

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 encode Linguistic Doubt Marks in the UCS**

**Source: Abteilung für Griechische und Lateinische Philologie der Ludwig-Maximilians-Universität München (Department of Greek and Latin Philology, Ludwig-Maximilians-University of Munich, Germany)**

**Authors: Martin Schrage, Karl Pentzlin**

**Status: Expert Contribution**

**Action: For consideration by JTC1/SC2/WG2 and UTC**

**Date: 2011-10-21**

*This document is based on an excerpt of WG2 N3913 and L2/10-358R (Proposal to encode Metrical Symbols and related characters), as it was decided to split that work and to propose the "related characters" separately by subject.*

## 1. Introduction

In linguistic texts, there is often the request to mark the reading or interpretation of a specific character (or the sound it denotes, or any other property of it which is discussed in the text) as doubtful, either if the author doubts the reading or interpretation itself, or to mark that the author is uncertain regarding the discussed property (e.g. the pronunciation of a given grapheme). The characters marked doubtful may be letters as well as e.g. metrical symbols.

This is often expressed by a combining question mark. This can be placed above or below the affected letter, depending of typographical considerations (which in turn may depend whether the text within it occurs contains high or low modifier letters to be marked).

Therefore, both versions of the combining question mark (above and below) are proposed here.

In theory, they could be considered as glyph variant of the same underlying character. However, there is no precedent of a combining character which has no fixed placement relative to the base letter, and especially there is no combining class indicating such a placement variation. Introducing such a combining class also would mean to extend the combining rules specified in Unicode as such.

Therefore, it is appropriate to simply propose two characters here, with existing and proven combining classes.

In addition, also a free-standing superscript question mark is used.

This character especially is used in critical apparatuses. There, it is used in contrast with the common question mark, to mark different doubts on the reading (see also the detailed explanations in the figure legends of fig.1992a-VIII and 2001a-50):

- The ordinary question mark indicates doubt on the reading on the character, while the fact that the character was corrected in the source is not subject of the doubt.
- The superscript question mark indicates doubt on the properties of the indicated correction, while the reading of the character itself is not subject of the doubt.

Also, the superscript question mark is found within texts or sequences of metrical symbols themselves in contrast to the ordinary question mark, to denote similarly different scopes of doubt.

## 2. Proposed Characters

Annotations in parentheses address special issues for a character, or reference to figures where such special issues are discussed. (These annotations are not intended to be retained in the character list when copied into the standard.)

### ***Block: Combining Diacritical Mark Supplement***

#### **Combining Marks for linguistic use**

◌<sup>?</sup>

U+1DF5 COMBINING QUESTION MARK ABOVE  
= combining doubt mark (linguistic and metrical)  
(see fig. 1982a-75, 1989a-35, 1989a-122)

◌<sub>?</sub>

U+1DF6 COMBINING QUESTION MARK BELOW  
= alternative combining doubt mark  
(see fig. 1896a-109)

### ***Block: Superscripts and Subscripts***

<sup>?</sup>

U+209D SUPERSCRIPT QUESTION MARK  
≈ <super> 003F  
= doubt mark (linguistic and metrical)  
(see fig. 1989a-122, 1992a-VIII, 1992a-42)

#### **Properties:**

U+1DF5 COMBINING QUESTION MARK ABOVE; Mn; 230; NSM; ; ; ; ; N; ; ; ;  
U+1DF6 COMBINING QUESTION MARK BELOW; Mn; 220; NSM; ; ; ; ; N; ; ; ;  
U+209D SUPERSCRIPT QUESTION MARK; So; 0; ON; <super> 003F; ; ; ; N; ; ; ;

