

CONFIRMED
MARCH 1984

Specification for

Optical character recognition —

Part 2: Character set OCR-B. Shapes and dimensions of the printed image

[ISO title: Alphanumeric character sets for optical recognition
Part II: Character set OCR-B — Shapes and dimensions of the
printed image]

UDC 681.327.12:681.3.04 + 003

NO COPYING WITHOUT BSI PERMISSION EXCEPT AS PERMITTED BY COPYRIGHT LAW

BSI
Documents

AVA

Document Sharing Hub

Cooperating organizations

The Data Processing Standards Committee, under whose supervision this British Standard was prepared, consists of representatives from the following Government departments and scientific and industrial organizations:

- British Airways
- British Computer Society*
- British Paper and Board Industry Federation*
- British Printing Industry Federation*
- Business Equipment Trade Association*
- Civil Service Department — Central Computer Agency*
- Committee of London Clearing Bankers*
- Department of Industry
- Department of Industry — National Physical Laboratory*
- Department of the Environment — Fire Research Station
- Electricity Supply Industry in England and Wales*
- Electronic Engineering Association*
- Engineering Equipment Users' Association*
- Government Communication Headquarters
- HM Customs and Excise
- Institute of Cost and Management Accountants
- Institute of Municipal Treasurers and Accountants (Incorporated)
- Institution of Electrical and Electronics Technician Engineers
- Institution of Electrical Engineers
- Institution of Mechanical Engineers
- London Transport Executive
- Ministry of Defence
- National Coal Board*
- National Research Development Corporation
- Post Office*
- Society of British Aerospace Companies

The Government departments and scientific and industrial organizations marked with an asterisk in the above list, together with the following, were directly represented on the committee entrusted with the preparation of this British Standard:

- Her Majesty's Stationery Office
- PIRA — Research Association for the Paper and Board Printing and Packaging Industries
- Production Engineering Research Association of Great Britain

This British Standard, having been prepared under the direction of the Data Processing Standards Committee, was published under the authority of the Executive Board on 29 April 1977

© BSI 09-1999

The following BSI references relate to the work on this standard:
Committee reference DPS/14
Draft for comment 75/72082

ISBN 0 580 09811 7

Amendments issued since publication

Amd. No.	Date of issue	Comments

Contents

	Page
Cooperating organizations	Inside front cover
Foreword	ii
1 General	1
1.1 Scope	1
1.2 Field of application	1
1.3 Definitions	1
2 Styles	1
3 OCR-B sizes	2
4 Typical dimensions of the nominal printed image	2
5 OCR-B character set	3
6 Index table	6
7 Use of diacritical signs	18
8 Use of the two underline characters	18
9 Space	19
10 Vertical line and long vertical mark	19
11 Character shape definition	19
12 Printing the letterpress and constant strokewidth fonts	21
13 Illustration of OCR-B	21
Annex A Old design of digit zero	33
Annex B Existing applications using former size II with former font design of the ten digits	37
Annex C Recommendation for the implementation of OCR-B on typewriters	47
Figure 1 — Heights above and below base line	3
Figure 2 — Erase characters	5
Figure 3 — Examples of accented letters	18
Figure 4 — Example of use of discontinuous underline	18
Table 1 — Typical dimensions <i>A</i> , <i>B</i> , <i>C</i> and <i>D</i>	3
Table 2 — Dimensions of erase characters	6
Table 3 — Dimensions of vertical line and long vertical mark	19

Foreword

This British Standard has been prepared under the authority of the Data Processing Standards Committee. It is identical with ISO 1073-II “*Alphanumeric character sets for optical recognition — Part II: Character set OCR-B — Shapes and dimensions of the printed image*”, which was prepared by subcommittee 3, Character and mark recognition, of Technical Committee 97, Computers and information processing. The United Kingdom is a participating member of both the subcommittee and the committee. Its delegations, appointed and briefed by the British Standards Institution Technical Committee DPS/14, Character and mark recognition, were involved in preparing this standard.

Part 1: is identical with ISO 1073-I, and deals with the character set OCR-A.

Part 3: deals with the encoding of character sets OCR-A and OCR-B.

Terminology and conventions. The text of the International Standard has been accepted as suitable for publication, without deviation, as a British Standard. Certain terminology and conventions are used, however, that are not identical with those used in British Standards. Attention is therefore drawn to the following.

