Source:	Tao Yang, Eiso Chan, Wang Yifan
Meeting:	IRG#49, San Jose, USA
	UTC#154
Title:	Proposal of 4 IDCs
Status:	Individual
Actions required:	To be considered by IRG & UTC
Distribution:	IRG
Medium:	Electronic
Pages:	12
Appendixes:	4

4 new IDCs are in need for describing the following characters. Using them

may reduce the quantity of unnecessary and irregular components for IDS.

1. Unicode Properties

2FFC;IDEOGRAPHIC DESCRIPTION CHARACTER SURROUND FROM

RIGHT;So;0;ON;;;;;N;;;;;

2FFD;IDEOGRAPHIC DESCRIPTION CHARACTER SURROUND FROM

LOWER RIGHT;So;0;ON;;;;;N;;;;;

2FFE;IDEOGRAPHIC DESCRIPTION CHARACTER HORIZONTAL

REFLECTION;So;0;ON;;;;;N;;;;;

2FFF;IDEOGRAPHIC DESCRIPTION CHARACTER ONE HUNDRED

EIGHTY DEGREE ROTATION;So;0;ON;;;;;N;;;;;

Script=Common

Proposed Code Points	2FFC	2FFD	2FFE	2FFF
PUA	A0000	A0001	A0002	A0003
Character Names	IDEOGRAPHIC DESCRIPTION CHARACTER SURROUND FROM RIGHT	IDEOGRAPHIC DESCRIPTION CHARACTER SURROUND FROM LOWER RIGHT	IDEOGRAPHIC DESCRIPTION CHARACTER HORIZONTAL REFLECTION	IDEOGRAPHIC DESCRIPTION CHARACTER ONE HUNDRED EIGHTY DEGREE ROTATION
Glyph				

2. IDS Syntax

IDS ::= <Unary_Symbol><IDS1> | <Binary_Symbol>< IDS1>< IDS1> |

<Ternary_Symbol>< IDS1>< IDS1>< IDS1>
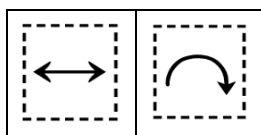
< IDS1> ::= <IDS> | <Ideograph_Component>

<Ideograph_Component> ::= coded_ideograph | coded_radical | coded_component

<Unary_Symbol> ::=

<Binary_Symbol> ::=

<Ternary_Symbol> ::=
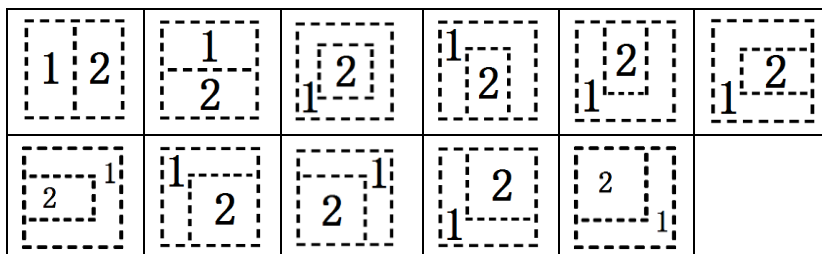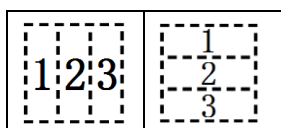
3. Ideographic Description Character Surround from Right

Code	Character	IDS
NC[1]	�	⿴叉丶
NC	�	⿴コ品
NC	㿭	⿴工釆
NC	𠃋	⿴コ乙
NC	�	⿴コ丙
NC	丁	⿴コ丁
NC	曰	⿴コ口
NC	番	⿴コ番
NC	咡	⿴耳口
NC	洰	⿴互冫

4. Ideographic Description Character Surround from Lower Right

Code	Character	IDS
NC	卣	⿷⿶⿸尸卜日
NC	斪	⿷⿶⿶尸卜日斤
U+7742	𥝂	⿷⿶⿸厂夂目
U+25171	𥅱	⿷⿶⿸厂屮目
U+2519E	𥆞	⿷⿶⿸厂罒罒
U+24401	𤐁	⿷⿶⿶⿸厂罒⿶日罒火
NC	実	⿶⿸宀⿶大丶
NC	屍	⿶尸⿶七丶
U+6C37	氷	⿷水丶
NC	𠂆	⿷𠂇丶
U+2669D	脄	⿷月⿷水冫

U+2097C	斗	□十十
U+20DF3	嶜	□□处日口
U+2298F	戋	□戈十
U+22992	戋	□戈丰
U+22994	戋	□戈才
NC	竷	□戈□亠丷
NC	戠	□戈兀
NC	戠	□戈中
U+39B0	戣	□戈从
U+8CAE	貳	□□弋一貝
U+5F10	弐	□□弋一二
U+6B66	武	□□弋一止

