

JIS

JAPANESE INDUSTRIAL STANDARD

Alphanumeric Character Sets for Optical Recognition

JIS X 9001—1976

Translated and Published

by

Japanese Standards Association

Printed in Japan

Translation without guarantee
In the event of any doubt arising, the original
standard in Japanese is to be evidence

JAPANESE INDUSTRIAL STANDARD

J I S

Alphanumeric Character Sets for
Optical RecognitionX 9001-1976
(Reaffirmed: 1985)1. Scope

This standard specifies the printed image shapes and sizes of alphanumeric characters and symbols designed for use in optical character recognition.

2. Meanings of Terms

The meanings of main terms used in this standard are as follows:

- (1) character A figure for denoting information. Alphabets, numerics and symbols are considered as characters.
- (2) character shape A shape of each character.
- (3) font A set of character shapes specified uniformly for a group of characters. Irrespective of variation of the size and the proportion of height and width for the whole of sets, the font is regarded as identical.
- (4) OCR-A, OCR-B Names of fonts to be specified in this standard. (See 3. Kinds of Fonts and Their Names)
- (5) centerline shape A shape drawn by a line or lines for showing the framework of a character.
- (6) printing frame A rectangle limiting dimensions of a centerline shape as a basic condition for font design. It depends upon the font and the size.

Remarks 1. The centerline shapes of Long Vertical Line, and some lower case alphabet letters and symbols of OCR-B may extend from the rectangle.

2. The centerline shapes of Character Erase and Group Erase are not defined in this standard.

- (7) strokewidth A width of character strokes forming a character shape.
- (8) base line A horizontal line used to determine the relative vertical position of characters printed on the same line.

Reference Standards:

JIS X 9010-Coding of OCR Character Sets for Information Interchange

JIS X 9003-KATAKANA Character Set for Optical Recognition

JIS X 9004-Printing Specifications for Optical Character Recognition

- (9) reference line A straight line perpendicular to the base line used to determine the relative horizontal position of printed characters.
- (10) constant-strokewidth character shape A character shape with a constant strokewidth.
- (11) variable-strokewidth character shape A character shape with a variable strokewidth.

Remark: For aesthetic reasons, the strokewidth of the variable-strokewidth character shape is varied deliberately.

3. Kinds of Fonts and Their Names

Two fonts are specified, and these are named OCR-A and OCR-B.

4. Character Sizes

The sizes of characters are indicated with the width "W" and the height "H" or printing frames, and are specified as shown in Table 1.

Table 1. Character Sizes

Unit: mm

Size designation		OCR-A		OCR-B	
		W	H	W	H
Size	I	1.40	2.40	1.40	2.40
	III	1.52	3.20	1.52	3.20
	IV	2.04	3.80	2.10	3.60

5. Character Set

5.1 OCR-A The full character set comprises 62 characters of 10 numerics 26 uppercase alphabets and 26 symbols as shown in Table 2.

5.2 OCR-B The full character set comprises 102 characters of 10 numerics, 26 uppercase alphabets, 26 lowercase alphabets and 40 symbols as shown in Table 2.

6. Kinds of Sub-set

6.1 OCR-A The numeric subset comprising 14 characters can be provided. (See Table 2)

6.2 OCR-B The following three sub-sets can be provided. (See Table 2)

- (1) Sub-set 1 (Numeric Sub-set comprising 22 characters)
- (2) Sub-set 2 (Alphanumeric Sub-set comprising 47 characters)
- (3) Sub-set 3 (Extended Alphanumeric Sub-set comprising 97 characters)

7. Character Shape

- 7.1 OCR-A The character shapes are specified in detail in Annex 1.
- 7.2 OCR-B The character shapes are specified in detail in Annex 2.

8. Relative Position of Printed Characters

- (1) The vertical relative position of characters printed on the same line is determined by the base line.
- (2) The horizontal relative position of printed characters is determined by the reference line.

9. Space (SP)

The shape (SP) is an intentionally blank position in a line of printing. With constant pitch printing its nominal width is equal to the printing pitch. With variable pitch printing its nominal width is equal to the largest character pitch available.

10. Character Erase and Group Erase

Character erase and group erase may use together with each sub-set of OCR-A or OCR-B.

- (1) Character erase may be used as it is printed correctly over any other printed character to erase it, or as it stands alone.
- (2) Group erase is used by being printed over all characters to be deleted in order to erase a string of printed characters.

Table 2. Character Set

Font	Name	OCR-A ⁽¹⁾				OCR-B						
		(2) Char. shape	(3) Page	(4) Sub- set	Remark	(2) Char. shape	Page ⁽³⁾			Subset ⁽⁴⁾	Size ⁽⁵⁾	Remark
							L	C	III			
1		1	11	N		1	35	63	91	1, 2, 3	I, III, IV	Drawing No. 1
2		2	11	N		2	35	63	91	1, 2, 3	I, III, IV	2
3		3	11	N		3	35	63	91	1, 2, 3	I, III, IV	3
4		4	11	N		4	35	63	91	1, 2, 3	I, III, IV	4
5		5	12	N		5	36	64	92	1, 2, 3	I, III, IV	5
6		6	12	N		6	36	64	92	1, 2, 3	I, III, IV	6
7		7	12	N		7	36	64	92	1, 2, 3	I, III, IV	7
8		8	12	N		8	36	64	92	1, 2, 3	I, III, IV	8
9		9	13	N		9	37	65	93	1, 2, 3	I, III, IV	9
0		0	13	N		0	37	65	93	1, 2, 3	I, III, IV	10
A		A	13	—		A	37	65	—	2, 3	I, IV	11
B		B	13	—		B	37	65	—	2, 3	I, IV	12
C		C	14	—		C	38	66	93	1, 2, 3	I, III, IV	13
D		D	14	—		D	38	66	—	2, 3	I, IV	14

Numerics



Documents

Document Sharing Hub

Uppercase alphabets



Documents

Document Sharing Hub

D	D	14	—		D	38	66	—	2, 3	I, IV	14
E	E	14	—		E	38	66	93	1, 2, 3	I, III, IV	15
F	F	14	—		F	38	66	—	2, 3	I, IV	16
G	G	15	—		G	39	67	—	2, 3	I, IV	17
H	H	15	—		H	39	67	—	2, 3	I, IV	18
I	I	15	—		I	39	67	—	2, 3	I, IV	19
J	J	15	—		J	39	67	—	2, 3	I, IV	20
K	K	16	—		K	40	68	—	2, 3	I, IV	21
L	L	16	—		L	40	68	—	2, 3	I, IV	22
M	M	16	—		M	40	68	—	2, 3	I, IV	23
N	N	16	—		N	40	68	94	1, 2, 3	I, III, IV	24
O	O	17	—		O	41	69	—	2, 3	I, IV	25
P	P	17	—		P	41	69	—	2, 3	I, IV	26
Q	Q	17	—		Q	41	69	—	2, 3	I, IV	27
R	R	17	—		R	41	69	—	2, 3	I, IV	28
S	S	18	—		S	42	70	94	1, 2, 3	I, III, IV	29
T	T	18	—		T	42	70	94	1, 2, 3	I, III, IV	30
U	U	18	—		U	42	70	—	2, 3	I, IV	31
V	V	18	—		V	42	70	—	2, 3	I, IV	32
W	W	19	—		W	43	71	—	2, 3	I, IV	33
X	X	19	—		X	43	71	94	1, 2, 3	I, III, IV	34
Y	Y	19	—		Y	43	71	—	2, 3	I, IV	35
Z	Z	19	—		Z	43	71	95	1, 2, 3	I, III, IV	36

Font Name		OCR-A (1)				OCR-B						
		(2) Char. shape	(3) Page	(4) Sub- set	Remark	(2) Char. shape	Page(3)			Subset(4)	Size(5)	Remark
							L	C	III			
a			—	—		a	44	72	—	3	I, IV	Drawing No. 37
b			—	—		b	44	72	—	3	I, IV	38
c			—	—		c	44	72	—	3	I, IV	39
d			—	—		d	44	72	—	3	I, IV	40
e			—	—		e	45	73	—	3	I, IV	41
f			—	—		f	45	73	—	3	I, IV	42
g			—	—		g	45	73	—	3	I, IV	43
h			—	—		h	45	73	—	3	I, IV	44
i			—	—		i	46	74	—	3	I, IV	45
j			—	—		j	46	74	—	3	I, IV	46
k			—	—		k	46	74	—	3	I, IV	47
l			—	—		l	46	74	—	3	I, IV	48
m			—	—		m	47	75	—	3	I, IV	49
m			—	—		m	59	87	—	—	I, IV	Alternative m See 4 of Annex 2 97
n			—	—		n	47	75	—	3	I, IV	50
o			—	—		o	47	75	—	3	I, IV	51
p			—	—		p	47	75	—	3	I, IV	52
q			—	—		q	48	76	—	3	I, IV	53

q	—	—	—	q	48	76	—	3	I, IV	53
r	—	—	—	r	48	76	—	3	I, IV	54
s	—	—	—	s	48	76	—	3	I, IV	55
t	—	—	—	t	48	76	—	3	I, IV	56
u	—	—	—	u	49	77	—	3	I, IV	57
v	—	—	—	v	49	77	—	3	I, IV	58
w	—	—	—	w	49	77	—	3	I, IV	59
x	—	—	—	x	49	77	—	3	I, IV	60
y	—	—	—	y	50	78	—	3	I, IV	61
z	—	—	—	z	50	78	—	3	I, IV	62
Asterisk	*	20	—	*	50	78	—	2, 3	I, IV	63
Plus sign	+	20	—	+	50	78	95	1, 2, 3	I, III, IV	64
Minus sign	—	20	—	—	51	79	—	2, 3	I, IV	65
Hyphen	-	27	—	Alternative shape see 2.1 of Annex 1						
Equal sign	=	20	—	=	51	79	—	2, 3	I, IV	66
Solidus	/	21	—	/	51	79	—	2, 3	I, IV	67
Full stop	•	21	—	•	51	79	—	2, 3	I, IV	68
	▪	27	—	Alternative shape see 2.1 of Annex 1						

Symbols



Documents

Document Sharing Hub

Font Name		OCR-A (1)				OCR-B							
		() Char. shape	() Page	() Sub- set	Remark	() Char. shape	Page()			Subset()	Size()	Remark	
							L	C	III				
	Comma	,	21	---		,	52	80	—		2, 3	I, IV	Drawing No. 69
	Colon	:	27	---	Alternative shape See 2.1 of Annex 1	:	52	80	—		3	I, IV	70
	Semi-colon	;	21	---		;	52	80	—		3	I, IV	71
	Quotation mark	"	22	---		"	52	80	—		3	I, IV	72
	Apostrophe	'	22	---	Alternative shape See 2.1 of Annex 1	'	53	81	—		3	I, IV	73
	Discontinuous underline		27	---		—	53	81	—		3	I, IV	74
	Question mark	?	—	---		?	53	81	—		3	I, IV	75
	Exclamation mark	?	22	---	Alternative shape See 2.1 of Annex 1	!	53	81	—		3	I, IV	76
	Left parenthesis		28	---		(	54	82	—		3	I, IV	77
	Right parenthesis		---	---		)	54	82	—		3	I, IV	78
	Less than sign		---	---		<	54	82	95		1, 2, 3	I, III, IV	79

Symbols

Font		OCR-A ⁽¹⁾					OCR-B						
		Name	(²) Char. shape	(³) Page	(⁴) Sub- set	Remark	(²) Char. shape	Page(³)			Subset(⁴)	Size(⁵)	Remark
			L	C	III								
Abstract symbols	Character erase	■	25	—			■	—	—	—	—	I, IV	See 1. of Annex 2
	Group erase	—	25	—			—	—	—	—	I, III, IV		
	Hook	フ	26	N				—	—	—	—		
	Fork	フ	26	N				—	—	—	—		
	Chair	フ	26	N				—	—	—	—		

Notes: (1) The character shapes for all "Size" of OCR-A are specified in this standard.