## 3. References

- [1896a] Thomsen, Vilh. – Inscriptions de l'Orkhon – Helsinki 1896
- [1982a] West, M. L. – Greek Metre - Oxford 1982 - ISBN 0-19-814018-5
- [1989a] Mahler, Hervicus (ed.) - Pindari Carmina cum Fragmentis, vol. II - Leipzig 1989 - ISBN 3-322-00 673-5
- [1992a] West, Martin L. - Aeschlyi Septem contra Thebas - Stuttgart 1992 - ISBN 3-519-01019-4
- [1993a] Mastronade, Donald J. – Euripides Phoenissae. Edited with introduction and commentary. – Cambridge 1993, ISBN 0 521 41071 1
- [1998a] West, Martin L. - Homeri Ilias., vol.I: rhapsodia I-XII continens - Stuttgart+Leipzig 1998 - ISBN 3-519-014301-9
- [2001a] Hutchinson, G. O. - Greek Lyric Poetry - New York 2001 - ISBN 0-19-924017-5



**Fig. 1982a-75:** Showing a specimen for COMBINING QUESTION MARK ABOVE applied to a metrical pause symbol.

It is interesting to note a certain parallelism between the development of the strophes in *O.* 12 (466?) and *P.* 4 (462):

<p><i>O.</i> 12   <math>e-D \parallel</math>  <math>e-D \mid -d \parallel E^2 \mid E-d \parallel</math>  <math>e-D- \mid E \mid</math>  <math>-D-E \parallel \parallel</math></p>	<p><i>P.</i> 4   <math>e-D \parallel</math>  <math>e-D \mid - \mid e-D \parallel</math>  <math>e-D-E- \parallel</math>  <math>D^2 \times e- \parallel D-E \parallel</math>  <math>E-D^2 \parallel E^2 e \parallel E- \parallel \parallel</math></p>
---	---

**Fig. 1982a-102:** Showing a specimen for COMBINING QUESTION MARK ABOVE applied to a metrical symbol.

DRAMA

102  
*-D-* (*OT* 196), 4*da* (*Tr.* 214, *Ant.* 879, *OC* 540). Only in *OC* 540 does the dactylic colon occupy the penultimate place.  
 The most unusual of Sophocles' iambic strophes is the epode *El.* 504-15:

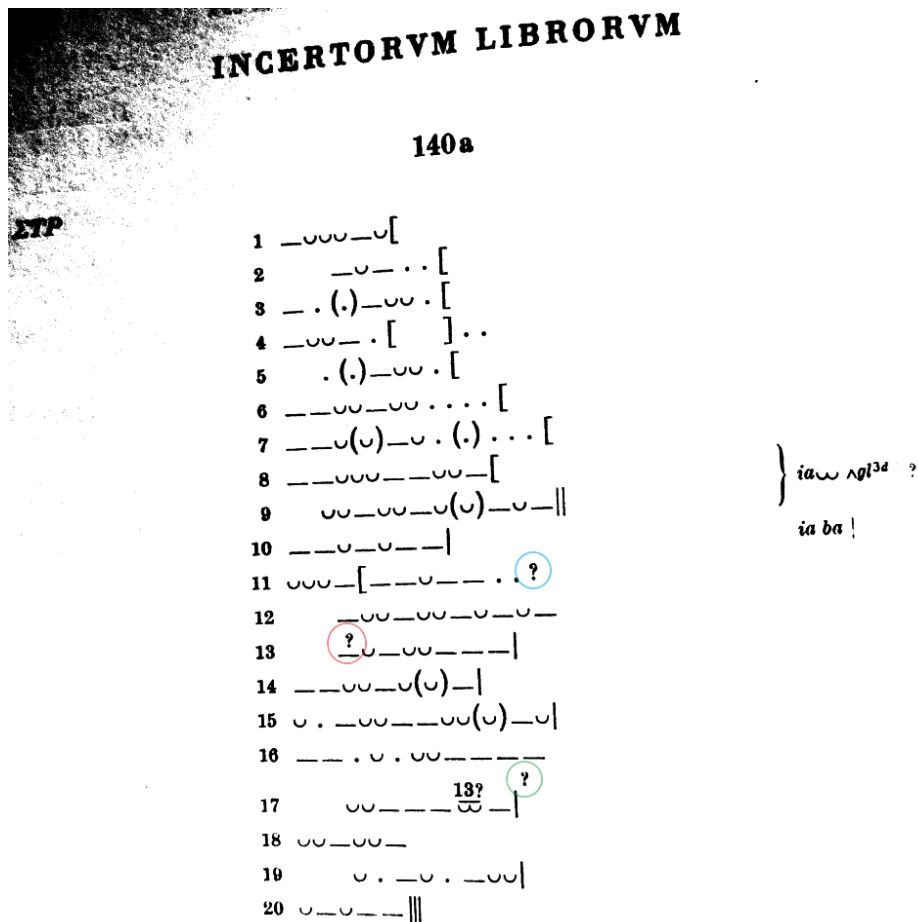
<p>ὦ Πέλοπος ἀ πρόσθεν          505 πολύπονος ἱππεΐα·          ὥς ἔμολες αἰανῆς ταῖδε γαῖ.          εὔτε γὰρ ὁ ποντισθεὶς          Μυρτίλος ἐκοιμάθη,          510 παγχρύσων δίφρων          δυστάνοις αἰκείαις          πρόρριζος ἐκρίφθεις, οὗ τί πω</p>	<p>---           ---           ---           ---           ---           ---           ---           ---           ---           --- </p>	<p><math>ia \wedge ia \wedge</math>  <math>ia \wedge ia \wedge</math>  <math>ia \wedge ia \wedge \wedge ia</math>  <math>ia \wedge ia \wedge</math>  <math>ia \wedge ia \wedge</math>  <math>\delta</math>  <math>(k\delta)</math>  <math>ia \wedge ia \wedge \wedge ia</math></p>
--	---	--

