obsolete
1. set changed

Editor's Note

Publication of the following proposed American Standard, developed by a Subcommittee of ASA Sectional Committee X3, has been authorized by the American Standards Association for the purpose of obtaining comment, criticism and general public reaction, with the understanding that such proposed American Standard has not been finally accepted by X3, [illegible] is subject to change, modification or withdrawal in whole or in part. Comments should be addressed to the Secretary X3, Business Equipment Manufacturers Association, 235 East 42 Street, New York, New York 10017.—E.L.

Proposed Revised American Standard Code for Information Interchange

1. Scope

This coded character set is to be used for the general interchange of information among information processing systems, communication systems and associated equipment.

2. Standard Code

b_4	b_3	b_2	b_1	Row ↓ / Column → 0	1	2	3	4	5	6	7
				b_7 b_6 b_5 → 0 0 0	0 0 1	0 1 0	0 1 1	1 0 0	1 0 1	1 1 0	1 1 1
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	SS	*	:	J	Z	j	z
1	0	1	1	11 VT	ESC	+	;	K	[	k	{
1	1	0	0	12 FF	FS	,	<	L	~	l	¬
1	1	0	1	13 CR	GS	-	=	M	]	m	}
1	1	1	0	14 SO	RS	.	>	N	^	n	|
1	1	1	1	15 SI	US	/	?	O	_	o	DEL

3. Character Representation

The standard 7-bit character representation, with b_7 the high-order bit and b_1 the low-order bit, is shown below.

Example. The bit representation for the character "K", positioned in column 4, row 11, is:

b_7 b_6 b_5 b_4 b_3 b_2 b_1
1 0 0 1 0 1 1

The code table position for the character "K" may also be represented by the notation "column 4, row 11" or alternately as "4/11." The decimal equivalent of the binary number formed by bits b_7, b_6 and b_5, collectively, forms the column number, and decimal equivalent of the binary number formed by bits b_4, b_3, b_2 and b_1, collectively, forms the row number.

4. Legend

4.1 *Control Characters*

NUL	Null	DC3	Device Control 3
SOH	Start of Heading (CC)	DC4	Device Control (stop)
STX	Start of Text (CC)	NAK	Negative Acknowledge (CC)
ETX	End of Text (CC)	SYN	Synchronous Idle (CC)
EOT	End of Transmission (CC)	ETB	End of Transmission Block (CC)
ENQ	Enquiry (CC)	CAN	Cancel
ACK	Acknowledge (CC)	EM	End of Medium
BEL	Bell (audible or attention signal)	SS	Start of Special Sequence
BS	Backspace (FE)	ESC	Escape
HT	Horizontal Tabulation (punched card skip) (FE)	FS	File Separator (IS)
LF	Line Feed (FE)	GS	Group Separator (IS)
VT	Vertical Tabulation (FE)	RS	Record Separator (IS)
FF	Form Feed (FE)	US	Unit Separator (IS)
CR	Carriage Return (FE)	DEL	Delete*
SO	Shift Out		
SI	Shift In		
DLE	Data Link Escape (CC)		
DC1	Device Control 1		
DC2	Device Control 2		

(CC) Communication Control. (FE) Format Effector. (IS) Information Separator.
* In the strict sense, DEL is not a control character. See 5.2.

4.2 *Graphic Characters*

Column/Row	Symbol	Name	Column/Row	Symbol	Name
2/0	SP	Space (normally nonprinting)	2/15	/	Slant
			3/10	:	Colon
2/1	!	Exclamation Point	3/11	;	Semicolon
2/2	"	Quotation Marks (diaeresis*)	3/12	<	Less Than
			3/13	=	Equals
2/3	#	Number Sign	3/14	>	Greater Than
2/4	$	Dollar Sign†	3/15	?	Question Mark
2/5	%	Percent	4/0	`	Grave Accent* (opening single quotation mark)†
2/6	&	Ampersand			
2/7	'	Apostrophe (closing single quotation mark; acute accent*)	5/11	[	Opening bracket†
			5/12	~	Tilde*†
			5/13	]	Closing bracket†
2/8	(	Opening Parenthesis	5/14	^	Circumflex*†
2/9	)	Closing Parenthesis	5/15	_	Underline
2/10	*	Asterisk	6/0	@	Commercial at
2/11	+	Plus	7/11	{	Opening brace
2/12	,	Comma (cedilla*)	7/12	¬	Overline
2/13	-	Hyphen (minus)	7/13	}	Closing brace
2/14	.	Period (decimal point)	7/14	|	Vertical line

* The use of the symbols in 2/2, 2/7, 2/12, 4/0, 5/12 and 5/14 as diacritical marks is described in Appendix A, paragraphs A5.2 and A5.3. When so used they should be preceded by an alphabetic character and a BS (Backspace) in that sequence.

