International Standard **ISO** 4873

Information processing — ISO 8-bit code for information interchange — Structure and rules for implementation

Traitement de l'information — Code ISO à 8 éléments pour l'échange d'information — Structure et règles de matérialisation

Second edition — 1986-07-15

ISO 4873-1986 (E)

Price based on 17 pages

Foreword

ISO (the International Organization for Standardization) is a worldwide federation of national standards bodies (ISO member bodies). The work of preparing International Standards is normally carried out through ISO technical committees. Each member body interested in a subject for which a technical committee has been established has the right to be represented on that committee. International organizations, governmental and non-governmental, in liaison with ISO, also take part in the work.

Draft International Standards adopted by the technical committees are circulated to the member bodies for approval before their acceptance as International Standards by the ISO Council. They are approved in accordance with ISO procedures requiring at least 75 % approval by the member bodies voting.

International Standard ISO 4873 was prepared by Technical Committee ISO/TC 97, *Information processing systems.*

This second edition cancels and replaces the first edition (ISO 4873-1979), of which it constitutes a technical revision.

Users should note that all International Standards undergo revision from time to time and that any reference made herein to any other International Standard implies its latest edition, unless otherwise stated.

Contents

Information processing — ISO 8-bit code for information interchange — Structure and rules for implementation

1 Scope and field of application

This International Standard specifies an 8-bit code derived from, and compatible with, the 7-bit coded character set specified in ISO 646.

The characteristics of this code are also in conformance with the code extension techniques specified in ISO 2022.

This International Standard specifies an 8-bit code with a number of options. It also provides guidance on how to exercise the options to define specific versions.

This character set is primarily intended for general information interchange within an 8-bit environment among data processing systems and associated equipment, and within data communication systems. The need for graphic characters and control functions in data processing has also been taken into account.

This character set includes the 52 small and capital letters of the basic Latin alphabet and may include accented letters, special Latin letters and/or the letters of one or several non-Latin alphabet(s).

2 References

ISO 646, *Information processing — ISO 7-bit coded character set for information interchange.*

ISO 1177, *Information processing — Character structure for start/stop and synchronous character-oriented transmission.*

ISO 2022, *Information processing — ISO 7-bit and 8-bit coded character sets — Code extension techniques.*

ISO 6429, *Information processing — Additional control functions for character-imaging devices.*

ISO 6937/2, *Information processing — Coded character sets for text communication — Part 2: Latin alphabetic and non-alphabetic graphic characters.*

3 Conformance and implementation

3.1 Conformance

An 8-bit code is in conformance with this International Standard if it is a version in accordance with clause 8. Equipment

claimed to implement this International Standard shall be able to interchange information by means of a version of the 8-bit code at a specified level according to clauses 8 and 9; this version and level shall be identified in any such claim.

3.2 Implementation

The use of this code requires definitions of its implementation in various media. For example, these could include magnetic and optical media and transmission channels, thus permitting interchange of data to take place either indirectly by means of an intermediate recording in a physical medium, or by means of data transmission equipment.

The implementation of this code in physical media and for transmission, taking into account the need for error checking, is the subject of other International Standards.

4 Definitions

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

4.1 bit combination: An ordered set of bits that represents a character or is used as part of the representation of a character.

4.2 character: A member of a set of elements used for the organization, control or representation of data.

4.3 coded character set; code: A set of unambiguous rules that establishes a character set and the one-to-one relationship between each character of the set and its coded representation by one or more bit combinations.

4.4 code extension: The techniques for the encoding of characters that are not included in the character set of a given code.

4.5 code table: A table showing the character allocated to each bit combination in a code.

4.6 control character: A control function the coded representation of which consists of a single bit combination.

4.7 control function: An action that affects the recording, processing, transmission or interpretation of data, and that has a coded representation consisting of one or more bit combinations.

4.8 graphic character: A *character*, *other than a control* function, that has a visual representation normally handwritten, printed or displayed, and that has a coded representation consisting of one or more bit combinations.

4.9 graphic symbol: A visual representation of a graphic character or of a control function.

4.10 position: That part of a code table identified by its column and row co-ordinates.

5 Notation, code table and names

5.1 Notation

The bits of the bit combinations of the 8-bit code are identified by b_8, b_7, b_6, b_5, b_4, b_3, b_2 and b_1, where b_8 is the highest-order, or most-significant bit and b_1 is the lowest-order, or least-significant bit.

