The I.S.O. character code

By H. McG. Ross*

A specific and fully detailed proposal has been prepared within the International Organization for Standardization on a character code for use in computer, data-processing and telecommunication systems. Although it is still subject to formal ratification, it is clear that this new character code and its repertoire of symbols should be considered for use in any future system. This affects designers of equipment and programming languages, systems planners, and those who have to choose the symbols and equipment used in commercial applications.

Introduction

For several years work has been in progress on developing standard character codes for use in computer, data-processing and communication systems. Ultimately, this has been under the auspices of the International Organization for Standardization, and during a recent meeting agreement was reached at the technical level on specific proposals. These have to go through a fairly long process of formal ratification, and it is impossible to forecast the outcome; but the overriding attitude of those taking part in the technical discussions was that the time has come to settle on these proposals, and thus to be able to begin to derive the benefits of using a common unified system.

The opportunity is presented, therefore, to enter a period of stability with respect to character codes; to escape from the present babel; to establish means for different users more readily to communicate with each other and interchange work; and for the development of a whole series of peripheral machines which, in this respect, are compatible and really suited for data-processing work.

The work covers three aspects: the choice of graphic symbols and control characters; the system for encoding them in binary form; and consideration of methods of exploiting the features of the overall code. Throughout the project very great care has been given to making the scheme as versatile and as widely useful as possible.

History of the project

This type of character code originated within Ferranti Limited (Mr. I. V. Idelson and Dr. C. H. Lindsey). It was first used on the punched tape for Orion and Atlas computers. This code is essentially a 6-bit code with two shifts, and when implemented on 7-track punched tape has the normal arrangement of using one track as a parity check on the data tracks. Similar character codes are used within Atlas and Orion and the various peripheral machines, and the final result is a group of eight minor variants or shifts; these are all closely related and conversion between one and another is easily affected, but it serves as an illustration of the limitations which are experienced in practice of any 6-bit code when used for anything other than the simplest equipment.

The work was developed within the British Standards Institution and in due course published as B.S. 3480: "Data processing codes for punched tape." This specification covers a series of character codes, all on punched tape, which are essentially derived from a basic 6-bit code. In the most sophisticated form, 8-track tape, the 8th track is used to determine two shifts, one containing capital letters and the other small letters. A number of British computer systems, particularly KDF 9 and Elliott 503, used punched tape and internal character codes based on B.S. 3480. The basic concepts of this type of code, as it then stood, were described in this *Journal* (Ross, 1961); most of these concepts have carried right through the project.

This type of code was used also in B.S. 3635: "Codes for punched tape for numerically controlled machine tools." As far as the character code and its implementation on punched tape is concerned, the importance of this standard lies in attempting to correlate the code for machine-tool control with that for computers. There are some small differences between this code and that of B.S. 3480 due to the developments in the international field referred to below.

This basic work was considered by workers in America (Mr. J. F. Auwaerter, and Bemer, Smith and Williams, 1961) who were developing advanced telecommunications systems—particularly those in which communication equipment and computers would be used together. Many valuable contributions were made in the American work, the more important being:

(*a*) A substantial number of characters were introduced for the control of telecommunication systems and for the associated equipment.

(*b*) The concept was introduced that the character code should be regarded in a generalized way for the interchange of data between one user or machine and another. Thus it has an abstract form, each printable symbol or control character being represented by a pattern of binary digits. Then, as a supplementary process, these are implemented on various media and in various machines, in different ways as appropriate. As a corollary, the basic code represents the form in which it is transferred from one user to another, and it is at this stage that the standard must be honoured; thereafter, each individual user is free

* *With Standard Telephones and Cables Ltd., Connaught House, Aldwich, London, W.C.2; formerly with I.C.T. (Ferranti Computer Department).*

to modify it in his own equipment as he wishes, thus giving substantial flexibility to designers of computers and peripheral machines.

(*c*) Major emphasis was placed on a code with 7-bit characters because, on the one hand, the limitations of any 6-bit code were clearly recognized and, on the other hand, there was a greater willingness to accept the consequences of handling the larger characters.