† These characters should not be used in international interchange without determining that there is agreement between sender and recipient. (See Appendix B4).

5. Definitions

5.1 *General*

(CC) Communication Control. A functional character intended to control or facilitate transmission of information over communication networks.

(FE) Format Effector. A functional character which controls the layout or positioning of information in printing or display devices.

(IS) Information Separator. A character which is used to separate and qualify information in a logical sense. There is a group of four such characters, which are to be used in a hierarchical order.

5.2 *Control Characters*

NUL (Null). The all zeros character which may serve to accomplish time fill and media fill.

SOH (Start of Heading). A communication control character used at the beginning of a sequence of characters which constitute a machine-sensible address or routing information. Such a sequence is referred to as the *heading*. An STX character has the effect of terminating a heading.

STX (Start of Text). A communication control character which precedes a sequence of characters that is to be treated as an entity and entirely transmitted through to the ultimate destination. Such a sequence is referred to as *text*. STX may be used to terminate a sequence of characters started by SOH.

ETX (End of Text). A communication control character used to terminate a sequence of characters started with STX and transmitted as an entity.

EOT (End of Transmission). A communication control character used to indicate the conclusion of a transmission, which may have contained one or more texts and any associated headings.

ENQ (Enquiry). A communication control character used in data communication systems as a request for a response from a remote station. It may be used as a "Who Are You" (WRU) to obtain identification, or may be used to obtain station status, or both.

ACK (Acknowledge). A communication control character transmitted by a receiver as an affirmative response to a sender.

BEL (Bell). A character for use when there is a need to call for human attention. It may control alarm or attention devices.

BS (Backspace). A format effector which controls the movement of the printing position one printing space backward on the same printing line (applicable also to display devices).

HT (Horizontal Tabulation). A format effector which controls the movement of the printing position to the next in a series of predetermined positions along the printing line (applicable also to display devices and the skip function on punched cards).

LF (Line Feed). A format effector which controls the movement of the printing position to the next printing line (applicable also to display devices).

VT (Vertical Tabulation). A format effector which controls the movement of the printing position to the next in a series of predetermined printing lines (applicable also to display devices).

FF (Form Feed). A format effector which controls the movement of the printing position to the first predetermined printing line on the next form or page (applicable also to display devices).

CR (Carriage Return). A format effector which controls the movement of the printing position to the first printing position on the same printing line (applicable also to display devices).

SO (Shift Out). A control character indicating that the code combinations which follow shall be interpreted as outside of the character set of the standard code table until a Shift In character is reached.

SI (Shift In). A control character indicating that the code combinations which follow shall be interpreted according to the standard code table.

DLE (Data Link Escape). A communication control character which will change the meaning of a limited number of contiguously following characters. It is used exclusively to provide supplementary controls in data communication networks.

DC1, DC2, DC3, DC4 (Device Controls). Characters for the control of ancillary devices associated with data processing or telecommunication systems, more especially switching devices "on" or "off". (If a single "stop" control is required to interrupt or turn off ancillary devices, DC4 is the preferred assignment.)

NAK (Negative Acknowledge). A communication control character transmitted by a receiver as a negative response to the sender.

SYN (Synchronous Idle). A communication control character used by a sychronous transmission system in the absence of any other character to provide a signal from which synchronism may be achieved or retained.

ETB (End of Transmission Block). A communication control character used to indicate the end of a block of data for communication purposes. ETB is used for blocking data where the block structure is not necessarily related to the processing format.

CAN (Cancel). A control character used to indicate that the data with which it is sent is in error or is to be disregarded.

EM (End of Medium). A control character associated with the sent data which may be used to identify the physical end of the medium, or the end of the used, or wanted, portion of information recorded on a medium. (The position of this character does not necessarily correspond to the physical end of the medium.)

SS (Start of Special Sequence). A control character used to indicate the start of a variable length sequence of characters which have special significance or which are to receive special handling.

ESC (Escape). A control character intended to provide code extension (supplementary characters) in general information interchange. The Escape character itself is a prefix affecting the interpretation of a limited number of contiguously following characters.

