SAH: Our first suggestion is Latin combining ur, U+1DD1. The Cyrillic titlo, U+0483, is a possibility if it is redefined to support Latin use. There would be even less possibility of conflict with titlo than with combining ur, and it should have broader font support. (Though it would still be a good idea to change the representative glyph of combining ur in the charts.) However, Denis comments that the shape of the titlo “would need to be different in Middle Vietnamese as it’s often too angular or too wavy and has the same hook turned on both ends instead of a more prominent one on the right.” A font that supports both Latin and Cyrillic is unlikely to accommodate a Cyrillic character to Vietnamese. Combining ur is more representative graphically. Another possibility is U+0342 COMBINING GREEK PERISPOMENI, but that is historically inappropriate.

We request an annotation, to the medievalist character U+1DD1 COMBINING UR ABOVE ⟨◌⟩, that it is the appropriate Unicode character for what is known in modern Vietnamese as dấu sóng or dấu lưỡi câu, a diacritic ⟨◌⟩ of Middle Vietnamese Latin orthography that marked a syllable-final nasal consonant. Since its description in Rhodes (1651), the diacritic has had the Latin name apex, from the Portuguese term for what is now called the til or tilde. The dấu sóng is distinct from the Middle and Modern Vietnamese tilde-shaped dấu ngã tone mark, ⟨◌̃⟩, which derives from the Greek perispomeni, ⟨◌͂⟩, not the tilde, nor is there any apparent connection to the Classical Latin apex, ⟨◌́⟩, which marked long vowels. In this document, we will call the dấu sóng nasalization mark ⟨◌⟩ the ‘apex,’ and both the dấu ngã tone mark ⟨◌̃⟩ and the Portuguese nasalization mark the ‘tilde.’

We also request a change of the representative glyph in the Unicode code chart to one that is closer to Latin sources (Figure 22 ff) and also not dissimilar to some typesetting of Middle Vietnamese (Figure 6).

In Classical Latin, the apex ⟨◌́⟩ marked long vowels (Figures 1–2) and was the ancestor of the acute. In modern typesetting it is generally replaced with the macron, which was borrowed from Greek poetic convention, though for graphic fidelity the acute is sometimes used in modern typesetting. We do not see need for a separate Unicode character for the Latin apex: a dedicated font should be able to handle its graphic peculiarities with the existing acute.

The Vietnamese apex marks a nasal consonant, and it may co-occur with the acute and tilde tone marks, dấu sắc and dấu ngã. Specifically, the Vietnamese apex transcribes a syllable-final [ŋm], an allophone of /ŋ/ found after rounded back vowels in the Hanoi area (thus ⟨ơ⟩ and ⟨ư⟩). It was
unified with orthographic ⟨ng⟩ in the mid-18th century with survivals as late as Binh (1822). The apex is of modern interest for reproducing and discussing seminal works of Vietnamese Latin orthography that almost any literate Vietnamese-speaker would know of. A modern description of the phonetics is given in Jacques (1998: 52), reproduced as Figure 11. De Rhodes’ contemporary description is reproduced as Figure 12.

Graphically, the Vietnamese apex resembles a tilde or macron with a curled right terminal, ⟨◌̍⟩ or ⟨◌̇⟩. The curled-tilde allograph ⟨◌̍⟩ is seen in some modern publications (Figure 6), though the curled-macron allograph ⟨◌̇⟩ is more common. In manuscript, the apex may be a simple curl joined directly to the letter o, resembling the horn on the letter ṭ but larger and contrasting with it (Figure 7). Indeed, in manuscript the apex may occur as an abbreviation for ng on the letters ṭ and u, for ⟨◌̍⟩ and ⟨◌̇⟩ (Figure 8 ff), though because ṭ and u are not rounded vowels, the apex in such cases is simply /ŋ/ rather than the [ŋ̃m] allophone of /ŋ/.