(*d*) A sophisticated and powerful arrangement was introduced for relating 6 and 7-bit codes.

This work, done within the various national standards organizations, passed as is normal practice to the International Organization for Standardization. The purpose of ISO is to produce internationally acceptable recommendations on which each national standards organization bases its national standard; the latter carry the ultimate authority, but this process ensures that the various national standards are identical or at least compatible. A number of other countries joined in this work, and in due course a first and then a second draft ISO proposal were produced. These proposals covered two codes in closely related form, with 6 and 7-bit characters.

The European Computer Manufacturers Association recognized the importance of having a common character code for use with the various computers in Europe, and in 1963 published a standard. Primary interest lay in a 6-bit code, and this ECMA standard follows the ISO 6-bit version. In addition ECMA published a very valuable discussion of the philosophy of this type of code, and the relation between the 6 and 7-bit forms.

The American Standards Association published in 1963 the American Standard Code for Information Interchange (ASCII). In this case primary interest lay in a 7-bit code, and therefore ASCII follows the 7-bit form of the ISO second draft proposal.

During 1963 C.C.I.T.T. – the standards organization for all international telecommunication matters – studied the ISO proposal with a view to its use for future message telecommunication systems. It was recognized that there was a need for a larger character code than the International Telegraph Alphabet No. 2—commonly called the Telex code—for future telecommunication systems and traffic although, because of the immense amount of Telex equipment currently established, the new code would complement, not displace, the Telex code. The C.C.I.T.T. introduced a particular need for adequate provision of accents and other diacritical signs which are required for any system which is to cope adequately with the many languages of the industrialized countries, even though so much of the previous computer work had been based on English.

Further study of the foregoing work, especially in telecommunications and in those fields in which IBM have exceptional experience, led to a number of suggestions for further improvements which, during May 1964, were embodied in a ISO third draft proposal. It is this which forms the basis of the present paper.

Present status of the ISO code

When a project has been worked out at the technical level, the normal procedure within ISO is for it to be ratified by the various national standards organizations. This usually takes some time, and it has to be gone through in connection with this character code; the whole matter must therefore be treated with a certain reserve until this has been completed. Then national standards may be created or modified in accordance with the ISO recommendation. B.S.I. have already announced their intention to develop B.S. 3480 to conform with the final ISO code, and it is of course to be hoped that ASCII will be similarly amended. ECMA have under discussion a 7-bit code to complement their published 6-bit code.

U.I.C. and I.C.A.O., the international organizations dealing with rail and air transport respectively, are known to be considering the use of this type of 7-bit character code for interchanging messages. An interesting development is that this character code, together with an extension to accommodate a further 64 of the special Japanese graphic symbols, has been found satisfactory for use in the Japanese computer industry.

ASCII is used in an ambitious new telecommunications service provided by the American Telephone and Telegraph Co., and, together with an extended BCD (punch-card oriented) code, it is adopted for the IBM System/360.

Explanation of ISO 7-bit code, and purpose of the characters

The ISO character code comprises two systems, one with 7-bit characters and 128 combinations, the other with 6-bit characters and 64 combinations. These are shown in Tables 1 and 2 respectively. The codes will be explained in terms of the 7-bit set, since this contains the 6-bit set.

Within each character the bits can be given the usual binary significance, the convention being to show the most-significant bit on the left. The code table is presented in a series of columns and rows, the columns being identified by the more-significant bits in each character, and the rows by the less-significant.

Considering the 128 characters as one continuous string in ascending order of significance, the characters are grouped together so that on the one hand they may be readily identified either by program or equipment, and on the other hand they collate satisfactorily for sorting. In general, for the latter purpose the sequence is controls, important separating and punctuating symbols, numerals, capital letters, and small letters; some of the less-commonly used symbols, particularly those not often involved in sorting, are interposed into gaps within this general sequence.