FS (File Separator), GS (Group Separator), RS (Record Separator) and US (Unit Separator). These information separators may be used within data in optional fashion, except that their hierarchical relationship shall be: FS is the most inclusive, then GS, then RS, and US is least inclusive. (The content and length of a File, Group, Record or Unit are not specified.)

DEL (Delete). This character is used primarily to "erase" or "obliterate" erroneous or unwanted characters in perforated tape. (In the strict sense, DEL is not a control character.)

5.3 *Graphic Characters*

SP (Space). A normally nonprinting graphic character used to separate words. It is also a format effector which controls the movement of the printing position, one printing position forward (applicable also to display devices).

6. General Considerations

6.1 This standard does not define the means by which the coded set is to be recorded in any physical medium, nor does it include any redundancy or define techniques for error control. Further, this standard does not define data communication character structure, data communication formats, code extension techniques, or graphic representation of control characters.

6.2 Deviations from the standard may create serious difficulties in information interchange and should be used only with full cognizance of the parties involved.

6.3 The relative sequence of any two characters, when used as a basis for collation, is defined by their binary values.

6.4 No specific meaning is prescribed for any of the graphics in the code table except that which is understood by the users. Furthermore, this standard does not specify a type style for the printing or display of the various graphic characters.

6.5 The appendices (A-D) to this standard contain additional information on the design and use of this code.

Appendices

(These Appendices are not a part of American Standard Code for Information Interchange, but are included to facilitate its use.)

Appendix A. Design Considerations for the Coded Character Set

A1. Introduction

The standard coded character set is intended for the interchange of information among information processing systems, communication systems and associated equipment.

A2. Considerations Affecting the Code

There were many considerations that determined the set size, set structure, character selection and character placement of the code. Among these were (not listed in order of priority):

(1) Need for adequate number of graphic symbols
(2) Need for adequate number of device controls, format effectors, communication controls and information separators
(3) Desire for a nonambiguous code, i.e. one in which every code combination has a unique interpretation
(4) Physical requirements of media and facilities
(5) Error control considerations
(6) Special interpretation of the all-zeros and all-ones characters
(7) Ease in the identification of classes of characters
(8) Data manipulation requirements
(9) Collating conventions
  (a) Logical
  (b) Historical
(10) Keyboard conventions
  (a) Logical
  (b) Historical
(11) Other set sizes
(12) International considerations
(13) Programming languages
(14) Existing coded character sets

A3. Set Size

A 7-bit set is the minimum size that will meet the requirements for graphics and controls in applications involving general information interchange.

A4. Set Structure

A4.1 In discussing the set structure it is convenient to divide the set into 8 columns and 16 rows as indicated in the standard.

A4.2 It was considered essential to have a dense subset which contained only graphics. For ease of identification this graphic subset was placed in six contiguous columns.

A4.3 The first two columns were chosen for the controls for three reasons:

(1) The character NUL by its definition has the location 0/0 in the code table. NUL is broadly considered a control character.

(2) The representations in column 7 were felt to be most susceptible to simulation by a particular class of transmission error—one which occurs during an idle condition on asynchronous systems.

(3) To permit the considerations of graphic subset structure described in A6 to be satisfied, the two columns of controls had to be adjacent.

A4.4 The character set was structured to enable the easy identification of classes of graphics and controls.

A5. Choice of Graphics

A5.1 Included in the set are the numerals 0 through 9, upper and lower cases of the alphabetic letters A through Z, and those punctuation, mathematical and business symbols considered most useful. The set includes a number of characters commonly encountered in programming languages. In particular, all the COBOL and FORTRAN graphics are included.

A5.2 In order to permit the representation of languages other than English, two diacritical (or accent) marks have been included, and provision has been made for the use of four punctuation symbols alternate as diacritical marks. The pairing of these punctuation symbols with their corresponding diacritical marks was done so as to facilitate the design of a typeface which would be acceptable for both uses.

These arrangements are:

Column/ Row	Code Table Symbol	Use: Punctuation	Use: Diacritical
2/2	"	Quotation Marks	Dialresis
2/7	'	Apostrophe	Acute Accent
2/12	,	Comma	Cedilla
4/0	`	Opening Single, Quotation Mark	Grave Accent
5/12	~	(none)	Tilde
5/14	^	(none)	Circumflex

A5.3 Diacritical marks should be applied in this sequence: alphabetic character, BS (Backspace), diacritical mark.
A5.4 The character *overline* is included because of its usefulness for such purposes as indicating negation or complementation. The "hook" on its code table representation is shown to suggest a means of avoiding confusion with *underline*.

