

JTC1/SC2/WG2 N4719
2016-04-21

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 symbols for Go game notation
Source: Andrew West
Status: Individual Contribution
Action: For consideration by JTC1/SC2/WG2 and UTC
Date: 2016-04-21

1. Introduction

The Unicode and ISO/IEC 10646 standards include four characters used for marking stones in the game of Go:

- U+2686 ◯ WHITE CIRCLE WITH DOT RIGHT
- U+2687 ◐ WHITE CIRCLE WITH TWO DOTS
- U+2688 ● BLACK CIRCLE WITH WHITE DOT RIGHT
- U+2689 ◐ BLACK CIRCLE WITH TWO WHITE DOTS

As discussed on the Unicode mailing list (see thread starting at <http://www.unicode.org/mail-arch/unicode-ml/y2016-m03/0022.html>), these four characters are used in mathematical research into the game of Go (see Fig. 2), but do not seem to be attested in actual Go game notation.

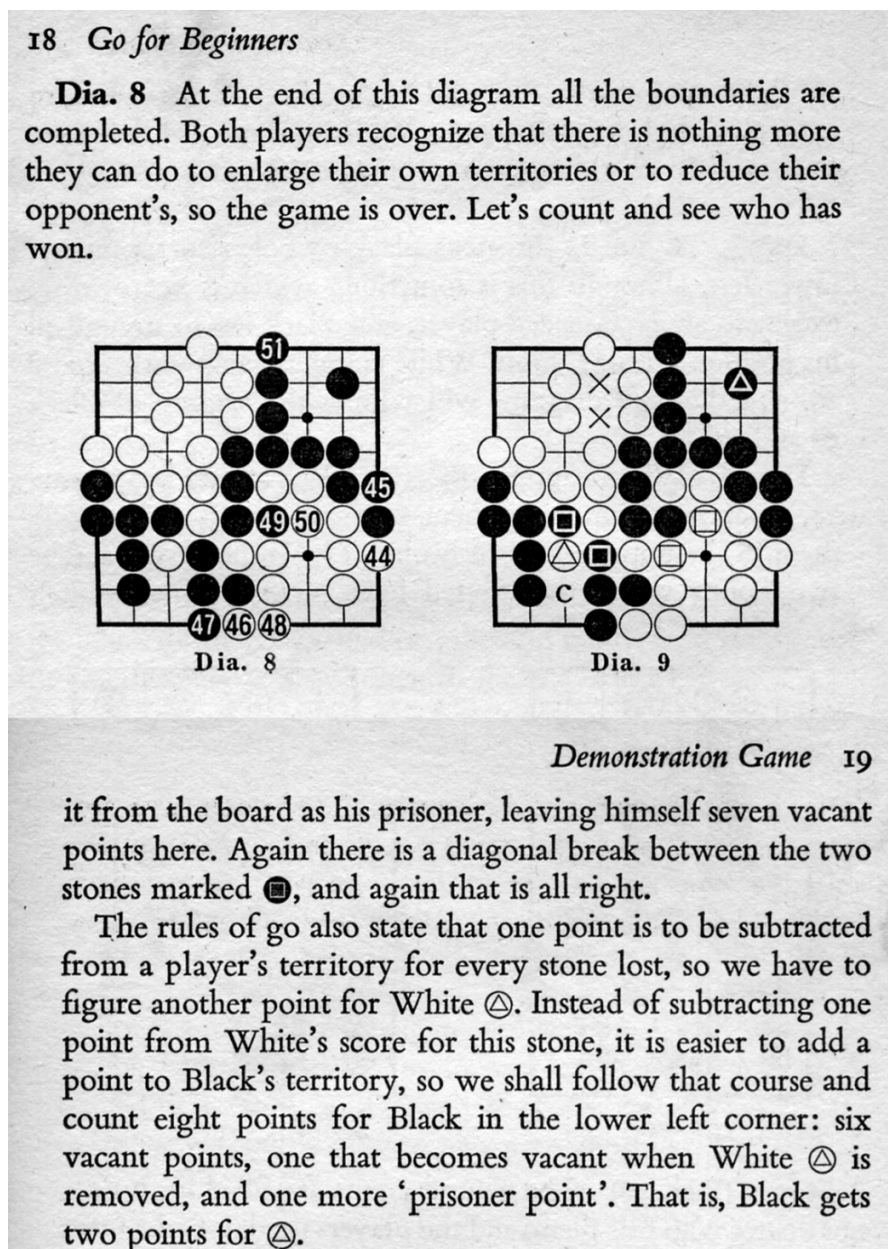
L2/16-021 “GAME PIECES to be considered for inclusion as emoji” proposes giving these four characters the emoji property. As these characters are intended for use in technical, mathematical discussion, and are not intended to visually represent actual Go stones, it would be inappropriate to treat them as emoji. If emoji for Go stones are required, then separate BLACK GO STONE and WHITE GO STONE characters should be encoded. Such characters are not proposed in this document as it is not clear emoji for Go stones are needed.

However, there are a number of unencoded symbols that are widely used in English, Japanese and Chinese sources for the notation of Go games, and these do need to be encoded.

2. Circles Containing Triangles and Squares

In commentary on Go game positions, it is customary to mark a particular stone or group of stones under discussion with a triangle. When an additional stone or group of stones needs to be marked for discussion in the same diagram in which a stone of the same colour marked with a triangle already occurs, then it is marked with a square. See Fig. 1 below for an example showing white and black stones marked with triangles and squares, both within the game diagram, and in text discussion of the game diagram. Further examples from books published in Japan and China are shown in Fig. 3 through Fig. 8 at the end of this document.





Fig. 1: Kaoru Iwamoto, *Go for Beginners* (Penguin Books, 1976) pp. 18-19



Sometimes a stone is marked with a circle or a filled circle (see Fig. 3), but these cases can be represented using existing characters such as U+25CE ☉ BULLSEYE, U+25C9 ● FISHEYE, U+1F78A ⦿ WHITE CIRCLE CONTAINING BLACK SMALL CIRCLE, and U+1F789 ◉ EXTREMELY HEAVY WHITE CIRCLE.

In order to represent the usage of triangles and squares to mark stones in Go notation, we propose encoding four symbol characters in the Geometric Shapes Extended block

Table 1: Proposed Characters for Marking Go Stones

Code Point	Glyph	Character Name
1F7D5		WHITE CIRCLE CONTAINING WHITE TRIANGLE
1F7D6		BLACK CIRCLE CONTAINING BLACK TRIANGLE
1F7D7		WHITE CIRCLE CONTAINING WHITE SQUARE
1F7D8		BLACK CIRCLE CONTAINING BLACK SQUARE

The names are suggested as WHITE CIRCLE CONTAINING WHITE TRIANGLE etc. instead of the simpler WHITE CIRCLE CONTAINING TRIANGLE etc. as it is not impossible that WHITE CIRCLE CONTAINING BLACK TRIANGLE etc. could be encoded in the future, and these names avoid ambiguity.