(2) The intentionally blank position in the "Character Shape" column means not to specify the corresponding character shape in this standard. The space character is designated by (SP).

(3) This column contains pages where the drawings of character shapes are inserted in this standard.

For OCR-B, the "L" column, the "C" column and the "III" column contain the drawings pages of the variable-strokewidth character shapes for Size I, the constant-strokewidth character shapes for Size I and the constant-strokewidth character shapes for size III respectively.

(4) This column indicates that the corresponding character belongs to the following sub-sets:

N: OCR-A Numeric sub-set

1: OCR-B Sub-set 1 (Numeric Sub-set)

2: OCR-B Sub-set 2 (Alphanumeric Sub-set)

3: OCR-B Sub-set 3 (Extended alphanumeric Sub-set)

(5) This column indicates "Size" for which the corresponding character shape is specified in this standard.

Annex 1 OCR-A

1. Character Shapes and Dimensions of OCR-A

The character shapes and their dimensions of OCR-A are specified by the drawings described in 11 to 28 pages and the table of values listed in 10 page.

The positions of the base line and the reference line are indicated on the drawings of all characters. The reference line is perpendicular to the base line and divides the printing frame of a character equally.

For characters whose nominal centerline height are not H , the shortest distance between the base line and the centerline shape of those characters is indicated by Y . The value of Y is indicated on the corresponding drawings.

2. Use of Symbols

2.1 Alternate Shapes For the characters of apostrophe, comma, full stop, hyphen and question mark, their alternate shapes may be used depending on requirements in some particular OCR applications.

2.2 Abstract Symbol Ψ and Letter Y The abstract symbol Ψ and the letter Y must not both be used in the same application.

10
X 9001-1976

OCR-A Table of Values

Unit: mm

Size	I	III	IV
H (Height of printing frame)	2.400	3.200	3.800
W (Width of printing frame)	1.400	1.520	2.040
T (Strokewidth)	0.350	0.380	0.510
$\frac{1}{2} T$	0.175	0.190	0.255
$\frac{3}{2} T$	0.525	0.570	0.765
$2 T$	0.700	0.760	1.020
$\frac{1}{8} W$	0.175	0.190	0.255
$\frac{3}{8} W$	0.263	0.285	0.383
$\frac{1}{4} W$	0.350	0.380	0.510
$\frac{5}{8} W$	0.525	0.570	0.765
$\frac{1}{2} W$	0.700	0.760	1.020
$\frac{5}{8} W$	0.875	0.950	1.275
$\frac{3}{4} W$	1.050	1.140	1.530
$\frac{1}{6} H$	0.150	0.200	0.238
$\frac{1}{6} H$	0.300	0.400	0.475
$\frac{1}{6} H$	0.450	0.600	0.713
$\frac{1}{4} H$	0.600	0.800	0.950
$\frac{5}{6} H$	0.750	1.000	1.188
$\frac{3}{8} H$	0.900	1.200	1.425
$\frac{1}{6} H$	1.050	1.400	1.663
$\frac{1}{2} H$	1.200	1.600	1.900
$\frac{5}{8} H$	1.500	2.000	2.375
$\frac{3}{4} H$	1.800	2.400	2.850
$\frac{7}{8} H$	2.100	2.800	3.325
$r_1^{(1)}$	0.635	1.022	1.084
$r_2^{(1)}$	0.283	0.283	0.398
$r_3^{(1)}$	0.255	0.266	0.366
$r_4^{(1)}$	0.223	0.310	0.360
$L^{(2)}$	3.7	5.0	5.9

Notes (1) There are the following relationship among r_1, r_2, r_3, r_4, W and H .

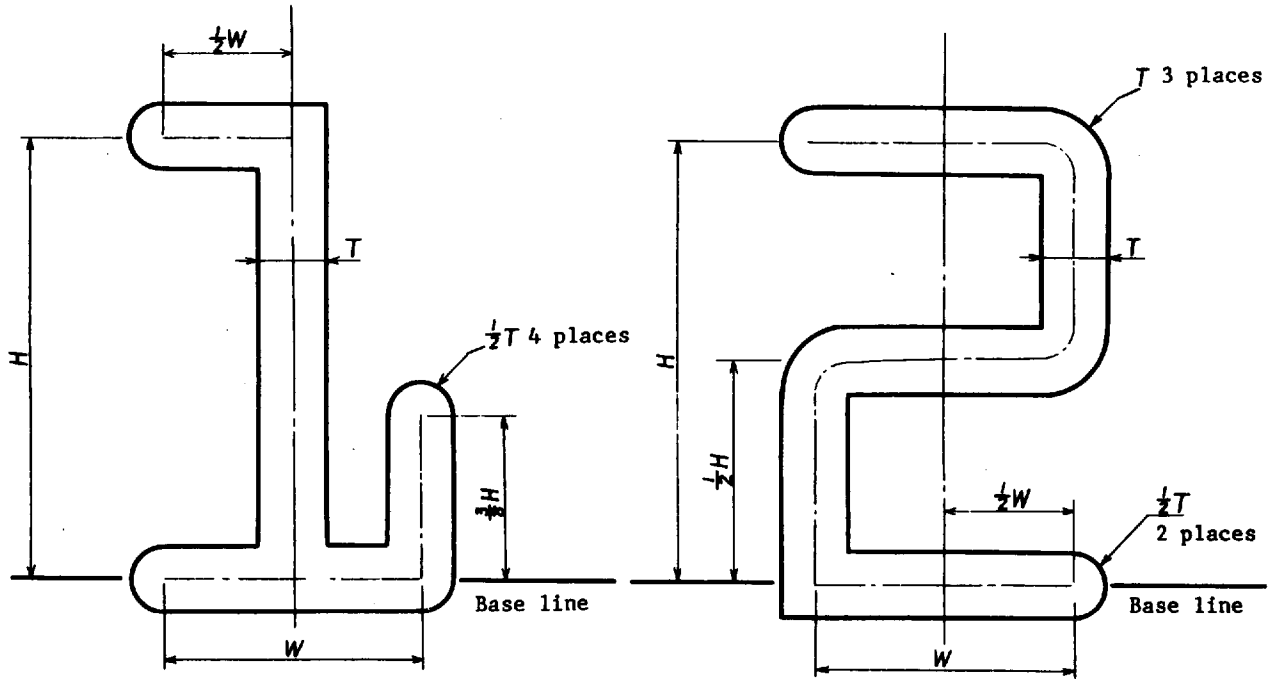
$$r_1 = \frac{H}{16} \left(\frac{7H}{6W} + \sqrt{\left(\frac{7H}{6W} \right)^2 + 1} \right)$$

$$r_2 = \frac{W}{8} \left(\frac{6W}{7H} + \sqrt{\left(\frac{6W}{7H} \right)^2 + 1} \right)$$

$$r_3 = \frac{W}{8} \left(1 - \frac{H}{2W} + \sqrt{\left(\frac{H}{2W} \right)^2 + 1} \right)$$

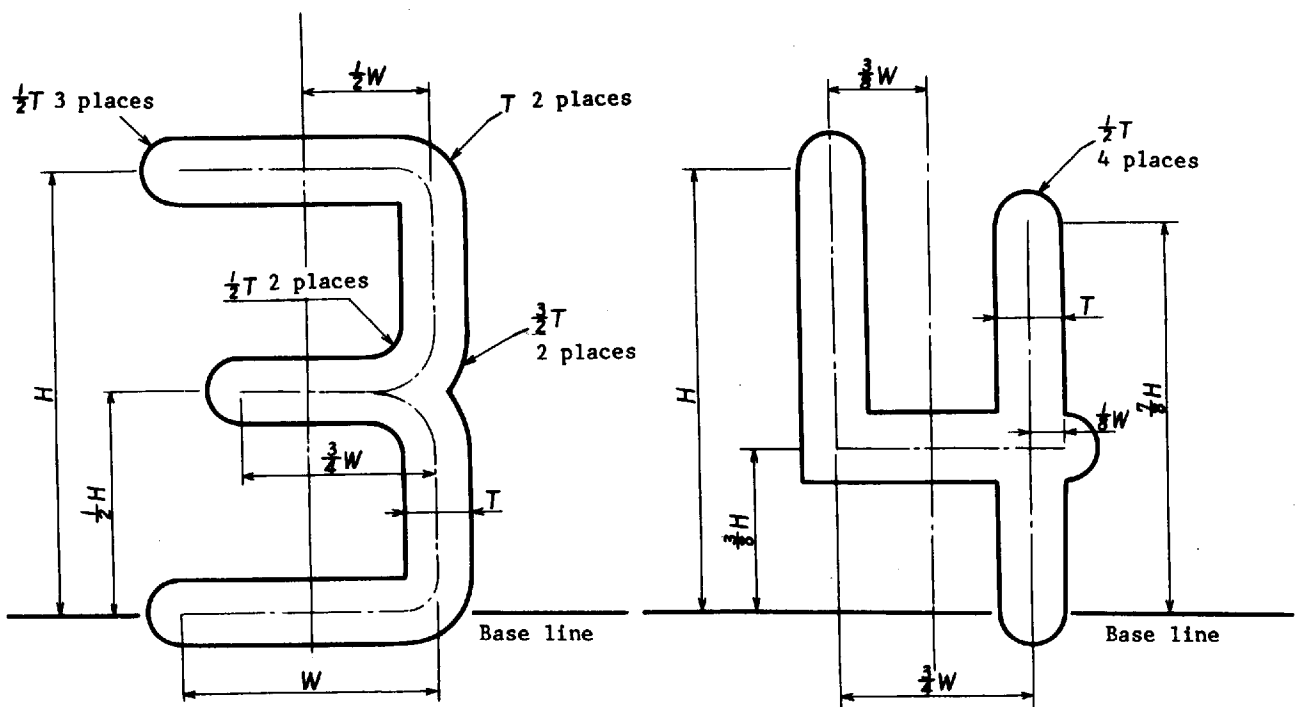
$$r_4 = \frac{H}{16} \left(1 - \frac{4W}{3H} + \sqrt{\left(\frac{4W}{3H} \right)^2 + 1} \right)$$

(2) L denotes the minimum length of the long vertical mark and corresponds to $\frac{14}{9} H$.