A6. Graphic Subset Structure

A6.1 The basic structure of the dense graphic subset was influenced by logical collating considerations, the requirements of simple related 6-bit sets, and the needs of typewriter-like devices. For information processing, it is desirable that the characters be arranged in such a way as to minimize both the operating time and the hardware components required for ordering and sequencing operations. This requires that the relative order of characters, within classes, be such that a simple comparison of the binary codes will result in information being ordered in a desired sequence.
A6.2 Conventional usage requires that SP (*space*) be ahead of any other symbol in a collatable set. This permits a name such as "JOHNS" to collate ahead of a name such as "JOHNSON." The requirement that punctuation symbols such as *comma* also collate ahead of the alphabet ("JOHNS, A" should also collate before "JOHNSON") establish the special symbol locations, including SP, in the first column of the graphic subset.
A6.3 To simplify the design of typewriter-like devices, it is desirable that there be only a common 1-bit difference between characters to be paired on keytops. This, together with the requirements for a contiguous alphabet, and the collating requirements outlined above, resulted in the placement of the alphabet in the last four columns of the graphic subset and the placement of the numerals in the second column of the graphic subset.
A6.4 It is expected that devices having the capability of printing only 64 graphic symbols will continue to be important. It may be desirable to arrange these devices to print one symbol for the bit pattern of both upper and lower case of a given alphabetic letter. To facilitate this, there should be a single bit difference between the upper and lower case representations of any given letter. Combined with the requirement that a given case of the alphabet be contiguous, this dictated the assignment of the alphabet, as shown, in columns 4 through 7.
A6.5 To minimize ambiguity caused by the use of a 64-graphic device as described above, it is desirable, to the degree possible, that each character in column 6 or 7 differ little in significance from the corresponding character in column 4 or 5. In certain cases, this was not possible.
A6.6 The assignment of *commercial at* (@) was dictated by the fact that, in certain countries, it will be replaced by a national symbol, most typically an accented lower case letter; therefore, it was located in column 6.
A6.7 The resultant structure of "specials" (S), "digits" (D), and "alphabetics" (A) does not conform to the most prevalent collating convention (S-A-D) because of other more demanding code requirements.
A6.8 The need for a simple transformation from the set sequence to the prevalent collating convention was recognized, and dictated the placement of some of the "specials" within the set. Specifically, those special symbols, viz., *ampersand* (&), *asterisk* (*), *comma* (,), *hyphen* (-), *period*(.) and *slant* (/), which are most often used as identifiers for ordering information and which normally collate ahead of both the alphabet and the numerals, were not placed in the column containing the numbers, so that the entire numeric column could be rotated via a relatively simple transformation to a position higher than the alphabet. The sequence of the aforementioned "specials" was also established to the extent practical to conform to the prevalent collating convention.
A6.9 The need for a useful 4-bit numeric subset also played a role in the placement of characters. Such a 4-bit subset, including the digits and the symbols *asterisk*, *plus* (+), *comma*, *hyphen*, *period* and *slant*, can easily be derived from the code.
A6.10 Considerations of other domestic code sets, including the Department of Defense former standard 8-bit data transmission code ["Fieldata," 1961] as well as international requirements, played an important role in deliberations that resulted in the code. The selection and grouping of the symbols *dollar sign* ($), *percent* (%), *ampersand* (&), *apostrophe* ('), *less than* (<), *equals* (=) and *greater than* (>) facilitate contraction to either a business or scientific 6-bit subset. The position of these symbols and of the symbols *comma*, *hyphen*, *period* and *slant* facilitates achievement of commonly accepted pairing on a keyboard. The historic pairing of *question mark* and *slant* is preserved and the *less than* and *greater than* symbols, which have comparatively low usage, are paired with *period* and *comma* so that in dual-case keyboard devices where it is desired to have *period* and *comma* in both cases, the *less than* and *greater than* symbols are the ones displaced. Provision was made for the accommodation of alphabets containing more than 26 letters and for 6-bit contraction by the location of low-usage characters in the area following the alphabet. In addition, the requirement for the digits 10 and 11 used in Sterling monetary areas was considered in the placement of the *asterisk*, *plus*, *semicolon*, and *colon*, so that the 10 and 11 could be substituted for the *semicolon* and *colon*.