The bit combinations may be interpreted to represent numbers in the range 0 to 255 in binary notation by attributing the following weights to the individual bits:

Bit	b_8	b_7	b_6	b_5	b_4	b_3	b_2	b_1
Weight	128	64	32	16	8	4	2	1

In this International Standard, the bit combinations are identified by notations of the form xx/yy, where xx and yy are numbers in the range 00 to 15. The correspondence between the notations of the form xx/yy and the bit combinations consisting of the bits b_8 to b_1, is as follows:

— xx is the number represented by b_8, b_7, b_6 and b_5 where these bits are given the weights 8, 4, 2 and 1 respectively;

— yy is the number represented by b_4, b_3, b_2 and b_1 where these bits are given the weights 8, 4, 2 and 1 respectively;

The notations of the form xx/yy are the same as the ones used to identify code table positions, where xx is the column number and yy is the row number (see 5.2).

5.2 Code table

An 8-bit code table consists of 256 positions arranged in 16 columns and 16 rows. The columns and rows are numbered 00 to 15.

The code table positions are identified by notations of the form xx/yy, where xx is the column number and yy is the row number.

The positions of the code table are in one-to-one correspondence with the bit combinations of the code. The notation of a code table position, of the form xx/yy, is the same as that of the corresponding bit combination.

5.3 Names

This International Standard assigns at least one name to each character. In addition, it specifies an acronym for each control character and for the character SPACE, and a graphic symbol for each graphic character. By convention, only capital letters, the graphic symbols for small letters and hyphens are used for writing the names of the characters. It is intended that the acronyms and this convention be retained in all translations of the text of this International Standard.

The names chosen to denote graphic characters are intended to reflect their customary meaning. However, this International Standard does not define and does not restrict the meanings of graphic characters. Neither does it specify a particular style or font design for imaging the graphic characters.

6 Structure of the 8-bit code

6.1 Elements of the 8-bit code

The 8-bit code consists of the following parts:

a) A C0 set: a set of 30 control characters represented by bit combinations 00/00 to 01/15, except 00/14 and 00/15 which shall be unused.

b) The character SPACE: a character represented by the bit combination 02/00, which may be interpreted as a control character, as a graphic character, or as both.

c) A G0 set: a set of up to 94 graphic characters represented by bit combinations 02/01 to 07/14.

d) The character DELETE: a control character represented by the bit combination 07/15.

e) A C1 set: a set of up to 32 control characters represented by bit combinations 08/00 to 09/15.

f) A G1 set: a set of up to 96 graphic characters represented by bit combinations 10/00 to 15/15.

g) A G2 set: a set of up to 96 graphic characters.

h) A G3 set: a set of up to 96 graphic characters.

6.2 Initial condition

At the beginning of any information interchange, a default C0 set containing only the character ESCAPE shall be assumed.

6.3 Designation and invocation

For the sake of simplicity, this International Standard shows escape sequences of three characters. Escape sequences with more characters may also be used or required. The format and use of escape sequences are described in detail in ISO 2022.

6.3.1 C0 set

The C0 set shall be allocated to columns 00 and 01 of the code table. It shall be designated and invoked by an escape sequence of the type ESC 02/01 F.

6.3.2 Character SPACE

The character SPACE shall be allocated to column 02 in position 02/00. It is not explicitly designated or invoked.

6.3.3 G0 set

The G0 set shall be allocated to columns 02 to 07 in positions 02/01 to 07/14. It shall be designated by an escape sequence of the type ESC 02/08 F and invoked as specified in clause 9.

6.3.4 Character DELETE

The character DELETE shall be allocated to column 07 in position 07/15. It is not explicitly designated or invoked.

6.3.5 C1 set

The C1 set shall be allocated to columns 08 and 09. It shall be designated and invoked by an escape sequence of the type ESC 02/02 F.

6.3.6 G1 set

The G1 set shall be allocated to columns 10 to 15 when invoked. It shall be designated by an escape sequence of the type ESC 02/09 F or ESC 02/13 F and invoked as specified in clause 9.

6.3.7 G2 set

The G2 set shall be designated by an escape sequence of the type ESC 02/10 F or ESC 02/14 F. Either the set as a whole shall be invoked by the locking-shift function LS2R into columns 10 to 15, or individual characters of it shall be invoked by means of the single-shift function SS2.

6.3.8 G3 set

The G3 set shall be designated by an escape sequence of the type ESC 02/11 F or ESC 02/15 F. Either the set as a whole shall be invoked by the locking-shift function LS3R into columns 10 to 15, or individual characters of the set shall be invoked by means of the single-shift function SS3.

