

ECMA

EUROPEAN COMPUTER MANUFACTURERS ASSOCIATION

STANDARD ECMA-6

7 - BIT CODED CHARACTER SET

5th Edition – March 1985

Free copies of this document are available from ECMA,
European Computer Manufacturers Association
114 Rue du Rhône – 1204 Geneva (Switzerland)

ECMA

EUROPEAN COMPUTER MANUFACTURERS ASSOCIATION

STANDARD ECMA-6

7 - BIT CODED CHARACTER SET

5th Edition – March 1985

BRIEF HISTORY

Technical Committee TC1 of ECMA met for the first time in December 1960 to prepare standard codes for Input/Output purposes. On April 30, 1965, Standard ECMA-6 was adopted by the General Assembly of ECMA.

Subsequent international activities necessitated three revisions of ECMA-6, which were issued in June 1967, July 1970 and August 1973. In the 3rd edition, TC1 introduced a new concept, that of versions of the code and proposed at the same time an International Reference Version. This concept and this specific version were eventually adopted by ISO for the new International Standard ISO 646-1972 and by CCITT for the International Telegraph Alphabet No. 5 (CCITT Rec. V.3).

In parallel to these developments, further work on extension of the 7-bit code was achieved. The techniques for extending the repertoire of the 7-bit code, remaining in a 7-bit environment, and for expanding it to 8-bit codes are described in Standard ECMA-35 (ISO 2022). In application of these techniques a specific 8-bit code is defined in Standard ECMA-43 (ISO 4873).

This 5th Edition corresponds to the 2nd edition of ISO 646 issued in 1983, a revision of the 1973 issue prepared by WG-7 of ISO/TC97/SC2 in which several members of ECMA/TC1 were major contributors. The differences between the 4th and the 5th editions, which are essentially editorial, are explained in Appendix B.

This 5th Edition, adopted by the General Assembly of ECMA in December 1984, supersedes the Version dated August 1983.

TABLE OF CONTENTS

	<u>Page</u>
1. SCOPE AND FIELD OF APPLICATION	1
2. CONFORMANCE AND IMPLEMENTATION	1
2.1 Conformance	1
2.2 Implementation	1
3. DEFINITIONS	2
3.1 Bit Combination	2
3.2 Character	2
3.3 Coded Character Set; Code	2
3.4 Code Extension	2
3.5 Code Table	2
3.6 Control Character	2
3.7 Control Function	2
3.8 Graphic Character	2
3.9 Graphic Symbol	2
3.10 Position	2
4. SPECIFICATION OF THE CODED CHARACTER SET	3
4.1 Control Characters	4
4.1.1 Transmission control characters	5
4.1.2 Format effectors	5
4.1.3 Code extension control characters	7
4.1.4 Device control characters	7
4.1.5 Information separators	7
4.1.6 Other control characters	8
4.2 Character SPACE	8
4.3 Graphic Characters	8
4.3.1 Unique graphic character allocations	8
4.3.2 Alternative graphic character allocations	11
4.3.3 National or application-oriented graphic character allocation	11
5. COMPOSITE GRAPHIC CHARACTERS	12
6. VERSIONS OF THE CODED CHARACTER SET	12
6.1 General	12
6.2 National Versions	13
6.3 Application-oriented Versions	13
6.4 International Reference Version (IRV)	13
7. CODE TABLES	14
8. DESCRIPTION OF THE CONTROL CHARACTERS	18
8.1 ACK ACKNOWLEDGE	18
8.2 BEL BELL	18
8.3 BS BACKSPACE	18
8.4 CAN CANCEL	18
8.5 CR CARRIAGE RETURN	18
8.6 DC1 DEVICE CONTROL ONE	18

TABLE OF CONTENTS (cont'd)

	<u>Page</u>
8.7 DC2 DEVICE CONTROL TWO	18
8.8 DC3 DEVICE CONTROL THREE	18
8.9 DC4 DEVICE CONTROL FOUR	19
8.10 DEL DELETE	19
8.11 DLE DATA LINK ESCAPE	19
8.12 EM END OF MEDIUM	19
8.13 ENQ ENQUIRY	19
8.14 EOT END OF TRANSMISSION	19
8.15 ESC ESCAPE	19
8.16 ETB END OF TRANSMISSION BLOCK	20
8.17 ETX END OF TEXT	20
8.18 FF FORM FEED	20
8.19 HT HORIZONTAL TABULATION	20
8.20 IS1 (US) INFORMATION SEPARATOR ONE (UNIT SEPARATOR)	20
8.21 IS2 (RS) INFORMATION SEPARATOR TWO (RECORD SEPARATOR)	20
8.22 IS3 (GS) INFORMATION SEPARATOR THREE (GROUP SEPA- RATOR)	20
8.23 IS4 (FS) INFORMATION SEPARATOR FOUR (FILE SEPARATOR)	20
8.24 LF LINE FEED	20
8.25 NAK NEGATIVE ACKNOWLEDGE	20
8.26 NUL NULL	21
8.27 SI SHIFT-IN	21
8.28 SO SHIFT-OUT	21
8.29 SOH START OF HEADING	21
8.30 STX START OF TEXT	21
8.31 SUB SUBSTITUTE CHARACTER	21
8.32 SYN SYNCHRONOUS IDLE	21
8.33 VT VERTICAL TABULATION	21
APPENDIX A - GUIDELINES FOR STANDARDS DERIVED FROM ECMA-6	22
APPENDIX B - MAIN DIFFERENCES BETWEEN THE 4th EDITION (August 1973) AND THE PRESENT 5th EDITION	24