More often, however, advantage will be taken of the grouping into columns within the code table. For example, the control characters are identified by the two most-significant bits being 00, and for all letters

Table 1: 7-bit code table

b7	b6	b5	b4	b3	b2	b1		0	0	0	0	1	1	1	1
								0	0	1	1	0	0	1	1
								0	1	0	1	0	1	0	1
								0	1	2	3	4	5	6	7
0	0	0	0	0	0	0	0	(TC_0) Null	(TC_7) DLE	Space	0	—	P	(@) ③	p
0	0	0	0	1	0	0	1	(TC_1) SOH	DC_1	!	1	A	Q	a	q
0	0	0	1	0	0	1	2	(TC_2) STX	DC_2	" ⑧	2	B	R	b	r
0	0	0	1	1	0	1	3	(TC_3) ETX	DC_3	# ⑥	3	C	S	c	s
0	0	1	0	0	1	0	4	(TC_4) EOT	DC_4 (Stop)	CS_1 ⑤	4	D	T	d	t
0	0	1	0	1	1	0	5	(TC_5) Enq	(TC_8) Nack	%	5	E	U	e	u
0	0	1	1	0	1	1	6	(TC_6) Ack	(TC_9) Sync	&	6	F	V	f	v
0	0	1	1	1	1	1	7	Bell	(TC_{10}) ETB	' ⑧	7	G	W	g	w
0	1	0	0	0	1	0	8	FE_0 (BS)	Cancel	(	8	H	X	h	x
0	1	0	0	1	1	0	9	FE_1 (HT)	EM	)	9	I	Y	i	y
0	1	0	1	0	1	1	10	FE_2 (LF) ①	SS	*	: ②	J	Z	j	z
0	1	0	1	1	1	1	11	FE_3 (VT)	Escape	+	; ②	K	([) ③	k	③
0	1	1	0	0	1	1	12	FE_4 (FF)	IS_4 (FS)	,	<	L	(CS_2) ③	l	③
0	1	1	0	1	1	1	13	FE_5 (CR) ①	IS_3 (GS)	–	=	M	(]) ③	m	③
0	1	1	1	0	1	1	14	SO	IS_2 (RS)	.	>	N	^ ⑦⑧	n	③
0	1	1	1	1	1	1	15	SI	IS_1 (US)	/	?	O	`	o	Delete

Explanatory notes about the 7-bit code table

1. The controls CR and LF are intended for printer equipment which requires separate combinations to return the carriage and to feed a line.
   For equipment which uses a single combination (called New Line) for combined carriage return and line feed operation, NL will be coded at FE_2. The use of New Line requires agreement between the sender and the recipient of the data.
2. If 10 and 11 as single characters are needed (for example, for Sterling currency subdivision), they should take the place of "colon" (:) and "semi-colon" (;) respectively.
3. "Reserved for National Use." These positions are primarily intended for alphabetic extensions. If they are not required for that purpose, they may be used for symbols and the recommended choice as shown in parentheses.
5. "Reserved for National Use." A currency sign will be assigned to this position.
6. The number sign (#) in position 2/3 may have an alternate graphical representation (N°).
7. It is acceptable to represent tilde (~) by circumflex (^) for international interchange of information in Spanish and Portuguese. In Spanish and Portugese speaking countries the tilde may replace the circumflex in position 5/14.
8. The graphics in positions 2/2, 2/7, 5/14 have the significance of quotation mark, apostrophe and upwards arrow, respectively; however, these characters take on the significance of diaeresis, acute accent and circumflex diacritical signs when they follow the Backspace code.

Table 2: 6-bit code table

b6	b5	b4	b3	b2	b1		0 0	0 1	1 0	1 1
							0	1	2	3
0	0	0	0	0	0	0	F_0 Space	0	Null	P
0	0	0	0	1	1	1	F_1 (HT)	1	A	Q
0	0	1	0	2	2	2	F_2 (LF) (1)	2	B	R
0	0	1	1	3	3	3	F_3 (VT)	3	C	S
0	1	0	0	4	4	4	F_4 (FF)	4	D	T
0	1	0	1	5	5	5	F_5 (CR) (1)	5	E	U
0	1	1	0	6	6	6	SO	6	F	V
0	1	1	1	7	7	7	SI	7	G	W
1	0	0	0	8	8	8	(	8	H	X
1	0	0	1	9	9	9	)	9	I	Y
1	0	1	0	10	10	*	: (2)	J	Z	
1	0	1	1	11	11	+	; (2)	K	([) (3)	
1	1	0	0	12	12	,	<	CS_1	L	(CS_2) (3)
1	1	0	1	13	13	–	= (4)	%	M	(]) (3)
1	1	1	0	14	14	.	>	&	N	Escape
1	1	1	1	15	15	/	'		O	Delete