3. Circles Enclosing Numbers

Modern Usage

In modern usage, Go games are notated using diagrams in which the move number of a stone is placed inside a black or white circle representing the Go stone (see Fig. 1 above). When a stone is played at the site of a previously-played stone (which was later taken and removed from the board), then the move cannot be recorded in the game diagram, but is recorded as a note below the diagram.

In some publications and web sites numbering of go stones in a single game diagram only goes up to 99, and then continues on a new diagram from “00” (representing a hundred or multiple of a hundred); the next nine moves are marked as “01” through “09” (in contrast with “1” through “9” for the absolute first nine moves). See Fig. 8 for an example of this usage on Wikipedia. However, numbering moves into the hundreds, with three circled digits, is equally common (see Fig. 9 and Fig. 10 for examples).

The existing encoded characters CIRCLED DIGIT ONE through CIRCLED NUMBER FIFTY, and DINGBAT NEGATIVE CIRCLED DIGIT ONE through NEGATIVE CIRCLED NUMBER TWENTY can be used for some Go diagrams representing game problems or opening positions (*fuseki*), but are insufficient for notating a complete game, which on average has between 150 and 250 moves.

It should be noted that circled and negative circled numbers occur in text below the game diagram as well as in the game diagram. The usage of these circled and negative circled numbers in running text indicates that these symbols should be dealt with at the character encoding level, and cannot be treated as images or dismissed as requiring some “higher level protocol” for correct rendering.

Historical Usage

Historically, Go game diagrams customarily used Chinese ideographs enclosed in black or white circles for the same purpose (see Fig. 11 and Fig. 12 for examples from Chinese Go manuals dating from the Ming and Qing dynasties). The ideographic numbers are stacked vertically within the circle, although for numbers greater than one hundred they may be in two columns, with hundreds on the right (see Fig. 13). Notice that in Fig. 11, where two or more stones were played at the same place (because the previously placed stone(s) had been taken) then the stones are indicated using narrow circled numbers squeezed into the space for one stone (see lower left and lower right corners of Fig. 11). It would be very difficult to reproduce these narrow circled numbers in plain text.

The earliest extant Go game diagrams, dating from the Southern Song and Yuan dynasty (12th through 14th centuries), mark moves on the Go board using diagonally-positioned rectangular black and white numbered tags (see Fig. 14 and Fig. 15). These diagonal tags do not occur in text, and we consider them to be outside the scope of character encoding.

Possible Solutions

Although most games of Go last between 150 and 250 moves, a game can theoretically have up to 361 moves (because there are 361 stones, 181 black and 180 white), although some professional games have lasted up to 400 moves. This means that there needs to be a mechanism for displaying circled numbers 1 through about 399, and negative circled numbers 1 through about 399. As mentioned above, currently encoded circled numbers range from 1 through 50, and negative circled numbers range from 1 through 20, which means that circled numbers greater than 50 and negative circled numbers greater than 20 cannot be represented in plain text.

A number of possible solutions to representing unencoded circled and negative circled numbers are outlined below.

A. Encode atomic characters for all circled numbers 00 through 09 and 51 through 399 (or 999 for good measure), and all negative circled numbers 00 through 09 and 21 through 399 (or 999 for good measure).

Advantages: Simple; convenient for end user.

Disadvantages: Requires a very large number of code points for a relatively specialized use case (but this could be mitigated by encoding the characters in Plane 14 [Supplementary Special-purpose Plane]); does not take into account historic use of circled ideographs.

B. Encode a limited set of atomic characters for circled numbers 51 through 99 and 00 through 09, and negative circled numbers 21 through 99 and 00 through 09.

Advantages: Simple; convenient for end user; only requires encoding 150 new characters.

Disadvantages: Sets a precedent for encoding circled numbers as atomic characters, and merely delays the inevitable request to encode circled numbers greater than 99; does not support three-digit circled numbers; does not take into account historic use of circled ideographs.

C. Use a TAG mechanism, along the lines of the mechanisms described in UTS #52 “Unicode Emoji Mechanisms”. For example, circled number 123 could be represented by the sequence <WHITE CIRCLE, TAG NUMBER SIGN, TAG DIGIT ONE, TAG DIGIT TWO, TAG DIGIT THREE, CANCEL TAG>, and the negative circled number 00 could be represented by the sequence <BLACK CIRCLE, TAG NUMBER SIGN, TAG DIGIT ZERO, TAG DIGIT ZERO, CANCEL TAG>.

Advantages: Does not require encoding any new characters; flexible (could allow numbers greater than 999, or decimal numbers such as 3.14159).

Disadvantages: Difficult for end users to construct tag sequences (especially as tag characters are normally invisible); does not take into account historic use of circled ideographs.

D. Use U+20DD COMBINING ENCLOSING CIRCLE and a new COMBINING ENCLOSING NEGATIVE CIRCLE character to dynamically compose arbitrary circled negative circled numbers. However, “[t]here is also no defined way of indicating the application of a combining enclosing mark to more than a single base character” (*The Unicode Standard Version 8.0* pp. 803–804). Therefore, in order to use combining enclosing circles with multi-digit numbers a mechanism would need to be defined to indicate the boundaries of the sequence of preceding characters that the circle should combine with. One possibility would be to use U+200D ZERO WIDTH JOINER to join the characters that should be

encircled together. For example, circled number 123 could be represented by the sequence <DIGIT ONE, ZWJ, DIGIT TWO, ZWJ, DIGIT THREE, COMBINING ENCLOSING CIRCLE>, and the negative circled number 00 could be represented by the sequence <DIGIT ZERO, ZWJ, DIGIT ZERO, COMBINING ENCLOSING NEGATIVE CIRCLE>. This mechanism could also work for circled ideographic numbers, for example <5345, ZWJ, 4E5D, COMBINING ENCLOSING CIRCLE> for circled vertically-stacked 卅九 (39). But extending this mechanism to letters may be problematic as ZWJ could cause ligaturing where no ligature was intended, so for example <LATIN SMALL LETTER S, ZWJ, LATIN SMALL LETTER T, COMBINING ENCLOSING CIRCLE> might produce circled st ligature rather than circled st.

Advantages: Only requires encoding one new character (COMBINING ENCLOSING NEGATIVE CIRCLE); flexible (could allow numbers greater than 999, or decimal numbers such as 3.14159); supports historic use of circled ideographs; easier to implement at the font level than Tag sequences; use of ZWJ should be relatively intuitive for end users.