**Fig. 1989a-35:** Showing a specimen for COMBINING QUESTION MARK ABOVE applied to a metrical symbol.

VIIa = fr. 52g (A)

-6    $\cup \cup [$   
       $- - [$   
       $\cdot - [$   
 -3    $\cup \cup [$   
       $\cup - [$   
       $\cup \cup \cup - . [ ? ] \parallel \parallel$

**Fig. 1989a-122:** Showing specimens for COMBINING QUESTION MARK ABOVE (red) and SUPERSCRIPT QUESTION MARK (green), in contrast to a common question mark (blue).



metrum: responsiones perspexit Sn.; aut corruptela aut alia colometria vid. fuisse  
 in v. 13

**Fig. 1992a-VIII:** *Showing specimens for SUPERSCRIPT QUESTION MARK (red), together with an ordinary question mark, explaining the different meaning of those characters when used in a critical apparatus. Right, an enlarged excerpt containing the question marks is shown. The abbreviations read:*

A<sup>a?</sup> – A, fortasse ante correctionem – manuscript A, the indicated reading is presumed to be the one before the correction (i.e. it is doubted that the otherwise undoubted identity of the character is in fact the one before the correction)

A<sup>a?</sup> – fortasse A, ante correctionem – the presumed reading in manuscript A, as it was before the correction (i.e. the identity of the corrected character itself is doubted)

## VIII INDEX SIGLORVM

### SIGLA CETERA

Ω	lectio omnium codicum vel utique archetypi
Σ	scholium vetus (Σ <sup>λ</sup> scholii lemma; Σ <sub>I</sub> scholium in cod. I)
Σ*	scholium incertae aetatis
Φ	commentarius paraphrasticus in triadem, saec. fere xii
Θ	Thomae Magistri textus
T <sup>σ</sup>	Triclinius in scholio suo
A <sup>1</sup> , A <sup>2</sup>	A a prima /ab altera manu correctus
A <sup>a</sup> , A <sup>c</sup>	A ante/post correctionem
A <sup>a?</sup> , A <sup>a?</sup>	A, fort. a.c./fort. A a.c.
A <sup>a2</sup> , A <sup>as</sup>	A ante correctionem ab altera manu/a scholiorum scriba factam
A <sup>x</sup>	A ante vel post correctionem (incertum utrum)
A <sup>gl</sup>	glossema in A, vel lectio quasi glossema adscripta
A <sup>γq</sup>	varia lectio cum γq. in A adscripta
A <sup>mg</sup>	A in margine
A <sup>r</sup>	A in rasura
A <sup>s</sup>	A tum correctus cum scholia addita sunt
A <sup>ss</sup>	lectio in A super lineam scripta
A <sup>t</sup>	A in textu, altera lectione inter lineas vel in margine adscripta
A <sup>ut</sup>	A ut videtur