Explanatory notes about the 6-bit code table

1. The controls CR and LF are intended for printer equipment which requires separate combinations to return the carriage and to feed a line.
   For equipment which uses a single combination (called New Line) for a combined carriage return and line feed operation, NL will be coded at F_2. This requires agreement between the sender and the recipient of the data. F_5 will then be regarded as Backspace (BS).
2. If 10 and 11 as single characters are needed (for example, for Sterling currency subdivision), they should take the place of "colon" (:) and "semi-colon" (;) respectively.
3. "Reserved for National Use." These positions are primarily intended for alphabetic extensions. If they are not required for that purpose they may be used for symbols and the recommended choice is shown in parentheses.
4. Either of the two sets of symbols shown in these positions in the table may be chosen; but this requires agreement between the sender and the recipient of the data.

bit-7 is 1. Another important group is the "dense graphic sub-set" comprising the four centre columns, in which the two most-significant bits differ from each other; this contains the graphic symbols regarded as the most important for data-processing work. For example, high-speed line printers might be confined to this character set. Capital and small letters differ only with respect to bit-6; this is one of the many points taken into account to simplify the design of associated keyboards.

Certain positions in the code table are shown as available for national use; these are intended for special letters, accents or other symbols which may be needed within a particular country but not for international purposes. These characters should only be used when there is some prior understanding between the sender and recipient as to what they mean. A particular example of this concerns the currency symbols: that in position 2/4 is regarded as the primary symbol within any one country, e.g. £ in U.K., or $ in the U.S.A.; if a second currency symbol is required it should be coded at 5/12.

The symbol # means the same as No., and it can be very useful. If a particular user requires a symbol not in the standard, it is a convention that the positions at the foot of column 5 (and perhaps 7) should first be used for such variants, but they require agreement between sender and recipient.

It is the control characters which are the most important and perhaps novel. There are certain groups of these, for example the Transmission Controls which are intended primarily for use in telecommunication systems, the Format Effectors which are used to control the layout of printed or other documents, the Device Controls which may be used to start and stop ancillary machines, and the Information Separators. The latter represent an important but difficult concept: they may be used to determine the structure of data, that is to say the way it is split up into large or small sections, or alternatively built up from small units into large. They comprise a hierarchy of increasing power; thus for example a *Space* might be used to separate the words within a name, the *Unit Separator* to separate items within an entry (as in the fields within a punched card), the *Record Separator* to separate each composite entry, and the *Group Separator* to identify the end of a series of entries.

In a number of controls the meaning is shown within parentheses. Then that meaning is regarded as being an example for a particular system or type of equipment; but it is permitted, for other systems, to use an alternative but related meaning. A particular example is 0/9 which means *Horizontal Tabulate* in a typewriter-like machine but *Skip* in a punched-card machine. This equivalence of the controls is specially discussed in the ECMA Philosophy document (1963).

The individual controls will now be described in general terms, in contradistinction to the formal definitions in the official documents.

Null "means nothing" (this is a very tricky thing to define precisely). It will be used as an idling condition to fill out time, or to fill out media as, for example, the leaders and trailers of a punched tape.

The next four characters are used to control and demarcate messages sent over transmission networks. *Start of Heading* indicates the beginning of an address or other preamble which may be used to establish the routing through the transmission network; *Start of Text* (STX) indicates the beginning of a message in the ordinary sense, and this is finished by *End of Text*; there may, of course, be any number of messages within one transmission, and the whole sequence is terminated by *End of Transmission.*

Enquiry is a general request for a remote station to indicate its identity or condition. The response may be an identification signal or it may be a positive *Acknowledge* or a *Negative Acknowledge* (NACK). These may be used not only to establish identity, but also to signal whether the remote station is in working condition, and they may also be used in connection with error-control systems.