NC	戝	□□戈大□口壬
U+38A4	㢤	□□弋十力
U+39B1	㦱	□□戈一土
U+2BEF2	𫻲	□□戈人欠
U+2192A	𡤪	□女□□戈竹韭
U+21B76	𡭶	□□□七丷且
U+22387	𢎇	□□□弋十手
U+229B1	𢦱	□□戈一日
U+229E4	𢧤	□□戈大呈
U+24CC0	𤳀	□□戈丝田
U+254DB	磎	□石□□戈一幺
U+276E6	𧛦	□□戈市衣

Code	Character	IDS
U+27D92	戳	▢▢戈口貝
U+27F6B	趰	▢走▢▢戈口日
NC	攦	▢才▢▢戈廿韭
NC	轖	▢車▢▢戈口月

5. Ideographic Description Character Horizontal Reflection

Code	Character	IDS
U+23944	玊	↔正
U+23957	歨	▢止↔止
U+23960	址	▢▢↔止止
U+20141	亠	↔亡
U+2009C	玊	▢丿↔止
U+2456A	目	↔目

U+25E28	𥸨	↔�
U+268E9	𦣩	□臣↔臣
U+81E6	臦	□↔臣臣
U+81E9	臩	□□↔臣臣乔
U+22374	𢍴	□□↔臣臣廾
U+28F5D	𨽝	□自□廿月↔自
U+28F61	𨽡	□自豖↔自
NC	𢎨	↔巴
U+28668	𨙨	↔邑
U+286DC	𨛜	□↔邑邑
NC	𢎪	↔弓
U+2BE2C	𫸬	□弓↔弓

6. Ideographic Description Character One Hundred Eighty Degree Rotation

Code	Character	IDS
U+20114	坐	⟲予
U+2010F	丄	⟲了
U+221B4	幻	幺⟲了
NC	彧	⟲或或
NC	彧	或⟲或
NC	彧	或⟲或
NC	彧	或⟲或
U+22A0B	彧	或⟲或
U+23952	正	⟲止止
NC	虐	⟲虎

NC	虓	☐ ↷ 虎 虎
NC	虝	☐ 虎 ↷ 虎
U+27E42	贙	⊟ ☐ 虎 ↷ 虎 貝
U+2C88C	覵	⊟ ☐ 虎 ↷ 虎 見
NC	椚	☐ 木 ↷ 彐 彐

<table>
<tr><td colspan="2" align="center">

ISO/IEC JTC 1/SC 2/WG 2
PROPOSAL SUMMARY FORM TO ACCOMPANY SUBMISSIONS
FOR ADDITIONS TO THE REPERTOIRE OF ISO/IEC 10646[2]
Please fill all the sections A, B and C below.
Please read Principles and Procedures Document (P & P) from http://std.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://std.dkuug.dk/JTC1/SC2/WG2/docs/summaryform.html**.**
See also http://std.dkuug.dk/JTC1/SC2/WG2/docs/roadmaps.html **for latest *Roadmaps*.**

</td></tr>
</table>

A. Administrative

1. **Title:**	*Proposal of 4 IDCs*
2. Requester's name:	*Tao Yang, Eiso Chan, Wang Yifan*
3. Requester type (Member body/Liaison/Individual contribution):	*Individual contribution*
4. Submission date:	*2018-01-04*
5. Requester's reference (if applicable):	*ISO/IEC JTC 1/SC 2/WG 2/IRG N2273*
6. Choose one of the following:	
This is a complete proposal:	*Yes*
(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):	*No*
Proposed name of script:	
b. The proposal is for addition of character(s) to an existing block:	*Yes*
Name of the existing block:	*Ideographic Description Characters*
2. Number of characters in proposal:	*4*

3. Proposed category (select one from below - see section 2.2 of P&P document):

A-Contemporary	**Y** B.1-Specialized (small collection)	B.2-Specialized (large collection)	
C-Major extinct	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?	*Yes*
a. If YES, are the names in accordance with the "character naming guidelines"	
in Annex L of P&P document?	*Yes*
b. Are the character shapes attached in a legible form suitable for review?	*Yes*
5. Fonts related:	
a. Who will provide the appropriate computerized font to the Project Editor of 10646 for publishing the standard?	
Dr. Ken Lunde	
b. Identify the party granting a license for use of the font by the editors (include address, e-mail, ftp-site, etc.):	
Dr. Ken Lunde(lunde@adobe.com)	
6. References:	
a. Are references (to other character sets, dictionaries, descriptive texts etc.) provided?	
b. Are published examples of use (such as samples from newspapers, magazines, or other sources)	
of proposed characters attached?	
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)?	