A<sup>a?</sup>, A<sup>a?</sup>

**Fig. 1992a-42:** Showing SUPERSCRIPT QUESTION MARK (red) in contrast to the ordinary question mark (green).

719 ἐκφύγοις M<sup>t</sup>: -οι fere cett. 720 ἔφριξα B<sup>ve</sup>, πέφριξα H<sup>t</sup>  
 ὠλεσί- MBO λ: οὐλεσί- H: ὀλεσι- cett. 721 post h.v. sch. ã  
 γὰρ νύκτωρ παρεκελεύσατο καὶ γέγονε in textu habent codd.  
 praeter MKT (in mg. K) 722 παναληθῆ Ω: corr. Dind.<sup>5</sup> (cf.  
 946) 725 βλαψίφρονος (-ας H<sup>a1</sup> N<sup>ss</sup> +) οἰδιπόδα (-ποδος V,  
 -πόδαο HB<sup>c2</sup> Y) Ω: corr. T 726 δ' om. I<sup>t</sup> V 727 κλήρους  
 M<sup>a2</sup> Θξ: -οις cett. 728 σκυθῶν Ω: corr. Dind.<sup>12</sup> 732 καὶ] ἄν  
 καὶ WεQ<sup>c</sup> λ φθιμένοισι Ω (-σιν T): φθιμένους (Bourdelot)  
 <ἄν> Blomf.<sup>4</sup>, <ἐγ>κατέχειν Headlam<sup>5</sup> 57 734 αὐτοκτόνως  
 κλ: αὐτόκτονοι Fb<sup>c</sup>: αὐτοκτόνωνσιν M<sup>a</sup>: αὐτοκτάνωσιν M<sup>s</sup> I:  
 αὐτοὶ κτάνωσι(v) cett. 736 γαῖα Dind.<sup>4</sup> 640: χθονία Ω: νεο-  
 τέρα Weil<sup>7</sup>: ἀγχωρία Newman<sup>1</sup> 56

**Fig. 1993a-556:** Showing SUPERSCRIPT QUESTION MARK (red).

1534	—υυ—υυ—υυ—υυ	4 da
1535	—ϣ— — —υ— —υ—    <sup>?</sup>	2 dochm
1536	υ— —υ— —	2 ba



**Fig. 1998a-156/157:** Showing *SUPERSCRIPT QUESTION MARK* (red) in a critical apparatus (lower picture from p.157), in contrast to an ordinary question mark in the same apparatus (blue; upper picture from p.156).

The third picture shows the same *SUPERSCRIPT QUESTION MARK* from p.157 by a higher resolution.