Bell is to attract the operator.

Within the Format Effectors, *Backspace* means move back the *printing* position one space; if it is being punched on tape it is punched just as any other character, and it is therefore *not* to be used to move the tape backwards.

Horizontal Tabulate corresponds to the normal tab action of a typewriter, moving to the next pre-set stopping position; and *Vertical Tabulate* gives the same facility with regard to moving the paper. *Form Feed* is used with pre-printed stationery and means move the paper to the first printing position on the next form. *Line Feed* moves the paper a unit distance and *Carriage Return* moves back to the left margin; if, however, a single character or key operation is needed for this combined operation (as on electric typewriters) *New Line* is coded at 0/10.

Five shift characters are provided. *Shift In* means that the following character(s) are as in the standard. The others are used to get out of the standard code in various ways and for various purposes, and in every case the meaning of the subsequent characters would require agreement between sender and recipient. *Shift Out* would be used to go into a different complete code, perhaps a new set of graphics. *Escape* and *Data Link Escape* (the letter confined for use within transmission systems) are in general intended to allow a *single* following character to be given a special meaning. *Start of Special Sequence* (SS) is used to indicate a sequence of characters which have special significance or require special handling, especially for transmission control; this sequence would be finished by *Shift In.*

The *Device Controls* are used to start and stop ancillary machines such as readers and punches.

Sync, which has a special bit-pattern, is used to synchronise transmission systems.

End of Transmission Block may be used with certain methods of error control in transmission systems to signify the end of a block (chosen to suit the design of the transmission system) of data, and that an error control character follows.

Cancel, on the other hand, is a character which can be included in data and means that the preceding block of data contains an error and may have to be ignored. There must be agreement between sender and recipient as to how far back to go.

End of Medium may be used to identify the end of the used portion of a medium such as punched or magnetic tape.

Delete is used particularly with punched tape to overpunch errors, being the all-holes character.

The 6-bit code

The 6-bit code of Table 2 represents a selection from the 7-bit table of the characters considered most important. The general concept for the relation between the two codes may be illustrated in terms of the conversion from the 7 to the 6-bit code, as follows. The characters in the lower half of column 0 are moved to the upper half (by inverting bit-4); then the columns are collapsed by ignoring bit-6 and moving bit-7 into the 6th position. An important consequence of this technique is that the small letters "map over" the capital letters. There are, however, some exceptions to these rules of conversion, the most important being for *Backspace* and *Escape*; and for some of the alternate symbols shown at the foot of column 1 of the 6-bit code table it may be necessary also to invert bit-4. In practice, therefore, it may be best to carry out such conversion by computer program or table look-up rather than by conversion hardware.

Use of Backspace

Certain proposals for the use of the Backspace character warrant special mention because of their influence on systems design and the writing of input programs.

It is visualized that Backspace would be used to prepare accented letters, composite symbols, and for underlining as in ALGOL. In all these cases the convention would be that the letter or ordinary symbol would be keyed (and therefore punched or transmitted) first, then the Backspace character, and then the diacritical sign. By diacritical sign is meant an accent, the underline symbol (4/0) or, for example, the solidus which might be used to make not-greater-than. This triplet of characters (interspersed of course with Delete) would need to be treated as an entity. Although some criticism has been made of the complication resulting from this use of Backspace, it is small compared with the very versatile use of Backspace which is common to all Orion input routines.

A subtle point about the ISO code is that some of the symbols will be considered as accents when used in this way with Backspace, but will have another meaning when used on their own: for example, diaeresis or quotation mark, acute accent or apostrophe, and circumflex accent or upwards arrow.