Some digital texts use custom fonts to repurpose assigned Unicode points for the apex, such as the character å that displays as ū in Bùi (2021) (Figure 17). Publications using typographic substitutions in non-custom fonts have been unsatisfactory. Nguyễn (1993), Nguyễn (1997) [no relation] and Jacques (1998), among others, use a tilde. This is unfortunate because both Middle and Modern Vietnamese orthography use the tilde as a tone mark, and this can co-occur with the apex in the same text and even on the same letter (Figure 3). An article by Lê & Nguyễn (2013) attempts a different substitution, but something went haywire during typesetting and the text was rendered gibberish (Figure 16).

Thanks to Denis Jacquesrye for research into the history of the Portuguese tilde, and to Harald Tveiten for research into the MUFI background of the Latin combining ur.

**Conflict with the Unicode tilde**

The Vietnamese apex dates to the Portuguese missionaries who devised the Vietnamese Latin alphabet in the early 1600s. Apex was the Latin term used at the time for the Portuguese tilde, as attested by dictionaries such as J. Cardoso (1592) Dictionarium Latino Lusitanicum, which equates Latin ‘apex’ with Portuguese ‘til da letra’ (tilde). The Portuguese tilde of the time was flatter than it is today (Figure 21), and was indistinguishable in form from the contemporary Vietnamese apex ⟨◌̍⟩. (The modern wavy shape of the tilde dates from the late 18th century.) Thus the til da letra, under its Latin name apex, was the source of the Vietnamese apex in form, name and sound value (Jacques 1998: 52). This creates a problem for Unicode, because U+0303 COMBINING TILDE is used for the Portuguese tilde, but in Vietnamese it is used for the unrelated tilde-shaped tone mark, dấu ngã, which derives instead from the Greek perispomeni ⟨◌̍⟩ that can have that shape. De Rhodes (1651: 8–10) called the dấu ngã ‘circumflexus’ (circumflex), a calque of the Greek perispomeni.
Vietnamese vowel mark found on â, ê, ơ that we now know as the ‘circumflex’ was called *accentus circumflexus latinorum* ‘Latin circumflex accent’ by de Rhodes.)

If Unicode had been based on 17th-century usage, the nasal Vietnamese *apex* would have been assigned U+0303 COMBINING TILDE along with its cognate in Portuguese orthography, and the tonic Vietnamese *circumflexus* would likely have been assigned to U+0342 COMBINING GREEK PERISPOMENI. Our problem is that it’s the *circumflexus* that was encoded as U+0303.

<table>
<thead>
<tr>
<th>Vietnamese mark</th>
<th>Encoding</th>
<th>Origin</th>
<th>Encoding</th>
</tr>
</thead>
<tbody>
<tr>
<td>dấu sóng, ⟨◌⟩ (apex)</td>
<td>(TBD)</td>
<td>Portuguese tilde, ⟨◌⃀⟩</td>
<td>U+0303</td>
</tr>
<tr>
<td>dấu ngã, ⟨◌⃀⟩ (circumflexus)</td>
<td>U+0303</td>
<td>Greek perispomeni, ⟨◌⃀⟩</td>
<td>U+0342</td>
</tr>
</tbody>
</table>

Table 1. Inconsistent use of U+0303 COMBINING TILDE for the Portuguese *til da letra / apex* and the unrelated Vietnamese dấu ngã / circumflexus ⟨◌⃀⟩, rather than for the tildes of both Portuguese and Vietnamese (dấu sóng / apex ⟨◌⃀⟩). Encoding the dấu sóng as the perispomeni (U+0342) would also be problematic, partly because that is the source of the dấu ngã but also because the perispomeni and tilde are shaped identically in many fonts.

The SAH has decided that creating a new character for the Vietnamese apex would be inadvisable, because doing so would call into question the identity of the 17th-century Portuguese ‘apex,’ which after all was the same character. It would thus be preferable for the UTC to declare some existing character to be the one to use for the Vietnamese apex, and to annotate it accordingly.

**Character to annotate for use as the Vietnamese apex**

The SAH is of the opinion that a distinct Unicode character for the Vietnamese apex is not justified, and that the the best solution would be to repurpose an existing character. Apart from the tilden itself and the Greek perispomeni, which is problematic for reasons mentioned in Table 1, the two most visually similar characters currently available in Unicode are the medievalist U+1DD1 COMBINING COMBINING UR ABOVE ⟨◌⟩ and U+0483 COMBINING CYRILLIC TITLO ⟨◌҃⟩.