Disadvantages: Requires the definition of a mechanism to allow combining enclosing characters to combine with more than a single base character or grapheme cluster (some would consider this to be an advantage); users cannot simply select characters from a character map or character picker application, but must either construct the sequence manually or rely on a specialist input method.

E. Encode partial circled numbers which can be combined together as a sequence to form a whole circled number. Sixty characters would be required: LEFT [NEGATIVE] HALF CIRCLE WITH DIGIT ZERO through NINE; RIGHT [NEGATIVE] HALF CIRCLE WITH DIGIT ZERO through NINE; and MIDDLE PART OF [NEGATIVE] CIRCLE WITH DIGIT ZERO through NINE. Left and right half circle characters would combine to form circled double digit numbers (e.g. <LEFT NEGATIVE HALF CIRCLE WITH DIGIT ZERO, RIGHT NEGATIVE HALF CIRCLE WITH DIGIT ZERO> would form negative circled number 00). Left, middle and right circle characters would combine to form circled triple digit numbers (e.g. <LEFT HALF CIRCLE WITH DIGIT ONE, MIDDLE PART OF CIRCLE WITH DIGIT TWO, RIGHT HALF CIRCLE WITH DIGIT THREE> would form circled number 123).

Advantages: Does not require encoding a huge number of new characters; relatively simple for users to construct the required numbers.

Disadvantages: Unintelligent implementations may produce triple digit numbers in an oval rather than a circle; may not work correctly in vertical text (many Japanese books about Go are typeset in vertical layout); does not take into account historic use of circled ideographs; seems too much like a font hack.

F. Encode format characters that allow the creation of arbitrary circled text. Four new format characters would be required: START OF ENCLOSING CIRCLE, END OF ENCLOSING CIRCLE, START OF ENCLOSING NEGATIVE CIRCLE, and END OF ENCLOSING NEGATIVE CIRCLE. The sequence <START OF ENCLOSING CIRCLE, DIGIT ONE, DIGIT TWO, DIGIT THREE, END OF ENCLOSING CIRCLE> would produce circled number 123. The sequence <START OF ENCLOSING NEGATIVE CIRCLE, DIGIT ZERO, DIGIT ZERO, END OF ENCLOSING NEGATIVE CIRCLE> would produce negative circled number 00.

Advantages: Only requires encoding four new characters; flexible.

Disadvantages: There is no precedent for encoding format characters with this sort of behaviour; would not work with ideographic numbers as they stack vertically inside a circle.

None of the above options are ideal, but our preferred solution is D. We therefore propose encoding COMBINING ENCLOSING NEGATIVE CIRCLE at U+20F1.

4. Unicode Properties

20F1;COMBINING ENCLOSING NEGATIVE CIRCLE;Me;0;NSM;;;;;N;;;;;
1F7D5;WHITE CIRCLE CONTAINING WHITE TRIANGLE;So;0;ON;;;;;N;;;;;
1F7D6;BLACK CIRCLE CONTAINING BLACK TRIANGLE;So;0;ON;;;;;N;;;;;
1F7D7;WHITE CIRCLE CONTAINING WHITE SQUARE;So;0;ON;;;;;N;;;;;
1F7D8;BLACK CIRCLE CONTAINING BLACK SQUARE;So;0;ON;;;;;N;;;;;

5. Figures

Fig. 2: Elwyn Berlekamp and Yonghoan Kim, "Where Is the "Thousand-Dollar Ko"?"; *Games of No Chance* vol. 29 (1996) p. 206

Values must also be computed for those regions that do not appear verbatim in the book. Such calculations are done as described in Chapter 2 of *Mathematical Go*. As an example of such a calculation, we consider the region Q . We have $Q = \{Q^L \mid Q^R\}$, where Q^R is the position after White has played there and Q^L is the position after black has played there. After appropriate changes of markings (which affects only the integer parts of the relative score), we may continue as follows:

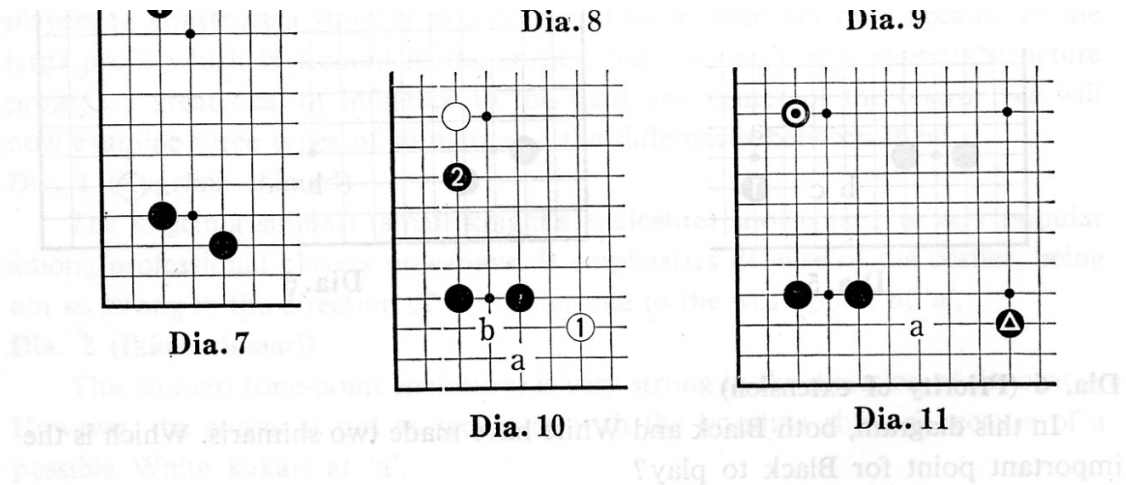
$$\begin{array}{c}
 \begin{array}{c} \text{Go board diagram} \\ \text{with stones and liberties} \end{array} = \left\{ \begin{array}{c} \text{Go board diagram 1} \\ \text{Go board diagram 2} \end{array} , \left| \begin{array}{c} \text{Go board diagram 3} \\ \text{Go board diagram 4} \end{array} \right. \right\} \\
 \\
 Q^R = \{ 0 \quad , \quad \text{---} \quad | \quad * \quad , \quad \text{---} \quad \}
 \end{array}$$

Although 0 and --- are both formal Black followers of Q^R , the miny option is dominated and we have

$$Q^R = \{0 \mid *, \text{---}\}.$$

The two formal White followers are * and ---, which are incomparable. However, White's move from Q^R to --- reverses through a hot position to 0, and we

Fig. 3: Haruyama and Nagahara, *Basic Techniques of Go* (Ishi press, 1973) p. 58



Dia. 10 (Weakness in the corner)

From the last two diagrams we saw that the ikken shimari is very strong towards the center. But because of the weakness at White 1, this shimari is not so strong with respect to the corner. After Black extends to 2, White will aim at 'a', to which Black will respond with 'b'. Depending upon the situation, White might choose to play at 'b' instead of 'a'.