1. SCOPE AND FIELD OF APPLICATION

- 1.1 This ECMA Standard specifies a set of 128 characters (control characters and graphic characters such as letters, digits and symbols) with their coded representation. Most of these characters are mandatory and unchangeable, but provision is made for some flexibility to accommodate national and other requirements.
- 1.2 This Standard specifies a 7-bit coded character set with a number of options. It also provides guidance on how to exercise the options to define specific national versions and application-oriented versions. Furthermore it specifies the International Reference Version (IRV) in which such options have been exercised.
- 1.3 This character set is primarily intended for the interchange of information 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 in determining this character set.
- 1.4 This character set is applicable to all alphabets of Latin letters.
- 1.5 This character set includes control characters for code extension where its 128 characters are insufficient for particular applications. Procedures for the use of these control characters are specified in Standard ECMA-35.
- 1.6 The definitions of some control characters in this Standard assume that data associated with them are to be processed serially in a forward direction. When they are included in strings of data which are processed other than serially in a forward direction or when they are included in data formatted for fixed-record processing they may have undesirable effects or may require additional special treatment to ensure that they result in their desired function.

2. CONFORMANCE AND IMPLEMENTATION

2.1 Conformance

A coded character set is in conformance with this Standard if it is a version in accordance with 6. Equipment claimed to implement this Standard shall be able to interchange information by means of a version of the 7-bit coded character set, this version shall be identified in any such claim.

2.2 Implementation

The use of this character set requires definitions of its implementation in various media. For example, these could include punched tapes, punched cards, magnetic 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 local connection of various units (such as input and output devices and computers) or by means of data transmission equipment.

The implementation of this coded character set in physical media and for transmission, taking into account the need for error checking, is the subject of other ECMA standards.

3. DEFINITIONS

For the purpose of this Standard the following definitions apply.

3.1 Bit Combination

An ordered set of bits used for the representation of characters.

3.2 Character

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

3.3 Coded Character set; Code

A set of unambiguous rules that establishes a character set and the one-to-one relationship between the characters of the set and their bit combinations.

3.4 Code Extension

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

3.5 Code Table

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

3.6 Control Character

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

3.7 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.

3.8 Graphic Character

A character, other than a control function, that has a visual representation normally handwritten, printed or displayed.

3.9 Graphic Symbol

A visual representation of a graphic character or of a control function.

3.10 Position

A character, other than a control function, that has a visual representation normally handwritten, printed or displayed.

4. SPECIFICATION OF THE CODED CHARACTER SET

The bits of the bit combinations of the 7-bit code are identified by b_7 , b_6 , b_5 , b_4 , b_3 , b_2 and b_1 , where b_7 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 integers in the range 0 to 127 in binary notation by attributing the following weights to the individual bits:

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

In this Standard, the bit combinations are identified by notation of the form x/y , where x is a number in the range 0 to 7 and y is a number in the range 0 to 15. The correspondence between the notations of the form x/y and the bit combinations consisting of the bits b_7 to b_1 is as follows:

- x is the number represented by b_7 , b_6 and b_5 where these bits are given the weights 4, 2 and 1 respectively;
- y 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 x/y are the same as those used to identify code table positions, where x is the column number and y the row number (see 7).

The 128 bit combinations of the 7-bit code represent control characters and graphic characters. The allocation of characters to bit combinations is based on the following principles:

- the bit combinations 0/0 to 1/15 represent 32 control characters;
- the bit combination 2/0 represents the character SPACE, which is interpreted both as a control character and as a graphic character;
- the bit combinations 2/1 to 7/14 represent up to 94 graphic characters as one or more of these bit combinations may be declared to be unused (see 4.3);
- the bit combination 7/15 represents the control character DELETE.

The allocation of individual characters to the bit combinations of the 7-bit code is specified in 4.1, 4.2 and 4.3 below.

This 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. For acronyms only capital letters and digits are used. It is intended that the acronyms and this convention be retained in all translations of the text.

