

Title: Background material for the proposal on the Hebrew vowel *HOLAM*
Source: Peter Kirk
Status: Individual Contribution
Action: As background for the UTC when considering the *HOLAM* proposal
Date: 2004-07-29

This document is background material for the separate proposal made to the August 2004 UTC meeting on the Hebrew vowel HOLAM, document L2/04-307 (also available as <http://qaya.org/academic/hebrew/Holam3.pdf>). This material is adapted from the proposal made to the June 2004 UTC meeting as document L2/04-193 (also available as <http://qaya.org/academic/hebrew/Holam.pdf>). The list of options in that proposal has been adjusted, extended and clarified, partly to meet the objections expressed at that meeting by UTC members. In further discussions among Hebrew experts a consensus has been reached on a single preferred option (which is Option A1b in the list below), and a new proposal document has been written recommending that option. This document is presented to the UTC to demonstrate that many other options have been considered and to outline the advantages and disadvantages of each. It is intended as reference material, to be consulted if any of these options are suggested as alternatives during UTC discussions of the proposal.

General Background

There are two ways of indicating vowels in Hebrew script, which may be used either separately or in combination. The ancient system, which does not fully distinguish the vowel sounds, is to insert the Hebrew letters ALEF, HE, VAV and YOD, which can therefore function as vowels as well as consonants. When "silent", i.e. used to indicate vowels, these letters are known *mothers of reading* (*imot qeri'a* or *ehevi* in Hebrew, *matres lectionis* in Latin). In the early mediaeval period several different systems of pointing were introduced to specify the vowel sounds more precisely. Only one of these systems, the Tiberian system, is in current use, and this is the only one currently encoded in Unicode. (Proposals for the other systems are currently being prepared.) This system is normally used for the biblical and other ancient texts (although not for synagogue scrolls, which are unpointed) and for some modern Hebrew texts. Most modern Hebrew is unpointed, but makes good use of *mothers of reading*.

One of the Tiberian vowel points, U+05B9 HEBREW POINT HOLAM, consists of a dot usually written above the left side of a Hebrew base character. This represents a long O sound pronounced after the base character. When there is no associated *mother of reading*, this way of writing a long O sound is known as *Holam Haser*, i.e. *Defective Holam*. In old manuscripts, the dot is often positioned over the space between the preceding and following base characters, and sometimes above the right side of the following (to the left) base character. In printed texts, the regular position of the dot is above the left side of the preceding base character.

In pointed Hebrew text the same vowel is often represented both by a vowel point and by a *mother of reading*. The latter has no vowel point of its own, because the vowel is associated with the preceding consonant. The commonest *mother of reading* for a long O sound is VAV. Therefore the combination of HOLAM with a VAV *mother of reading* is common in pointed texts. This combination is known as *Holam Male* (*Male* is pronounced as two syllables, *mah-leh*), i.e. *Full Holam*. The HOLAM dot is logically associated with the preceding base character, the consonant for which it indicates the vowel sound; the VAV is redundant because the vowel is fully indicated by the HOLAM. Thus the VAV may be considered silent, corresponding to the general rule for pointed texts that a non-final base character with no point is silent; an alternative analysis is that the VAV and the

HOLAM together indicate the vowel sound. In the oldest manuscripts which use this pointing scheme, dating from the 10th century CE, the dot was positioned above the space between the preceding base character and the VAV, but it has gradually shifted on to the redundant VAV. In modern typography the dot is positioned above the VAV, usually above its right edge or its centre. However, the HOLAM dot is not shifted on to a following VAV when the VAV is not silent but consonantal, except sometimes in rendering the divine name.

The difficulty arises because VAV can also be a consonant, and as such can be followed, like every other consonant, by *Holam Haser* (or by *Holam Male*, but this causes no special difficulty). Therefore the HOLAM dot can combine in two logically different ways with VAV. The combination of VAV with *Holam Haser* is known as *Vav Haluma*, and is pronounced VO (or in some traditions WO). A combination of VAV with HOLAM could be *Holam Male*, where the VAV is silent and the letter VAV and the point HOLAM together represent the vowel; or it could be the letter VAV with *Holam Haser*, where the VAV is a consonant and the HOLAM point is a vowel. There is no difference in pronunciation between *Holam Male* and *Holam Haser*.

In more exact typography, especially of the Hebrew Bible and other religious and liturgical texts, of educational materials, and of poetry, a careful distinction is made between *Holam Male* and *Vav Haluma*: in *Holam Male*, the HOLAM dot is positioned above the right side of the VAV, or sometimes centred above the VAV; but in *Vav Haluma*, *Holam Haser* is rendered in its normal position above the left side of VAV. This seems to have been the original practice, as witnessed in manuscripts and printed editions from the 10th to 19th centuries CE. But, because VAV is a rather narrow letter, and because *Vav Haluma* is rare in modern Hebrew (in which long O is usually written as *Holam Male*), most modern typographers of general texts make no distinction, rendering both *Holam Male* and *Vav Haluma* by VAV with a HOLAM dot usually centred above it.

The distinction between *Holam Male* and *Vav Haluma* is an important and semantically significant one. This is especially true for religious texts; the distinction is made in many Hebrew Bible editions, and in texts quoting from the Bible. It is also important in educational materials and in poetry, wherever the exact pronunciation must be marked unambiguously. See the examples in the figures below, in which *Holam Male* and *Vav Haluma* are distinguished in several Hebrew Bible editions and in various other works.

This distinction is not a rare one. *Holam Male* is very common in the Hebrew Bible, occurring about 34,808 times or in about 13% of all words. *Vav Haluma* is much less common, occurring about 421 times.

Samples

(These figures are the same as in the proposal except that Figure 5 is not in the proposal.)

Codex Leningradensis (1006-7)	Lisbon Bible (1492)	Rabbinic Bible (1524-5)
Ginsburg/BFBS edition (1908)	Biblia Hebraica Stuttgartensia (1976)	Stone edition of Tanach (1996)

Figure 1: *Holam Male* (marked in red) and *Vav Haluma* (marked in blue) distinguished in ancient and modern editions of the Hebrew Bible - these words are from Genesis 4:13. (If the colours are not visible: In each image, the third base character from the right, with the dot above its right side or its centre, is *Holam Male*; the third base character from the left, with the dot above its left side, is *Vav Haluma*.)



Figure 2: *Holam Male* (left, twice, red, from p.529) and *Vav Haluma* (right, blue, from p.528) contrasted in Keil & Delitzsch *Commentary on the Old Testament*, vol.1, reprint by Hendrickson, 1996 (Hebrew words quoted in English text).



Figure 3: *Holam Male* (right Hebrew word, red) and *Vav Haluma* (left word, blue) contrasted in Langenscheidt's *Pocket Hebrew Dictionary*, p.243.