With the acceptance of L2/06-027 by Unicode, the MUFI character M+F1CC COMBINING CURLY BAR ABOVE was conflated with M+F1C2 COMBINING ABBREVIATION MARK SUPERSCRIPT UR LEMNISkate FORM into the Unicode character U+1DD1 COMBINING UR ABOVE ⟨◌⟩ and U+0483 COMBINING CYRILLIC TITLO ⟨◌҃⟩.

With the acceptance of L2/06-027 by Unicode, the MUFI character M+F1CC COMBINING CURLY BAR ABOVE was conflated with M+F1C2 COMBINING ABBREVIATION MARK SUPERSCRIPT UR LEMNISkate FORM into the Unicode character U+1DD1 COMBINING UR ABOVE ⟨◌⟩ and U+0483 COMBINING CYRILLIC TITLO ⟨◌҃⟩. Janusz Bień (author of Bień 2021) comments on GitHub issue #11 (concerning Junicode font) that the lemniscate form of M+F1C2 should be deprecated. As noted in that GitHub discussion, the figures supplied by L2/06-027 match M+F1CC rather than M+F1C2. (L2/06-027 figures 60 and 61 are from Balbi (1460), which is shown at higher resolution in Figure 23 below. The other item claimed to justify U+1DD1 was figure 32, but the symbol doesn’t actually occur there.) It is therefore unfortunate that the representative glyph
in the Unicode chart is based on M+F1C2 rather than on M+F1CC, and in our opinion it should be replaced. M+F1CC as attested in historical Latin sources such as Zaborowski (1514) and Balbi (1460), as well as in modern fonts such as Junocode, closely approaches the desired Vietnamese shaping. The glyphs for M+F1CC in Junicode and Cardo fonts are:

![Glyphs for M+F1CC in Junicode and Cardo fonts](image)

The Cyrillic titlo, U+0483 COMBINING CYRILLIC TITLO ⟨⪸⟩, has the advantage of broader font support, and a negligible chance of conflict in a Vietnamese Latin context, but the disadvantage of being too angular and having a turn at both ends rather than only on the right as the Vietnamese apex does. Specialized fonts that would give it its Vietnamese shape would be unlikely, especially online, given how widely it is supported for a Cyrillic context. The glyphs for the titlo in Brill and Noto Serif fonts are:

![Glyphs for titlo in Brill and Noto Serif fonts](image)

Wikimedia wikis substitute the IPA pitch-contour character U+1DC4 COMBINING MACRON-ACUTE ⟨◌᷄⟩. This workaround is present on 43 pages across three books in the Vietnamese Wikisource, 11 entries in the English Wiktionary and six in the Vietnamese Wiktionary (Figure 20). It would be trivial to replace these instances with whichever character the UTC decides on, and there is as yet no known usage of U+1DC4 for the apex outside Wikimedia. One problem with Unicode adopting U+1DC4 as a permanent solution is that a text about Middle Vietnamese could well include IPA transcription, since Middle Vietnamese orthography differs nontrivially from the modern and makes distinctions that have since disappeared, and pitch contours such as the one encoded in U+1DC4 are an essential element of the tones that Vietnamese is famous for. Using an IPA tone diacritic for an unrelated orthographic convention could thus be a problem with forward compatibility. Another problem is that U+1DC4 is clearly identified as a ligature of a macron and an acute, and that identity – combined with its broad font support as IPA – makes it unlikely that it will be given the proper Vietnamese form in fonts. The glyph in Gentium font is:
We therefore request that the UTC annotate \texttt{U+1DD1 COMBINING COMBINING UR ABOVE} for use as the Vietnamese apex. If the UTC should decide instead on the Cyrillic \textit{titlo}, \texttt{U+0483}, it will require redefinition in \texttt{ScriptExtensions.txt} for use as a Latin character.