Dia. 11 (Double-wing formation)

When Black has made an ikken shimari, the extensions of Black \blacktriangle and \bullet are ideal placements. This configuration is known as the double-wing formation.

— 58 —

Fig. 4: James Davies, *Tesuji* (Ishi Press, 1975) p. 89

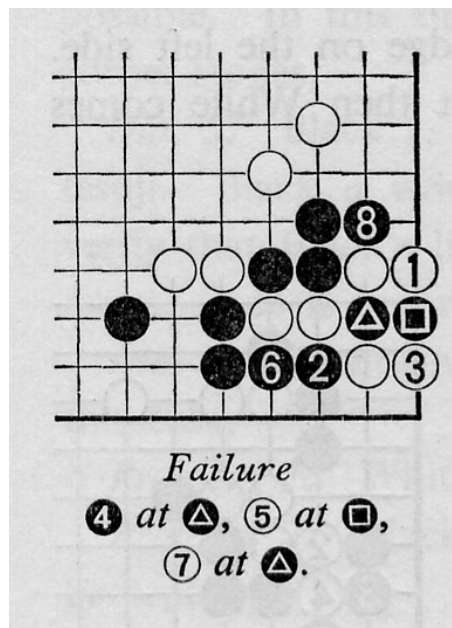


Fig. 5: James Davies, *Life and Death* (Ishi Press, 1975) p. 9

Dia. 6. Two-stage ko, (favorable to Black).

White 1 starts a two-stage ko. While the ko is being fought at 1 and 2, it is direct for Black, (all he has to do is to capture White \triangle), and indirect for White, (he has to play *a* and then *b* to win it). If White ignores a ko threat, however, the ko will move over and be fought at \triangle and *a*, and in that stage it will be direct for White and indirect for Black.

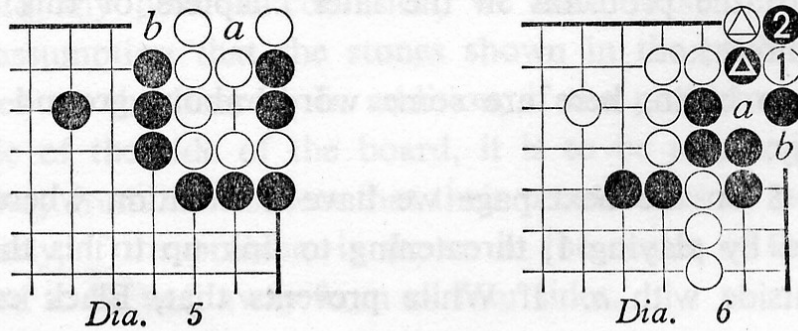
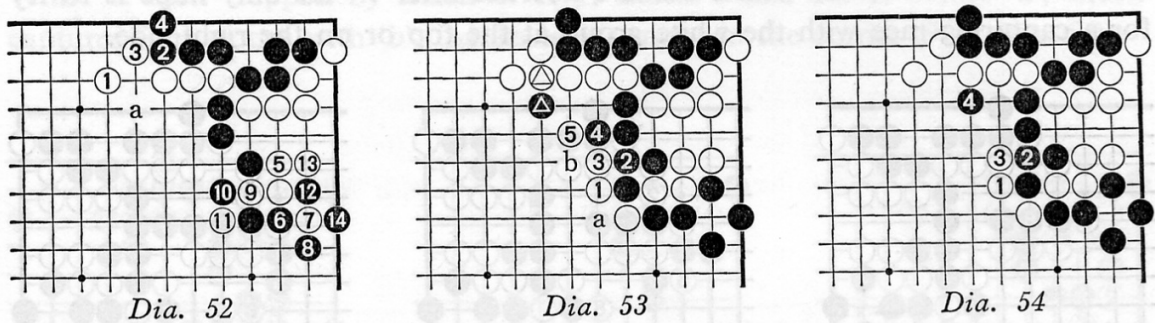


Fig. 6: Kaoru Iwamoto, *Dictionary of Basic Joseki* (Ishi Press, 1977) vol. 3 p. 31

Dia. 53 (the meaning of the peep). If the \triangle \triangle exchange has been made, there is no problem. White captures the black group with 1 to 5. Trying to escape by cutting at 'a' and 'b' is of course useless. This is why Black omits \triangle .

Dia. 54 (tricked). By omitting the peep, Black is able to escape by attaching at 4. There is now no way White can avoid collapse.