Wherever the words “International Standard”, relating to this publication, appear, they should be interpreted as “British Standard”.

Except in the case of imperial equivalents, the comma has been used throughout as a decimal marker. In British Standards it is current practice to use a full point (a full stop on the baseline) as a decimal marker.

Cross reference (See 1.3.1 and 1.3.2.) BS 4730 “*The United Kingdom 7-bit data code (ISO-7-UK)*” is the UK national version of ISO 646 “*7-bit coded character set for information processing interchange*”. It is based on and is fully in conformity with ISO 646 but exercises some of the possible variations and options in a national sense.

A British Standard does not purport to include all the necessary provisions of a contract. Users of British Standards are responsible for their correct application.

Compliance with a British Standard does not of itself confer immunity from legal obligations.

Summary of pages

This document comprises a front cover, an inside front cover, pages i and ii, pages 1 to 47 and a back cover.

This standard has been updated (see copyright date) and may have had amendments incorporated. This will be indicated in the amendment table on the inside front cover.

1 General

1.1 Scope

This International Standard for character shapes and sizes is intended to facilitate and foster the use of Optical Character Recognition (OCR) in data processing, by defining character shapes suitable for both human and machine reading.

It establishes a common basis for printing equipment and optical scanning equipment for OCR interchange applications.

Additional International Standards will cover the print quality and the relevant characteristics of the formats needed to satisfy interchange requirements.

1.2 Field of application

This International Standard specifies the printed image shapes and sizes of alphanumeric characters, graphics and symbols designed for use in Optical Character Recognition. They are also suitable for general purposes.

In order to satisfy present requirements and encourage the wide extension of OCR applications, two sets of characters are specified. These are named OCR-A and OCR-B.

Character set OCR-A includes the numeric sub-set which was recommended in draft ISO Recommendation No. 890 (now part of this International Standard). The shapes of the characters have been designed to be suitable for use in many applications of OCR. Dimensions of OCR-A are given in three sizes. (See part I.)

The shapes of the OCR-B characters have been designed for use in OCR systems without undue sacrifice of their suitability for general purposes in a wide range of applications. Dimensions of OCR-B are given in three sizes.

1.3 Definitions

For the purpose of this International Standard the following definitions apply:

1.3.1

OCR-A

a repertoire of 69 characters of which 56 are graphics included in the ISO 7-bit coded character set (ISO 646:1973). It comprises digits, capital letters, capital national letters and other graphics (See part I.)

1.3.2

OCR-B

a repertoire of 121 characters comprising digits, capital and small letters, all the graphics specified in the ISO 7-bit coded character set (ISO 646:1973), national letters, diacritical signs and further graphics

NOTE 1 For applications which involve circulation of documents across boundaries between areas in which different national characters are in use, agreement between the sender and the recipient of the documents is required.

NOTE 2 The metric and inch dimensions in this International Standard are rounded and therefore consistent but not exactly equal. Either system may be used but the two should not be intermixed.

NOTE 3 It is recognized that some type-making and printing processes will not be able to produce sharp corners. Corners not specified as having a specific radius should be as sharp as practicable. However, it is not necessary for OCR purposes that the radii of the corners of the nominal printed image be less than 0,08 mm (0.0035 in).

2 Styles

The OCR-B font (see clause 13) comprises 121 characters, but, in general, only a subset will be used for a specific application.

The character shapes and dimensions are specified by reference drawings on a reference grid. The nominal strokewidth is constant for each character of the standard set entitled “constant-strokewidth font”; the centreline of each character is indicated on the reference grid.

A second style of characters entitled “letterpress font” may be used with printing equipment which can reproduce fine details with sufficient accuracy. For many classes of printers, however, the strokewidth is less controllable and therefore for these printers the constant-strokewidth font shall be used.

For aesthetic reasons, the strokewidth of the letterpress font characters is varied deliberately and the stroke endings are specially designed. But the centrelines are the same for both fonts and these centrelines, as defined for the constant-strokewidth font, are the definitive part of this standard.

3 OCR-B sizes

3.1 Three sizes are specified for OCR-B characters in order to provide for use with a wide range of printing equipment processing differing print quality characteristics. Devices such as typewriters, cash registers, numbering machines, high-speed printers, and credit card imprinters, besides printing processes such as letterpress and offset lithography, are all suitable.