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

4.1 Control Characters

The control characters of the 7-bit coded character set are classified in the following categories:

i) Transmission control characters

Transmission control characters are intended to control or facilitate transmission of information over telecommunication networks. Procedures for the use of the transmission control characters on telecommunication networks are the subject of other standards.

ii) Format effectors

Format effectors are mainly intended for the control of the layout and positioning of information on character-imaging devices such as printing and display devices.

iii) Control characters for code extension

Code extension control characters are used to extend the character set of the code. They may alter the meaning of one or more bit combinations that follow them in the data stream. Procedures for the use of the code extension control characters are specified in ECMA-35.

iv) Device control characters

Device control characters are intended for the control of local or remote devices or ancillary devices connected to a data processing or data communication system. These control characters are not intended to control data communication systems; this should be achieved by the use of transmission control characters.

v) Information separators

Information separators are used to separate and qualify data logically. There are four such characters. They may be used either in hierarchical order or non-hierarchically; in the latter case, their specific meanings depend on the application.

vi) Other control characters

These are the control characters that fall outside the preceding categories.

The composition of each category, and the allocation of the individual control characters in each category to bit combinations of the 7-bit code are specified in 4.1.1 to 4.1.6, each containing a table consisting of three columns. The first column specifies the acronym of the control characters, the second column specifies their standard name and the third column, labelled "Coded representation", specifies the bit combination representing the control character concerned.

Detailed functional descriptions of all control characters are given in 8.

4.1.1 Transmission control characters

Acronym	Name	Coded Representation
SOH	START OF HEADING	0/1
STX	START OF TEXT	0/2
ETX	END OF TEXT	0/3
EOT	END OF TRANSMISSION	0/4
ENQ	ENQUIRY	0/5
ACK	ACKNOWLEDGE	0/6
DLE	DATA LINK ESCAPE	1/0
NAK	NEGATIVE ACKNOWLEDGE	1/5
SYN	SYNCHRONOUS IDLE	1/6
ETB	END OF TRANSMISSION BLOCK	1/7

4.1.2 Format effectors

Acronym	Name	Coded Representation
BS	BACKSPACE	0/8
HT	HORIZONTAL TABULATION	0/9
LF	LINE FEED	0/10
VT	VERTICAL TABULATION	0/11
FF	FORM FEED	0/12
CR	CARRIAGE RETURN	0/13

4.1.2.1 Concepts

The definitions of the format effectors use the following concepts:

- i) A page is composed of a number of lines, each being composed of a number of character positions.
- ii) Each character position is capable of imaging SPACE or a graphic symbol.
- iii) The graphic symbol imaged at a character position represents a graphic character, a control function, or a combination of one or more graphic characters and/or control functions.
- iv) The active position is the character position at which the action required by the next character in the data stream is to be effected. If the next character is a graphic character, its graphic symbol is imaged at that position; if it is a control character, the corresponding function is performed relative to that position.
- v) Movements of the active position are effected as follows:
 - The active position is advanced one character position immediately after imaging a SPACE or a graphic symbol, and upon the execution of the function corresponding to a control character for which a graphic symbol is required to be imaged.
 - The active position is moved to a specified character position upon the execution of the function corresponding to a control character that is defined to cause a movement of the active position (i.e. a format effector).
- vi) The active position is not moved upon execution of the function corresponding to a control character that is neither required to be imaged by a graphic symbol nor defined to cause a movement of the active position.
- vii) The effect of an attempt to move the active position beyond the boundaries of a line or a page is not defined by this Standard.

4.1.2.2 Combined horizontal and vertical movements of the active position

The format effectors are defined for applications in which horizontal and vertical movements of the active position are effected separately. If a single control character is required to effect the action of CARRIAGE RETURN in combination with a vertical movement, the format effector for that vertical movement shall be used. For

example, if the function "new line" (equivalent to the combination of CARRIAGE RETURN and LINE FEED) is required as a single control character, bit combination 0/10 shall be used to represent it. This substitution requires agreement between the sender and the recipient of the data, and the format effectors (LINE FEED, VERTICAL TABULATION and/or FORM FEED) that are affected shall be identified (see 6).

In order to avoid the need for such prior agreement, to facilitate interchange and to avoid conflicts with specifications in other standards, the use of format effectors for vertical movement to effect combined horizontal and vertical movements by the format effector for the vertical movement only is deprecated. It is strongly recommended to use two control characters, for example CARRIAGE RETURN (CR) and LINE FEED (LF) to obtain the effect of "new line".