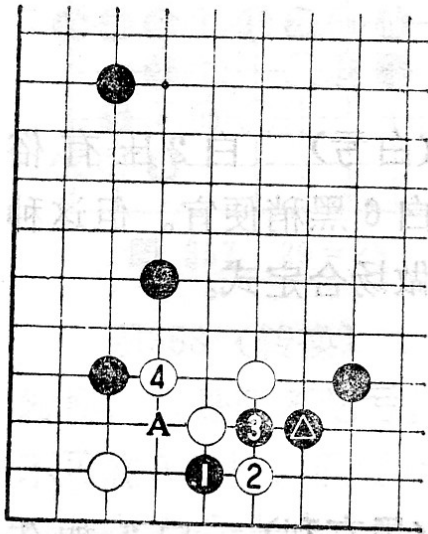


图 351

图351 (手筋?) 黑▲尖时, 白如应一手是可以确保活棋的, 但在实战中也常脱先。

黑1托, 是普通场合的手筋, 白如A退, 白形不好。但白2扳、4虎好, 抵挡住黑1的进攻。

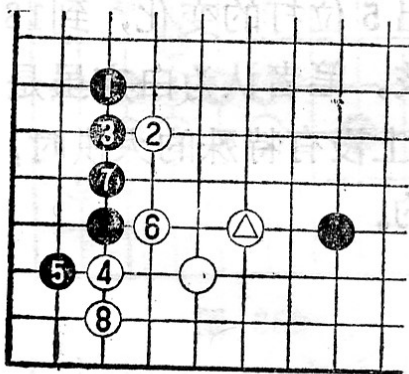
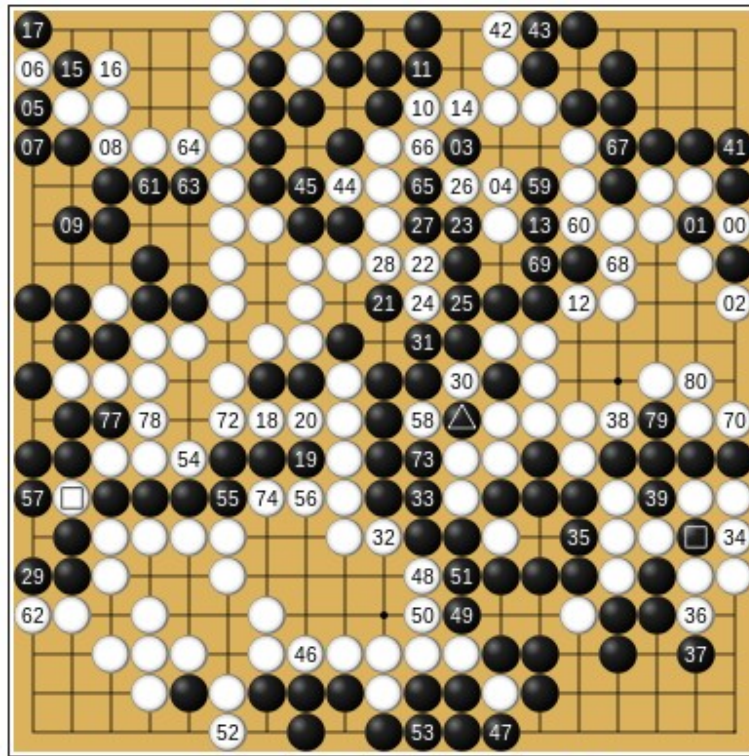


图 352

图352 (白好形) 黑1拆二, 在本形不适用, 白2点后到6时因有▲关系黑不能反打, 到白8后白形很挺, 黑处低位, 和图347相比得失可以了然。



Fig. 8: AlphaGo vs Lee Sedol Game 5
(https://en.wikipedia.org/wiki/AlphaGo_versus_Lee_Sedol)






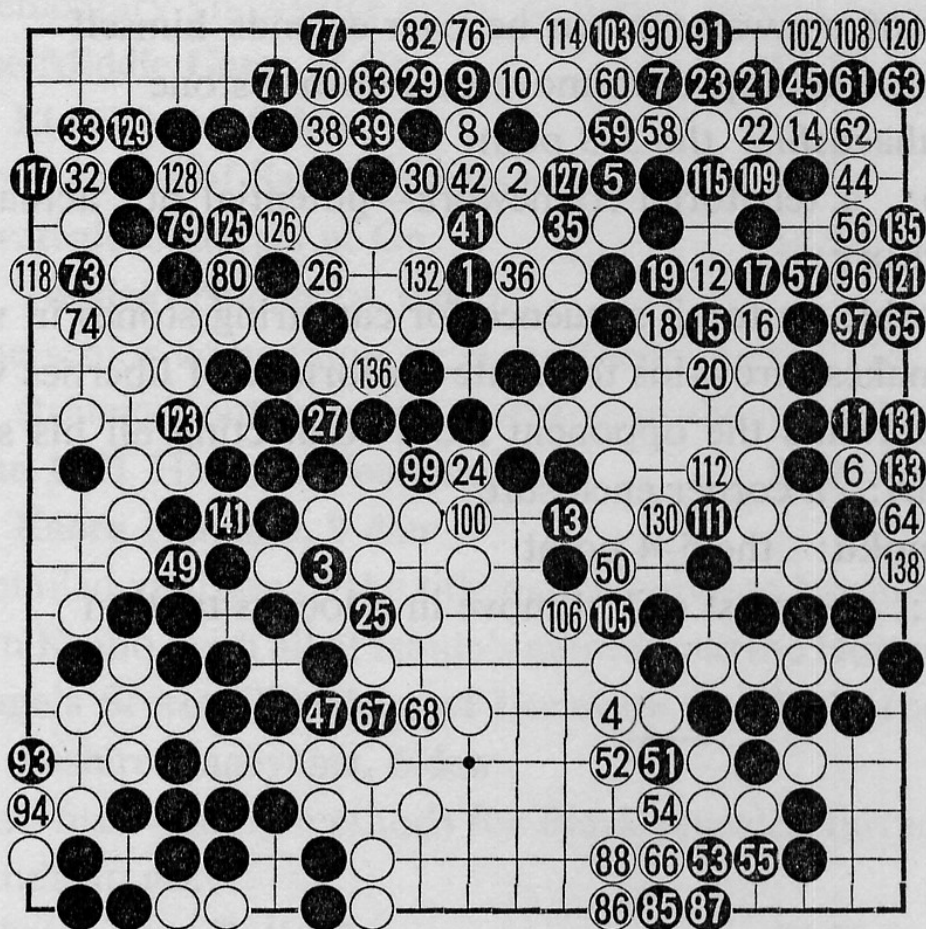
Moves 200-280 (240 at ,
275 at , 276 at )

Fig. 9: Shuzo Ohira, *Appreciating Famous Games* (Ishi Press, 1977) p. 279



28 takes ko (left of 25) 31, 34, 37, 40,
 43, 46 take ko. 48 takes ko (left of 3)
 69, 72, 75, 78, 81, 84, 89, 92, 95, 98, 101, 104, 107,
 110, 113, 116, 119, 122 take ko. 124 connects at
 3. 134 takes ko at 90. 137 takes below
 6. 139 fills at 6. 140 fills (right of
 8) 142 fills at 103.

Fig. 3 (201-342)

342 moves. Black wins by 3 points.

Fig. 10: Fan Hui vs AlphaGo - Game 5
 (https://commons.wikimedia.org/wiki/File:FHvAG5.jpg)