6.3.9 Summary of structure, designation and invocation

Figure 1 summarizes, in a schematic form, the structure of the 8-bit code and designation and allocation of its different parts.

7 Specification of the characters of the 8-bit code

7.1 C0 set

The requirements for the C0 set are

— bit combinations 00/14 and 00/15 shall not be used (see annex D);

— the control character ESCAPE shall be allocated to bit combination 01/11;

— any control characters can be allocated to the other bit combinations subject to the restrictions specified in annex A.

NOTE — A C0 set comprising only ESCAPE allocated to position 01/11 has been registered, and is identified by ESC 02/01 04/07.

7.2 Character ESCAPE

ESCAPE is a control character used to form escape sequences. In this International Standard escape sequences are used to announce the level at which a version is defined and to designate and invoke character sets.

Acronym	Name	Coded representation
E S C	ESCAPE	01/11

7.3 Character SPACE

This character may be interpreted as a graphic character, as a control character, or as both.

As a graphic character, it has a visual representation consisting of the absence of a graphic symbol.

As a control character, it acts as a format effector that causes the active position to be advanced one character position.

Acronym	Name	Coded representation
S P	SPACE	02/00

7.4 G0 set

The 94 bit combinations 02/01 to 07/14 are used to represent graphic characters. All graphic characters allocated to bit combinations in the range 02/01 to 07/14 are spacing characters, i.e. they cause the active position to advance. Sub-clause 7.4.1 specifies the graphic characters allocated by this International Standard to 82 of these 94 bit combinations. Sub-clauses 7.4.2, 7.4.3 and 7.4.4 specify how further graphic characters may be allocated to the remaining twelve bit combinations.

Each of the sub-clauses 7.4.1, 7.4.2 and 7.4.4 contains a table consisting of three columns. The first column specifies the graphic symbol of each graphic character, the second column specifies the standard name of the graphic character and the third column specifies the bit combination representing the graphic character concerned.

7.4.1 Unique graphic character allocations

A unique graphic character is allocated to each of the 82 bit combinations 02/01, 02/02, 02/05 to 03/15, 04/01 to 05/10, 05/15 and 06/01 to 07/10. These characters are specified in table 1.
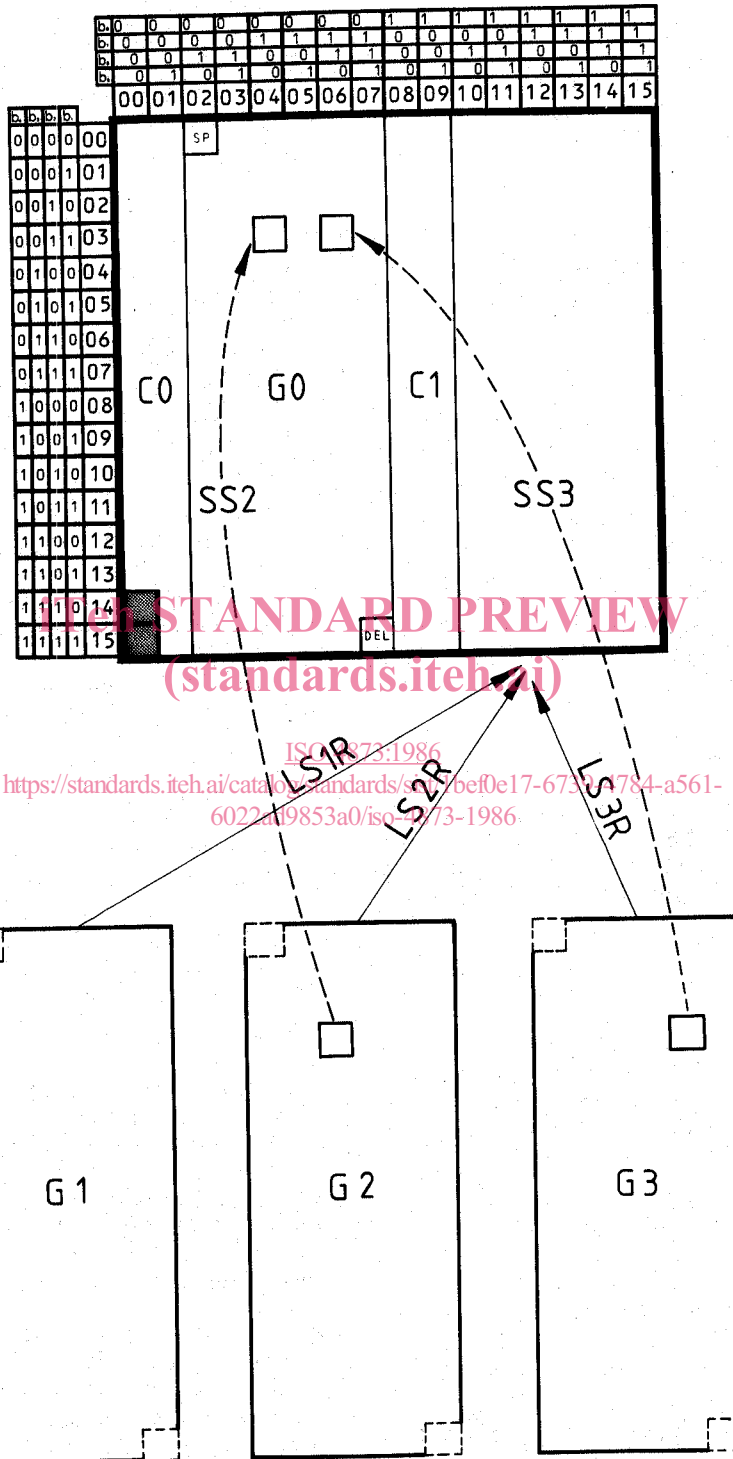
Figure 1 — Structure of the 8-bit code