Numeral 1

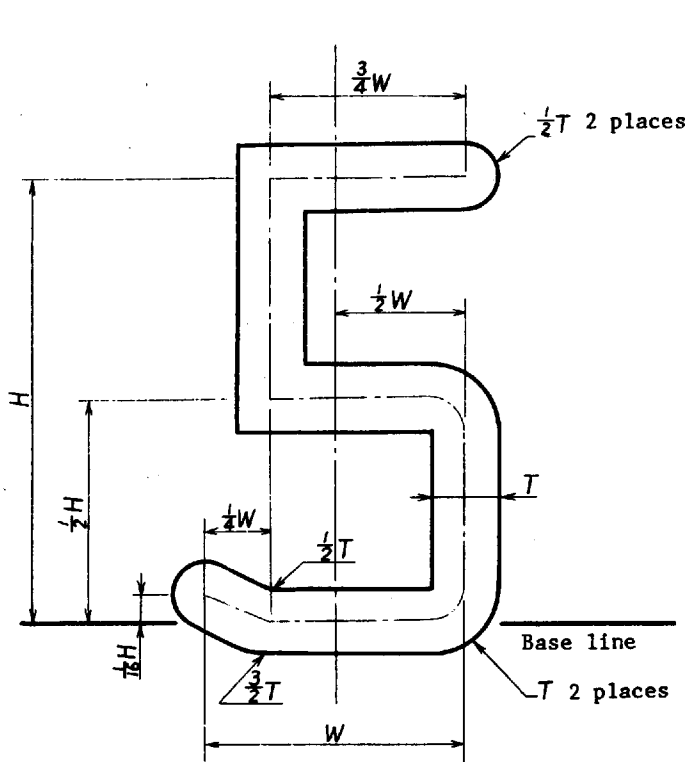
Numeral 2



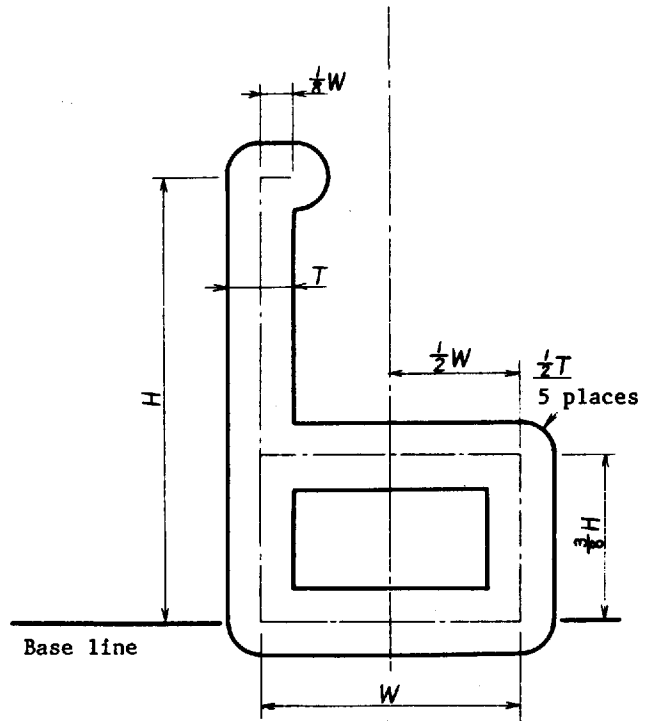
Numeral 3

Numeral 4

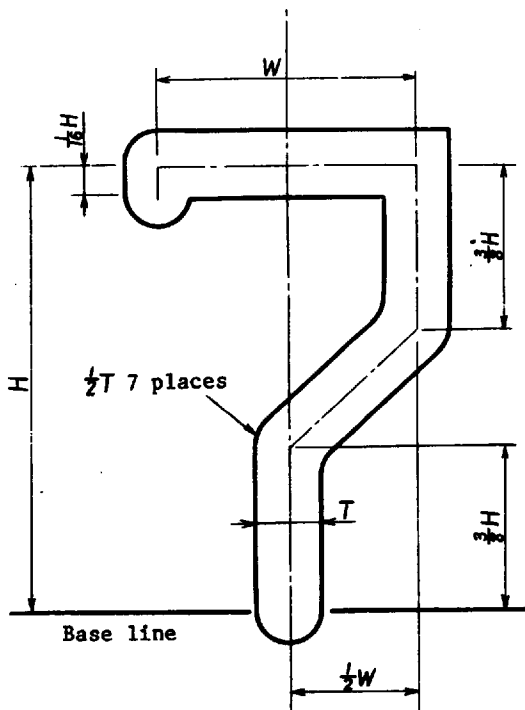
12
X 9001-1976



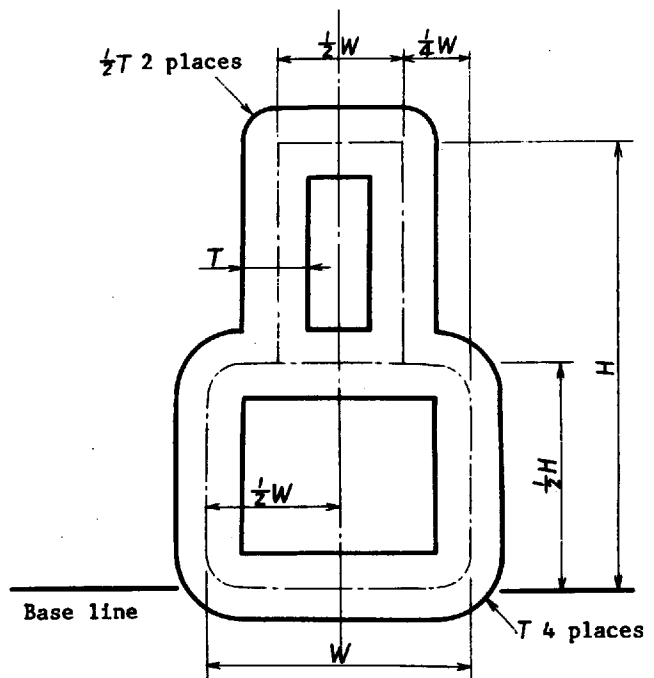
Numeral 5



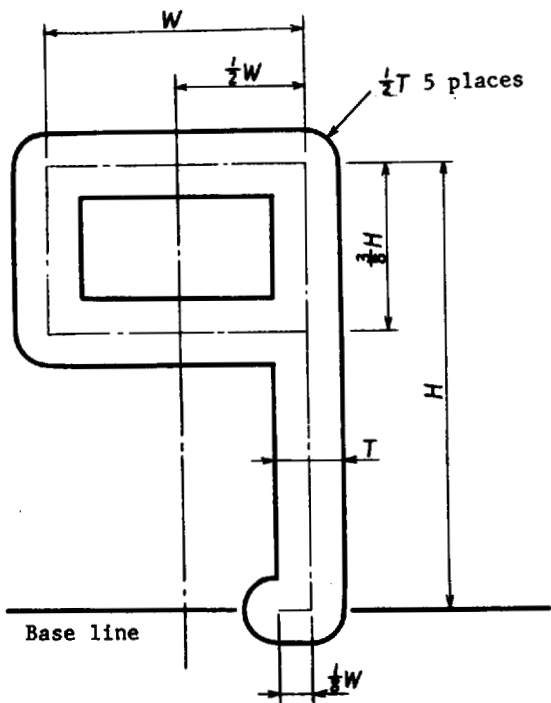
Numeral 6



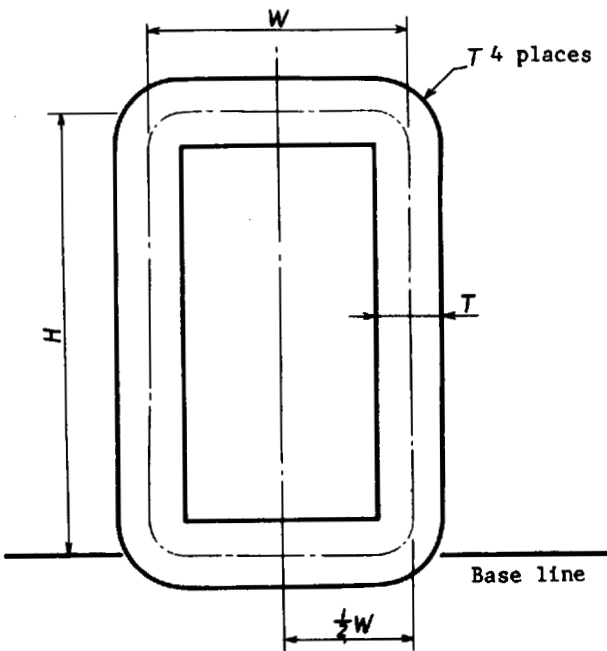
Numeral 7



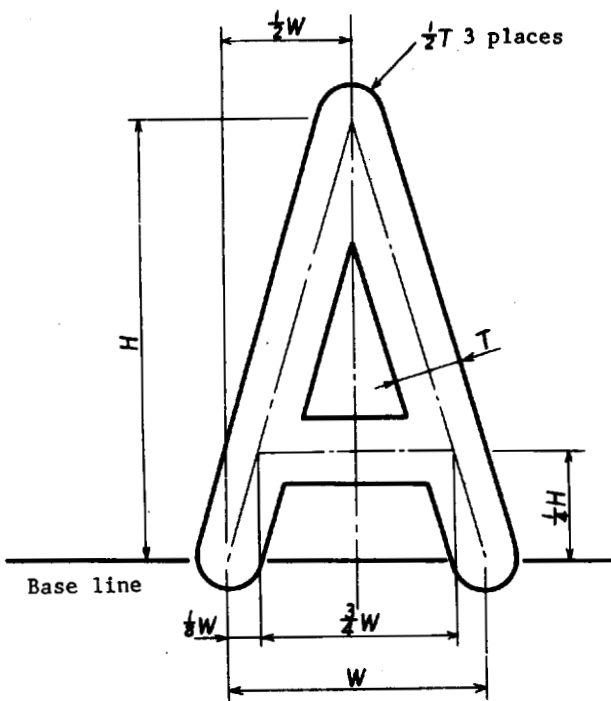
Numeral 8



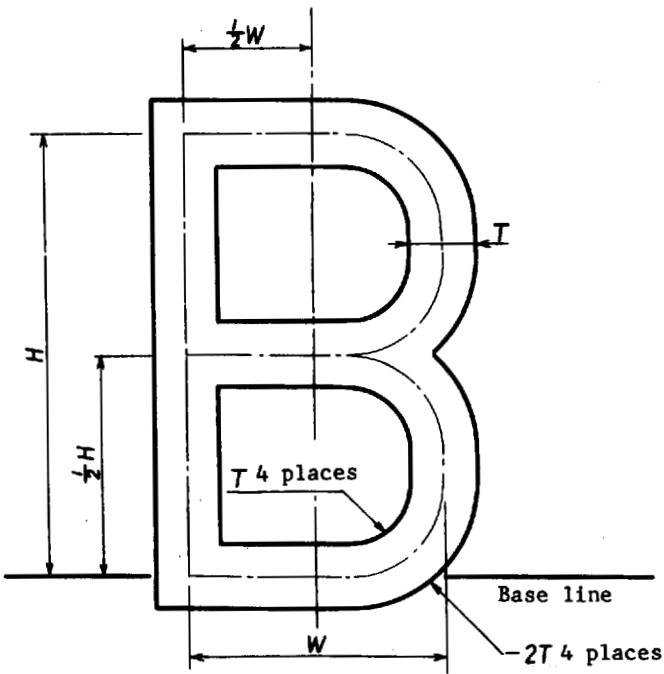
Numeral 9



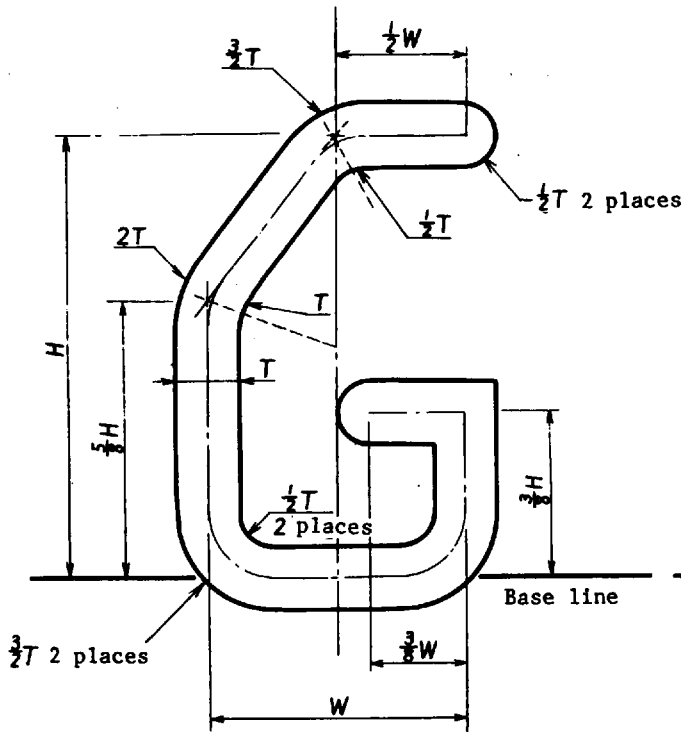
Numeral 0



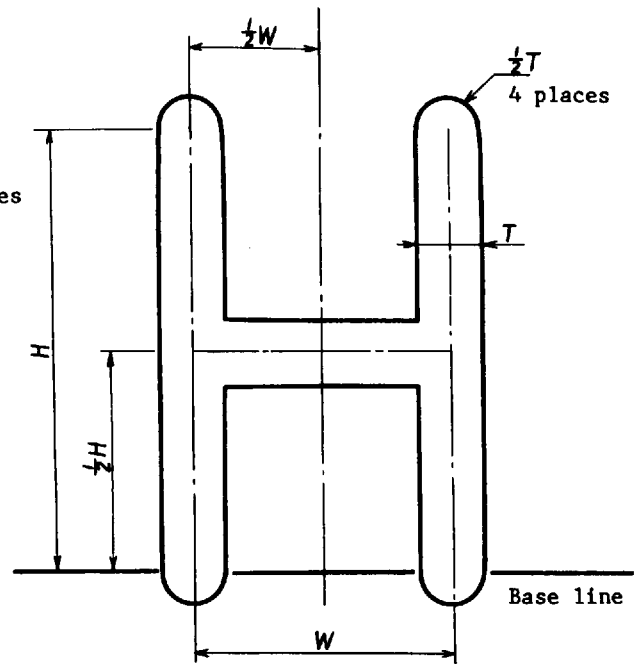
Uppercase alphabet A



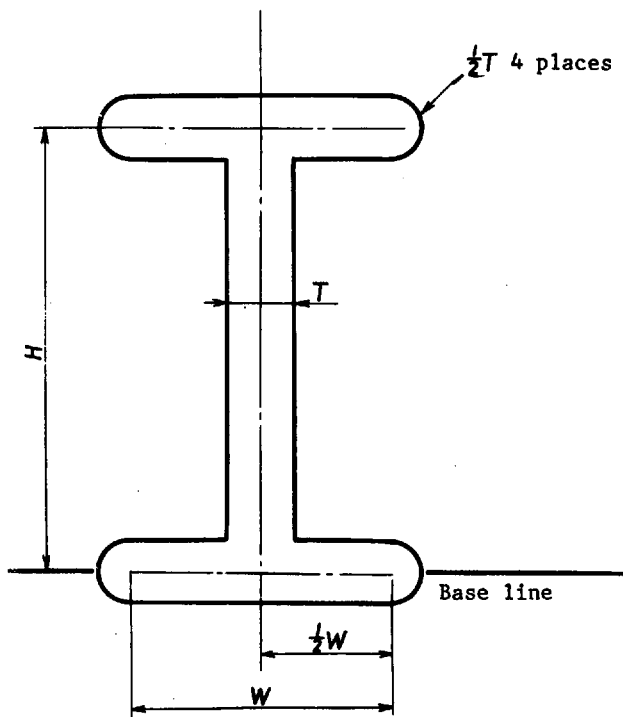
Uppercase alphabet B