A7. Control Subset Content and Structure

A7.1 The control characters included in the set are those required for the control of terminal devices, input and output devices, format, or communication transmission and switching on a general enough basis to justify inclusion in a standard set.
A7.2 Many control characters may be considered to fall into the following categories: (1) Communication Controls, (2) Format Effectors, (3) Device Controls, and (4) Information Separators. To the extent practical controls of each category were grouped, the structure chosen also facilitates the contraction of the set to a logically related 6-bit set.
A7.3 The information separators (FS, GS, RS, US) identify boundaries of various elements of information, but differ from punctuation in that they are primarily intended to be machine sensible. They were arranged in accordance with an expected hierarchical use, and the lower end of the hierarchy is contiguous in binary order with SP (space) which is sometimes used as a machine-sensible separator. Subject to this hierarchy the exact nature of their use within data is not specified.
A7.4 The character SYN (Synchronous Idle) was located so that its binary pattern in serial transmission was unambiguous as to character framing, and also to optimize certain other aspects of its communication usage.
A7.5 ACK (Acknowledge) and NAK (Negative Acknowledge) were located so as to gain the maximum practical protection against mutation of one into the other by transmission errors.

Appendix B. Notes on Application

B1. Introduction

B1.1 The standard code was developed to provide for information interchange among information processing systems, communications systems and associated equipment. In a system consisting of equipment with several local or native codes, maximum flexibility will be achieved if each of the native codes is translated to the standard whenever information interchange is desired.
B1.2 Within any particular equipment or closed system it may be necessary to substitute characters. For example, some systems may require special graphic symbols and some devices may require special control codes. Design efforts on the standard code included consideration of these types of adaptations.

B2. Character Substitutions

B2.1 Any character substitution will result in a coded character set which does not conform to this standard.
B2.2 It is recommended that when a character is substituted in the code table for a standard character, the standard character should not be reassigned elsewhere in the table. Such a substitute character, once assigned, should not be subsequently reassigned elsewhere.
B2.3 It is recommended that graphic substitutions be made only in the graphic area and control substitutions only in the control area. Any substitution involving a control should be made only with full cognizance of all possible operational effects.
B2.4 It should be noted that this standard specifies, for each position of the code table, the information represented by the character and not necessarily the precise action taken by the recipient when the character is received. In the case of graphics, considerable variation in the actual shape printed or displayed may be appropriate to different units, systems, or fields of application. In the case of controls, the action performed is dependent upon the use for which the particular system is intended, the application to which it is being put, and a number of conventions established by the user or designer—some system-wide and some unique to a particular unit.
B2.5 Typical examples of diversity in execution not necessarily contrary to this standard are:

(1) A number of graphic symbols, other than that used in the code table, are used for *ampersand* in various type styles; still other symbols may be more appropriate to electronic display devices. The use of such alternate symbols does not in itself constitute deviation from the standard as long as *ampersand* is the concept associated with the character. Note that this does not necessarily restrict the use of such an alternate symbol design to mean "and"; in any type style *ampersand* may, of course, be used with arbitrary meaning.

(2) A card punch in one application may "skip" when the character HT (Horizontal Tabulation: used as skip) is presented to it; in another application the character HT may be recorded in the card without further action.

B3. Related Larger and Smaller Sets

Consideration has been given to the relationship between the standard set and sets of other sizes. A number of straightforward logical transformations are possible which result in a variety of sets related to the standard. None of the transformed sets are recognized by this standard.

B4. International Considerations

This standard conforms to the anticipated recommendations of the International Organization for Standardization (ISO) and the International Telephone and Telegraph Consultative Committee (CCITT) for a 7-bit code. The characters of this code are identical to those specified for international standardization by these bodies. Their recommendations are expected to permit national standardization by the various countries in eleven code table positions. These "national usage" positions and their assignments in this standard are as follows:

Column/Row	*Character (U.S.)*	*Column/Row*	*Character (U.S.)*
2/4	$	6/0	@
4/0	`	7/11	{
5/11	[	7/12	¬
5/12	~	7/13	}
5/13	]	7/14	|
5/14	^		

In international interchange of information these 11 characters should not be used except where it is determined that there is agreement between sender and recipient.

In addition, in other countries, the Number Sign "#" (in position 2/3) may be represented as "No".