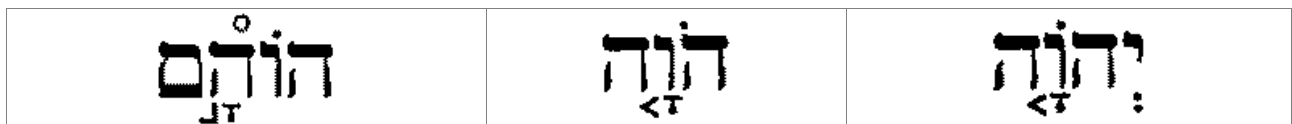


Figure 4: Comparison of positions of HOLAM after HE and with VAV in *Biblia Hebraica Stuttgartensia*. Left: regular *Holam Male*, from Joshua 10:3. Centre: HOLAM dot not shifted on to consonantal VAV, as this is not *Holam Male*, from Ezekiel 7:26. Right: HOLAM dot shifted to *Holam Male* position on a consonantal VAV in the divine name, although this is not *Holam Male*, from Exodus 13:15.



Figure 5: *Holam Male* (red) written with a different glyph from a regular VAV (blue), from *Siddur Tikkun Meir Hashalem*, R. Greenfield, 1982.

<p>Yose ben Yose (5th century), from <i>sidrei avodah</i> for <i>yom hakipurim</i> ("etain <i>tehila</i>"), in Goldschmidt, <i>Mahzor L'yamim Nora'im</i>, Koren publishing 1970, p464</p>	<p>R. Elazar Hakalir (poetry of the late 6th century), from <i>piyyut</i> for <i>Shavuot</i>, "<i>eret mateh</i>", in Shulamit Elizur, <i>Kedushtaot l'yom matan torah</i>, Meketzzei Nirdamim, 2000, p116</p>	<p><i>Midrash Tanchuma</i> (8th century), <i>Or haHayim</i>, v1, 1998, p185</p>
<p>Yannai (poet of the early 6th century), from <i>kedushta piyyut</i> "<i>ashrei mo'asei alrla</i>", in Zaulai, <i>Piyyute Yannai</i>, Shocken Publishing, 1938, p32</p>		

Figure 6: *Holam Male* (red) and *Vav Haluma* (blue) distinguished in modern editions of mediaeval Hebrew poetry and midrashic literature.

<p><i>Mahzor Yom Hakippurim</i>, Israel Ariel, ed., Makhon Hamikdash / Carta Publishing, 1995, p92</p>	<p><i>Siddur Tefila</i>, Koren Publishing, 1996, p60</p>	<p><i>Hagada Shel Pesach</i>, Torat Chaim series, Mosad Harav Kook, 1998, p142</p>

Figure 7: *Holam Male* (red) and *Vav Haluma* (blue) distinguished in modern editions of liturgical texts. Note the larger and higher HOLAM dots in *Vav Haluma* in the right hand two examples; other idiosyncratic distinctions are made especially in Koren Publishing editions of such liturgical texts.

Holam Male and Unicode

The Unicode Hebrew block is based on the Israeli national standard SI 1311. This standard was originally designed for unpointed modern Hebrew texts, although later extended to cover points (SI 1311.1) and accents (SI 1311.2) (see <http://qsm.co.il/Hebrew/stdisr.htm> for further details), but was not designed for full support of biblical Hebrew. As a result there are some minor inadequacies in the Unicode support for biblical Hebrew.

The most significant of these inadequacies, because it is the only one which affects the vowel points rather than only the accents, is that there is no support for the distinction between *Holam Male* and *Vav Haluma*. There is a single VAV character and a single HOLAM character, and only one way of combining these two, the sequence <VAV, HOLAM>, which is apparently intended to be used for both *Holam Male* and *Vav Haluma*. There is thus no defined way of distinctively encoding either *Holam Male* or *Vav Haluma*.

The alphabetic presentation form U+FB4B HEBREW LETTER VAV WITH HOLAM cannot be used for *Holam Male* distinct from *Vav Haluma*, because it is canonically equivalent to the sequence <VAV, HOLAM>, i.e. it has a canonical decomposition (which cannot be changed) to 05D5 05B9. It is included in Unicode for compatibility purposes.

Because there is a real need to distinguish between *Holam Male* and *Vav Haluma*, but there is no standard way of doing so, various ad hoc solutions have been used by text providers and by font developers. The Hebrew Bible text from Mechon Mamre (at Genesis 4:13, <http://www.mechon-mamre.org/c/ct/c0104.htm#13>) uses <VAV, HOLAM> for *Holam Male* and <VAV, ZWJ, HOLAM> for *Vav Haluma*. The "alpha release" text at <http://whi.wts.edu/WHI/Members/klowery/eL/leningradCodex-alpha.zip> and the text at <http://users.ntplx.net/~kimball/Tanach/Genesis.xml> use (again at Genesis 4:13) <HOLAM, VAV> (actually <HOLAM, accent, VAV> according to canonical ordering) for *Holam Male* and <VAV, HOLAM> for *Vav Haluma*, and this is also the encoding recommended in the documentation for the fonts SBL Hebrew and Ezra SIL. There is however a larger body of existing data, including pointed modern Hebrew and some biblical texts (e.g. the one at <http://www.anastesontai.com/b-cantilee/encant.asp?A=1&listeB=4>), in which *Holam Male* and *Vav Haluma* are not distinguished but are both encoded as <VAV, HOLAM>.

Design Goals

To avoid this inconsistency and potential confusion, a proposal has been made that the UTC should specify distinctive character sequences for representation of *Holam Male* and *Vav Haluma*, for use when these two need to be distinguished. Various options for these distinctive sequences are discussed below. It is noted that although Option B1 can technically be chosen without UTC involvement, because it involves only a spelling rule, the other options do require UTC approval as they involve sequences with ZWJ or ZWNJ, or variation sequences, or new characters.

The options for distinctive sequences have been chosen in an attempt to meet the following design goals and preferences which have been expressed:

- Representations should conform to the general rules and principles of Unicode, as specified in *The Unicode Standard (TUS)*, and not require any extension to these rules and principles.

In this regard there is a specific issue concerning use of ZWJ and ZWNJ. These characters were not permitted within combining character sequences in *TUS* version 4.0.0, but this restriction has been lifted at least partially in *TUS* version 4.0.1 (<http://www.unicode.org/versions/Unicode4.0.1/>), as stated clearly in the approved minutes of the February 2004 UTC meeting (<http://www.unicode.org/consortium/utc-minutes/UTC-098-200402.html>):

[98-C33] **Consensus:** Allow U+200D ZERO WIDTH JOINER and U+200C ZERO WIDTH NON-JOINER in combining character sequences. The interpretation of a joiner or a nonjoiner between two combining marks is not yet defined.