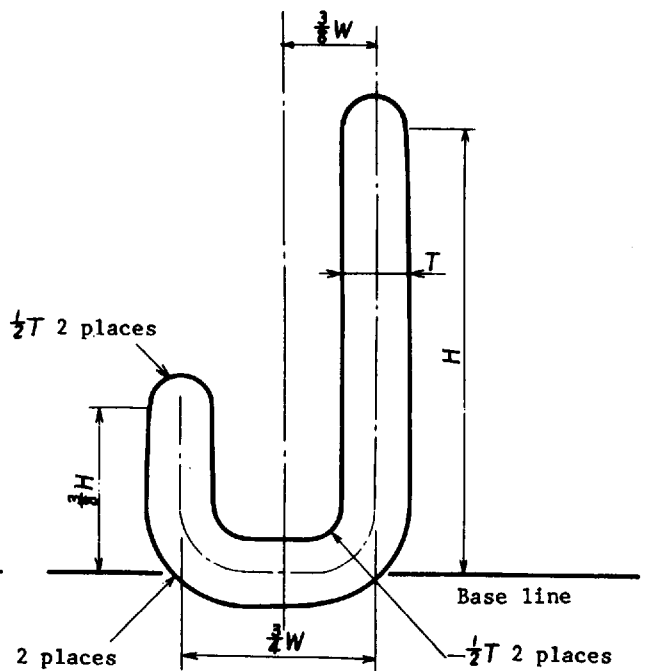
Uppercase alphabet G



Uppercase alphabet H

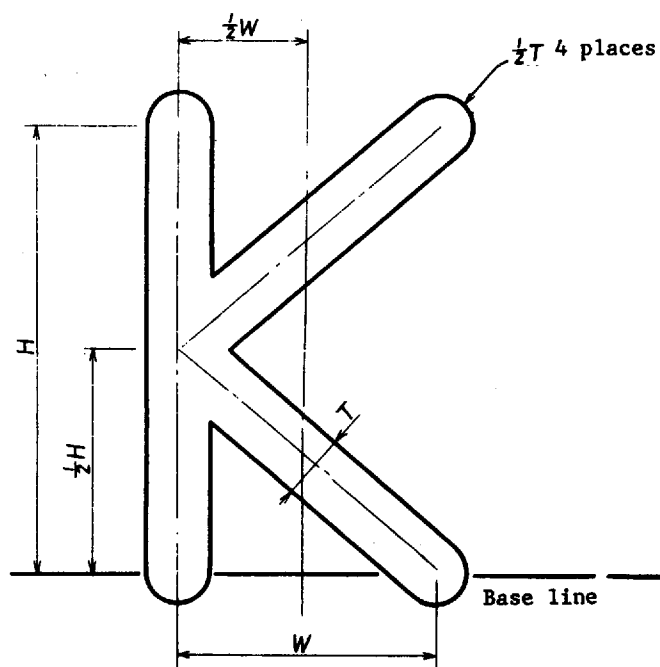


Uppercase alphabet I

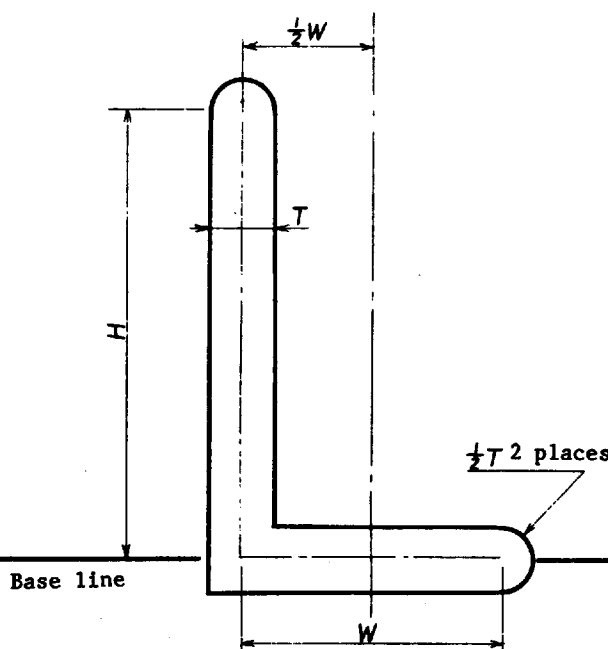


Uppercase alphabet J

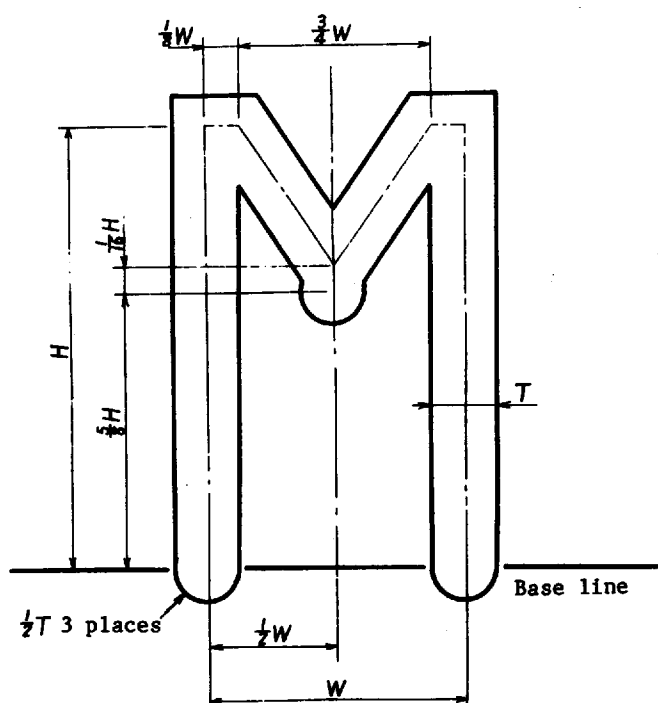
16
X 9001-1976



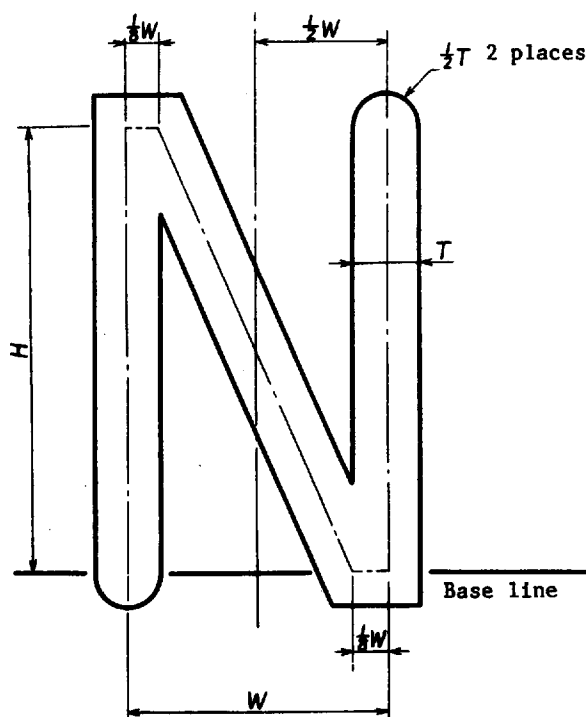
Uppercase alphabet K



Uppercase alphabet L



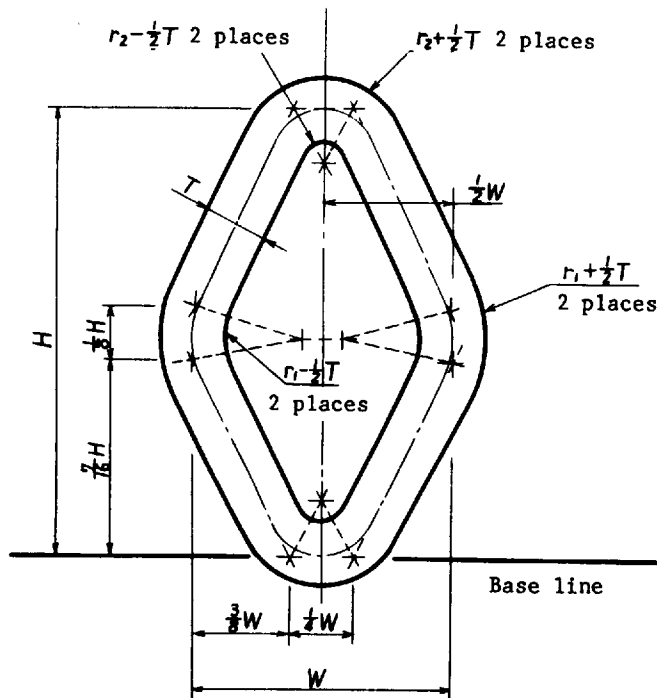
Uppercase alphabet M



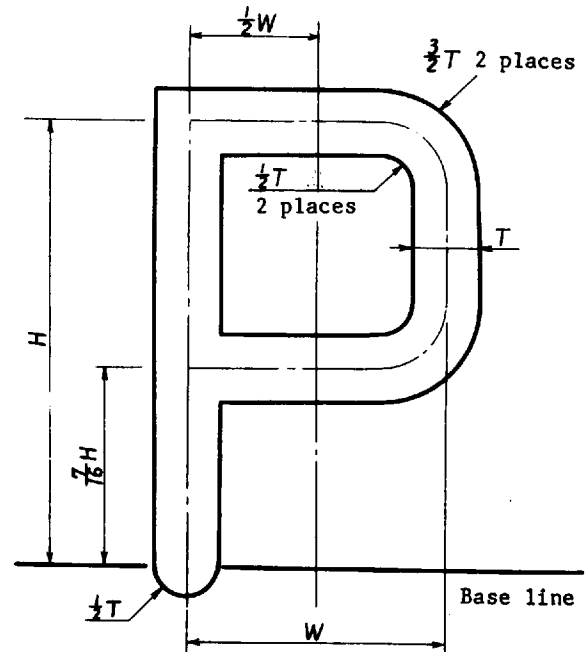
Uppercase alphabet N

$$r_1 = \frac{H}{16} \left(\frac{7H}{6W} + \sqrt{\left(\frac{7H}{6W} \right)^2 + 1} \right)$$

$$r_2 = \frac{W}{8} \left(\frac{6W}{7H} + \sqrt{\left(\frac{6W}{7H} \right)^2 + 1} \right)$$