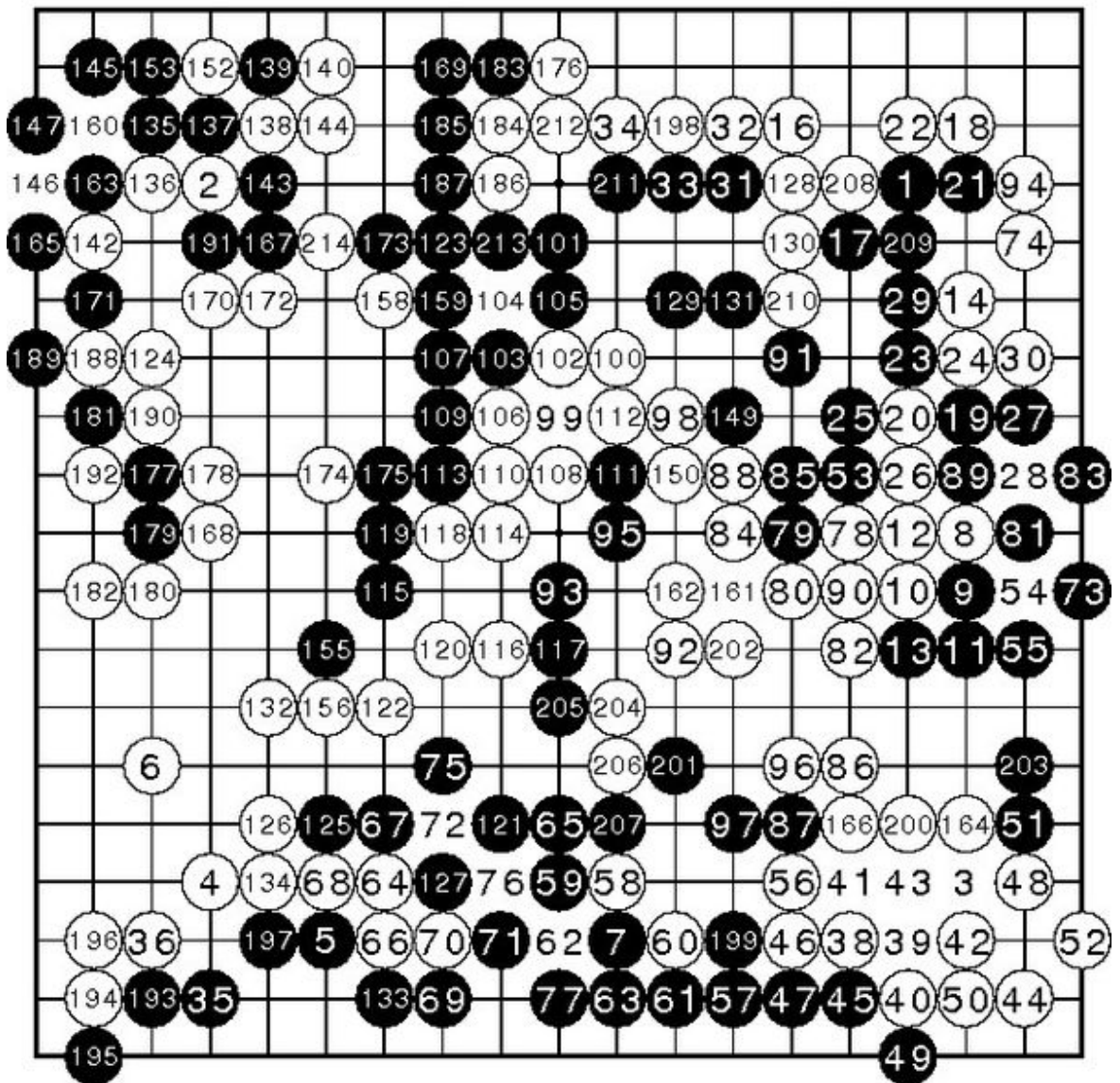


Fig. 11: Wànbǎo Quánshū 萬寶全書 (1610) v. 10 p. 13a

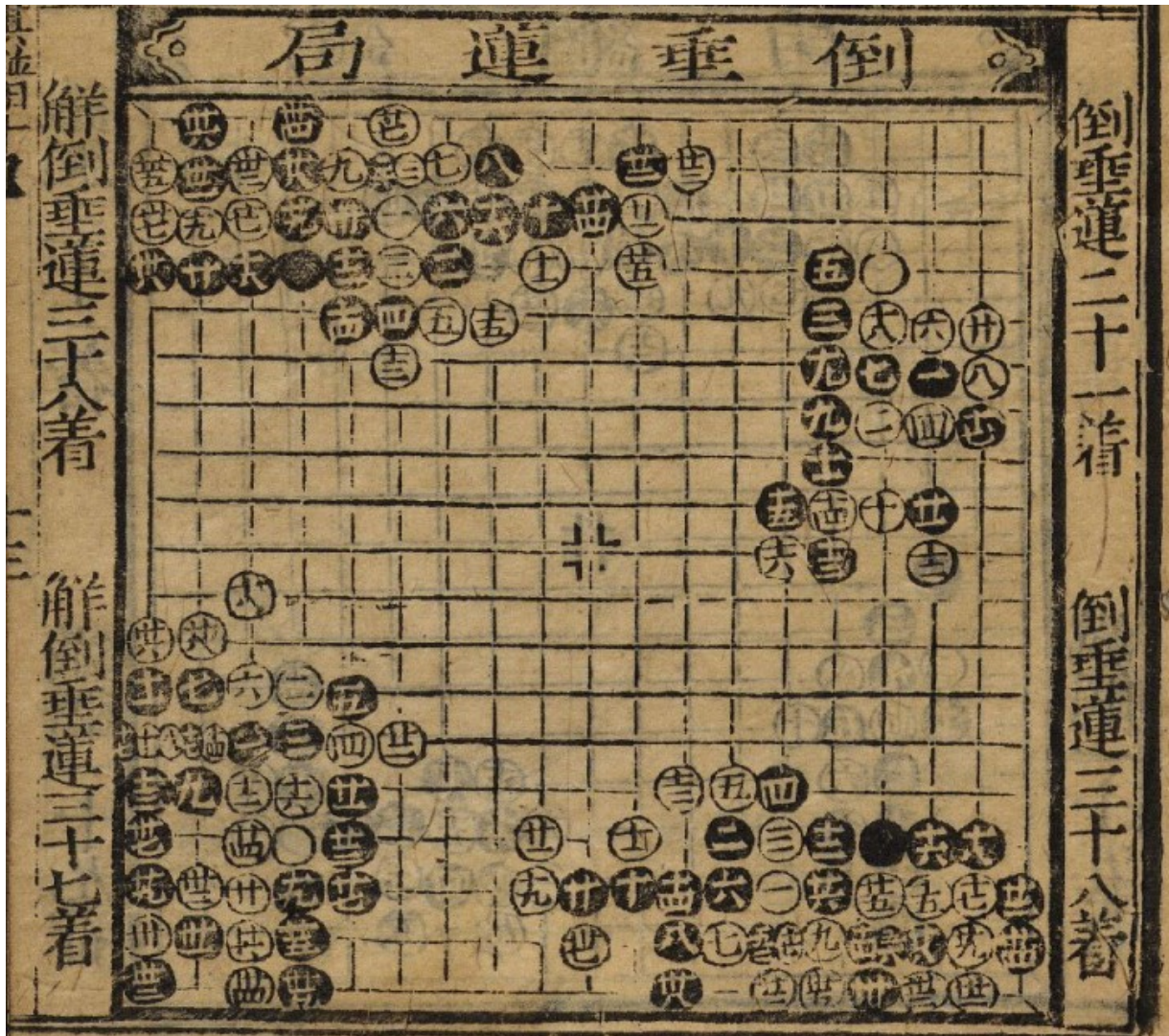


Fig. 12: Táo Shiyù 陶式玉 et al., *Guānzǐ Pǔ* 官子譜 (1694) [1981 reprint p.549]