394 καί Ar<sup>a</sup> 1008 b F<sup>c</sup>: κέν Ar<sup>b</sup> 474c t Z C<sup>yp</sup> Ω\*, κέ F<sup>a</sup> μιν Ar Z C<sup>yp</sup> Ω\*: περ 404 1008  
t b F 396 ωντὸς 474c 1008 h126 tt Z Ω: αὐτὸς Ernesti (cf. M 225, Ψ 480): οὗτος Wack.  
KS 1587 397 πύλῳι tamquam ('Αίδου) πύλῃι interpr. Ar 398–402 damn. Koechly:  
hab. 16 400 1008 Z Ω 399 κῆρ h126 D B F R: κῆρ' 16 Z Ω\* 400 ἐνι Barnes: ἐνι (vel.  
ἐνι) Hdn Ω 401 ἐπὶ (vel ἐπι) Hdn Ω: ἐπι D<sup>a</sup>? 402 γ' ἐτέτυκτο Z Ω: γε τέτ- O 403  
αἰσυλοεργός Ar tt\* (ad figuram cf. E 63, I 124, M 295 sq., P 5, α 299 sq.): ὀβριμοεργός 1008  
Eudoc. Z Ω ὅ τ' Ar b<sup>s</sup>: ὅς 400 tt\* Z Ω 404 damn. Heyne: hab. 16 400 1015 Z  
Ω 405 ἐπὶ Hdn Z Ω: ἐπι rr 406 Τυδέος υἱός 16 Ω: καὶ κατὰ θυμόν 474:  
Eust. 407 μάχεται 1 Ω\*: -οίτο t b W G 408 τί Ar<sup>a</sup> 1 w5 t Ω: τέ Ar<sup>b</sup> παππάζουσιν  
1 tt Z Ω: [πα]παζ- h126: παμπαζ- w5 410 τὼ D<sup>a</sup> F<sup>x</sup>: τῶ(ι) 1 Ω

415 post 414 ferunt 1 Z Ω, ante eum 400: post 412 trai. Cobet Misc. Crit. 369: noverat  
Zen quoquo loco: del. Christ 416 ἰχῶ sch 1 t B E T<sup>s</sup> W<sup>c</sup>: ἰχῶρ R W<sup>a</sup> G<sup>a</sup>, -ὦρ' Z F<sup>c</sup> T,  
-ὦρ C D F<sup>a</sup> G<sup>c</sup> χειρὸς Ar 1 Ω: χερσὶν Zen (cf. Σ 123) 417 χεῖρ tt Ω\*: χεῖρ D, χεῖρ'  
T<sup>a</sup> R<sup>a</sup> 422 ἀνιείσθαι 400 tt Z Ω: ενιεισα 1.-cf. ad I 700 423 ἄμα σπέσθαι Plut. Z Ω:  
ἄμ' ἔσπεσθαι 1 (αμ' ἐσπέσθαι) t\* rr.-cf. Praef. III 1 ἔκπαγλ' ἐφίλησε 1 400 t\* Ω: ἔκπα-  
γλα φίλ- Plut. b.-cf. ad Γ 415 424 εὐπέπλων 400 D T R, εὐπ- 1 h126 tt\* Z Ω\*: βα-  
θυκόλπων Plut. 425 χρυσεῖ 400: χρυσῇ 1 tt Z (-ση) Ω καταμύξατο Ar C<sup>s</sup> T R<sup>c</sup>:  
κατεμ- (nov. Did) 1 400 tt Z Ω\* ἀραιήν Ar Hdn Z B F<sup>c</sup>: ἀρ- 1 tt Ω\*: βαρεῖαν W 427  
χρυσῇ 1 16 1016 h126 Ω (-σην D): -σέην Barnes.-cf. Praef. III 10 428 οὐ τοι τέκνον  
ἐμόν Zen (οὐ τοί) 1 400 1017 tt\* Ω: τέκνον ἐμόν, οὐ τοι Cic. Plut. Zenodoto contradi-  
xit Hdn post Ar, v. Epm. 433 γινώσκων Ω: γιγν- h126 434 ἄλλ' ὃ γ' ἄρ W O: ἄλλ'

Ar C<sup>s</sup> T R<sup>c</sup>?



**Fig. 2001a-50:** Showing SUPERScript QUESTION MARK (red).

Here (for Bacchylides 17: carmen [canto] 3, line 9 to 14 shown in the excerpt), a manuscript "A" is considered which had been corrected by two scribes.