4.1.3 Code extension control characters

Acronym	Name	Coded Representation
SO	SHIFT-OUT	0/14
SI	SHIFT-IN	0/15
ESC	ESCAPE	1/11

4.1.4 Device control characters

Acronym	Name	Coded Representation
DC1	DEVICE CONTROL ONE	1/1
DC2	DEVICE CONTROL TWO	1/2
DC3	DEVICE CONTROL THREE	1/3
DC4	DEVICE CONTROL FOUR	1/4

4.1.5 Information separators

Acronym	Name	Coded Representation
IS4	INFORMATION SEPARATOR FOUR (FILE SEPARATOR)	1/12
IS3	INFORMATION SEPARATOR THREE (GROUP SEPARATOR)	1/13
IS2	INFORMATION SEPARATOR TWO (RECORD SEPARATOR)	1/14
IS1	INFORMATION SEPARATOR ONE (UNIT SEPARATOR)	1/15

Each information separator is given two names. The names INFORMATION SEPARATOR FOUR, INFORMATION SEPARATOR THREE, INFORMATION SEPARATOR TWO and INFORMATION SEPARATOR ONE are the general names. The names FILE SEPARATOR, GROUP SEPARATOR, RECORD SEPARATOR and UNIT SEPARATOR are the specific names and are intended mainly for applications where the information separators are used hierarchically. The ascending order is then US, RS, GS, FS. In this case, data normally delimited by a particular separator cannot be split by a higher-order separator but will be considered as delimited by any higher-order separator.

4.1.6 Other control characters

Acronym	Name	Coded Representation
NUL	NULL	0/0
BEL	BELL	0/7
CAN	CANCEL	1/8
EM	END OF MEDIUM	1/9
SUB	SUBSTITUTE CHARACTER	1/10
DEL	DELETE	7/15

4.2 Character SPACE

The acronym of the character SPACE is **SP** and its coded representation is 2/0.

This character is interpreted both as a graphic character and as a control character. 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.

4.3 Graphic Characters

The 94 bit combinations 2/1 to 7/14 are used for the representation of graphic characters as specified in 4.3.1, 4.3.2 and 4.3.3 below, each containing a table consisting of three columns. The first column is labelled "Graphic Symbol" and specifies the graphic symbol of each graphic character, the second column specifies its standard name and the third column, labelled "Coded Representation" specifies the bit combination representing the graphic character concerned.

All graphic characters of any version of the 7-bit coded character set are spacing characters, i.e. they cause the active position to advance.

4.3.1 Unique graphic character allocations

A unique graphic character is allocated to each of the 82 bit combinations 2/1, 2/2, 2/5 to 3/15, 4/1 to 5/10, 5/15 and 6/1 to 7/10.

Graphic Symbol	Name	Coded Representation
!	EXCLAMATION MARK	2/1
"	QUOTATION MARK	2/2
%	PERCENT SIGN	2/5
&	AMPERSAND	2/6
'	APOSTROPHE	2/7
(LEFT PARENTHESIS	2/8
)	RIGHT PARENTHESIS	2/9
*	ASTERISK	2/10
+	PLUS SIGN	2/11
,	COMMA	2/12
-	HYPHEN, MINUS SIGN	2/13
.	FULL STOP	2/14
/	SOLIDUS	2/15
0	DIGIT ZERO	3/0
1	DIGIT ONE	3/1
2	DIGIT TWO	3/2
3	DIGIT THREE	3/3
4	DIGIT FOUR	3/4
5	DIGIT FIVE	3/5
6	DIGIT SIX	3/6
7	DIGIT SEVEN	3/7
8	DIGIT EIGHT	3/8
9	DIGIT NINE	3/9
:	COLON	3/10
;	SEMICOLON	3/11
<	LESS-THAN SIGN	3/12
=	EQUALS SIGN	3/13
>	GREATER-THAN SIGN	3/14
?	QUESTION MARK	3/15
A	CAPITAL LETTER A	4/1
B	CAPITAL LETTER B	4/2
C	CAPITAL LETTER C	4/3

Graphic Symbol	Name	Coded Representation
D	CAPITAL LETTER D	4/4
E	CAPITAL LETTER E	4/5
F	CAPITAL LETTER F	4/6
G	CAPITAL LETTER G	4/7
H	CAPITAL LETTER H	4/8
I	CAPITAL LETTER I	4/9
J	CAPITAL LETTER J	4/10
K	CAPITAL LETTER K	4/11
L	CAPITAL LETTER L	4/12
M	CAPITAL LETTER M	4/13
N	CAPITAL LETTER N	4/14
O	CAPITAL LETTER O	4/15
P	CAPITAL LETTER P	5/0
Q	CAPITAL LETTER Q	5/1
R	CAPITAL LETTER R	5/2
S	CAPITAL LETTER S	5/3
T	CAPITAL LETTER T	5/4
U	CAPITAL LETTER U	5/5
V	CAPITAL LETTER V	5/6
W	CAPITAL LETTER W	5/7
X	CAPITAL LETTER X	5/8
Y	CAPITAL LETTER Y	5/9
Z	CAPITAL LETTER Z	5/10
-	LOW LINE, UNDERLINE	5/15
a	SMALL LETTER a	6/1
b	SMALL LETTER b	6/2
c	SMALL LETTER c	6/3
d	SMALL LETTER d	6/4
e	SMALL LETTER e	6/5
f	SMALL LETTER f	6/6
g	SMALL LETTER g	6/7
h	SMALL LETTER h	6/8