3.2 The letterpress font is specified in size I (the smallest) only. It provides the option of a variable pitch between characters as is usual with letterpress.

3.3 The constant-strokewidth font is specified in three sizes, I, III and IV. Mechanisms using the constant-strokewidth font will usually maintain a fixed pitch.

3.4 Size II which was in ISO/R 1073:1969 has been deleted. (See Annex B.)

3.5 The centrelines for the three sizes are simply related by appropriate horizontal and vertical scale factors. The factors for size III and size IV referred to size I are:

for size III

Vertical: 1,333 Horizontal: 1,086

for size IV

Vertical: 1,500 Horizontal: 1,500

This scale relationship does not apply to the outline shapes, since nominal strokewidth is not strictly proportional to centreline dimensions. The strokewidths for each size are shown in the reference drawings.

3.6 The character with the greatest height in each size is digit EIGHT. It is the character which extends farthest above the base line for capital letters. The longest character is small letter j, because of its descender.

The centreline heights of the character EIGHT are:

for size I: 2,40 mm (0.094 in)

for size III: 3,20 mm (0.126 in)

for size IV: 3,60 mm (0.142 in)

3.7 The widest character in each size (except for the alternative small letter m) is digit ZERO. Its centreline widths are:

for size I: 1,40 mm (0.055 in)

for size III: 1,52 mm (0.060 in)

for size IV: 2,10 mm (0.083 in)

3.8 Constant-pitch printing

In constant-pitch printing for OCR applications, the following minimum nominal pitches are appropriate:

size I: 2,54 mm (0.100 in) min.

size III: 2,54 mm (0.100 in) min.

size IV: 3,63 mm (0.143 in) min.

4 Typical dimensions of the nominal printed image

4.1 Constant-strokewidth font

Typical dimensions for the nominal printed image of the constant-strokewidth font in size I are given below. These dimensions are the heights above and below the horizontal base line of digits, capital and small letters, ascenders and descenders (see Figure 1). These dimensions are for general information only. The values for individual characters are obtainable from the reference drawings.

4.2 Letterpress font

The shapes of the letterpress characters are similar except that the stroke ends are not rounded.

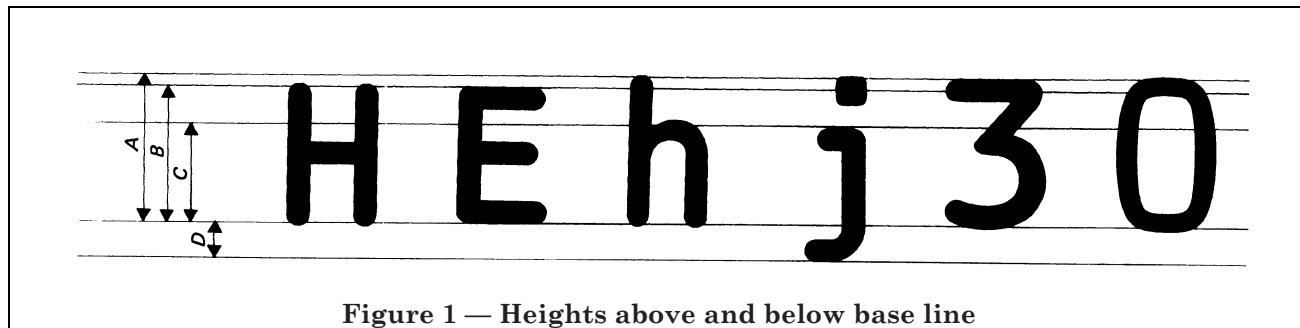


Figure 1 — Heights above and below base line

Table 1 — Typical dimensions *A*, *B*, *C* and *D*

Size	millimetres				inches			
	<i>A</i>	<i>B</i>	<i>C</i>	<i>D</i>	<i>A</i>	<i>B</i>	<i>C</i>	<i>D</i>
1	2,66	2,46	1,83	0,60	0.105	0.097	0.072	0.024

5 OCR-B character set

The full character set comprises 121 characters.

The following sub-sets can be distinguished.