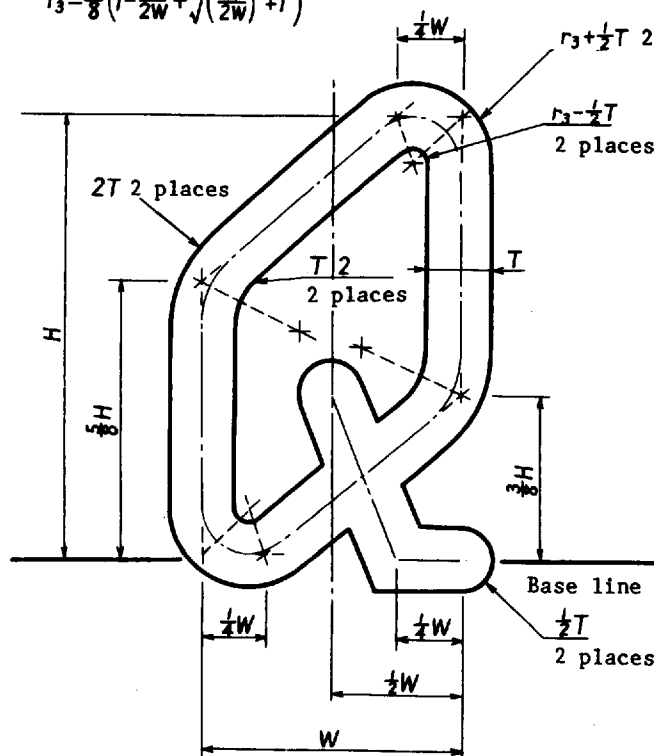


Uppercase alphabet O

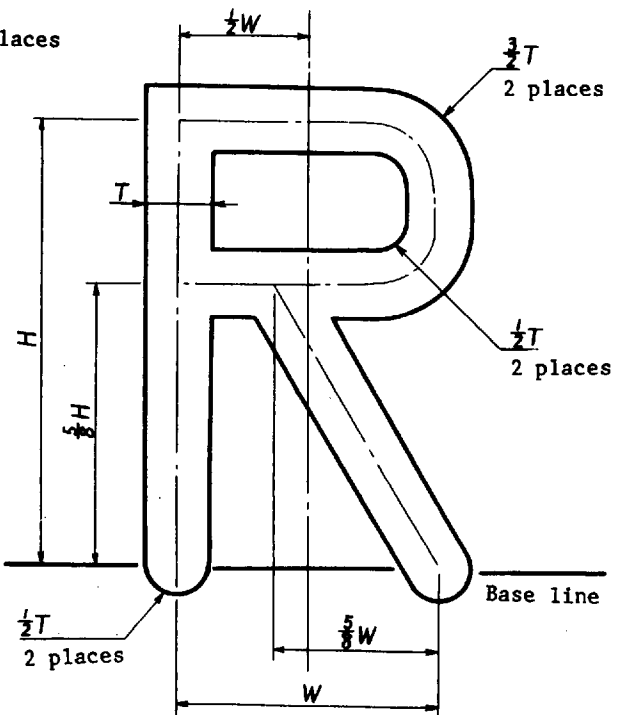


Uppercase alphabet P

$$r_3 = \frac{W}{8} \left(1 - \frac{H}{2W} + \sqrt{\left(\frac{H}{2W} \right)^2 + 1} \right)$$



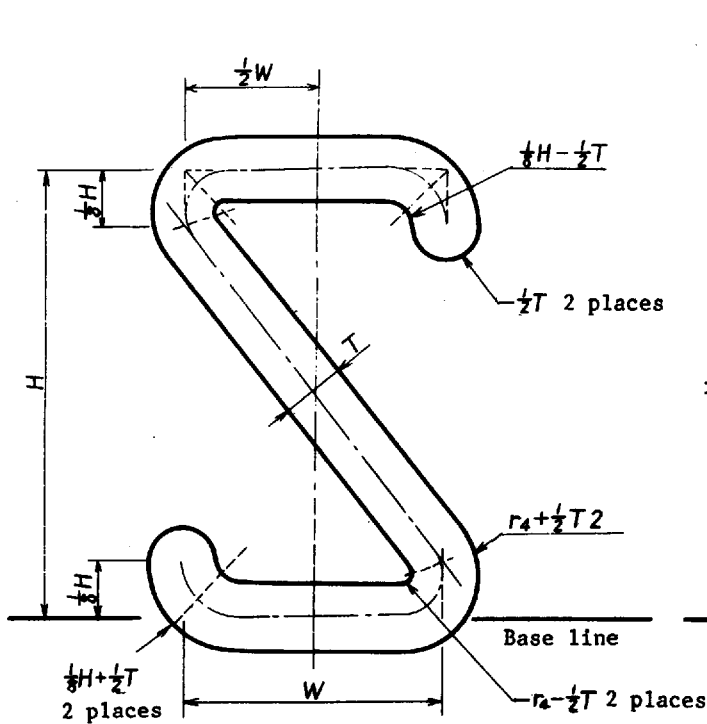
Uppercase alphabet Q



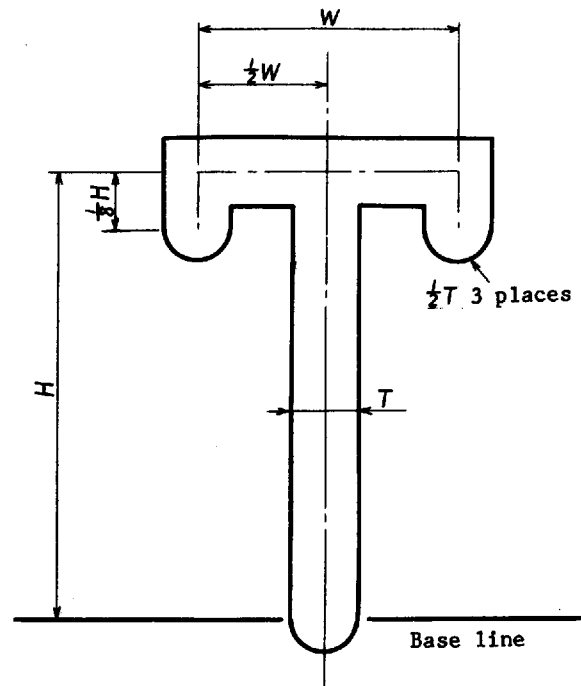
Uppercase alphabet R

18
X 9001-1976

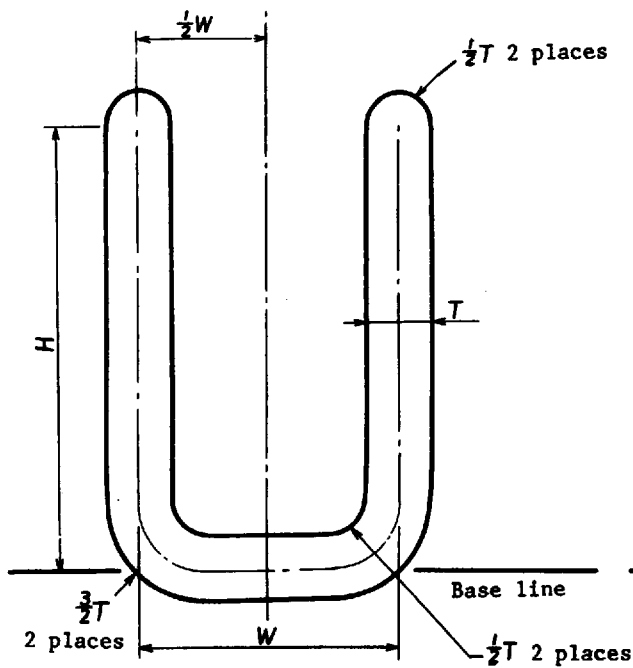
$$r_4 = \frac{H}{16} \left(1 - \frac{4W}{3H} + \sqrt{\left(\frac{4W}{3H} \right)^2 + 1} \right)$$