\textbf{Annotation}

\begin{enumerate}
\item \texttt{0303 COMBINING TILDE} \rightarrow \texttt{1DD1 combining ur above}
\item \texttt{1DD1 COMBINING UR ABOVE} \rightarrow \texttt{Vietnamese apex (đầu sóng)}
\end{enumerate}

\texttt{ScriptExtensions.txt} [if the \textit{titlo} is chosen]

\texttt{U+0483 COMBINING CYRILLIC TITLO} is defined only for Cyrillic script, with \texttt{scx = Cyrl Perm}. \texttt{Latn} will need to be added if it is to be used for the Vietnamese apex. (\texttt{Glag} should also be added, per GitHub issue \#375.)

\textbf{Representative glyph in the code chart}

We request that the glyph for \texttt{U+1DD1 COMBINING COMBINING UR ABOVE} in the Unicode chart for Combining Diacritical Marks Supplement (1DC0–1DFF) be changed to \textless {\textasciitilde} \textgreater{}. This is both more representative of the medieval Latin form (Figure 22 ff) than is the current glyph on the code chart, and more appropriate for Middle Vietnamese. We will be happy to provide the font.

\textbf{References}

Balbi, Giovanni (1460) \textit{Catholicon}. High-resolution scans available at: \url{www.loc.gov/resource/rbc0001.2012rosen0031/?st=gallery}


Bình, Philippê (1822). \textit{Sách số sang chép mọi việc} [Book recording all things]. Page images available online at Vietnamese Wikisource.

Bùi Thị Minh Thùy (2021) \textit{Từ điển VIỆT – BỘ – LA và các cự liệu liên quan.} [\textit{Dictionarium Annamiticum}}


——— (1651b) Phép giảng tám ngày cho kẻ muốn chịu phép rủa tội, ma šěào đạo thánh đức Chúa bối / Cathechismus pro ijs, qui volunt suscipere baptismum in octo dies divisus [Catechism for those who want to be baptized and enter the faith in eight days]. Sacrae Congr. de Propaganda Fide [Congregation for the Propagation of the Faith], Rome. High-resolution images available online at the Bayerische Staatsbibliothek.

Figures

Figure 1. Late 1st – early 2nd-century Latin inscription with apices, ⟨ האתר⟩. The diacritic is light-weight and has a serif at the end; the serif becomes a round hook in handwriting. The Latin apex is used on the letters A E O V but not I, which instead of taking the apex is substituted with i longa, U+A7FE ⟨இ⟩ (blue). The apex tends to bear to the right, but in handwriting it may attach to the left stroke in the case of the letter V. CIL XII 3203 in the Musée archéologique de Nîmes. Image from Wikimedia Commons.

Figure 2. In this Latin example the apex connects to the top-right corner of the letter E and overlies the following letter. ‘Pilate stone’ (cast), 1st century, Caesarea Maritima. Image from Wikimedia Commons.
Figure 3. de Rhodes (1651b: 14; 1651: 147). Some clear examples of the common word củ “also” that combines the Vietnamese apex and tilde.

Figure 4. de Rhodes (1651b: 14). The apex on củ, khoû and thú (red) and the horn on bloti and những (blue) appear to be typeset with the same element, but the apex lies above the letter whereas the horn connects to the side. Note also the recently unicodified letter b in bêy.
Figure 5. de Rhodes (1651: 170). Apex alongside other diacritics in dếũ and dếú. The apex here is similar in shape to the hooks in nước người (blue).

Figure 6. de Rhodes (1991 [1651]: 105). The Vietnamese apex entered by hand on the words giáo tlóú in a modern reprinting. In some tokens its origin in a Portuguese tilde is apparent.
Figure 7. Bình (1822: 2). A late manuscript with the apex in tơi củ chưng (together with a grave); ọn traọ Đức ; phá Đào D.C.J. ở traọ địa ; Saọ le trước ; Cụ Tito củ Cụ Bảo and various other tokens of Đào (red), where it joins with the letter o. It contrasts with the tilde in, for example, Vĩnh lại huyền, Ngãi (blue). Transliteration from Sach so sang chep cac viec.pdf/23 at Vietnamese Wikisource.
Figure 8. Bính (1822: TOC, 23). Manuscript tokens of ⟨ơ⟩, the apex on the horned letter ơ, in the word phương ‘direction.’ Phương is repeatedly abbreviated phươ in this document. Here the apex is a simple abbreviation for final ng and does not mark the allophonic value [ŋ̃m] that it would on o or u. The first context is Thầy Dào D.C.J. sang giảng Dao Phươń bac [page] 239, where the apex also appears on ⟨ơ⟩, and following an unusual token of phương spelled out in the preceding line. Note that the horns of the letters учитыва are detached and resemble a typeset apex. The second context is ngủ Phương Tây, short for người Phương Tây ‘Westerner.’