5.1 Sub-set 1: Numeric sub-set

This sub-set comprises 22 characters:

0 1 2 3 4 5 6 7 8 9

< + >

C E N S T X Z

| SPACE

NOTE 1 The character ZERO is the only digit which had to be modified in this revision of ISO/R 1073:1969. The use of the original design is tolerated in numeric applications implemented before 1976. OCR reading of both old and revised design is subject to special agreement between OCR equipment supplier and user. For any application implemented after 1976, only the new design is standard.

NOTE 2 The characters C E N S T X Z should preferably not be used in document reading applications.

5.2 Sub-set 2: Initial alphanumeric sub-set

This sub-set comprises 47 characters:

0 1 2 3 4 5 6 7 8 9
A B C D E F G H I J K L M
N O P Q R S T U V W X Y Z
< + > * - = / . ,
| SPACE

5.3 Sub-set 3: Extended alphanumeric sub-set

This sub-set comprises 98 characters, in particular those of the ISO 7-bit coded character set (ISO 646:1973):

! " # £ ¤ \$ % & ' () * + , - . /
0 1 2 3 4 5 6 7 8 9 : ; < = > ?
@ A B C D E F G H I J K L M N O
P Q R S T U V W X Y Z [\] ^ _
` a b c d e f g h i j k l m n o
p q r s t u v w x y z { | } ~
| SPACE

5.4 Sub-set 4: Options sub-set

This sub-set comprises 8 capital national letters, 5 small national letters, 4 diacritical signs and 4 further characters:

À Á Â Ã Ä Å Æ Ç È É

Ê Ë Ì Í Î Ï Ñ Ò Ó Ô Õ

Ö × Ø Ù Ú Û Ü

Ý

à á â ã

5.5 Sub-set 5: Erase characters

This sub-set comprises 2 characters:

CHARACTER ERASE



GROUP ERASE



The dimensions of these two characters are as shown below:

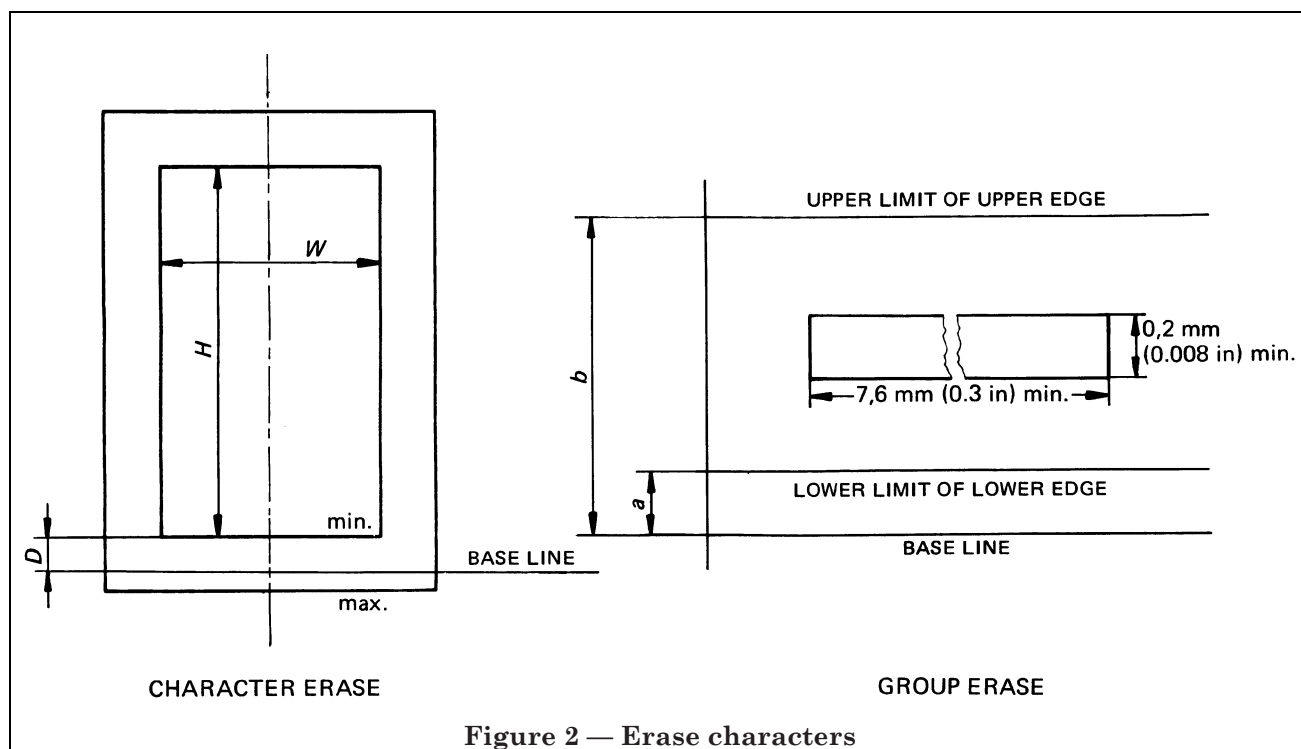


Table 2 — Dimensions of erase characters

	millimetres			inches		
	Size I	Size III	Size IV	Size I	Size III	Size IV
CHARACTER ERASE:						
min. <i>H</i>	2,4		3,8	0.094		0.149
max. <i>H</i>	2,9		4,6	0.115		0.181
min. <i>W</i>	1,4		2,0	0.055		0.079
max. <i>W</i>	1,9		2,8	0.075		0.110
<i>D</i>	0,13		0,20	0.005		0.008
GROUP ERASE:						
minimum length	7,6	7,6	10,9	0.300	0.300	0.430
minimum width	0,2	0,2	0,2	0.008	0.008	0.008
<i>a</i>	0,4	0,5	0,6	0.016	0.020	0.024
<i>b</i>	2,0	2,7	3,0	0.077	0.106	0.118

6 Index table

6.1 All characters are available in size I as constant-strokewidth font and as letterpress font.

Only the characters of the numeric sub-set (sub-set 1) and the character GROUP ERASE are available in size III as constant-strokewidth font.

All characters are available in size IV as constant-strokewidth font, with the exception of VERTICAL LINE.

6.2 In the following index table each character is given with the indication of the reference drawing or drawings and the sub-set or sub-sets in which it is comprised.

The drawings are identified as follows:

L: for letterpress font, size I

C: for the constant-strokewidth font, size I

III: for the constant-strokewidth font, size III

6.3 As stated in 11.6, the character shapes for size IV are derived from those of size I for the constant-strokewidth font (designated by C).

6.4 Application advice is given in the column “Remarks”, where it is indicated, *inter alia*, which characters are included for general-purpose use only and should not be used for OCR purposes.

It is recommended that prospective users of this standard consult manufacturers before deciding on a particular character set.

Index table

Ref. No.	Shape	Drawing(s) No.	Name	Sets	Remarks
1	1	1 L, C, III	DIGIT ONE	1 2 3	
2	2	2 L, C, III	DIGIT TWO	1 2 3	
3	3	3 L, C, III	DIGIT THREE	1 2 3	
4	4	4 L, C, III	DIGIT FOUR	1 2 3	
5	5	5 L, C, III	DIGIT FIVE	1 2 3	
6	6	6 L, C, III	DIGIT SIX	1 2 3	
7	7	7 L, C, III	DIGIT SEVEN	1 2 3	
8	8	8 L, C, III	DIGIT EIGHT	1 2 3	
9	9	9 L, C, III	DIGIT NINE	1 2 3	
10	0	10 L, C, III	DIGIT ZERO	1 2 3	The character ZERO is the only digit which had to be modified in this revision of ISO/R 1073:1969. The use of the original design is tolerated in numeric applications implemented before 1976. OCR reading of both old and revised design is subject to special agreement between OCR equipment supplier and user. For any application implemented after 1976, only the new design is standard.
11	A	11 L, C	CAPITAL LETTER A	2 3	

Index table

Ref. No.	Shape	Drawing(s) No.	Name	Sets	Remarks
12	B	12 L, C	CAPITAL LETTER B	2 3	
13	C	13 L, C, III	CAPITAL LETTER C	1 2 3	
14	D	14 L, C	CAPITAL LETTER D	2 3	
15	E	15 L, C, III	CAPITAL LETTER E	1 2 3	
16	F	16 L, C	CAPITAL LETTER F	2 3	
17	G	17 L, C	CAPITAL LETTER G	2 3	
18	H	18 L, C	CAPITAL LETTER H	2 3	
19	I	19 L, C	CAPITAL LETTER I	2 3	
20	J	20 L, C	CAPITAL LETTER J	2 3	
21	K	21 L, C	CAPITAL LETTER K	2 3	
22	L	22 L, C	CAPITAL LETTER L	2 3	