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 Unicode Character Database (http://www.unicode.org/reports/tr44/) and associated Unicode Technical Reports for information needed for consideration by the Unicode Technical Committee for inclusion in the Unicode Standard.

C. Technical - Justification

1. Has this proposal for addition of character(s) been submitted before?   *No*
    If YES explain

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.)?   *No*
        If YES, with whom?
        If YES, available relevant documents:

3. Information on the user community for the proposed characters (for example:
    size, demographics, information technology use, or publishing use) is included?   *Yes*
    Reference: *ISO/IEC JTC 1/SC 2/WG 2/IRG N2273*

4. The context of use for the proposed characters (type of use; common or rare)   *common*
    Reference:

5. Are the proposed characters in current use by the user community?   *Yes*
    If YES, where?   Reference:

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:

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, or could be confused with, an existing character?   *No*
        If YES, is a rationale for its inclusion provided?
        If YES, reference:

11. Does the proposal include use of combining characters and/or use of composite sequences?   *No*
    If YES, is a rationale for such use provided?
        If YES, reference:
    Is a list of composite sequences and their corresponding glyph images (graphic symbols) provided?
        If YES, reference:

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 characters?   *No*
    If YES, are the equivalent corresponding unified ideographic characters identified?
        If YES, reference:

Doc Type:	Ideographic Rapporteur Group Document	ISO/IEC JTC1/SC2/WG2/**IRG N2273 Feedback**
Title:	Feedback on Proposals of 4 IDCs (IRG N2273R)	
Source:	Kushim Jiang (姜兆勤)	
Status:	Individual Contribution to IRG #55	
Action:	For consideration by IRG	
Date:	2020-3-18	

IRG N2273R shows 4 IDCs which are indeed necessary to be encoded. However, the proposed code points and blocks may need to be changed.

The security of generating IDS programmatically requires that an IDS must be structurally finite. This needs to be ensured by two factors:

- Characters are finite in stroke count;
- IDS are finite by definition.

The first factor is obvious. However, due to the recursive definition of IDS, any unary, binary and ternary operators must be not idempotent. Conversely, an idempotent operator cannot be encoded as an IDC, but only as an IDM (Ideographic Description Mark).

For proposition always holds that

$$GetStrokeCount(IDS) > 0$$

and

$$GetStrokeCount(BinaryOperator\ IDS1\ IDS2) = GetStrokeCount(IDS1) + GetStrokeCount(IDS2),$$

then

$$GetStrokeCount(BinaryOperator\ IDS1\ IDS2) > max\ \{GetStrokeCount(IDS1), GetStrokeCount(IDS2)\},$$

and similarly

$$GetStrokeCount(TernaryOperator\ IDS1\ IDS2\ IDS3) > max\ \{GetStrokeCount(IDS1), GetStrokeCount(IDS2),$$
$$GetStrokeCount(IDS3)\},$$

which shows that in IDS definition, any binary and ternary operators are not idempotent.

For any sequences in the form of "⍰ IDS", there is always:

$$⍰(⍰\ IDS) = ⍰\ IDS.$$

That is why U+303E ɪᴅᴇᴏɢʀᴀᴘʜɪᴄ ᴠᴀʀɪᴀᴛɪᴏɴ ɪɴᴅɪᴄᴀᴛᴏʀ (IVI) is not considered a part of the IDS and does not invalidate the sequence (*Core Specification* v13.0, p. 734).

For any IDSes, there is always:

$$⟷ IDS = IDS, \quad ⟲ IDS = IDS.$$

Then the operator "⟷" and "⟲" are idempotent, this will lead to insecurity in generating IDS. So ⟷ and ⟲ cannot be encoded as an IDC, but an IDM like ⍰.

For "U+303E ⍰" is encoded separate from "U+2246 ≆", and to make semantic distinctions, ⟷ and ⟲ had better be encoded separate from "U+2194 ↔" and "U+21B7 ↷".

The IVI ⍰ need to be put ahead of an IDS. Due to the recursive definition of IDS, the IVI can also be put inside an IDS if the IDS is composed of sub-IDSes. Therefore, ⟷ and ⟲ can also be put ahead of an IDS or a sub-IDS, which means that ▢▢虎⟲虎貝 is a legal pseudo-IDS expression of "U+27E42 贙".

In conclusion, ▢ and ▢ can be encoded as IDCs in maybe U+2FFC and U+2FFD, but ⟷ and ⟲ should be encoded as IDMs in maybe U+16FE5 and U+16FE6.

Eiso Chan (陈永聪) and Tao Yang (陶扬) also contributed to this article. #