B5. Communications Considerations

Certain control characters are designated as *communication controls*. They are:

SOH	(Start of Heading)	ACK	(Acknowledge)
STX	(Start of Text)	DLE	(Data Link Escape)
ETX	(End of Text)	NAK	(Negative Acknowledge)
EOT	(End of Transmission)	SYN	(Synchronous Idle)
ENQ	(Enquiry)	ETB	(End of Transmission Block)

These may be used by communication systems for their internal signaling, or for the exchange of information relating to the control of the communication system between that system and its end terminals. Some such systems may impose restrictions on the use of these communication control characters by the end terminals. For example, the use of some of them may be completely prohibited while others may be restricted to use in conformity with the formats and procedures required by the communication system for its operation.

Appendix C. Original Criteria

C1. Introduction

C1.1 This Appendix contains the original criteria upon which the design of the code was based. Not all criteria have been entirely satisfied. Some are conflicting, and the characteristics of the set represent accepted compromises of these divergent criteria.

C1.2 The criteria were drawn from communication, processing, and media recording aspects of information interchange.

C2. Criteria

C2.1 Every character of the code set shall be represented by the same number of bits (i.e. binary digits).

C2.2 The standard set shall be so structured as to facilitate derivation of logically related larger or smaller sets.

C2.3 In a code of n bits, all possible 2^n patterns of ones and zeros will be permitted and considered valid.

C2.4 The number of bits, n, shall be sufficient to provide for the alphabetic and numeric characters, commonly used punctuation marks, and other special symbols, along with those control characters required for interchange of information.

C2.5 The numerals 0 through 9 shall be included within a 4-bit subset.

C2.6 The numerals 0 through 9 shall be so represented that the four low-order bits shall be the binary-coded-decimal form of the particular numeral that the code represents. In the selection of the two characters immediately succeeding the numeral 9, consideration shall be given to their replacement by the graphics 10 and 11 to facilitate the adoption of the code in the Sterling monetary area.

C2.7 The interspersion of control characters among the graphic characters shall be avoided. The characters devoted to controls shall be easily separable from those devoted to graphics.

C2.8 Within the standard set, each character shall stand by itself and not depend on surrounding characters for interpretation.

C2.9 An entire case of the alphabet (A through Z) shall be included within a 5-bit subset. Consideration shall be given to the need for more than 26 characters in some alphabets.

C2.10 The letters of each case of the alphabet shall be assigned, in conventional order (A through Z), to successive, increasing binary representations.

C2.11 Suitable control characters required for communication and information processing shall be included.

C2.12 Escape functions that provide for departures from the standard set shall be incorporated.

C2.13 A simple binary comparison shall be sufficient to determine the order within each class of characters. (In this regard, the special graphics, the numerals and the alphabet are each defined as distinct classes.) Simple binary rules do not necessarily apply between classes when ordering information.

C2.14 Space (i.e. the space between words) must collate ahead of all other graphics.

C2.15 Special symbols used in the ordering of information must collate ahead of both the alphabet and the numerals.

C2.16 Insofar as possible, the special symbols shall be grouped according to their functions; for example, punctuation and mathematical symbols. Further, the set shall be so organized that the simplest possible test shall be adequate to distinguish and identify the basic alphabetic, numeric and special symbol subsets.

C2.17 Special symbols shall be placed in the set so as to simplify their generation by typewriters and similar keyboard devices. This criterion means, in effect, that the codes for pairs of characters that normally appear on the same keytops on a typewriter shall differ only in a common single-bit position.

C2.18 The set shall contain the graphic characters of the principal programming languages.

C2.19 The codes for all control characters shall contain a common, easily recognizable, bit pattern.

C2.20 The Null (000···) and Delete (111···) characters shall be provided.

Appendix D. Terminology

This appendix is intended to clarify the sense in which certain terms are used.

Bit— Contraction of "binary digit".

Bit pattern—The binary representation of a character.

Character— A member of a coded character set; the binary representation of such a member and its graphic symbol or control function.

Code— A system of discrete representation of a set of symbols and functions.

Expository Remarks

1. Introduction

1.1 This document relates to the proposed revision of the American Standard Code for Information Interchange (ASCII-ASA X3.4-1963) and discusses the difference between the proposed revision and the current standard, and explains why the changes were made.