Figure 9. Bính (1822: 50, 36). Manuscript tokens of ⟨ư⟩, the apex on the horned letter ư. contrasting with the tilde (on ngữ) in vì chú ngữ kẻ, where vì chú is short for vì chúng ‘because of’; and in cho nó xư tỏi khi, where xư tỏi is short for xưng tỏi ‘to confess.’
Figure 10. de Rhodes (1991 [1651]: 11). Description of the diacritic in Vietnamese, with examples. A tilde is also used (blue).

Dans la version du quoc ngu dont est témoin le dictionnaire d’Alexandre de Rhodes, les consonnes nasales finales dorso-prépalatale, dorso-vélaire et labio-vélaire font l’objet de trois notations graphiques différentes, respectivement “nh” (par ex. “lành,” “bon, doux”); “ng” (par ex. “làng”, “village”), and a signe dénommé apex, dérivé du til (tilde) portugais “~”, placé au-dessus de la voyelle (par ex. “làõ”, aujourd’hui “lông”, “ventre, cœur”). Ces notations reflètent bien la prononciation standard de la région de Hà Nội, tel qu’un auditeur portugais attentif peut l’analyser dans le système de notation auquel il est habitué. Dans la version fixée par le dictionnaire de Pigneau de Béhaine et Taberd (cf. note 110), les deux derniers types sont notés uniformément “ng”, du fait qu’il ne sont pas en opposition, leur appariation dépendant de la voyelle précédente (dans le jargon de la phonologie, ils sont en « distribution complémentaire »). La distinction entre les deux est donc superflue pour un locuteur vietnamien. La simplification aurait pu aller jusqu’au bout, en confondant les notations “nh” et “ng” : leur distinction est peu rentable pour le système vietnamien, voire sans utilité si on note différemment les voyelles /a/ et /æ/.

Figure 11. Jacques (1998: 52). Background information on the Vietnamese apex:
In the version of Quoc Ngữ that the dictionary of Alexander de Rhodes witnesses, the dorso-prepalatal, dorso-velar and labio-velar final nasal consonants are the target of three different spellings, respectively “nh” (for example “lành,” “good, sweet”); “ng” (for example “làng,” “village”), and a symbol called apex, derived from the Portuguese til (tilde) “~”, placed above the vowel (for example “lãõ,” today “lòng,” “belly, heart”). These spellings nicely reflect the standard pronunciation of the region of Hanoi, so that an attentive Portuguese listener can interpret them per the orthographic system that he is accustomed to. In the version established by the dictionary of Pigneau de Béhaine and Taberd (see note 110), the last two are uniformly transcribed “ng”, due to the fact that they are not in opposition, their appearance depending on the preceding vowel (in the jargon of phonology, they are in “complementary distribution”). A distinction between the two is therefore not necessary for a Vietnamese-speaker. The reduction could have been completed by conflating the spellings “nh” and “ng”: their differentiation is not helpful for the Vietnamese [orthographic] system, and would be unnecessary if we wrote the vowels /a/ and /æ/ differently.
Finally, the third sign is the apex, which is absolutely necessary in this language because of the difference in the ending that the apex makes, which is entirely distinct from the ending that m or n makes, with a completely different meaning in the words in which it is used. However, this sign, namely the apex, only affects ō and ū at the end of a word, as in aō, ‘bee,’ oū, ‘grandfather’ or ‘lord.’ It is pronounced such that neither the lips touch each other nor the tongue touches the palate.