Table 1 — Unique graphic character allocations

Graphic symbol	Name	Coded representation	Graphic symbol	Name	Coded representation
!	EXCLAMATION MARK	02/01	M	CAPITAL LETTER M	04/13
"	QUOTATION MARK	02/02	N	CAPITAL LETTER N	04/14
%	PERCENT SIGN	02/05	O	CAPITAL LETTER O	04/15
&	AMPERSAND	02/06	P	CAPITAL LETTER P	05/00
'	APOSTROPHE	02/07	Q	CAPITAL LETTER Q	05/01
(	LEFT PARENTHESIS	02/08	R	CAPITAL LETTER R	05/02
)	RIGHT PARENTHESIS	02/09	S	CAPITAL LETTER S	05/03
*	ASTERISK	02/10	T	CAPITAL LETTER T	05/04
+	PLUS SIGN	02/11	U	CAPITAL LETTER U	05/05
,	COMMA	02/12	V	CAPITAL LETTER V	05/06
−	HYPHEN, MINUS SIGN	02/13	W	CAPITAL LETTER W	05/07
.	FULL STOP	02/14	X	CAPITAL LETTER X	05/08
/	SOLIDUS	02/15	Y	CAPITAL LETTER Y	05/09
0	DIGIT ZERO	03/00	Z	CAPITAL LETTER Z	05/10
1	DIGIT ONE	03/01	_	LOW LINE, UNDERLINE	05/15
2	DIGIT TWO	03/02	a	SMALL LETTER a	06/01
3	DIGIT THREE	03/03	b	SMALL LETTER b	06/02
4	DIGIT FOUR	03/04	c	SMALL LETTER c	06/03
5	DIGIT FIVE	03/05	d	SMALL LETTER d	06/04
6	DIGIT SIX	03/06	e	SMALL LETTER e	06/05
7	DIGIT SEVEN	03/07	f	SMALL LETTER f	06/06
8	DIGIT EIGHT	03/08	g	SMALL LETTER g	06/07
9	DIGIT NINE	03/09	h	SMALL LETTER h	06/08
:	COLON	03/10	i	SMALL LETTER i	06/09
;	SEMICOLON	03/11	j	SMALL LETTER j	06/10
<	LESS-THAN SIGN	03/12	k	SMALL LETTER k	06/11
=	EQUALS SIGN	03/13	l	SMALL LETTER l	06/12
>	GREATER-THAN SIGN	03/14	m	SMALL LETTER m	06/13
?	QUESTION MARK	03/15	n	SMALL LETTER n	06/14
A	CAPITAL LETTER A	04/01	o	SMALL LETTER o	06/15
B	CAPITAL LETTER B	04/02	p	SMALL LETTER p	07/00
C	CAPITAL LETTER C	04/03	q	SMALL LETTER q	07/01
D	CAPITAL LETTER D	04/04	r	SMALL LETTER r	07/02
E	CAPITAL LETTER E	04/05	s	SMALL LETTER s	07/03
F	CAPITAL LETTER F	04/06	t	SMALL LETTER t	07/04
G	CAPITAL LETTER G	04/07	u	SMALL LETTER u	07/05
H	CAPITAL LETTER H	04/08	v	SMALL LETTER v	07/06
I	CAPITAL LETTER I	04/09	w	SMALL LETTER w	07/07
J	CAPITAL LETTER J	04/10	x	SMALL LETTER x	07/08
K	CAPITAL LETTER K	04/11	y	SMALL LETTER y	07/09
L	CAPITAL LETTER L	04/12	z	SMALL LETTER z	07/10