The original proposal, presented to the June 2004 UTC meeting, relied on the new definitions in *TUS* version 4.0.1, and in several of its options (equivalent to A1a/b/c and

B2a in the list of options below), ZWJ and ZWNJ were used in combining character sequences, and in one case (B2a) between two combining marks. However, at the June meeting UTC members seemed reluctant to accept use of ZWJ and ZWNJ within combining character sequences, even when not between two combining marks. Therefore in this document new options (A3a/b/c, A4a/b/c and B3a) have been added in which ZWJ and ZWNJ are used strictly as defined in *TUS* version 4.0.0, as well as an option (A2a) in which a variation selector is used instead of ZWJ and according to the *TUS* rules for use of variation selectors.

- Representations should be based as far as possible on existing Unicode and Israeli national standard encodings of Hebrew. They should be compatible with the principles expressed in the quotation from Israeli Standard SI 4281 in Unicode document L2/04-213 (also available as <http://www.qsm.co.il/Hebrew/Responses%20to%20Several%20Hebrew%20Items.pdf>), especially that software may choose not to render or transmit Hebrew points, and as far as possible with existing implementations which ignore a defined list of points.
- The sequence <VAV, HOLAM> should continue to be a valid representation of both *Holam Male* and *Vav Haluma* when there is no need to distinguish them, as commonly in modern Hebrew text.
- It is recognised that introduction of a new representation of any character or graphical form, in a situation where an existing representation is already widely used and must remain valid, will inevitably result in ambiguities in the representation of that character or form and inconsistencies in data. However, the level of ambiguity and inconsistency should be kept to a minimum. It is considered highly undesirable to define two incompatible representations of Hebrew data, one for use in biblical, liturgical, educational and poetic texts and another for general use. The Hebrew user community has already decisively rejected a proposal for separate encoding of vowels for biblical Hebrew; and now, almost all Hebrew users involved in discussions of the current proposal have expressed a clear preference against encoding any new characters for support of *Holam Male* or *Vav Haluma*. Means of minimising the problem include choosing a new representation which automatically falls back or folds to the existing representation, and defining a new representation only for a rarely used character or graphical form.
- Rendering processes should provide sensible fallback renderings of *Holam Male* and *Vav Haluma* when a font is applied which does not have special features to display these two correctly. One option is to fall back to treating *Holam Male* and *Vav Haluma* as identical, following the practice of many modern typographers. An alternative fallback for *Holam Male* is to render it as *Holam Haser* followed by unpointed VAV; this may be preferable, especially for educational materials and poetry, because it preserves the pronunciation distinction.
- Specifically, the new sequences for *Holam Male* and, if applicable, *Vav Haluma* should be displayed legibly and as far as possible correctly, although necessarily without every fine typographical distinction, by existing rendering systems and fonts which currently display Hebrew without distinguishing *Holam Male* from *Vav Haluma*.

This design goal is feasible if careful use is made of default ignorable characters such as ZWJ, ZWNJ and variation selectors, which according to existing Unicode principles should be ignored by rendering systems and fonts which do not recognise specific sequences using these characters.

This unusual design goal requires some justification. An important motivation for bringing this issue to the UTC as a matter of some urgency is the proliferation of ad hoc solutions described above. These have been developed to meet a perceived need to

make available to the general public, on the Internet and by other means, standard electronic texts of the Hebrew Bible and other ancient Hebrew texts. The proposers consider it important that this proliferation is stopped as quickly as possible. The first requirement for stopping such proliferation is that a standard representation of distinctive *Holam Male* is agreed, and that is the main purpose of this proposal. However, proliferation will be halted only when the new standard representation becomes widely supported, at least to a sufficiently close approximation to satisfy most users. There is also a strong resistance to solutions which formalise distinctions between ancient and modern Hebrew; modern Hebrew readers are likely to reject on these grounds any solution which makes the Hebrew Bible text unreadable on their existing systems. Therefore priority is given in this proposal to options which can already be rendered at least approximately by existing rendering systems and fonts, without the need to wait for a number of years for updated fonts to be installed worldwide.

Note that this distinction differs from some others made in Unicode, for example between HYPHEN and MINUS, in that it must be made not only for special typographical purposes. For example, whereas HYPHEN and MINUS do not need to be distinguished in general purpose electronic texts, *Holam Male* does need to be distinguished from *Vav Haluma* in some such texts because the distinction affects the interpretation and pronunciation of the text; therefore with certain texts it is not an optional matter. There are also special issues of integrity and authority with the biblical text which makes it undesirable that different versions of the text should be distributed for different groups of end users. Modern Hebrew readers require the ability to view the full Hebrew text with all the points, accents and fine distinctions, even though they are not able to understand all of these distinctions.

- Processes other than rendering should fall back to treating *Holam Male* and *Vav Haluma* as identical when no deliberate distinction is being made. Thus, for example, the new sequences for *Holam Male* and, if applicable, *Vav Haluma* should by default collate together with <VAV, HOLAM> except at the binary level.

It is noted that because in the current *Default Unicode Collation Element Table (DUCET)* VAV and HOLAM have weights at different levels, for practical purposes <VAV, HOLAM> and <HOLAM, VAV> collate together, and ZWJ, ZWNJ and variation selectors are ignored, except at the binary level. Therefore with all of the options using only these characters *Holam Male* and *Vav Haluma* collate together except at the binary level as desired. If a new character is defined, it should be given an appropriate default collation weight to meet this design goal.

- It is *not* a general design goal to allow a full three-way distinction between *Holam Male*, *Vav Haluma*, and undifferentiated VAV with HOLAM. It has been suggested by some that this might be necessary, but no evidence has been presented that any typesetters make a three-way distinction. The options in the Appendix with a final "c" allow a three-way distinction to be made if required.
- The choice of sequence should meet the objections of UTC members to the options in the original June 2004 proposal.

There is no single solution which ideally meets all of these design goals. For this and other reasons many options have been considered for representation of *Holam Male* and *Vav Haluma*, offering various trade-offs between the design goals. All the options considered worthy of serious consideration are listed below, with a comparative table of their advantages and disadvantages.

Full List of Options

In the options with a final "a", and in Options B1 and C1, the recommended encoding for *Vav Haluma* is simply <VAV, HOLAM>. Thus *Vav Haluma* is identified with undifferentiated VAV with HOLAM. Although *Vav Haluma* is less common than *Holam Male*, this corresponds to the regular use of HOLAM with other Hebrew consonants; this is the reason for proposing the specially marked, and in most cases longer, encoding for more common case.

In the options with a final "b" *Holam Male* is identified with undifferentiated VAV with HOLAM. These have the advantage that the longer and more complex sequence or the new character is used for the less common combination, *Vav Haluma* (or with the logical structure options *Holam Haser* followed by consonantal VAV), but the disadvantage that *Holam Haser* is treated differently when adjacent to consonantal VAV from when adjacent to all other Hebrew consonants. Most of these options also have the advantage that the representation of the very common *Holam Male* continues to be <VAV, HOLAM>.