Figure 13. Nguyễn (1993: 39). The apex (substituted with a tilde) on Middle Vietnamese aō (modern aong, ong), cū (cūng), maō (maóng, móng), oū (oung, ōng), sóū (soúng, sŏng), tô tū (tô tung, tô tông). An orthographic tilde occurs on Ngũ and cūng (blue), so the substitution of tilde for apex conflates the diacritics on cū and cūng.
Middle Vietnamese [Gregerson 1969], as recorded in the trilingual dictionary by Alexandre de Rhodes (Rome, 1651), displayed the above distinctions through the use of vowel markers and tone markers. However, modern ń was written by means of ơ, ũ --- ơác, ũác for vác; ơẹai, ũai for vai. Ông was written ông; trống was written trăng; học was written haoc, haopr; and cường was written cuang; etc.

Figure 14. Nguyễn (1997: 33). A well-cited linguistics book on Vietnamese that also substitutes a simple tilde for the apex.

<table>
<thead>
<tr>
<th>name</th>
<th>code</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>la</td>
<td>Cocincina</td>
<td></td>
</tr>
<tr>
<td>lzh</td>
<td>Nội</td>
<td></td>
</tr>
<tr>
<td>mkh-mvi</td>
<td>dăng tlaõ</td>
<td></td>
</tr>
<tr>
<td>nl</td>
<td>Quinan</td>
<td></td>
</tr>
<tr>
<td>pt</td>
<td>Cochinina</td>
<td></td>
</tr>
<tr>
<td>vi</td>
<td>Đàng Trong</td>
<td></td>
</tr>
<tr>
<td>vi-Hani</td>
<td>唐冲</td>
<td></td>
</tr>
</tbody>
</table>

Figure 15. The entry for Cochinchina (2753942) at OpenHistoricalMap.org. The Middle Vietnamese name is given as ⟨dàng tlaõ⟩, based on sources such as Đỗ (2008) and Thụy (2023). The only indication to a naive reader that the tilde in tlaõ is not the tilde diacritic that marks tone is that it appears over the ơ rather than the a; the other clue is that it corresponds to -ng in modern orthography, which is ⟨Đàng Trong⟩ (blue).
Figure 16. Lê & Nguyễn (2013: 4). A modern attempt to reproduce the apex, approximated as a dấu ngắn (‘’), in a Vietnamese-language text without Unicode support. It comes out variously as && and (.)

Figure 17. Bùi (2021: 21). Typesetting of the apex (red), in the same text as the tilde (blue), in the publisher’s preview of a recent publication. Copying and pasting the word trư from the PDF, we see it is encoded as ⟨trư⟩.
Figure 18. Bùi (2021: 76–77). Although the resolution is a bit low, we can see the apex on the words **blở laoğ**, **đaød boi dau** and **tũ böň**, along with the tilde on **ngũ** and the recently unicodified Middle Vietnamese letter `<ŋ>`.

Figure 19. Đỗ (2008 [1972]: 74). Substitution of the apex with a tilde (**dàng tlaõ**, red) in a text that also uses the tilde as a tone mark (**nghĩa**, blue).
Figure 20. Entries for xaõ on English Wiktionary and for lào dư on Vietnamese Wiktionary, with U+1DC4 ◌᷄ COMBINING MACRON-ACUTE substituted for the apex.

Figure 21. Flat tildes in 17th- and 18th-century Portuguese publication:
(a) nação in D. Nunes de Leão (1610) Descripção do Reino de Portugal.
(b) defensaõ in G. Estaço (1625) Varias Antiquidades de Portugal.
(c) embarçaõ and mençaõ in L. de Menezes (1759) Historia De Portugal Restaurado, vol. 2.
These are the apex ⟨⟩ of Middle Vietnamese.
Figure 22. Bień (2021: 34–35). Latin combining ur in accipit\textsuperscript{ur} from the 16th-century Orthographia seu modus recte scribendi et legendi Polonicum idioma quam utilissimus (Stanisław Zaborowski, 1514). The form is close to the Vietnamese apex seen in Figure 6.

Figure 23. Balbi (1460: 1–2 [of text; p. 9–10 of LOC pagination]). Baldi is a primary source for combining ur in L2/06-027, which requested it. The form is close to the Vietnamese apex seen in Figure 6.