Graphic Symbol	Name	Coded Representation
i	SMALL LETTER i	6/9
j	SMALL LETTER j	6/10
k	SMALL LETTER k	6/11
l	SMALL LETTER l	6/12
m	SMALL LETTER m	6/13
n	SMALL LETTER n	6/14
o	SMALL LETTER o	6/15
p	SMALL LETTER p	7/0
q	SMALL LETTER q	7/1
r	SMALL LETTER r	7/2
s	SMALL LETTER s	7/3
t	SMALL LETTER t	7/4
u	SMALL LETTER u	7/5
v	SMALL LETTER v	7/6
w	SMALL LETTER w	7/7
x	SMALL LETTER x	7/8
y	SMALL LETTER y	7/9
z	SMALL LETTER z	7/10

4.3.2 Alternative graphic character allocations

Two alternative graphic characters are allocated to each of the bit combinations 2/3 and 2/4.

Graphic Symbol	Name	Coded Representation
£	POUND SIGN	2/3
#	NUMBER SIGN	2/3
\$	DOLLAR SIGN	2/4
¤	CURRENCY SIGN	2/4

Either the character POUND SIGN or the character NUMBER SIGN shall be allocated to bit combination 2/3 and either the character DOLLAR SIGN or the character CURRENCY SIGN shall be allocated to bit combination 2/4 (see 6).

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

4.3.3 National or application-oriented graphic character allocation

No specific graphic character is allocated to the ten bit combinations 4/0, 5/11 to 5/14, 6/0 and 7/11 to 7/14. These bit combinations are available for national or application-oriented use. A unique graphic character shall be allocated to each of these bit combinations, or the bit combination shall be declared unused (see 6).

5. COMPOSITE GRAPHIC CHARACTERS

In any version of the 7-bit coded character set specified according to this Standard, all graphic characters are spacing characters which cause the active position to move forward. However, by using BACKSPACE or CARRIAGE RETURN, it is possible to image two or more graphic characters at the same character position.

For example, SOLIDUS and EQUALS SIGN can be combined to image "not equals". The character LOW LINE, that may be used as a free-standing character, can also be associated with other character(s) to represent the graphic rendition "underlined".

Diacritical marks may be allocated to the bit combinations specified in 4.3.3 and be available for composing accented letters. For such composition, it is recommended to use a sequence of three characters, the first or last of which is the letter to be accented and the second of which is BACKSPACE. Furthermore, QUOTATION MARK, APOSTROPHE or COMMA can be associated with a letter by means of BACKSPACE for the composition of an accented letter with a diaeresis, an acute accent or a cedilla, respectively.

6. VERSIONS OF THE CODED CHARACTER SET

6.1 General

In order to use the 7-bit coded character set for information interchange, it is necessary to exercise the options left open in 4:

- to each of the bit combinations 2/3 and 2/4 one of the alternative graphic characters specified in 4.3.2 shall be allocated;
- each of the bit combinations 4/0, 5/11 to 5/14, 6/0 and 7/11 to 7/14 shall have a unique graphic character allocated to it, or be declared unused;
- the format effectors, if any, to which the facility of 4.1.2.2 applies, shall be identified.

A graphic character allocated to a bit combination specified in 4.3.1 and 4.3.2 shall not be allocated to any other bit combination. For example the POUND SIGN, if not allocated to bit combination 2/3, shall not be allocated to any other bit combination.

A character set completed in this way is called a "version of ECMA-6" (see Appendix A).

6.2 National Versions

- 6.2.1 The responsibility for defining national versions lies with the national standardization bodies. These bodies shall exercise the options available and make the required selection (see Appendix A).
- 6.2.2 If so required, more than one national version can be defined within a country. The different versions shall be separately identified. In particular when for a given bit combination, for example 5/12, alternative graphic characters are required, two different versions shall be identified, even if they differ only by this single character.
- 6.2.3 If there is in a country no special demand for specific graphic characters, it is strongly recommended that the characters of the International Reference Version (IRV) (see 6.4) be selected and allocated to the same bit combinations as in the IRV.

However, when graphic characters that are different from the characters of the IRV are required, they shall have distinct forms and be given distinctive names which are not in conflict with any of the forms or the names of any of the graphic characters in the IRV.

6.3 Application-oriented Versions

Within national or international industries, organizations or professional groups, application-oriented versions can be used. They require precise agreement among the interested parties, who will have to exercise the options available and to make the required selection.

6.4 International Reference Version (IRV)

This version is available for use when there is no requirement to use a national or an application-oriented version. In information interchange, the IRV is assumed unless a particular agreement exists between sender and recipient of the data. The graphic characters allocated to the IRV are specified below.