The options with a final "c" allow typesetters to make a three-way distinction, distinguishing undifferentiated VAV with HOLAM both from *Holam Male* and from *Vav Haluma* (or both from *Holam Male* and from *Holam Haser* followed by consonantal VAV). It seems unlikely that this is ever necessary, and so that the extra complexity of these options can be justified.

A. Graphical Structure Solutions

These options are called "graphical structure solutions" because they represent the dot in *Holam Male* according to its graphical association with the VAV.

A1a. *Holam Male* = <VAV, ZWJ, HOLAM>, *Vav Haluma* and undifferentiated VAV with HOLAM = <VAV, HOLAM>

[This was Option A1 in the June 2004 proposal.]

This option effectively takes *Holam Male* as a variant of <VAV, HOLAM> with "a more connected rendering" (to quote from *The Unicode Standard*, version 4.0.0, section 15.3, p.390). This more connected rendering is indicated by inserting U+200D ZERO WIDTH JOINER (ZWJ) between VAV and HOLAM. This option was earlier rejected because ZWJ and ZWNJ were not permitted between a base character and a combining character. But this restriction was partially relaxed at the February 2004 UTC meeting.

This encoding has the advantage that the fallback behaviour should be automatically as required. One disadvantage is that as a layout control character ZWJ is intended for making rendering distinctions which have no other semantic significance. However, there are already several defined uses of ZWJ and ZWNJ with Arabic and Indic scripts which do have other semantic significance. There are similar objections to any possible variant of this option using Variation Selectors.

There are no known existing implementations of this option. However, it would be simple to support in fonts.

This option, as well as Options B1 and B2, implies that undifferentiated VAV with HOLAM will be rendered like *Vav Haluma*, not like *Holam Male*. In fact it seems that many typesetters who do not generally distinguish *Vav Haluma* from *Holam Male* render the HOLAM dot above VAV further to the right than the HOLAM dot indicating *Holam Haser* when used with other letters, for example with YOD whose upper part is usually the same as that of VAV. This suggests that if in a particular text these typesetters did need to distinguish *Vav Haluma* from *Holam Male*, the glyph they would use for *Vav Haluma* would not be the one which they used for undifferentiated VAV with HOLAM.

Another disadvantage of this option is that each *Holam Male* consists of three Unicode characters, including ZWJ which takes three bytes in UTF-8. This increases the size of the encoded Hebrew Bible, relative to Options A1b and B1 (in which *Holam Male* consists of two characters), by 34,000 characters and more than 100,000 UTF-8 bytes, i.e. around 2% of its total length.

A1b. *Holam Male* and undifferentiated VAV with HOLAM = <VAV, HOLAM>, *Vav Haluma* = <VAV, ZWNJ, HOLAM>

[This was Option A2 in the June 2004 proposal.]

This option differs from Option A1a in that the simple sequence <VAV, HOLAM> is used for *Holam Male*, rather than for *Vav Haluma*. The proposed sequence for *Vav Haluma* uses U+200C ZERO WIDTH NON-JOINER (ZWNJ), because *Vav Haluma* is a *less* connected rendering than *Holam Male*. This option has the advantage that the longer and more complex sequence is used for the less common *Vav Haluma*, but the disadvantage that consonantal VAV is treated differently from all other Hebrew consonants in how it combines with *Holam Haser*. The fallback behaviour of this option should be as required.

This sequence was rejected earlier for the same theoretical reasons as Option A1a, but for the same reasons it can now be considered acceptable.

This option implies that undifferentiated VAV with HOLAM will be rendered like *Holam Male*, not like *Vav Haluma*. It may therefore represent more closely than Options A1a, B1 or B2 the practice of typesetters who do not normally distinguish *Vav Haluma* from *Holam Male* but may have to for certain special texts. This option shares with A2b, A3b and A4b the advantage of minimising the incompatibility between new and existing texts: the existing widely used sequence <VAV, HOLAM> continues to be used for *Holam Male* and is changed for *Vav Haluma* only by the addition of a default ignorable control character.

The encoding already used by Mechon Mamre is similar to this option except that ZWNJ is replaced by ZWJ. This encoding is apparently supported by existing some fonts and rendering engines, but this support may be largely accidental, because the ZWJ unintentionally breaks a rule to position HOLAM centrally over VAV. The long term encoding of text should not be determined in this way by unintended features of current implementations.

A1c. *Holam Male* = <VAV, ZWJ, HOLAM>, *Vav Haluma* = <VAV, ZWNJ, HOLAM>, undifferentiated VAV with HOLAM = <VAV, HOLAM>

[This was Option A3 in the June 2004 proposal.]

This option differs from Options A1a and A1b in that explicit sequences with ZWJ or ZWNJ are used to distinguish both *Holam Male* and *Vav Haluma* from the undifferentiated VAV with HOLAM. Again, the fallback behaviour of this option should be as required. Otherwise, this option seems to have the disadvantages of both Options A1a and A1b.

A2a. *Holam Male* = <VAV, variation selector, HOLAM>, *Vav Haluma* and undifferentiated VAV with HOLAM = <VAV, HOLAM>

This option differs from Option A1a in that a variation selector is used in place of ZWNJ. The point has been made that in Options A1a-A1c ZWJ and ZWNJ are used where a variation selector is more appropriate. This option is intended to respond to that point. On the one hand, it can be argued that the variation sequence <VAV, variation selector> should indicate a variant form of VAV rather than a variant positioning of the HOLAM dot. On the other hand, the logical difference between *Holam Male* and *Vav Haluma* is not so much in the HOLAM as in the VAV. At the glyph level, the VAV in

Holam Male differs from a regular consonantal VAV in having a different attachment point for the HOLAM dot. There is also occasional use of a slightly different VAV glyph in *Holam Male*, as in Figure 5 above.

Arguably it would make more sense to use a variation selector with HOLAM rather than with VAV, but the definition of variation selectors does not allow them to be used with combining characters.

A2b. *Holam Male* = <VAV, HOLAM>, *Vav Haluma* and undifferentiated VAV with HOLAM = <VAV, variation selector, HOLAM>

This option relates to Option A2a in the same way that Option A1b relates to Option A1a. It has no real advantages over Option A2a, and again the disadvantage that consonantal VAV is treated differently from all other Hebrew consonants in how it combines with *Holam Haser*.

A2c. *Holam Male* = <VAV, variation selector, HOLAM>, *Vav Haluma* = <VAV, another variation selector, HOLAM>, undifferentiated VAV with HOLAM = <VAV, HOLAM>

This option relates to Option A2a in the same way that Option A1c relates to Option A1a. It has no real advantages over Option A2a, and the same disadvantage as Option A2b.