Implementation of the code on various media

The basic codes being regarded in an abstract sense, work is in progress to establish standards for their implementation on the following media:

(*a*) punched paper tape
(*b*) punched cards
(*c*) edge-punched cards
(*d*) magnetic tape for data interchange
(*e*) telecommunication systems
(*f*) keyboards
(*g*) printers

On *punched paper tape*, the tracks are identified by being numbered 1, 2, 3, feed hole, 4, 5, 6, 7 and 8 (if used). Each character in the code is punched in one row across the tape, the bit numbers corresponding to the track numbers. In addition the next higher numbered track is used for a parity bit; for the 7-bit code, the parity is made even overall, and for the 6-bit code odd overall. Prime emphasis is being placed on the use of 1 inch wide tape. This is the obvious thing to use for the 7-bit code but is rather more unconventional for a 6-bit code, in which case parity goes in track 7 and the 8th track is unpunched. This is part of a move to discourage the need for future tape machines to exist for both 7/8ths and 1 inch widths.

On *punched cards* consideration is confined to cards with 80 columns and 12 rows (identified from top to bottom as *A*, *B*, 0, 1. . .9) and rectangular holes. It is emphasized that the use of punched cards with this character code in no way mitigates against their continuing use with any of the well-established card codes, such as that in B.S. 3174: "Alpha-numeric punching codes for data-processing cards," and it is to be anticipated that the two will co-exist for a substantial time.

Considerable support has been given for a very subtle arrangement whereby the ordinary single hole punching in rows 0 to 9 for the numerals is retained. With this arrangement card-rows *A*, *B* and 0 are used to identify the columns of the code table as follows:

ISO column #	0	1	2	3	4	5	6	7
Hole punched in card-row #	*A*	*A*		N o n e	*A*		*A*	
	B	*B*				*B*		*B*
	0		0				0	0

and card-rows 1 to 9 are used to represent the rows of the code table as follows:

ISO row #	Hole punched in card-row #
0	None
1	1
.	.
.	.
.	.
9	9
10	8 + 2
.	.
.	.
.	.
15	8 + 7

In addition, an exception to these rules is made by transposing the punching for *Space* and *zero*. This gives a hole in card-row 0 for zero, and also the very desirable feature of an unpunched column for Space. (There may ultimately be other transpositions for special purposes.)

For *edge-punched cards*, although no specific proposals have yet been made, it is anticipated that the punching would be equivalent to a length of punched tape along one long-side of the card, with row 1 and bit 1 nearest the edge, and a feed-hole track between tracks 3 and 4.

Implementation of the code on *magnetic tape* presents much more severe problems than the other media, and although substantial progress has been made the work is not yet complete. The following main aspects have to be covered:

(*a*) Dimensions and quality of the unrecorded tape.
(*b*) The dimensions and other properties of the spool.
(*c*) The layout and magnetic parameters of the recorded tracks.
(*d*) The format of the recorded information with respect to the way the character is placed within the tracks and the way the record is broken up into blocks, and the means for checking.
(*e*) Any labels or heading and terminating records which may be needed to allow a computer to recognize and correctly interpret the recorded data.

Work on the other media is not yet sufficiently advanced to warrant a useful summary.

References

"American Standard Code for Information Interchange (1963)." *A.C.M. Communications*, Vol. 6, p. 422.

BEMER, R. W., SMITH, H. J., and WILLIAMS, F. A. (1961). "Design of an Improved Transmission/Data Processing Code." *A.C.M. Communications*, Vol. 4, p. 212.

"Bit Sequencing of ASCII in Serial-by-Bit Data Transmission" (1964). *A.C.M. Communications*, Vol. 7, p. 333.

European Computer Manufacturers Association (1963). *ECMA Standard for a 6-Bit Input/Output Character Code.*

European Computer Manufacturers Association (1963). *EMCA Philosophy on Codes.*

International Organization for Standardization (1964). "Third Draft Proposal on ISO 6 Bit and 7 Bit Codes." Working document ISO/TC 97/SC 2/125.

"Perforated Tape Code for Information Interchange" (1964). *A.C.M. Communications*, Vol. 7, p. 336.

Ross, H. McG. (1961). "Considerations in Choosing a Character Code for Computers and Punched Tapes," *The Computer Journal*, Vol. 3, p. 202.