Graphic Symbol	Name	Coded Representation
#	NUMBER SIGN	2/3
¤	CURRENCY SIGN	2/4
@	COMMERCIAL AT	4/0
[LEFT SQUARE BRACKET	5/11
\	REVERSE SOLIDUS	5/12
]	RIGHT SQUARE BRACKET	5/13
^	CIRCUMFLEX ACCENT	5/14
`	GRAVE ACCENT	6/0
{	LEFT CURLY BRACKET	7/11
	VERTICAL LINE	7/12
}	RIGHT CURLY BRACKET	7/13
~	TILDE, OVERLINE	7/14

It should be noted that no substitution is allowed when using the IRV and that the facility of 4.1.2.2 does not apply to any format effector.

According to 5 it is permitted to use composite graphic characters and there is no limit to their number. Because of this freedom, their processing and imaging may cause difficulties at the receiving end. Therefore agreement between sender and recipient of the data is recommended if composite characters are used.

7. CODE TABLES

A 7-bit code table consists of 128 positions arranged in 8 columns and 16 rows. The columns are numbered 0 to 7, and the rows are numbered 0 to 15.

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

The 128 positions of the code table are in one-to-one correspondence with the bit combinations of the 7-bit code. The notation of a code table position, of the form x/y, is the same as that of the corresponding bit combination (see 4).

Each code table position contains a symbol and/or a reference to a clause of this Standard. When a code table position corresponds to a bit combination that represents a control character or the character SPACE, the symbol is the acronym of the character allocated; otherwise it is the graphic symbol representing the character allocated, if any. A reference to 4.1.2.2, 4.3.2 or 4.3.3 is denoted by ①, ② or ③ respectively.

Table 1 is the Basic 7-bit Code Table. It shows the 7-bit coded character set specified in 4 and indicates the options related to format effectors (4.1.2.2), alternative graphic characters (4.3.2) and national or application-oriented use (4.3.3).

Table 2 is the code table for the IRV of the 7-bit coded character set. It shows the result of exercising the three identified options in the manner specified in 6.4.

Table 1 - Basic 7-Bit Code Table

					b7	0	0	0	0	1	1	1	1
					b6	0	0	1	1	0	0	1	1
					b5	0	1	0	1	0	1	0	1
						0	1	2	3	4	5	6	7
b4	b3	b2	b1										
0	0	0	0	0	NUL	DLE	SP	0	③	P	③	p	
0	0	0	1	1	SOH	DC1	!	1	A	Q	a	q	
0	0	1	0	2	STX	DC2	"	2	B	R	b	r	
0	0	1	1	3	ETX	DC3	# ② £	3	C	S	c	s	
0	1	0	0	4	EOT	DC4	¤ ② \$	4	D	T	d	t	
0	1	0	1	5	ENQ	NAK	%	5	E	U	e	u	
0	1	1	0	6	ACK	SYN	&	6	F	V	f	v	
0	1	1	1	7	BEL	ETB	'	7	G	W	g	w	
1	0	0	0	8	BS	CAN	(8	H	X	h	x	
1	0	0	1	9	HT	EM)	9	I	Y	i	y	
1	0	1	0	10	LF ①	SUB	*	:	J	Z	j	z	
1	0	1	1	11	VT ①	ESC	+	;	K	③	k	③	
1	1	0	0	12	FF ①	IS4	,	<	L	③	l	③	
1	1	0	1	13	CR ①	IS3	-	=	M	③	m	③	
1	1	1	0	14	SO	IS2	.	>	N	③	n	③	
1	1	1	1	15	SI	IS1	/	?	0	_	o	DEL	

① See 4.1.2.2.
 ② See 4.3.2.
 ③ See 4.3.3 and 6.2.3.

Table 2 - International Reference Version (IRV)

					b7	0	0	0	0	1	1	1	1
					b6	0	0	1	1	0	0	1	1
					b5	0	1	0	1	0	1	0	1
						0	1	2	3	4	5	6	7
b4	b3	b2	b1										
0	0	0	0	0	NUL	DLE	SP	0	@	P	`	p	
0	0	0	1	1	SOH	DC1	!	1	A	Q	a	q	
0	0	1	0	2	STX	DC2	"	2	B	R	b	r	
0	0	1	1	3	ETX	DC3	#	3	C	S	c	s	
0	1	0	0	4	EOT	DC4	␣	4	D	T	d	t	
0	1	0	1	5	ENQ	NAK	%	5	E	U	e	u	
0	1	1	0	6	ACK	SYN	&	6	F	V	f	v	
0	1	1	1	7	BEL	ETB	'	7	G	W	g	w	
1	0	0	0	8	BS	CAN	(8	H	X	h	x	
1	0	0	1	9	HT	EM)	9	I	Y	i	y	
1	0	1	0	10	LF	SUB	*	:	J	Z	j	z	
1	0	1	1	11	VT	ESC	+	;	K	[k	{	
1	1	0	0	12	FF	IS4	,	<	L	\	l		
1	1	0	1	13	CR	IS3	-	=	M]	m	}	
1	1	1	0	14	SO	IS2	.	>	N	^	n	~	
1	1	1	1	15	SI	IS1	/	?	O	_	o	DEL	

NOTE

The graphic symbol represented in position 7/14 in ISO 646 is a straight line.