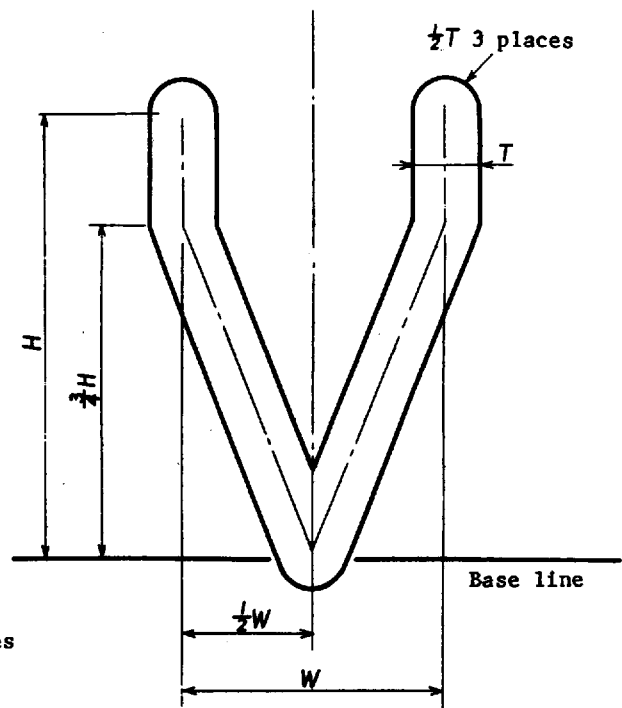
Uppercase alphabet S



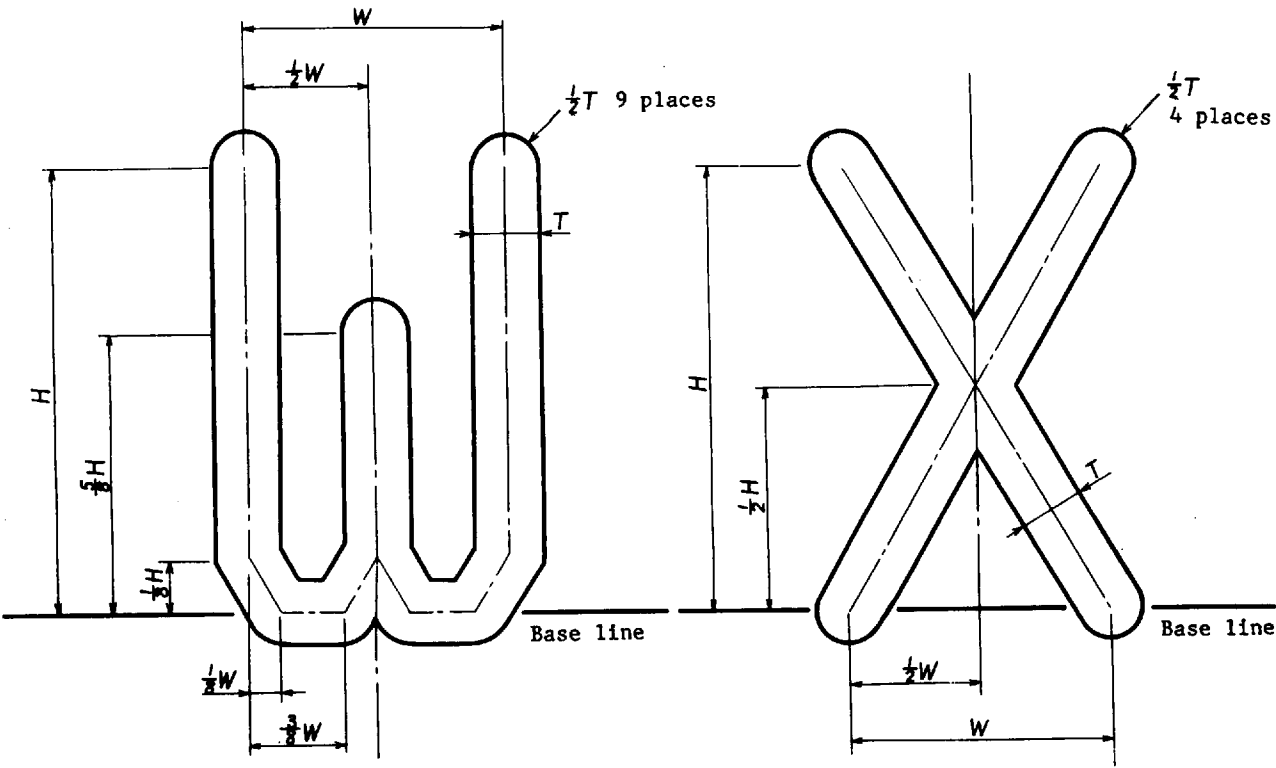
Uppercase alphabet T



Uppercase alphabet U

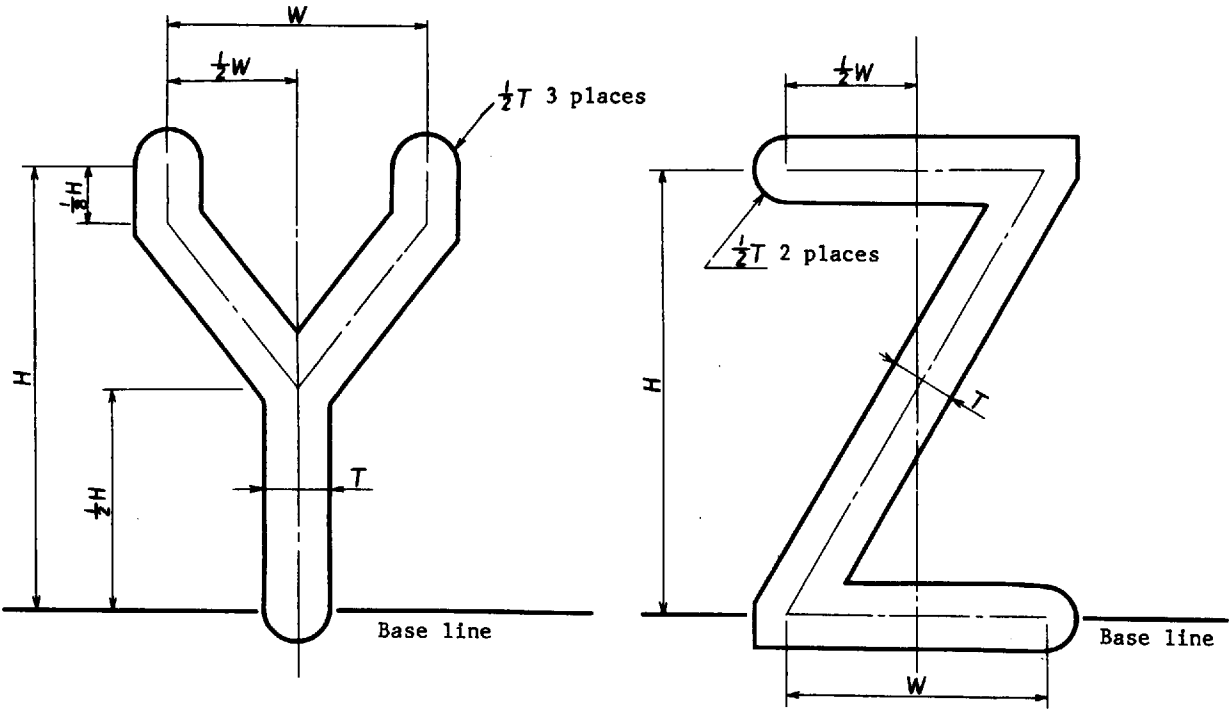


Uppercase alphabet V



Uppercase alphabet W

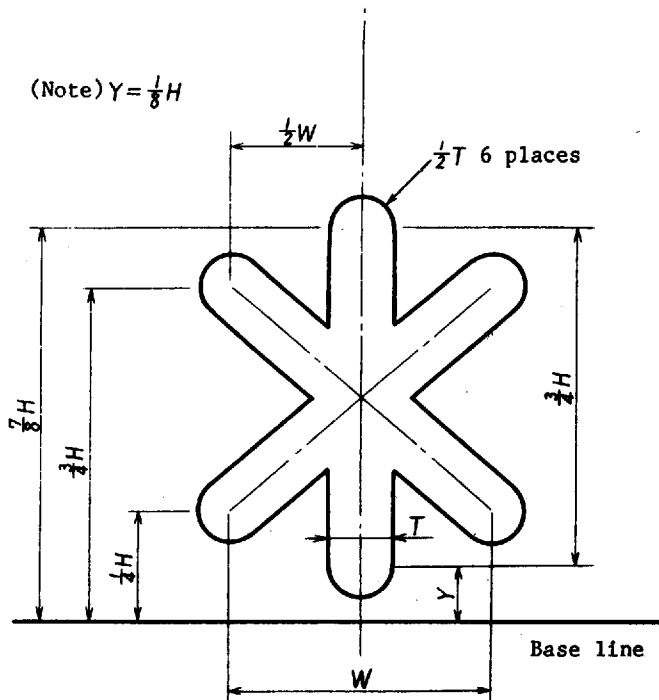
Uppercase alphabet X



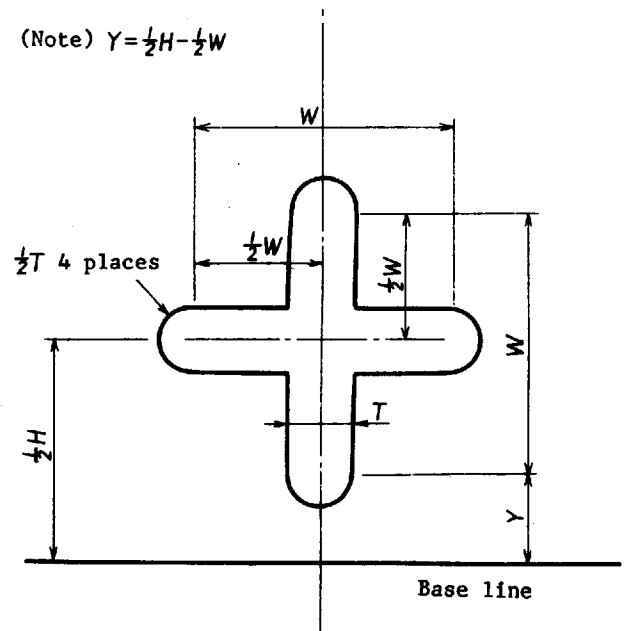
Uppercase alphabet Y

Uppercase alphabet Z

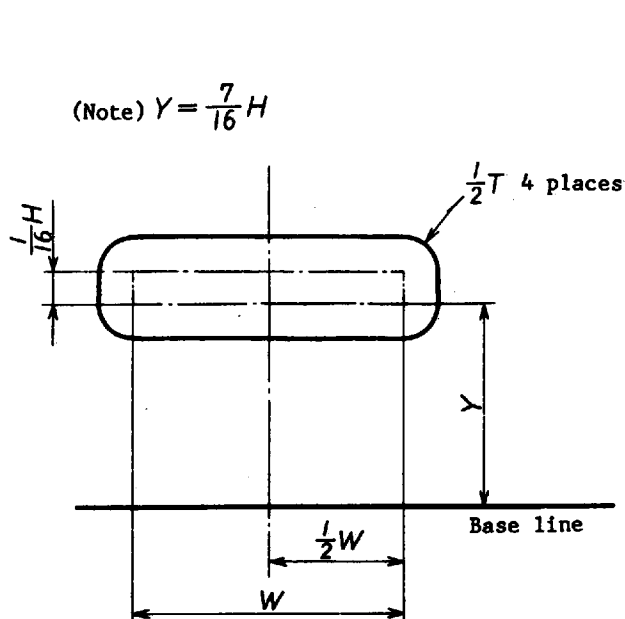
20
X 9001-1976



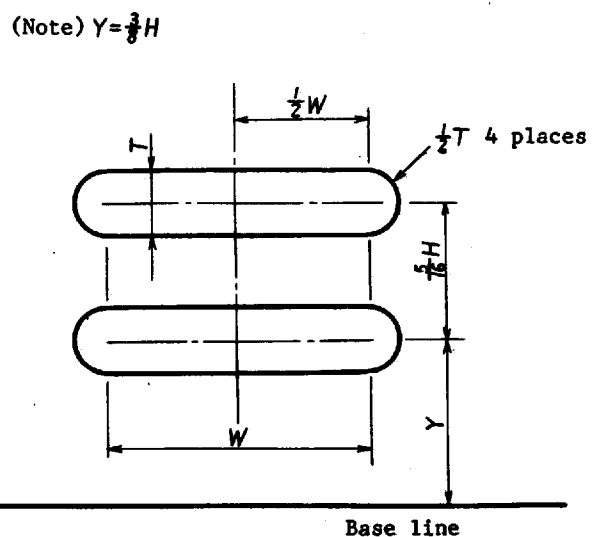
Asterisk



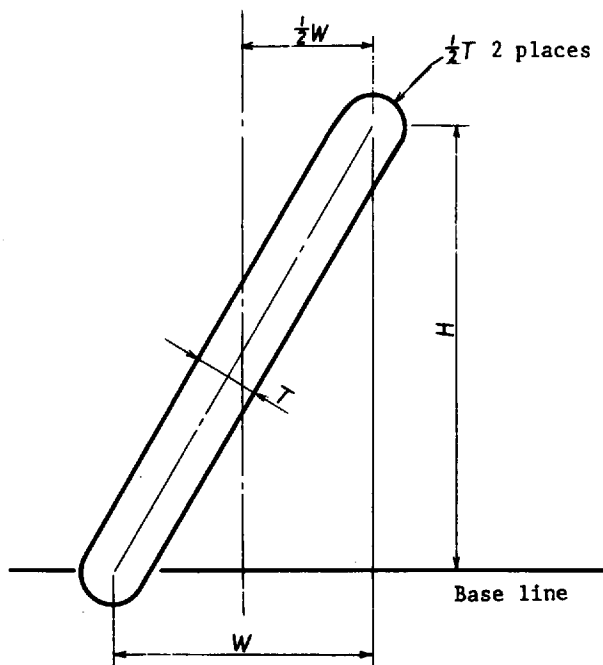
Plus sign



Minus sign, hyphen

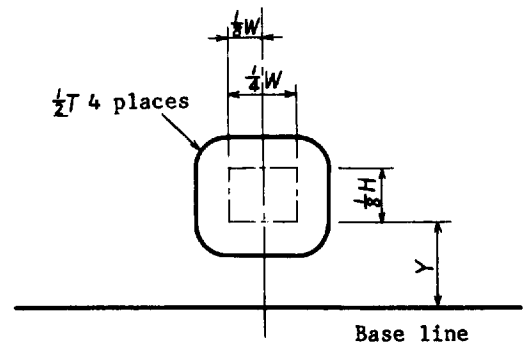


Equal sign



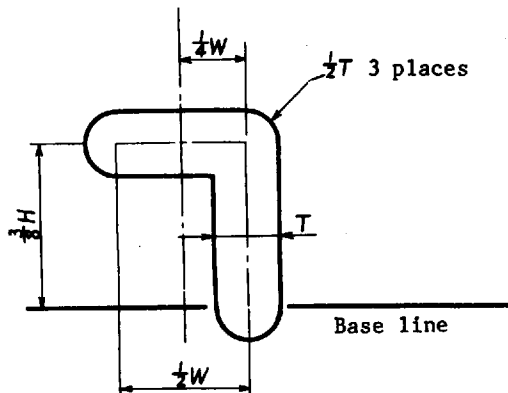