7.4.2 Alternative graphic character allocations

Two alternative graphic characters are allocated to each of the bit combinations 02/03 and 02/04. These characters are specified in table 2.

Table 2 — Alternative graphic character allocations

Graphic symbol	Name	Coded representation
£	POUND SIGN	02/03
#	NUMBER SIGN	02/03
$	DOLLAR SIGN	02/04
¤	CURRENCY SIGN	02/04

Either the character POUND SIGN or the character NUMBER SIGN shall be allocated to bit combination 02/03 and either the character DOLLAR SIGN or the character CURRENCY SIGN shall be allocated to bit combination 02/04.

Unless otherwise agreed between sender and recipient, the graphic symbols £, $ and ¤ do not designate the currency of a specific country.

7.4.3 National or application-orientated graphic character allocations

No specific graphic characters are allocated to the ten bit combinations 04/00, 05/11 to 05/14, 06/00 and 07/11 to 07/14. A unique graphic character shall be allocated to each of these bit combinations, or the bit combination shall be declared unused.

7.4.4 Recommendation for the allocation of graphic characters

It is recommended that the bit combinations listed in 7.4.2 and 7.4.3 have allocated to them the characters in existing national or International Standards. Those of the International Reference Version (IRV) of ISO 646, are shown in table 3.

Table 3 — IRV graphic character allocations

Graphic symbol	Name	Coded representation
#	NUMBER SIGN	02/03
¤	CURRENCY SIGN	02/04
@	COMMERCIAL AT	04/00
[	LEFT SQUARE BRACKET	05/11
\	REVERSE SOLIDUS	05/12
]	RIGHT SQUARE BRACKET	05/13
^	CIRCUMFLEX ACCENT	05/14
`	GRAVE ACCENT	06/00
{	LEFT CURLY BRACKET	07/11
|	VERTICAL LINE	07/12
}	RIGHT CURLY BRACKET	07/13
~	TILDE, OVERLINE	07/14

7.5 Character DELETE

DELETE is a character that was originally used to erase or obliterate an erroneous or unwanted character in punched tape. DEL characters may serve to accomplish media-fill or time-fill. They may be inserted into, or removed from, a stream of data without affecting the information content of that stream, but such action may affect the information layout and/or the control of equipment.

Acronym	Name	Coded representation
D E L	DELETE	07/15

7.6 C1 set

The C1 set is available for up to 32 control characters in addition to those provided by the C0 set. It shall not include any of the control characters listed in annex A.

No specific control characters are allocated to positions 08/00 to 08/13 and 09/00 to 09/15.

When the single-shift functions SS2 and SS3 are used, they shall be allocated to positions 08/14 and 08/15, respectively, otherwise these positions shall not be used.

7.7 G1 set

The G1 set shall be either a 94-character or a 96-character set of graphic characters.

This set is available for graphic characters in addition to those provided by the G0 set.

A unique graphic character shall be allocated to each position or the position shall be declared unused.

The characters of the G1 set are represented by bit combinations 10/01 to 15/14 if the G1 set is a 94-character set, or by bit combinations 10/00 to 15/15 if the G1 set is a 96-character set.

7.8 G2 set

The G2 set shall be either a 94-character or a 96-character set of graphic characters.

This set is available for graphic characters in addition to those provided by the G0 and the G1 sets.

A unique graphic character shall be allocated to each position or the position shall be declared unused.

If the G2 set is a 94-character set, then no characters shall be allocated to positions 10/00 and 15/15.

The characters of the G2 set shall be invoked either by the single-shift function SS2 or by the locking-shift function LS2R.

— When invoked by SS2, each character is represented by the bit combination of SS2 followed by one of the bit combinations in the range 02/01 to 07/14 if the G2 set is a 94-character set, or 02/00 to 07/15 if the G2 set is a 96-character set.