8. DESCRIPTION OF THE CONTROL CHARACTERS

The control characters are listed below in the alphabetical order of their acronyms.

8.1 ACK ACKNOWLEDGE

A transmission control character transmitted by a receiver as an affirmative response to the sender.

8.2 BEL BELL

A control character that is used when there is a need to call for attention; it may control alarm or attention devices.

8.3 BS BACKSPACE

A format effector which causes the active position to move one character position backwards.

8.4 CAN CANCEL

A character, or the first character of a sequence, indicating that the data preceding it is in error. As a result, this data shall be ignored. The specific meaning of this character shall be defined for each application and/or between sender and recipient.

8.5 CR CARRIAGE RETURN

A format effector which causes the active position to move to the first character position on the same line.

8.6 DC1 DEVICE CONTROL ONE

A device control character which is primarily intended for turning on or starting an ancillary device. If it is not required for this purpose, it may be used to restore a device to the basic mode of operation (see also DC2 and DC3), or for any other device control function not provided by other DCs.

8.7 DC2 DEVICE CONTROL TWO

A device control character which is primarily intended for turning on or starting an ancillary device. If it is not required for this purpose, it may be used to set a device to a special mode of operation (in which case DC1 is used to restore the device to the basic mode), or for any other device control function not provided by other DCs.

8.8 DC3 DEVICE CONTROL THREE

A device control character which is primarily intended for turning off or stopping an ancillary device. This function may be a secondary level stop, for example wait, pause, stand-by or halt (in which case DC1 is used to restore normal operation). If it is not required for this purpose, it may be used for any other ancillary device control function not provided by other DCs.

8.9 DC4 DEVICE CONTROL FOUR

A device control character which is primarily intended for turning off, stopping or interrupting an ancillary device. If it is not required for this purpose, it may be used for any other device control function not provided by other DCs.

8.10 DEL DELETE

A character used primarily to erase or obliterate an erroneous or unwanted character in punched tape. DEL characters may also 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.

8.11 DLE DATA LINK ESCAPE

A transmission control character which will change the meaning of a limited number of contiguously following bit combinations. It is used exclusively to provide supplementary transmission control functions. Only graphic characters and transmission control characters can be used in DLE sequences.

8.12 EM END OF MEDIUM

A control character that may be used to identify the physical end of a medium, or the end of the used portion of a medium, or the end of the wanted portion of data recorded on a medium. The position of this character does not necessarily correspond to the physical end of the medium.

8.13 ENQ ENQUIRY

A transmission control character used as a request for a response from a remote station - the response may include station identification and/or station status. When a "Who are you" function is required on the general switched transmission network, the first use of ENQ after the connection is established shall have the meaning "Who are you" (station identification). Subsequent use of ENQ may, or may not, include the function "Who are you", as determined by agreement.

8.14 EOT END OF TRANSMISSION

A transmission control character used to indicate the conclusion of the transmission of one or more texts.

8.15 ESC ESCAPE

A control character which is used to provide additional characters. It alters the meaning of a limited number of contiguously following bit combinations. The use of this character is identified in ECMA-35.

8.16 ETB END OF TRANSMISSION BLOCK

A transmission control character used to indicate the end of a transmission block of data where data is divided into such blocks for transmission purposes.

8.17 ETX END OF TEXT

A transmission control character which terminates a text.

8.18 FF FORM FEED

A format effector which causes the active position to advance to the corresponding character position on a pre-determined line of the next form or page.

8.19 HT HORIZONTAL TABULATION

A format effector which causes the active position to advance to the next pre-determined character position.

8.20 IS1 (US) INFORMATION SEPARATOR ONE (UNIT SEPARATOR)

A control character used to separate and qualify data logically; its specific meaning has to be defined for each application. If this character is used in hierarchical order as specified in the general definition of IS, it delimits a data item called a unit.

8.21 IS2 (RS) INFORMATION SEPARATOR TWO (RECORD SEPARATOR)

A control character used to separate and qualify data logically; its specific meaning has to be defined for each application. If this character is used in hierarchical order as specified in the general definition of IS, it delimits a data item called a record.

8.22 IS3 (GS) INFORMATION SEPARATOR THREE (GROUP SEPARATOR)

A control character used to separate and qualify data logically; its specific meaning has to be defined for each application. If this character is used in hierarchical order as specified in the general definition of IS, it delimits a data item called a group.