A3a. *Holam Male* = <VAV, HOLAM, ZWNJ>, *Vav Haluma* and undifferentiated VAV with HOLAM = <VAV, HOLAM>

The A3 and A4 options differ from the A1 options in that ZWJ and ZWNJ are used only between combining character sequences, and not within them. This corresponds to the usage of these characters defined in *The Unicode Standard* version 4.0.0, and avoids use of the extended mechanisms accepted at the February 2004 UTC meeting and incorporated into *TUS* version 4.0.1. In these options ZWJ and ZWNJ are used, according to the definitions in *TUS* 4.0.0 section 15.2, to indicate renderings in which whole combining character sequences are respectively more or less closely connected in rendering.

Option A3a is based on an understanding of *Holam Male* as a rendering of VAV with HOLAM which is less connected with the following base character than *Vav Haluma*. It is therefore distinguished from *Vav Haluma* by insertion of ZWNJ before the following base character.

A3b. *Holam Male* and undifferentiated VAV with HOLAM = <VAV, HOLAM>, *Vav Haluma* = <VAV, HOLAM, ZWJ>

This option differs from Option A3a in that *Holam Male* is taken as the default case, and *Vav Haluma* as a special case in which the VAV with HOLAM is taken as more closely connected with the following base character. One advantage of this is that *Vav Haluma* is not normally used word finally, at least in the Hebrew language, whereas *Holam Male* is commonly word final; and so a theoretically problematic common use of word final ZWNJ is avoided. This option also has the same advantages and disadvantages relative to Option A3a as Option A1b does relative to Option A1a.

A3c. *Holam Male* = <VAV, HOLAM, ZWNJ>, *Vav Haluma* = <VAV, HOLAM, ZWJ>, undifferentiated VAV with HOLAM = <VAV, HOLAM>

This option relates to Options A3a and A3b in the same way as Option A1c relates to Options A1a and A1b.

A4a. *Holam Male* = <ZWJ, VAV, HOLAM>, *Vav Haluma* and undifferentiated VAV with HOLAM = <VAV, HOLAM>

The A3 options are based on VAV with HOLAM being either more or less connected with the **following** base character and its combining character sequence, but this connection difference is not a real one. But there is a real difference in how *Holam Male* and *Vav Haluma* are connected with the **preceding** base character and its combining character sequence. Within the logical structure of the Hebrew *abjad*, *Holam Male* acts as the vowel for the preceding base character and as part of the same syllable; indeed, if it were a separate character (as in Option C1) a good case could be made for defining it as a spacing combining mark, comparable to such marks in Indic scripts. It thus has a closer logical connection with the preceding base character than does *Vav Haluma*, which represents a separate syllable. Graphically, the closer connection is commonly indicated by the positioning of the HOLAM dot over the space between the base characters.

Option A4a is based on this understanding of *Holam Male* as a rendering of VAV with HOLAM which is more connected with the preceding base character and its combining character sequence than *Vav Haluma*. It is therefore distinguished from *Vav Haluma* by insertion of ZWJ between this and the preceding combining character sequence.

A4b. *Holam Male* and undifferentiated VAV with HOLAM = <VAV, HOLAM>, *Vav Haluma* = <ZWNJ, VAV, HOLAM>

This option differs from Option A4a in that *Holam Male* is taken as the default case, and *Vav Haluma* as a special case in which the VAV with HOLAM is taken as less closely connected with the preceding combining character sequence. This seems to accord less well with the logical structure of the script. This option also has the same advantages and disadvantages relative to Option A4a as Option A1b does relative to Option A1a.

A4c. *Holam Male* = <ZWJ, VAV, HOLAM>, *Vav Haluma* = <ZWNJ, VAV, HOLAM>, undifferentiated VAV with HOLAM = <VAV, HOLAM>

This option relates to Options A4a and A4b in the same way as Option A1c relates to Options A1a and A1b.

B. Logical Structure Solutions

These options are called "logical structure solutions" because they represent the dot in *Holam Male* according to its logical association with the preceding base character. In all of these solutions *Vav Haluma* and undifferentiated VAV with HOLAM are represented as <VAV, HOLAM>.

B1. *Holam Male* and *Holam Haser* followed by consonantal VAV = <HOLAM, VAV>

[This was Option B1 in the June 2004 proposal.]

In this option *Holam Male* is distinguished from *Vav Haluma* in that HOLAM is encoded before VAV. This appears to be a breach of the Unicode rule that combining characters must follow their associated base characters. But it is not really a breach of the rule, because the HOLAM in *Holam Male* can be understood as logically associated with the preceding base character, for which it is the associated vowel, and the VAV is a separate silent letter. On this analysis *Holam Male* is analogous to *Hiriq Male*, i.e. HIRIQ followed by silent YOD, in which the HIRIQ is written below the preceding base character; also to the sequence of HOLAM with silent ALEF, which is encoded unambiguously in this order although the HOLAM is often rendered above the top right side of the ALEF.

With this encoding, the HOLAM is for Unicode purposes linked with the preceding base character in a combining character sequence. The HOLAM will often become separated from the VAV by DAGESH and/or an accent character, because within a combining character sequence DAGESH and accents are sorted after vowel points in canonical ordering and also in the specific orderings recommended for certain fonts.

The fallback behaviour of this encoding, with a font which has not been set up to work with it, is not ideal but still legible: the *Holam Male* will be broken up, with the HOLAM being rendered above the left side of the preceding base character.

Some existing texts use this encoding, and it is supported in OpenType fonts such as SBL Hebrew and Ezra SIL, with Microsoft Windows only. However, this implementation proved to be very complex, and may be beyond the capabilities of other rendering systems.

The complicating factor is the rule that *Holam Male* is not formed, and so HOLAM is not shifted on to a following VAV, if the VAV is consonantal and followed by a vowel, except in the divine name. This rule, which is illustrated in Figure 4 above, is complex and not entirely conditioned by the immediate glyph or character environment. In most cases it is possible in principle, although rather complex, to determine within the font which VAVs are silent and so may form *Holam Male*; the rule is that if VAV is followed by any Hebrew point or accent it is not silent. But there are two cases where this is not possible. Firstly, a VAV followed by *Holam Male* or by *Vav Shruqa* (i.e. VAV with DAGESH acting as a vowel; but this combination may also be consonantal) is consonantal and so cannot form *Holam Male*, but any attempt to distinguish these cases within a font is potentially recursive and well beyond the capabilities of existing rendering systems. (This situation does not occur in the Hebrew Bible, but it can do in modern Hebrew.) Secondly, in at least one major edition of the Hebrew Bible, when the divine name is written with HOLAM (which is in a small minority of cases) the HOLAM dot is positioned over the VAV as in *Holam Male* although the VAV is consonantal and carries another vowel point and usually an accent; this case can be distinguished from a similar word in which the HOLAM is not positioned as in *Holam Male* only from the remote context, in a way which is clearly outside the scope of any rendering system - see the centre and right hand images in Figure 4.