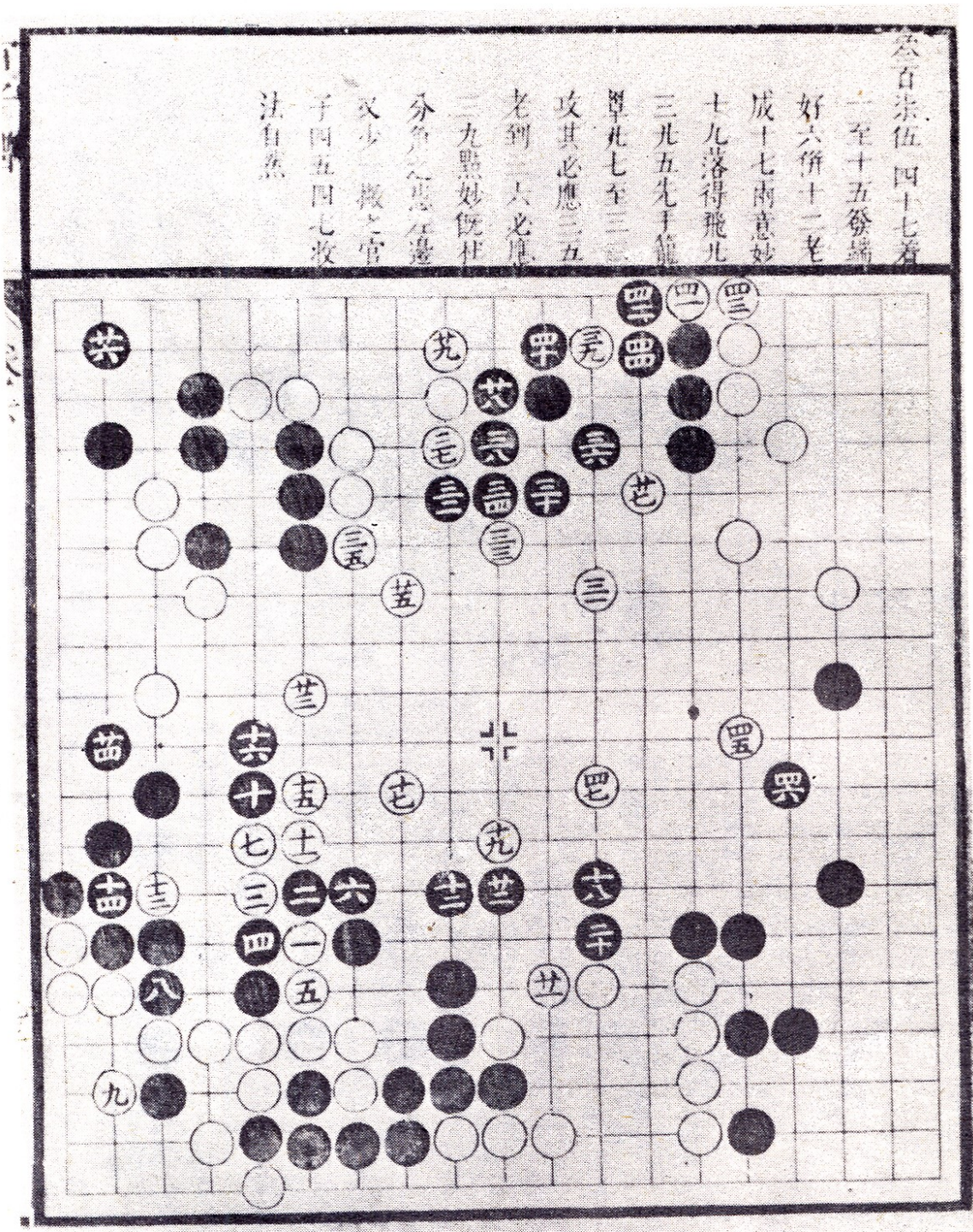


Fig. 13: Dèng Yuánhùi 鄧元鏞, *Yiqiánzhāi Jípǔ* 弈潛齋集譜 (1881 - 1889 ed.)

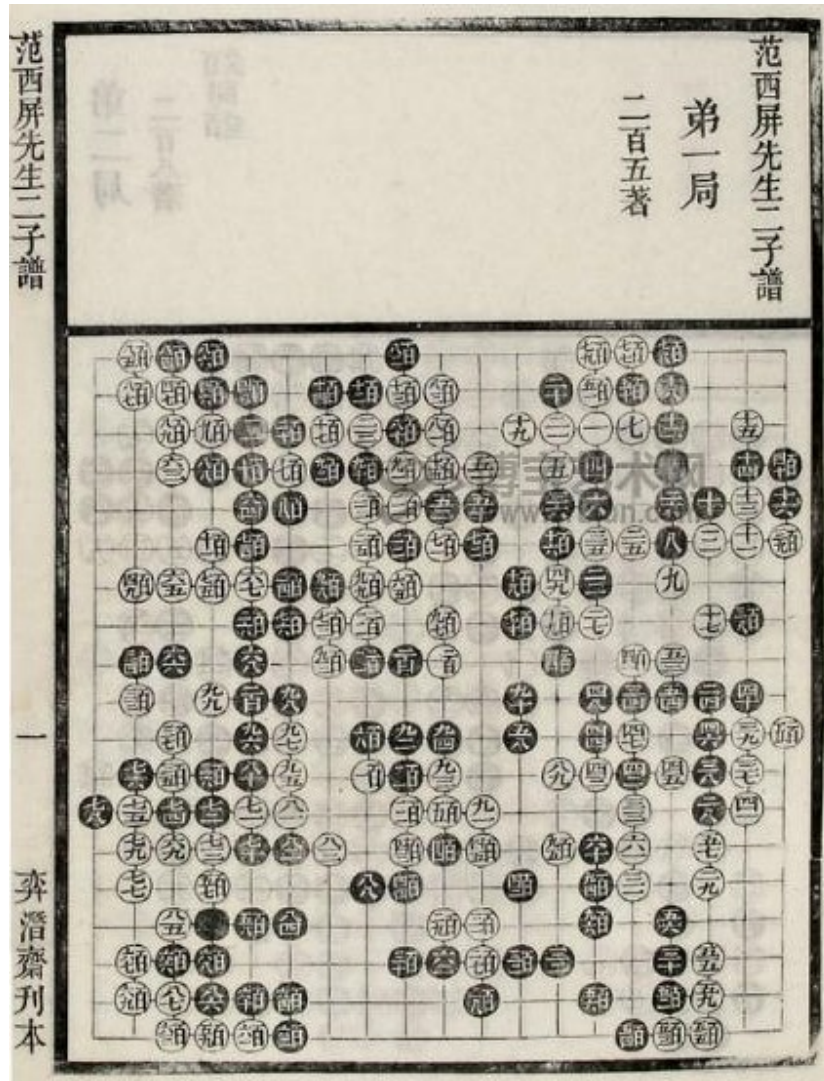


Fig. 14: Lǐ Yìmín 李逸民, Wàngyōu Qīnglè Jí 忘憂清樂集 (12th century ed.)