Solidus

(Note) $Y = \frac{3}{16} H$



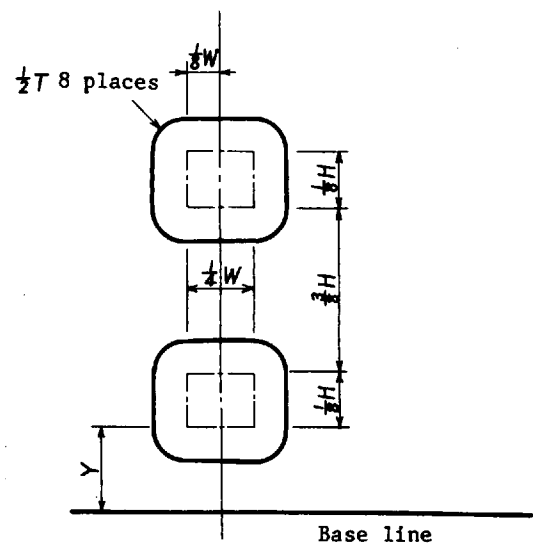
Full stop

(Note) $Y = 0$



Comma

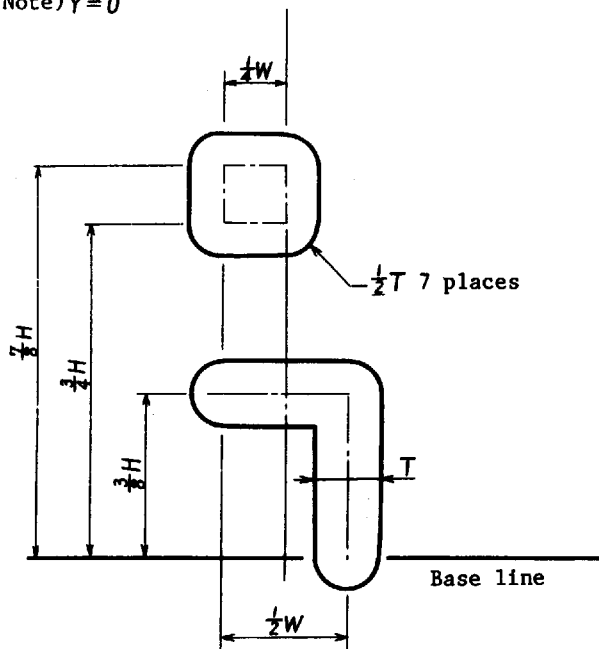
(Note) $Y = \frac{3}{16} H$



Colon

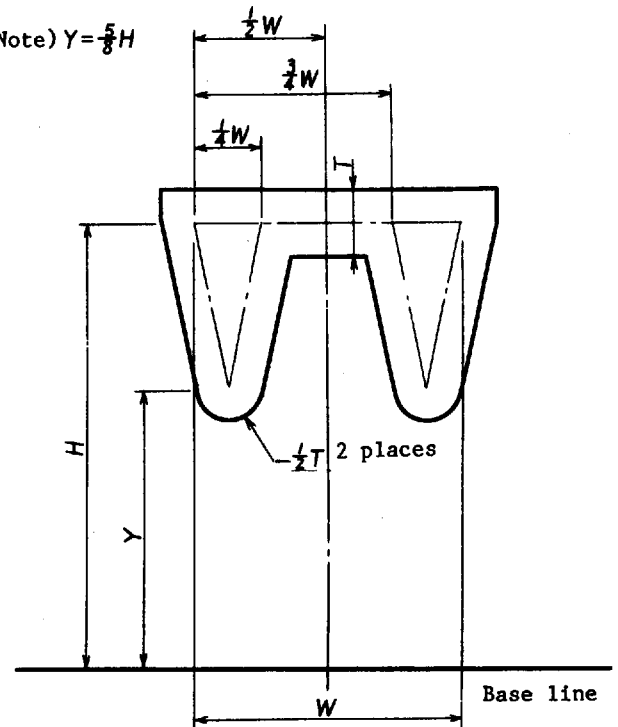
22
X 9001-1976

(Note) $Y=0$



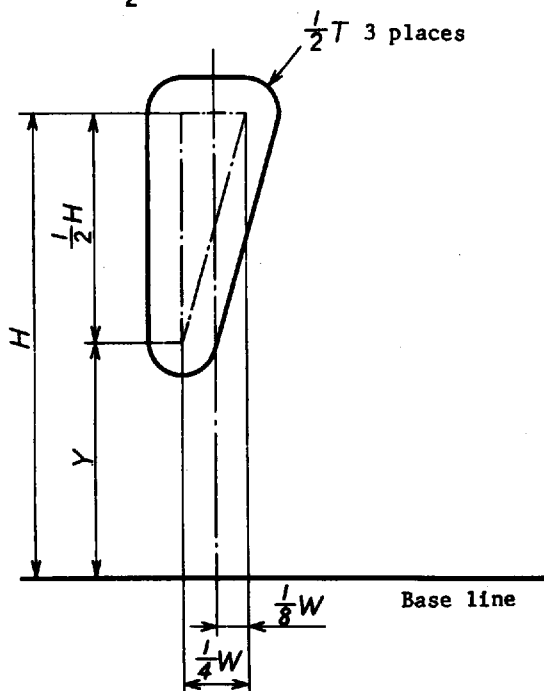
Semi-colon

(Note) $Y=\frac{5}{8}H$

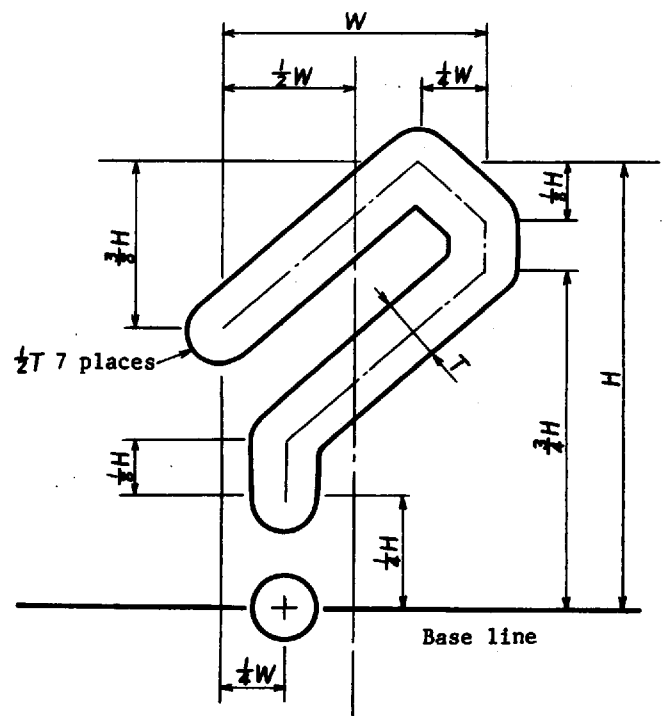


Quotation mark

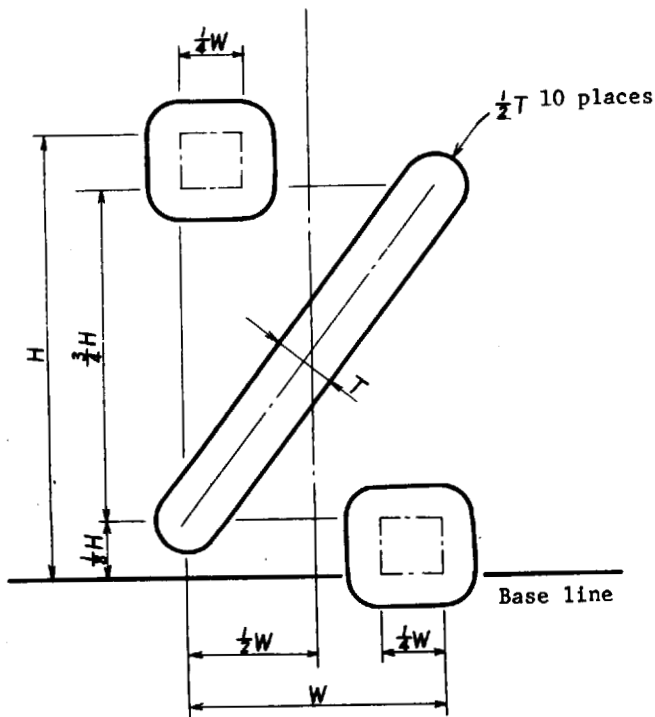
(Note) $Y=\frac{1}{2}H$



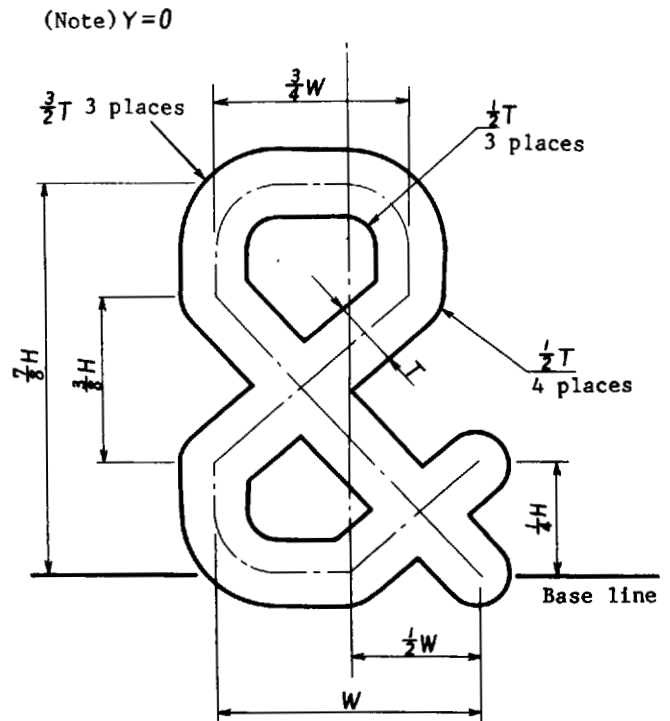
Apostrophe



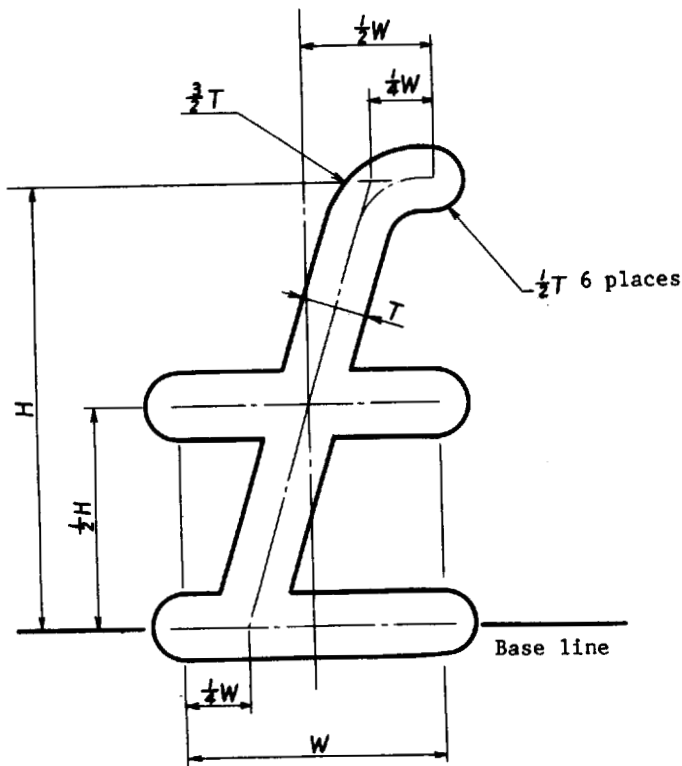
Question mark



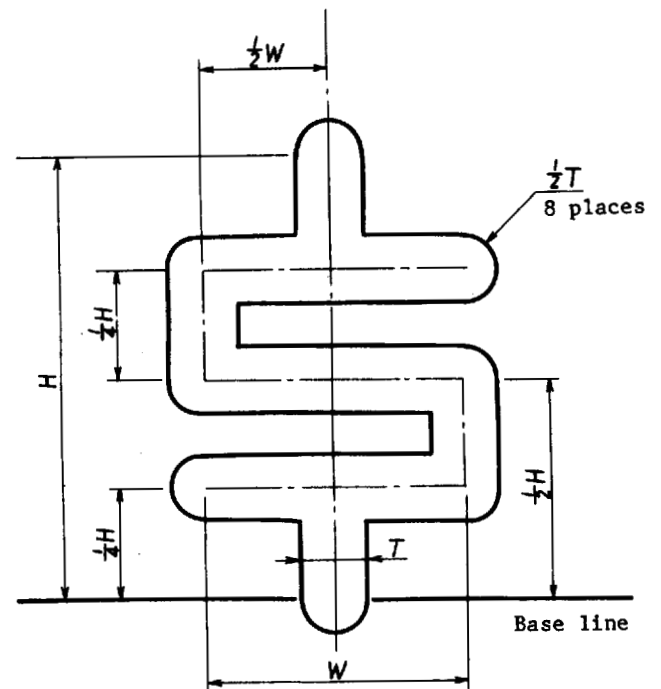
Percent