— When invoked by LS2R, the characters of the G2 set are represented by bit combinations 10/01 to 15/14 if the G2 set is a 94-character set, or by bit combinations 10/00 to 15/15 if the G2 set is a 96-character set.

7.9 G3 set

The G3 set shall be either a 94-character or a 96-character set of graphic characters. This set is available for graphic characters in addition to those provided by the G0, the G1 and the G2 sets.

A unique graphic character shall be allocated to each position or the position shall be declared unused.

If the G3 set is a 94-character set, then no character shall be allocated to positions 10/00 and 15/15.

The characters of the G3 set shall be invoked either by the single-shift function SS3 or by the locking-shift function LS3R.

— When invoked by SS3, each character is represented by the bit combination of SS3 followed by one of the bit combinations in the range 02/01 to 07/14 if the G3 set is a 94-character set, or 02/00 to 07/15 if the G3 set is a 96-character set.

— When invoked by LS3R, the characters of the G3 set are represented by bit combinations 10/01 to 15/14 if the G3 set is a 94-character set, or by bit combinations 10/00 to 15/15 if the G3 set is a 96-character set.

7.10 Summary of the specification of the 8-bit code

Figure 2 summarizes, in a schematic form, the specification of the elements of the 8-bit code.

8 Versions of the 8-bit code

A version of the 8-bit code is a coded character set comprising a C0, a G0, a C1 and a G1 set and, optionally a G2 and a G3 set, with a specification of the necessary escape sequences to designate them. In a version the same character shall not be allocated to more than one of the G0, G1, G2 or G3 sets. In specifying the G0 set the following options shall be exercised:

a) one of the alternative graphic characters specified in 7.4.2 shall be allocated to each of the bit combinations 02/03 and 02/04, and

b) up to ten unique graphic characters shall be allocated to the bit combinations 04/00, 05/11 to 05/14, 06/00 and 07/11 to 07/14 specified in 7.4.3, and any of these bit combinations to which no character is allocated shall be declared unused.

9 Levels

This International Standard specifies three nested levels of implementation each of which may be identified by an announcer sequence.

ESC 02/00 04/12 shall identify Level 1

ESC 02/00 04/13 shall identify Level 2

ESC 02/00 04/14 shall identify Level 3

When announcer sequences are used (see clauses 10 and 11) the announcer sequence of a given level shall be correctly interpreted at any other level.

9.1 Level 1

Level 1, see figure 3, comprises the following facilities:

— the C0 set, designated and invoked by ESC 02/01 F;

— the character SPACE in position 02/00;

— the G0 set, designated by ESC 02/08 F;

— the character DELETE in position 07/15;

— a C1 set, designated and invoked by ESC 02/02 F;

— a G1 set, designated by ESC 02/09 F or ESC 02/13 F.

At Level 1 no shift functions shall be used and the G0 and G1 sets are assumed to be invoked permanently in columns 02 to 07 and 10 to 15, respectively. Therefore an escape sequence that designates the G0 or the G1 set also implicitly invokes that set.

At Level 1 the C1 set and/or the G1 set may be empty if there is no requirement for control characters in addition to those provided by the C0 set and/or for graphic characters in addition to these provided by the G0 set. An empty C1 set shall be designated and invoked by ESC 02/02 07/14 and an empty G1 set by ESC 02/09 07/14 or ESC 02/13 07/14.

At Level 1, G2 and G3 sets shall not be designated.

9.2 Level 2

Level 2, see figure 4, comprises the facilities of Level 1 and in addition to them

— a G2 set, designated by ESC 02/10 F or ESC 02/14 F, the characters of which shall be invoked individually by SS2;

— a G3 set, designated by ESC 02/11 F or ESC 02/15 F, the characters of which shall be invoked individually by SS3.

At Level 2 no other shift functions shall be used.

The G1 set shall not be empty, either the G2 or the G3 set can be designated as empty but not both. The G2 set may be empty if there is no requirement for graphic characters in addition to those provided by the G0, G1 and G3 sets. In this case the G2 set shall be designated by ESC 02/10 07/14 or ESC 02/14 07/14. The G3 set may be empty if there is no requirement for graphic characters in addition to those provided by the G0, G1 and G2 sets. In this case the G3 set shall be designated by ESC 02/11 07/14 or ESC 02/15 07/14.

The C1 set shall not be empty, it shall contain at least the single-shift functions SS2 and SS3 in positions 08/14 and 08/15, respectively.

NOTE — A C1 set comprising only SS2 and SS3 allocated to these positions has been registered, and is identified by ESC 02/02 04/07.