Thus, the critical apparatus remark for line 14 following the colon reads:

φαρθι in manuscript A, »<sup>2</sup>« *ab altera manu correctus* (corrected by second hand) »<sup>?</sup>« *fortasse* (perhaps) »<sup>ac</sup>« *ante correctionem* (before correction): φαρθιν

– i.e.: there is a correction to φαρθι in the manuscript A, possibly by the second scribe, but it is doubtful that this in fact was done by the second scribe.

ep. i

θρόῃσε δὲ λ[αὸς  
 10 ἄ τρις εὐδαίμ[ων  
 col. vii ὄς, παρὰ Ζηνὸς λαχὼν  
 πλείσταρχον Ἑλλάνων γέρας,  
 οἶδε πυργωθέντα πλοῦτον μὴ μελαμ-  
 φαρεῖ κρύπτειν σκότῳ.

**17.** inscr. (in marg.) Ἰέρωνι Κυρακοσίῳ ἵπποις [Ὀλύ]μπια 1 (κάρ) 2 (ἰο-  
 στεφανόν) 3 (κύ) Κλειοῖ Blass: Κλειοῖ A (ᾱς) 4 Ὀ-: prima syllaba  
 uersus quarti strophae antistrophaeque in papyro semper tertium potius uersum  
 terminat (δρό) (ἰε) 5 *ceύον]το* e.p. (Kenyon), *φέρον]το* Platt, al., *ἰεν]το*  
 Edmonds, quorum uerborum nullum spatio bene conuenit (όχ, acc. tum deleto,  
 ut uid.) (καῖ) Νίκαι scribunt 6 Ἀγ]λαῖαι scribunt (ῖαι) (δίν) 7 τόθι]  
 Palmer (μέν) 8 *τρέκος* Edmonds, *στεφάνων* e.p. (σαι·) 9 [ ἀπείρων |  
 Blass 10 (ά) [ἀνήρ | e.p. 11 (ό) 12 (πλεί) γέρας A<sup>1pc</sup> (corr.  
 i.l.): γένος A<sup>1ac</sup> 13 μελαμ- A<sup>1?pc</sup> (λη in αμ corr.): μελλη A<sup>1ac</sup> 14 φαρεῖ e.p.,  
 tamquam lectionem A<sup>pc</sup>: φαρθι A<sup>2?pc</sup>: φαρθιν A<sup>1ac</sup> (τωι·)

**ISO/IEC JTC 1/SC 2/WG 2  
PROPOSAL SUMMARY FORM TO ACCOMPANY SUBMISSIONS  
FOR ADDITIONS TO THE REPERTOIRE OF ISO/IEC 10646<sup>1</sup>**

Please fill all the sections A, B and C below.

Please read Principles and Procedures Document (P & P) from <http://www.dkuug.dk/JTC1/SC2/WG2/docs/principles.html> for guidelines and details before filling this form.

Please ensure you are using the latest Form from <http://www.dkuug.dk/JTC1/SC2/WG2/docs/summaryform.html>.

See also <http://www.dkuug.dk/JTC1/SC2/WG2/docs/roadmaps.html> for latest Roadmaps.

**A. Administrative**

1. Title:	<b>Proposal to encode Linguistic Doubt Marks in the UCS</b>	
2. Requester's name:	<i>Martin Schrage; Karl Pentzlin</i>	
3. Requester type (Member body/Liaison/Individual contribution):	<i>Expert Contribution</i>	
4. Submission date:	<i>2011-10-21</i>	
5. Requester's reference (if applicable):	<i>University of Munich, Germany (M. S.)</i>	
6. Choose one of the following:		
This is a complete proposal:	<i>Yes</i>	
(or) More information will be provided later:		