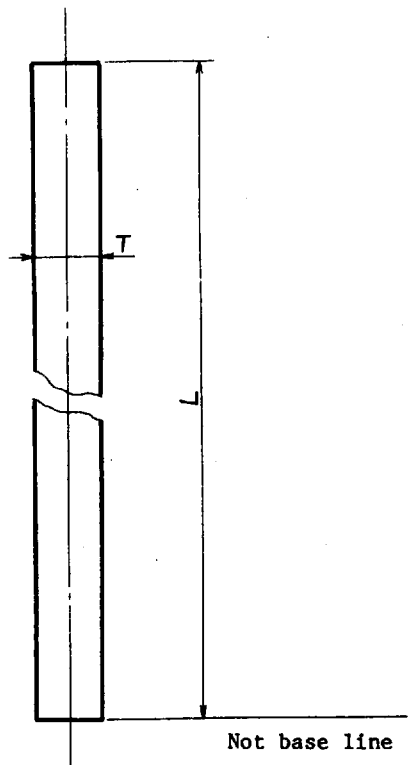
Ampersand



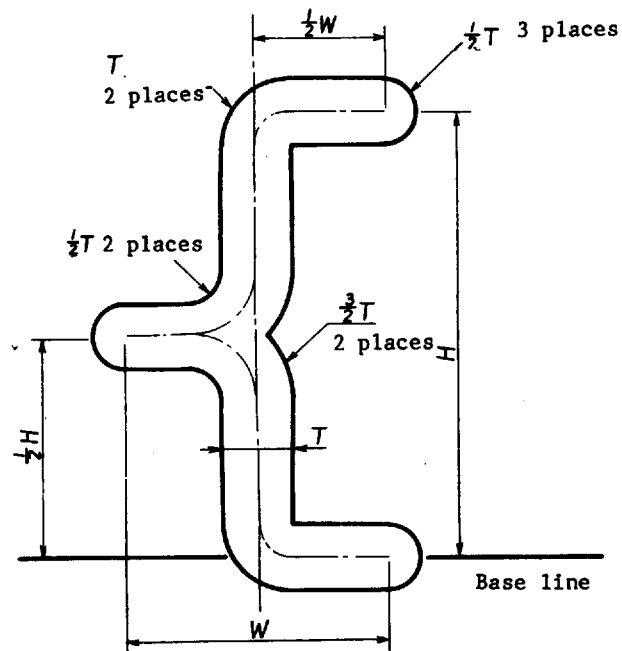
Pound sign



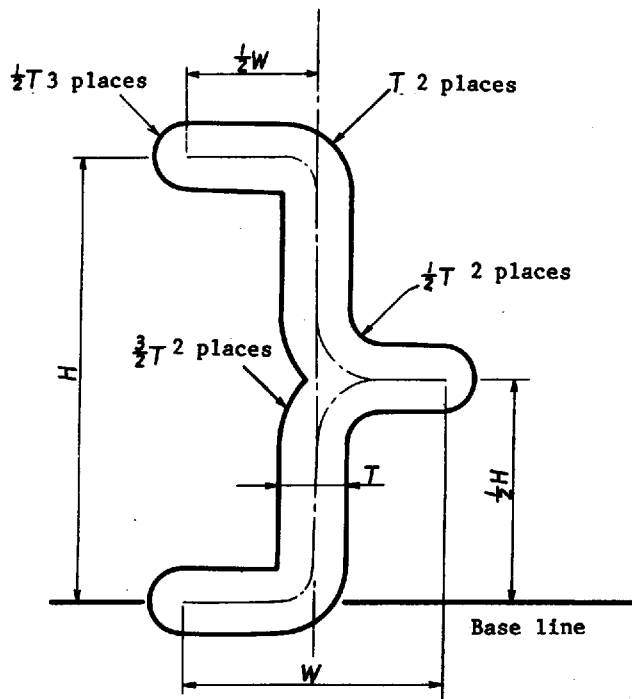
Dollar sign



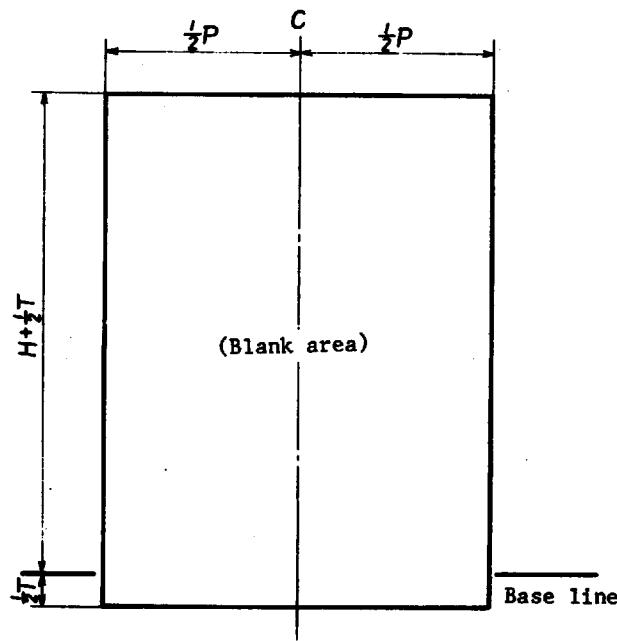
Long vertical mark



Left curly bracket

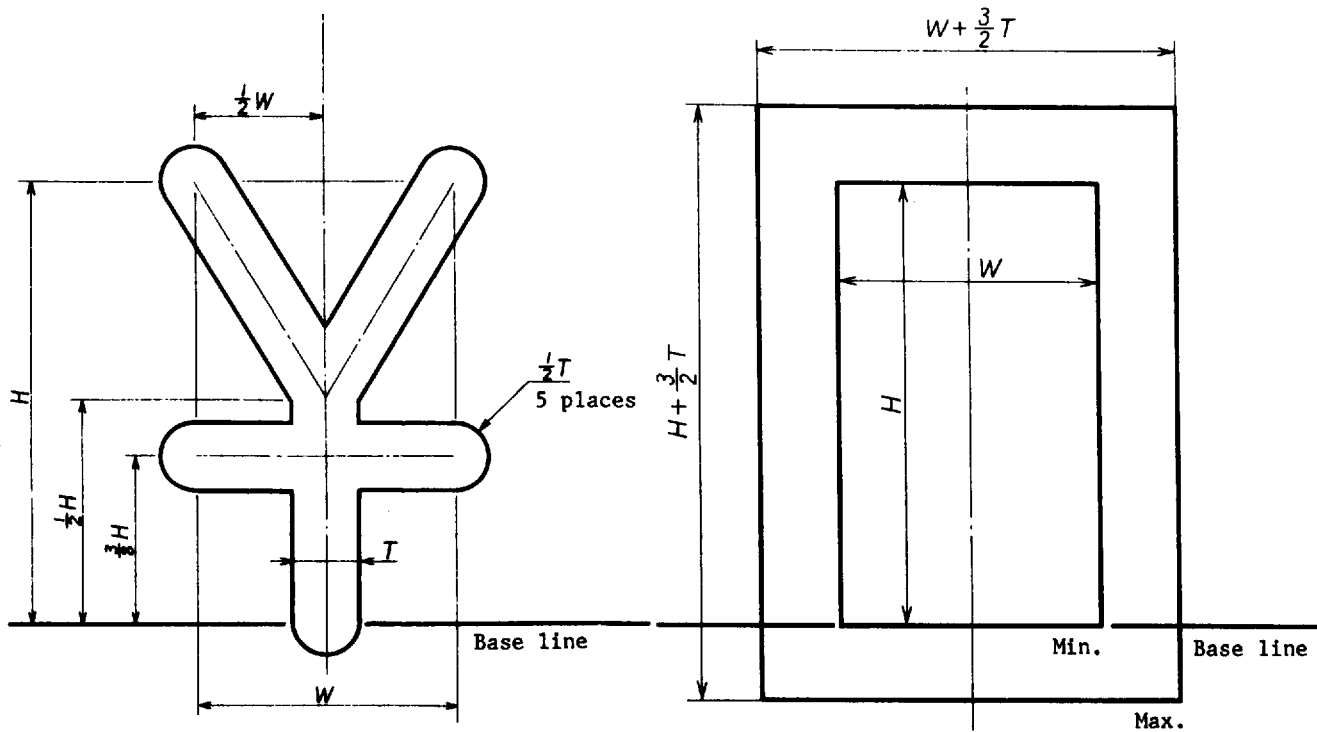


Right curly bracket



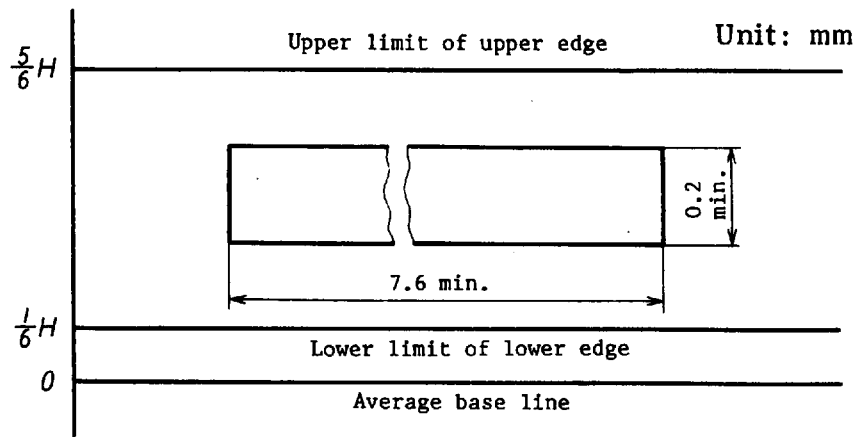
Note: C is the nominal centerline position for a character as defined by the pitch of the printing mechanism and characters on the line. P is the nominal character spacing (pitch of printing mechanism) and its minimum value is $W + 2T$.

Space (SP)



Yen sign

Character erase



Group erase

26
X 9001-1976

