

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

Doc Type: Working Group Document**Title: Disunification costs regarding HOLAM and VAV in Hebrew****Source: Michael Everson & Mark Shoulson****Status: Individual Contribution****Action: For consideration by JTC1/SC2/WG2 and UTC****Date: 2004-08-18**

In document N2840 we proposed to encode HEBREW POINT HOLAM HASER FOR VAV as a disunification solution to be distinguished from the existing HEBREW POINT HOLAM also used with the letter VAV. The present document elaborates on that proposal, in particular the reason for the choice of the new character we recommended. In N2840, we stated the following:

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.

It is apparent that further statistics will help understand the kinds of choices which could be made in terms of disunifying HEBREW POINT HOLAM from another character in order to make the particular distinction required. Returning to the *eBHS*, we note that the total number of *holams* in the text is 74,415. Of these, 46.63% represent *holam male*, that is, a vav with a dot above or slightly to the right of it. This means that 53.37% of the *holams* in the *eBHS* represent *holam haser* on VAV and on other letters.

While one problem has been how to represent the relevant distinctions in text, another has been to address the disunification costs. There are basically three disunification scenarios. Before discussing them, it should be noted that currently, most fonts draw HEBREW POINT HOLAM on the top left of the letters, with the exception of VAV, where the dot will be drawn above or above and slightly to the right. (Whether most fonts actually *do* this may be debatable; but it is probably reasonable to say that they *should*.) The glyph which cannot be distinguished by such fonts is that of VAV with *holam haser* on its top left.

1. **Retain HEBREW POINT HOLAM** for all instances of *holam male* (46.63% of all *holams*) and for all instances of *holam haser* on letters other than VAV (52.81% of all *holams*), and **add HEBREW POINT HOLAM HASER FOR VAV** (0.56% of all *holams*). Fonts which support this solution will draw HEBREW POINT HOLAM on the top left of the letters, with the exception of VAV, where the dot will be drawn above or above and slightly to the right; fonts will also draw HEBREW POINT HOLAM HASER FOR VAV on the top left of the VAV. Users who do not graphically distinguish *holam male* from VAV with *holam haser* do not have to do anything (need do nothing, except possibly choose fonts that accord to their tastes, when working with texts that use the new encoding). Users who do wish to make this distinction must only change 0.56% of the data. The use of HEBREW POINT HOLAM HASER FOR VAV with any other letter would be considered a spelling error. 421 of the VAVs with *holam* in the *eBHS* would need to be changed following this disunification.

2. **Retain HEBREW POINT HOLAM** for all instances of *holam haser* (53.37% of all *holams*), and **add HEBREW POINT HOLAM MALE DOT** (46.63% of all *holams*). Fonts which support this solution will draw HEBREW POINT HOLAM on the top left of all of the letters, with the exception of VAV, where the dot will be drawn above or above and slightly to the right; fonts will also draw HEBREW POINT HOLAM MALE DOT above or above and slightly to the right of the VAV. Users who do not graphically distinguish *holam male* from VAV with *holam haser* do not have to do anything (need do nothing, except possibly choose fonts that accord to their tastes, when working with texts that use the new encoding). Users who do wish to make this distinction must change 46.63% of the data. The use of HEBREW POINT HOLAM MALE DOT with any letter other than VAV would be considered a spelling error. 34,699 of the VAVs with *holam* in the *eBHS* would need to be changed following this disunification.

3. **Retain HEBREW POINT HOLAM** for all instances of *holam male* (46.63% of all *holams*), and **add HEBREW POINT HOLAM HASER** (53.37% of all *holams*). Fonts which support this solution will draw HEBREW POINT HOLAM above or above and slightly to the right of VAV, on the top left of the letters, with the exception of VAV, where the dot will be drawn above or above and slightly to the right; fonts will also draw HEBREW POINT HOLAM HASER on the top left of all of the letters. Users who do not graphically distinguish *holam male* from VAV with *holam haser* do not have to do anything (need do nothing, except possibly choose fonts that accord to their tastes, when working with texts that use the new encoding). Users who do wish to make this distinction must change 53.37% of the data in order to get 0.56% of it to be distinguished from *holam male*. 39,716 of the letters with *holam* in the *eBHS* would need to be changed following this disunification.

From a technical standpoint, all three solutions will permit *holam male* to be distinguished from VAV with *holam haser*. When we proposed HEBREW POINT HOLAM HASER FOR VAV, it was on the basis that people interested in this issue had insisted that the stability of existing data was a primary consideration. Our solution respects existing data; the other two solutions are far more expensive.

Option	% of holams to change	Holams in <i>eBHS</i> affected	% of points in <i>eBHS</i> affected
HHFV	0.56%	421	0.054%
HMD	46.63%	34,699	4.473%
HH	53.37%	39,716	5.119%
Total Hataf-Patah:		25,740	
Total Hataf-Qamats:		645	
Total Hataf-Segol:		4,483	
Total Hiriq:		109,656	
Total Holam:		74,415	
Total Qamats:		152,669	
Total Qubuts:		4,383	
Total Patah:		117,870	
Total Segol:		75,372	
Total Shewa:		151,316	(does not include hatafs)
Total Tsere:		59,270	
Total points in <i>eBHS</i>		775,819	
Total Consonants:		1,206,338	
Total cons. & pts. in <i>eBHS</i>		1,982,157	