**B. Technical – General**

1. Choose one of the following:		
a. This proposal is for a new script (set of characters):	<i>No</i>	
Proposed name of script:		
b. The proposal is for addition of character(s) to an existing block:	<i>Yes</i>	
Name of the existing block:	<i>Combining Diacritical Marks Supplement; Superscripts and Subscripts</i>	
2. Number of characters in proposal:	<i>3</i>	
3. Proposed category (select one from below - see section 2.2 of P&P document):		
A-Contemporary	<i>X</i>	B.1-Specialized (small collection)
C-Major extinct		B.2-Specialized (large collection)
D-Attested extinct		E-Minor extinct
F-Archaic Hieroglyphic or Ideographic		G-Obscure or questionable usage symbols
4. Is a repertoire including character names provided?	<i>Yes</i>	
a. If YES, are the names in accordance with the "character naming guidelines" in Annex L of P&P document?	<i>Yes</i>	
b. Are the character shapes attached in a legible form suitable for review?	<i>Yes</i>	
5. Fonts related:		
a. Who will provide the appropriate computerized font to the Project Editor of 10646 for publishing the standard?	<i>The authors (if requested)</i>	
b. Identify the party granting a license for use of the font by the editors (include address, e-mail, ftp-site, etc.):	<i>The authors (if requested)</i>	
6. References:		
a. Are references (to other character sets, dictionaries, descriptive texts etc.) provided?	<i>Yes</i>	
b. Are published examples of use (such as samples from newspapers, magazines, or other sources) of proposed characters attached?	<i>Yes</i>	
7. Special encoding issues:		
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)?	<i>No</i>	

**8. Additional Information:**

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 <http://www.unicode.org/Public/UNIDATA/UCD.html> and associated Unicode Technical Reports for information needed for consideration by the Unicode Technical Committee for inclusion in the Unicode Standard.

<sup>1</sup> Form number: N3702-F (Original 1994-10-14; Revised 1995-01, 1995-04, 1996-04, 1996-08, 1999-03, 2001-05, 2001-09, 2003-11, 2005-01, 2005-09, 2005-10, 2007-03, 2008-05, 2009-11)

### C. Technical - Justification

1. Has this proposal for addition of character(s) been submitted before?	Yes
If YES explain	<i>They are contained in WG2 N3913 = L2/10-358R and are separated here from its revision</i>
2. 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
If YES, with whom?	<i>One of the authors (M. S.) is a member of the scientific community himself</i>
If YES, available relevant documents:	<i>See text</i>
3. Information on the user community for the proposed characters (for example: size, demographics, information technology use, or publishing use) is included?	Yes
Reference:	<i>See text</i>
4. The context of use for the proposed characters (type of use; common or rare)	Common scientific
Reference:	<i>See text</i>
5. Are the proposed characters in current use by the user community?	Yes
If YES, where? Reference:	<i>See text</i>
6. After giving due considerations to the principles in the P&P document must the proposed characters be entirely in the BMP?	Yes
If YES, is a rationale provided?	Yes
If YES, reference:	<i>To keep them in line with related characters</i>
7. Should the proposed characters be kept together in a contiguous range (rather than being scattered)?	Yes
8. Can any of the proposed characters be considered a presentation form of an existing character or character sequence?	No
If YES, is a rationale for its inclusion provided?	
If YES, reference:	
9. Can any of the proposed characters be encoded using a composed character sequence of either existing characters or other proposed characters?	No
If YES, is a rationale for its inclusion provided?	
If YES, reference:	
10. Can any of the proposed character(s) be considered to be similar (in appearance or function) to an existing character?	Yes
If YES, is a rationale for its inclusion provided?	Yes
If YES, reference:	<i>See text</i>
11. Does the proposal include use of combining characters and/or use of composite sequences?	Yes
If YES, is a rationale for such use provided?	Yes
If YES, reference:	<i>See text</i>
Is a list of composite sequences and their corresponding glyph images (graphic symbols) provided?	n/a
If YES, reference:	<i>The proposal contains combining characters but no composite sequences</i>
12. Does the proposal contain characters with any special properties such as control function or similar semantics?	No
If YES, describe in detail (include attachment if necessary)	
13. Does the proposal contain any Ideographic compatibility character(s)?	No
If YES, is the equivalent corresponding unified ideographic character(s) identified?	
If YES, reference:	