Since it is beyond the reasonable scope of a rendering system to determine in every case whether *Holam Male* should be formed or not, there is a need to define more specific encodings at least for certain marginal cases. Thus, for example, formation of *Holam Male* could be inhibited by the sequence <ZWJ, HOLAM, VAV> or <HOLAM, ZWNJ, VAV>, which would indicate *Holam Haser* followed by consonantal VAV; but this formation could be promoted by the sequence <ZWNJ, HOLAM, VAV> or <HOLAM, ZWJ, VAV>, which would indicate the rendering of the divine name as in the right hand image in Figure 4. The implication of this is that Option B1 does not in fact have the simplicity which it appears to have at first sight.

B2a. *Holam Male* = <ZWNJ, HOLAM, VAV>, *Holam Haser* followed by consonantal VAV = <HOLAM, VAV>

[This was Option B2 in the June 2004 proposal.]

This option differs from Option B1 in that HOLAM is preceded by ZWNJ to separate it from the preceding combining character sequence. Again, this is a sequence which was rejected earlier for the same theoretical reasons as Option A1a, but for the same reasons it can now be considered acceptable. The HOLAM is technically and logically combined with the preceding base character as in Option B1, but the intervening ZWNJ can be understood as indicating that it should not be combined graphically.

With this proposal, any accents and other combining characters which are graphically as well as

logically associated with the preceding base character should be encoded before the ZWNJ. The ZWNJ, which is in combining class 0, inhibits canonical reordering, and so these other combining characters will never be moved to between HOLAM and VAV. The ZWNJ also explicitly signals that the HOLAM is to be shifted to form *Holam Male* or as in the divine name, and so distinguishes this from the cases in which the HOLAM dot remains on the preceding base character before consonantal VAV. This implies that it is significantly simpler to implement Option B2 than Option B1.

This option has the same disadvantage as Options A1a and A1c that the length of a text is significantly increased. Its fallback behaviour should be the same as that of Option B1.

B2b. *Holam Male* = <HOLAM, VAV>, *Holam Haser* followed by consonantal VAV = <ZWJ, HOLAM, VAV>

This option differs from Option B2a in that *Holam Haser* followed by consonantal VAV is treated as the marked case. This option has the advantage that the longer and more complex sequence is used for the less common case, but the disadvantage that *Holam Haser* is treated differently when followed by consonantal VAV from when followed by other Hebrew consonants.

B2c. *Holam Male* = <ZWNJ, HOLAM, VAV>, *Holam Haser* followed by consonantal VAV = <ZWJ, HOLAM, VAV>

In this option marked sequences are used for both cases. It has no real advantages over Options B2a and B2b.

B3a. *Holam Male* = <HOLAM, ZWJ, VAV>, *Holam Haser* followed by consonantal VAV = <HOLAM, VAV>

This option differs from Option B1 in that HOLAM is followed by ZWJ. This sequence has the advantage over the one in Option B2 that ZWJ is used between combining character sequences, according to the definitions in *TUS* version 4.0.0. ZWJ is properly used to indicate a more closely connected rendering of the two combining character sequences, in that the HOLAM dot which logically belongs to the former is graphically shifted on to the latter. ZWJ can be omitted where the HOLAM dot is not to be shifted, but included in the anomalous cases of the divine name. Therefore, again, this option is significantly simpler to implement than Option B1. But it does not have the advantage of Option B2 of inhibiting canonical reordering, and so the implementation advantage is less.

This option has the same disadvantage as Options A1a and A1c that the length of a text is significantly increased. Its fallback behaviour should be the same as that of Option B1.

B3b. *Holam Male* = <HOLAM, VAV>, *Holam Haser* followed by consonantal VAV = <HOLAM, ZWNJ, VAV>

This option differs from Option B3a in that *Holam Haser* followed by consonantal VAV is treated as the marked case. This option has the advantage that the longer and more complex sequence is used for the less common case, but the disadvantage that *Holam Haser* is treated differently when followed by consonantal VAV from when followed by other Hebrew consonants.

B3c. *Holam Male* = <HOLAM, ZWJ, VAV>, ***Holam Haser*** followed by consonantal VAV = <HOLAM, ZWNJ, VAV>

In this option marked sequences are used for both cases. It has no real advantages over Options B3a and B3b.

C. New Character Solutions

The common factor with these options is that one or more new Unicode characters is proposed, for use only when *Holam Male* is to be distinguished from *Vav Haluma*. They have the common disadvantage that they have very poor fallback behaviour when used with fonts which do not support the new character. Some experts have commented that any of these solutions have the effect of making existing uses of HOLAM illegal. In fact the definitions could be carefully written so that existing uses are not made illegal but only deprecated. Nevertheless, this effect on existing texts is a significant argument against any of these new character solutions.

C1a. New character HEBREW LETTER HOLAM MALE

[This was Option C1 in the June 2004 proposal.]

In some ways the simplest option of all is to define a new Unicode character HEBREW LETTER HOLAM MALE, which might have a compatibility decomposition to <VAV, HOLAM>. This would certainly be simple to implement, and would reduce the size of the encoded text. But it would have no suitable fallback behaviour with fonts which do not support this new character. This solution also loses the essential identity of the HOLAM and the VAV in *Holam Male* with HOLAM and VAV in other contexts. There is also a significant complication, shared by all the options involving new base characters, that conversion to and collation with unpointed text becomes more complex than simply stripping off combining marks.

This option shares with all of the options in which *Holam Male* is represented by a sequence including a new character (i.e. also C1c, C2a, C2c and C4; also C3 in which there is comparable disruption to *Holam Haser*) the serious disadvantage that it introduces a second and incompatible representation for a form which is already widely represented as <VAV, HOLAM>.

C1b. New character HEBREW LETTER VAV HALUMA

This alternative of defining a new character for *Vav Haluma* is equally simple to implement, and has the advantage that its fallback behaviour is good except for the relatively rare *Vav Haluma*. But it introduces an entirely illogical distinction between *Vav Haluma* and other combinations of consonants with *Holam Haser*, which is justified neither by character semantics nor by typography.

C1c. Two new characters HEBREW LETTER HOLAM MALE and HEBREW LETTER VAV HALUMA

In this option two new characters are defined, one for *Holam Male* and the other for *Vav Haluma*. The fallback behaviour is uniformly bad for all cases of VAV with HOLAM, and it introduces the same illogical distinctions as Option C1b. The only advantage of defining a second new character is that it would make possible support for a three-way distinction in HOLAM positioning for which no requirement has been demonstrated.

C2a. New character HEBREW POINT RIGHT HOLAM

[This was Option C2 in the June 2004 proposal.]