Index table

Ref. No.	Shape	Drawing(s) No.	Name	Sets	Remarks
23	M	23 L, C	CAPITAL LETTER M	2 3	
24	N	24 L, C, III	CAPITAL LETTER N	1 2 3	
25	O	25 L, C	CAPITAL LETTER O	2 3	
26	P	26 L, C	CAPITAL LETTER P	2 3	
27	Q	27 L, C	CAPITAL LETTER Q	2 3	
28	R	28 L, C	CAPITAL LETTER R	2 3	
29	S	29 L, C, III	CAPITAL LETTER S	1 2 3	
30	T	30 L, C, III	CAPITAL LETTER T	1 2 3	
31	U	31 L, C	CAPITAL LETTER U	2 3	
32	V	32 L, C	CAPITAL LETTER V	2 3	
33	W	33 L, C	CAPITAL LETTER W	2 3	

Index table

Ref. No.	Shape	Drawing(s) No.	Name	Sets	Remarks
34	X	34 L, C, III	CAPITAL LETTER X	1 2 3	
35	Y	35 L, C	CAPITAL LETTER Y	2 3	
36	Z	36 L, C, III	CAPITAL LETTER Z	1 2 3	
37	a	37 L, C	SMALL LETTER a	3	Smaller strokewidth, see clause 11.
38	b	38 L, C	SMALL LETTER b	3	Smaller strokewidth, see clause 11.
39	c	39 L, C	SMALL LETTER c	3	Smaller strokewidth, see clause 11.
40	d	40 L, C	SMALL LETTER d	3	Smaller strokewidth, see clause 11.
41	e	41 L, C	SMALL LETTER e	3	Smaller strokewidth, see clause 11.
42	f	42 L, C	SMALL LETTER f	3	Smaller strokewidth, see clause 11.
43	g	43 L, C	SMALL LETTER g	3	Smaller strokewidth, see clause 11.
44	h	44 L, C	SMALL LETTER h	3	Smaller strokewidth, see clause 11.

Index table

Ref. No.	Shape	Drawing(s) No.	Name	Sets	Remarks
45	i	45 L, C	SMALL LETTER i	3	Smaller strokewidth, see clause 11.
46	j	46 L, C	SMALL LETTER j	3	Smaller strokewidth, see clause 11.
47	k	47 L, C	SMALL LETTER k	3	Smaller strokewidth, see clause 11.
48	l	48 L, C	SMALL LETTER l	3	Smaller strokewidth, see clause 11.
49	m	49 L, C	SMALL LETTER m	3	Smaller strokewidth, see clause 11.
50	n	50 L, C	SMALL LETTER n	3	Smaller strokewidth, see clause 11.
51	o	51 L, C	SMALL LETTER o	3	Smaller strokewidth, see clause 11.
52	p	52 L, C	SMALL LETTER p	3	Smaller strokewidth, see clause 11.
53	q	53 L, C	SMALL LETTER q	3	Smaller strokewidth, see clause 11.
54	r	54 L, C	SMALL LETTER r	3	Smaller strokewidth, see clause 11.
55	s	55 L, C	SMALL LETTER s	3	Smaller strokewidth, see clause 11.

Index table

Ref. No.	Shape	Drawing(s) No.	Name	Sets	Remarks
56	t	56 L, C	SMALL LETTER t	3	Smaller strokewidth, see clause 11.
57	u	57 L, C	SMALL LETTER u	3	Smaller strokewidth, see clause 11.
58	v	58 L, C	SMALL LETTER v	3	Smaller strokewidth, see clause 11.
59	w	59 L, C	SMALL LETTER w	3	Smaller strokewidth, see clause 11.
60	x	60 L, C	SMALL LETTER x	3	Smaller strokewidth, see clause 11.
61	y	61 L, C	SMALL LETTER y	3	Smaller strokewidth, see clause 11.
62	z	62 L, C	SMALL LETTER z	3	Smaller strokewidth, see clause 11.
63	*	63 L, C	ASTERISK	2 3	
64	+	64 L, C, III	PLUS SIGN	1 2 3	
65	—	65 L, C	HYPHEN (MINUS SIGN)	2 3	
66	=	66 L, C	EQUALS SIGN	2 3	