8.23 IS4 (FS) INFORMATION SEPARATOR FOUR (FILE SEPARATOR)

A control character used to separate and qualify data logically; its specific meaning has to be defined for each application. If this character is used in hierarchical order as specified in the general definition of IS, it delimits a data item called a file.

8.24 LF LINE FEED

A format effector which causes the active position to advance to the corresponding character position of the next line.

8.25 NAK NEGATIVE ACKNOWLEDGE

A transmission control character transmitted by a receiver as a negative response to the sender.

8.26 NUL NULL

A control character used to accomplish media-fill or time-fill. NUL characters 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.

8.27 SI SHIFT-IN

A control character which is used in conjunction with SO and ESC to extend the graphic character set of the code. It may re-instate the standard meanings of the bit combinations which follow it. The effect of this character when using code extension techniques is described in ECMA-35.

8.28 SO SHIFT-OUT

A control character which is used in conjunction with SI and ESC to extend the graphic character set of the code. It may alter the meaning of the bit combinations 2/1 to 7/14 which follow it until a SI character is reached. The effect of this character when using code extension techniques is described in ECMA-35.

8.29 SOH START OF HEADING

A transmission control character used as the first character of a heading of an information message.

8.30 STX START OF TEXT

A transmission control character which precedes a text and which is used to terminate a heading.

8.31 SUB SUBSTITUTE CHARACTER

A control character used in the place of a character that has been found to be invalid or in error. SUB is intended to be introduced by automatic means.

8.32 SYN SYNCHRONOUS IDLE

A transmission control character used by a synchronous transmission system in the absence of any other character (idle condition) to provide a signal from which synchronism may be achieved or retained between data terminal equipment.

8.33 VT VERTICAL TABULATION

A format effector which causes the active position to advance to the corresponding character position on the next predetermined line.

APPENDIX A

GUIDELINES FOR STANDARDS DERIVED FROM ECMA-6

A.1 General

When national or application-oriented standards based on ECMA-6 are drafted, it is recommended to take account of the following considerations.

A.2 Structure of a Standard

It is recommended to adopt the same structure and editorial style as implemented for ECMA-6. All facilities, restrictions and specifications of the standard should be stated clearly in sentences using plain language, rather than be summarized by tables with notes.

A.2.1 Control functions

The standard should contain explicit descriptions of the control functions. Even where these descriptions are identical with the descriptions in 8, they should be explicit descriptions, not just referred to ECMA-6. For application-oriented standards specific meanings of the Information Separators and of the Device Controls should be defined.

A.2.2 Graphic characters (see 6.2.3)

Where there is no need for particular characters, the graphic characters of the International Reference Version (IRV) should be allocated to the same positions and with the same name as in ECMA-6.

A.2.3 Composite graphic characters and repertoire (see 5)

ECMA-6 permits the construction of composite graphic characters by using the control characters BACKSPACE and CARRIAGE RETURN so as to image two or more graphic characters at the same character position.

The total number of graphic characters which can be obtained from any version of the character set, with or without using this facility, is called the repertoire. ECMA-6 does not define a particular repertoire. However, as the interpretation and/or the imaging of composite characters may cause difficulties, agreement between sender and recipient of the data may be required. In order to avoid the necessity of such agreement and to facilitate interchange, national or application-oriented standards may specify a standard repertoire of graphic characters and thus recognize only a limited

number of composite graphic characters. Such limitations are considered fully compatible with ECMA-6.

A.2.4 Versions

In a standard one or more versions can be specified. It should be noted that a version is not a standard but only part of a standard. The standard itself consists of the well defined version or versions and a set of clauses as mentioned above. The definition of a version requires that the options mentioned in 6.1 be accurately exercised.

APPENDIX B

MAIN DIFFERENCES BETWEEN THE 4th EDITION (August 1973)
AND THE PRESENT 5th EDITION

B.1 Editing and Presentation

The text has been edited so as to improve its intelligibility. All facilities are explicitly described in the text rather than in notes or by implication. Unnecessary abbreviations have been suppressed.

B.2 Conformance

A conformance clause 2.1 has been introduced.

B.3 Definitions

A clause 3 defining the major terms used in this ECMA Standard has been introduced. These definitions are reproduced in ECMA-35.

B.4 National-use Positions

In the 4th edition of ECMA-6, the ten national-use positions were divided into two categories, characters being allocated to three of them. In this edition all ten national-use positions are treated equally and, as a result, the Basic Code Table has ten positions to which no graphic character is allocated.

B.5 International Reference Version (IRV)

The graphic characters of the IRV are unchanged. The name of the character in position 5/14 is now CIRCUMFLEX ACCENT.

B.6 Composite Graphic Characters

Facilities for the superimposition of graphic characters which were implicit in the 4th edition are explicitly described in this edition.