This is the first of four options based on defining one or two new combining characters for variant of HOLAM. Thus one variant of HOLAM can be used for the dot in *Holam Male*, and another variant can be used in *Vav Haluma*. These options are reasonably simple to implement. They have the small advantage over Option C1 that the identity of VAV, though not of HOLAM, is preserved.

In this option, the new combining character is HEBREW POINT RIGHT HOLAM, and is to be used only in combination with VAV to form *Holam Male*. The existing HOLAM character is to be used only for *Holam Haser*, when combined with any Hebrew consonant, and for undifferentiated VAV with HOLAM. The fallback behaviour is good for *Holam Haser* but not for *Holam Male*.

C2b. New character HEBREW POINT LEFT HOLAM or HEBREW POINT HOLAM FOR VAV HALUMA

[This was Option C4 in the June 2004 proposal.]

In this option, the new combining character is HEBREW POINT LEFT HOLAM, or more specifically HEBREW POINT HOLAM FOR VAV HALUMA, and is to be used only in combination with VAV to form *Vav Haluma*. The existing HOLAM character is to be used in combination with VAV to form *Holam Male*, and for *Holam Haser* in combination with consonants other than VAV, and for undifferentiated VAV with HOLAM. The fallback behaviour is good except for the relatively rare *Vav Haluma*, i.e. *Holam Haser* with VAV. But this option introduces an entirely illogical distinction between *Holam Haser* with VAV and *Holam Haser* with other consonants, which is justified neither by character semantics nor by typography.

C2c. Two new characters HEBREW POINT RIGHT HOLAM and HEBREW POINT LEFT HOLAM

[This was the June 2004 UTC ad hoc committee's suggestion.]

In this option two new combining characters are defined: HEBREW POINT RIGHT HOLAM to be used as in Option C2a and HEBREW POINT LEFT HOLAM to be used as in Option C2b. The existing HOLAM character is to be used with VAV only for undifferentiated VAV with HOLAM. The fallback behaviour is uniformly bad for all cases of VAV with HOLAM, and it introduces the same illogical distinctions as Option C2b. The only advantage of defining a second new combining character is that it would make possible support for a three-way distinction in HOLAM positioning for which no requirement has been demonstrated.

C3. New character HEBREW POINT HOLAM HASER

[This was Option C3 in the June 2004 proposal.]

This option differs from the C2 options, and indeed from all the other options in this proposal, in proposing a change in the representation of HOLAM even when not associated with VAV. In this option, the new combining character is HEBREW POINT HOLAM HASER, and is to be used for *Holam Haser* when combined with any Hebrew consonant, not only with VAV. The existing HOLAM character is to be used only in combination with VAV to form *Holam Male*, and for every HOLAM if *Holam Male* is not differentiated from *Vav Haluma*. The fallback behaviour is good for *Holam Male* but not for *Holam Haser*; this may be preferable to the fallback behaviour of Option C2a because *Holam Male* is commoner than *Holam Haser* in modern Hebrew.

C4. New character HEBREW LETTER VAV VOWEL

This option is based on the observation that *Holam Male* differs from *Vav Haluma* not in the HOLAM but in the VAV. Therefore it is theoretically preferable to define a new VAV character rather than a new HOLAM character. Unicode does not encode distinctions between consonants and vowels

when there is no graphical distinction; thus there is only one LATIN SMALL LETTER Y. However, there is a graphical distinction between the VAVs in *Holam Male* and *Vav Haluma*, in that they are positioned differently relative to HOLAM; also a distinctive VAV glyph is occasionally used in *Holam Male*, as shown in Figure 5. There is thus justification for encoding a separate character HEBREW LETTER VAV VOWEL, for use primarily as the base character in *Holam Male*, and possibly also as the base character in *Vav Shruqa*. However, this option shares with all the new character solutions the disadvantage of bad fallback behaviour. It also shares the complication that conversion to and collation with unpointed text becomes more complex than simply stripping off combining marks.

Summary of Advantages and Disadvantages

<i>Option</i>	<i>Summary</i>	<i>Fallback Behaviour</i>	<i>Advantages</i>	<i>Disadvantages</i>
A1a	<i>Holam Male</i> = <VAV, ZWJ, HOLAM>	<i>Holam Male</i> = <i>Vav Haluma</i>	Best fit to the graphical structure of Hebrew script	ZWJ used within combining character sequence and with semantic significance; long sequence for a common character
A1b	<i>Vav Haluma</i> = <VAV, ZWNJ, HOLAM>	<i>Holam Male</i> = <i>Vav Haluma</i>	Best fit to the graphical structure of Hebrew script; long sequence only for a rare combination; minimal incompatibility between existing and new texts	ZWNJ used within combining character sequence and with semantic significance; arbitrary use of different sequence for <i>Holam Haser</i> in the context of VAV
A1c	<i>Holam Male</i> = <VAV, ZWJ, HOLAM> and <i>Vav Haluma</i> = <VAV, ZWNJ, HOLAM>	<i>Holam Male</i> = <i>Vav Haluma</i>	Best fit to the graphical structure of Hebrew script; support for conjectured three-way HOLAM positioning distinction	ZWJ and ZWNJ used within combining character sequence and with semantic significance; long sequence for a common character; arbitrary use of different sequence for <i>Holam Haser</i> in the context of VAV
A2a	<i>Holam Male</i> = <VAV, variation selector, HOLAM>	<i>Holam Male</i> = <i>Vav Haluma</i>	Doesn't use ZWJ or ZWNJ	Variation selector used with semantic significance; long sequence for a common character
A2b	<i>Vav Haluma</i> = <VAV, variation selector, HOLAM>	<i>Holam Male</i> = <i>Vav Haluma</i>	Doesn't use ZWJ or ZWNJ; minimal incompatibility between existing and new texts	Variation selector used with semantic significance; arbitrary use of different sequence for <i>Holam Haser</i> in the context of VAV
A2c	<i>Holam Male</i> = <VAV, variation selector, HOLAM> and <i>Vav Haluma</i> = <VAV, another variation selector, HOLAM>	<i>Holam Male</i> = <i>Vav Haluma</i>	Doesn't use ZWJ or ZWNJ	Variation selector used with semantic significance; long sequence for a common character; arbitrary use of different sequence for <i>Holam Haser</i> in the context of VAV