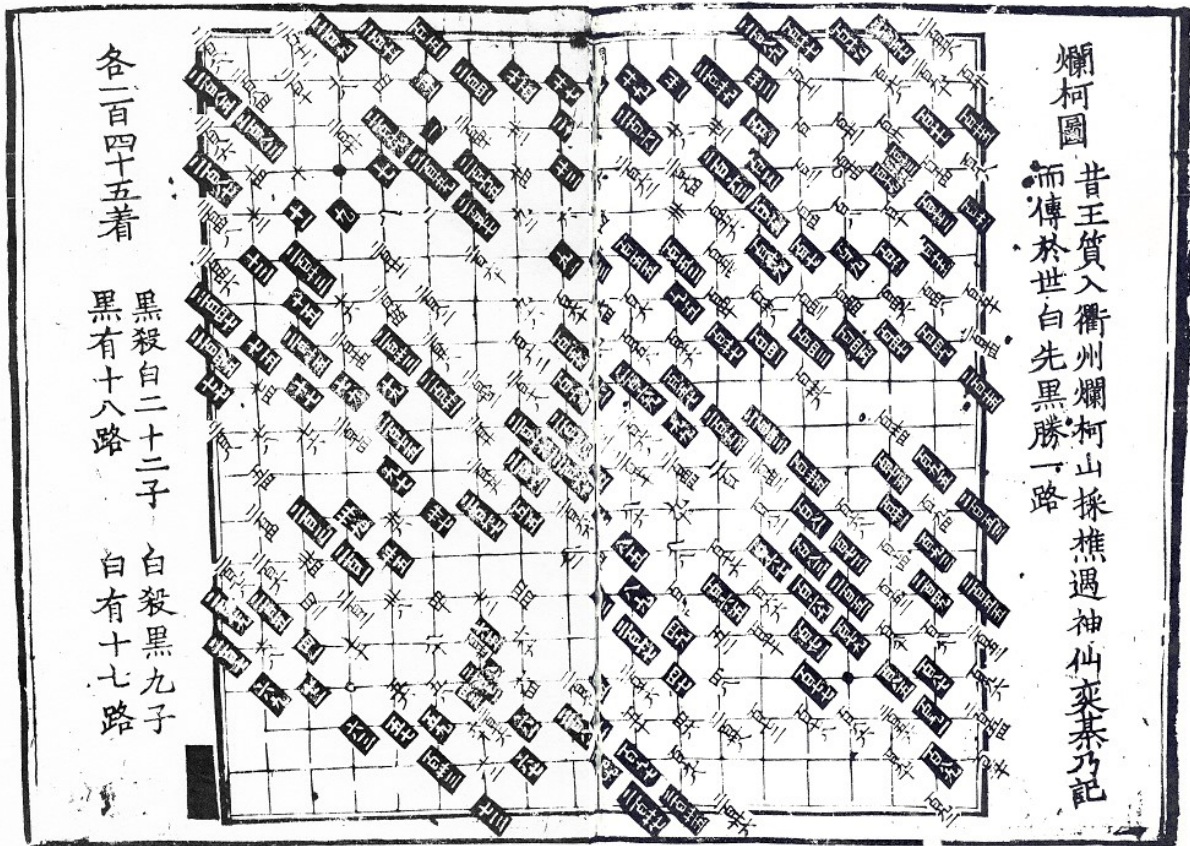


Fig. 15: *Shilín Guǎngjì* 事林廣記 (1330 - 1333 ed.) Xu v. 4 p. 14a



6. Proposal Summary Form

SO/IEC JTC 1/SC 2/WG 2
PROPOSAL SUMMARY FORM TO ACCOMPANY SUBMISSIONS
FOR ADDITIONS TO THE REPERTOIRE OF ISO/IEC 10646¹
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: *Proposal to encode symbols for Go game notation*

2. Requester's name: *Andrew West*

3. Requester type (Member body/Liaison/Individual contribution): *Individual contribution*

4. Submission date: *2016-04-21*

5. Requester's reference (if applicable):

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: *Geometric Shapes Extended, Combining Diacritical Marks for Symbols*

2. Number of characters in proposal: *5*

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

A-Contemporary 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? *Andrew West*

b. Identify the party granting a license for use of the font by the editors (include address, e-mail, ftp-site, etc.): *Andrew West*

6. References:

a. Are references (to other character sets, dictionaries, descriptive texts etc.) provided? YES

b. Are published examples of use (such as samples from newspapers, magazines, or other sources) of proposed characters attached? YES

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)? YES

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.

¹ Form number: N4102-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, 2011-03, 2012-01)

C. Technical - Justification

1. Has this proposal for addition of character(s) been submitted before? If YES explain		NO
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.)? If YES, with whom? If YES, available relevant documents:		YES
	Unicode mailing list	
3. Information on the user community for the proposed characters (for example: size, demographics, information technology use, or publishing use) is included? Reference:		NO
4. The context of use for the proposed characters (type of use; common or rare) Reference:		common
5. Are the proposed characters in current use by the user community? If YES, where? Reference:		YES
6. After giving due considerations to the principles in the P&P document must the proposed characters be entirely in the BMP? If YES, is a rationale provided? If YES, reference:		NO
7. Should the proposed characters be kept together in a contiguous range (rather than being scattered)?		NO
8. Can any of the proposed characters be considered a presentation form of an existing character or character sequence? If YES, is a rationale for its inclusion provided? If YES, reference:		NO
9. Can any of the proposed characters be encoded using a composed character sequence of either existing characters or other proposed characters? If YES, is a rationale for its inclusion provided? If YES, reference:		NO
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? If YES, is a rationale for its inclusion provided? If YES, reference:		NO
11. Does the proposal include use of combining characters and/or use of composite sequences? 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:		NO
12. Does the proposal contain characters with any special properties such as control function or similar semantics? If YES, describe in detail (include attachment if necessary)		NO
13. Does the proposal contain any Ideographic compatibility characters? If YES, are the equivalent corresponding unified ideographic characters identified? If YES, reference:		NO