְטְעָחָאֹוֹר

The dots of Shewa and of the Hatafs, as well as of the Holam, the Shuruk, the Dagesh, the Mappik and the diacritical points of Shin and Sin may be somewhat smaller than the other dots. The Holam on the Waw should be placed above the centre of its 'head', while the Holam on other letters should be placed to the left of their upper corners. The Patah and Kamatz should be narrower than the narrow letters. The punctuation marks should also fit the script-style and be in proportion to the size of the letters. Their spacing should be calculated according to the spacing of the letter-signs.

בְּשָׂא   שׁשֹׂוּוֹפכבנ
שׁשֹׂוּוֹפכב
שׁשֹׂוּוֹפכב

Figure 14. From Yardeni 2002. This text is in a presecriptive section describing best practice for designing inscriptions and typefaces. While Yardeni does not refer to *holam* when used with consonantal VAV, it is clear that she considers the position of *holam* in its more common usage to have a positioning behaviour which differs from that of *holam* with other letters.

Bibliography

Feyerabend, Karl. [1961]. *Langenscheidt's pocket Hebrew dictionary to the Old Testament. Hebrew-English*. Langenscheidt. ISBN 3 468 97082 X

Fisch, Harold, ed. 1997. *The Holy Scriptures* (*The Jerusalem Bible*). Jerusalem: Koren Publishers.

Horev Publishing House. 1999. *Humash ha-Menuqad*. Jerusalem: Horev Publishing House.

Joseph ben Hayyim. 1524-25. *Biblia Rabbinica*. Venice: Daniel Bomberg.

Letteris, Meir, ed. 1866. *Sefer Torah, Nevi'im u-Khetuvim*. London: British and Foreign Bible Society.

Schenker, A., ed. 1977. *Biblia Hebraica Stuttgartensia*. 5th edition. Stuttgart: Deutsche Bibelgesellschaft. ISBN 3 438 05218 0

Scherman, Nosson, ed. 1996. *The Stone Edition Tanach*. New York: Mesorah Publications Ltd.

Snaith, Norman Henry, ed. [1995?]. *Sefer Torah, Nevi'im u-Khetuvim*. London: British and Foreign Bible Society. ISBN 0 564 00029 9

Yardeni, Ada. 2002. *The book of the Hebrew script: history, palaeography, script styoles, calligraphy & design*. New Castle, Delaware: Oak Knoll Press; London: British Library. ISBN 1-58456-087-8 (Oak Knoll) and 0-7123-4793-3 (British Library)