A3a	<i>Holam Male</i> = <VAV, HOLAM, ZWNJ>	<i>Holam Male</i> = <i>Vav Haluma</i>		ZWNJ used arbitrarily with semantic significance; long sequence for a common character
A3b	<i>Vav Haluma</i> = <VAV, HOLAM, ZWJ>	<i>Holam Male</i> = <i>Vav Haluma</i>	Long sequence only for a rare combination; minimal incompatibility between existing and new texts	ZWJ used arbitrarily with semantic significance; arbitrary use of different sequence for <i>Holam Haser</i> in the context of VAV
A3c	<i>Holam Male</i> = <VAV, HOLAM, ZWNJ> and <i>Vav Haluma</i> = <VAV, HOLAM, ZWJ>	<i>Holam Male</i> = <i>Vav Haluma</i>	Support for conjectured three-way HOLAM positioning distinction	ZWJ and ZWNJ used arbitrarily with semantic significance; long sequence for a common character; arbitrary use of different sequence for <i>Holam Haser</i> in the context of VAV
A4a	<i>Holam Male</i> = <ZWJ, VAV, HOLAM>	<i>Holam Male</i> = <i>Vav Haluma</i>	Use of ZWJ corresponds to logical structure of script	ZWJ used with semantic significance; long sequence for a common character
A4b	<i>Vav Haluma</i> = <ZWNJ, VAV, HOLAM>	<i>Holam Male</i> = <i>Vav Haluma</i>	Long sequence only for a rare combination; minimal incompatibility between existing and new texts	ZWNJ used arbitrarily with semantic significance; arbitrary use of different sequence for <i>Holam Haser</i> in the context of VAV
A4c	<i>Holam Male</i> = <ZWJ, VAV, HOLAM> and <i>Vav Haluma</i> = <ZWNJ, VAV, HOLAM>	<i>Holam Male</i> = <i>Vav Haluma</i>	Support for conjectured three-way HOLAM positioning distinction	ZWJ and ZWNJ used arbitrarily with semantic significance; long sequence for a common character; arbitrary use of different sequence for <i>Holam Haser</i> in the context of VAV
B1	<i>Holam Male</i> = <HOLAM, VAV>	<i>Holam Male</i> = <i>Holam Haser</i> + VAV	Best fit to the logical structure of Hebrew script; doesn't use ZWJ, ZWNJ or variation sequence; existing implementations and texts	Most complex implementation; difficulties with unusual combinations e.g. the divine name; difficulties with canonical reordering
B2a	<i>Holam Male</i> = <ZWNJ, HOLAM, VAV>	<i>Holam Male</i> = <i>Holam Haser</i> + VAV	Best fit to the logical structure of Hebrew script; implementation much easier than Option B1	ZWNJ used within combining character sequence, but with only graphical significance; long sequence for a common character
B2b	<i>Holam Haser</i> + VAV = <ZWJ, HOLAM, VAV>	<i>Holam Male</i> = <i>Holam Haser</i> + VAV	Implementation much easier than Option B1; long sequence only for a rare combination	ZWJ used within combining character sequence, but with only graphical significance; arbitrary use of different sequence for <i>Holam Haser</i> in the context of VAV

B2c	<i>Holam Male</i> = <ZWNJ, HOLAM, VAV> and <i>Holam Haser</i> + VAV = <ZWJ, HOLAM, VAV>	<i>Holam Male</i> = <i>Holam Haser</i> + VAV	Implementation much easier than Option B1	ZWJ and ZWNJ used within combining character sequence, but with only graphical significance; long sequence for a common character; arbitrary use of different sequence for <i>Holam Haser</i> in the context of VAV
B3a	<i>Holam Male</i> = <HOLAM, ZWJ, VAV>	<i>Holam Male</i> = <i>Holam Haser</i> + VAV	Best fit to the logical structure of Hebrew script; ZWJ used as defined in TUS 4.0.1; implementation easier than Option B1	Long sequence for a common character; difficulties with canonical reordering
B3b	<i>Holam Haser</i> + VAV = <HOLAM, ZWNJ, VAV>	<i>Holam Male</i> = <i>Holam Haser</i> + VAV	ZWNJ used as defined in TUS 4.0.1; implementation easier than Option B1; long sequence only for a rare combination	Difficulties with canonical reordering; arbitrary use of different sequence for <i>Holam Haser</i> in the context of VAV
B3c	<i>Holam Male</i> = <HOLAM, ZWJ, VAV> and <i>Holam Haser</i> + VAV = <HOLAM, ZWNJ, VAV>	<i>Holam Male</i> = <i>Holam Haser</i> + VAV	ZWJ and ZWNJ used as defined in TUS 4.0.1; implementation easier than Option B1	Difficulties with canonical reordering; long sequence for a common character; arbitrary use of different sequence for <i>Holam Haser</i> in the context of VAV
C1a	New character HOLAM MALE	<i>Holam Male</i> illegible	Doesn't use ZWJ, ZWNJ or variation sequence; simplest implementation	Bad fallback behaviour; unity of HOLAM lost; complicated conversion to and collation with unpointed text; serious incompatibility between existing and new texts
C1b	New character VAV HALUMA	<i>Vav Haluma</i> illegible	Doesn't use ZWJ, ZWNJ or variation sequence; simplest implementation; few characters affected by bad fallback behaviour	Unity of HOLAM lost; arbitrary use of different character for <i>Vav Haluma</i> ; complicated conversion to and collation with unpointed text
C1c	Two new characters HOLAM MALE and VAV HALUMA	All VAV with HOLAM combinations illegible	Doesn't use ZWJ, ZWNJ or variation sequence; simplest implementation; support for conjectured three-way HOLAM positioning distinction	Worst fallback behaviour; unity of HOLAM lost; arbitrary use of different character for <i>Vav Haluma</i> ; complicated conversion to and collation with unpointed text; unnecessary new character defined; serious incompatibility between existing and new texts

C2a	New character RIGHT HOLAM	<i>Holam Male</i> illegible	Doesn't use ZWJ, ZWNJ or variation sequence	Bad fallback behaviour; unity of HOLAM lost; serious incompatibility between existing and new texts
C2b	New character LEFT HOLAM	<i>Vav Haluma</i> illegible	Doesn't use ZWJ, ZWNJ or variation sequence; few characters affected by bad fallback behaviour	Unity of HOLAM lost; arbitrary use of different character for <i>Holam Haser</i> in the context of VAV
C2c	Two new characters RIGHT HOLAM and LEFT HOLAM	All VAV with HOLAM combinations illegible	Doesn't use ZWJ, ZWNJ or variation sequence; support for conjectured three-way HOLAM positioning distinction	Worst fallback behaviour; unity of HOLAM lost; arbitrary use of different character for <i>Holam Haser</i> in the context of VAV; unnecessary new character defined; serious incompatibility between existing and new texts
C3	New character HOLAM HASER	<i>Holam Haser</i> illegible	Doesn't use ZWJ, ZWNJ or variation sequence	Bad fallback behaviour; unity of HOLAM lost; serious incompatibility between existing and new texts
C4	New character VAV VOWEL	<i>Holam Male</i> illegible	Doesn't use ZWJ, ZWNJ or variation sequence	Bad fallback behaviour; complicated conversion to and collation with unpointed text; serious incompatibility between existing and new texts