1.2 The activity of Subcommittee X3.2, since the development of ASCII, has functioned within the scope and framework of ASA Document X3.4-1963. In addition, Subcommittee X3.2 has engaged in close liaison and participation in the development of a 7-bit international information interchange code in an attempt to make this code as effective as possible and as compatible with ASCII as is feasible. This international information interchange code has been developed, concurrently with the ASCII revision, by ISO/TC97/SC2 and reviewed by the CCITT, and serious efforts have been made to minimize the differences.

1.3 Specific activity was directed toward the unassigned area as prescribed in paragraphs A1.2 and B3 of document X3.4-1963. The assignment of the lower case of the alphabet to the unassigned area precipitated a review of the previous code assignments and in particular a review and reevaluation of the control functions meriting inclusion in the code table. This resulted in a limited number of rearrangements, additions, and deletions of graphics and controls.

2. ASCII Tables-X3.4-1963 and Proposed Revision

2.1 As an aid in following the development of the revision of ASCII, Figure 1 shows the proposed revised code table and the ASCII table (ASA X3.4-1963), in which changes are denoted by shaded areas.

3. ISO 7-Bit Code Tables (First, Second, Third and Fourth Draft Proposals)

3.1 The parallel international developments are shown in Figures 2 and 3, where Figure 2 shows the changes between the first and second draft ISO 7-Bit Code Tables and Figure 3 illustrates the changes between the second and third draft ISO 7-Bit Code Tables.

The following notes pertain to the ISO Code Tables (Figures 2 and 3). [I, II, III are used here to indicate First, Second and Third Drafts, respectively, for section numbers shown.—Ed. Note]

Section Number

II1 The controls CR and LF are intended for printer equipment which requires separate combinations to return the carriage and feed a line.

As an alternative, for equipment which uses a single combination for a combined carriage-return and line-feed operation (called New-Line), NL will be coded at FE 2. Then FE 5 will be regarded as Backspace BS.

If the latter type of equipment has to interwork with the former, it may be necessary to take steps to introduce the CR character.

III1 The controls CR and LF are intended for printer equipment which requires separate combinations to return the carriage and feed a line.

As an alternative, for equipment which uses a single combination (called New Line), for a combined carriage-return and line-feed operation. NL will be coded at FE 2.

If the latter type of equipment has to interwork with the former, it may be necessary to take steps to introduce the CR character.

III1 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 control for combined carriage return and line feed operation, the function FE 2 will have the meaning New Line (NL). This requires agreement between the sender and the recipient of the data.

I2, II2 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.

III2 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. These substitutions require agreement between the sender and the recipient of the data.

III3 "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 a recommended choice is shown in parenthesis in some cases.

I4 In those countries in which the alphabet includes more than 26 letters, but no more than 29, the symbols located in the position indicated by 4 may be replaced by the additional characters of the extended alphabet.

II4 "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 parenthesis.

III4 The graphical representation and meaning of the two currency signs CS_1 (position 2/4) and CS_2 (position 5/12) may vary according to national use.

I5, II5 If a second currency sign is required, it should take the place of "reverse slash" (\).

I6 Reserved for future standardization work i.e, not available for individual choice.

II6 "Reserved for National use". A currency sign will be assigned to this position.

III6 The number sign (#) in position 2/3 may have an alternate graphical representation (No).

I7 Reserved for future standardization for use as an acknowledge signal.

II7 If an "Acknowledge" (ACK) signal is required it should be coded in this position and the "Underline" sign becomes its graphical representation.

III7 It is acceptable to represent tilde (~) by circumflex accent (^) for international interchange of information in Spanish and Portuguese languages. In Spanish and Portuguese speaking countries, the tilde may replace the circumflex accent in positions 5/14.

III8 The graphics in positions 2/2, 2/7, 5/14 have respectively the significance of quotation mark, apostrophe and upwards arrow; however these characters take on the significance of diaeresis, acute accent and circumflex accent diacritical signs when they follow the Backspace character. The symbol shown in position 5/15 did not receive any denomination other than grave accent which corresponds to its use as a diacritical sign.

3.2 Figures 2 and 3 are for reference only.

3.3 Figure 4 illustrates the anticipated ISO Fourth Draft Proposal and the proposed revision of ASCII.

4. Unassigned Area (Columns 6 and 7)

4.1 *General Considerations*

4.1.1 The published American Standard, X3.4-1963 contains 28 unassigned positions in the last two columns reserved for future standardization. After X3.4-1963 was published, Subcommittee X3.2 adopted the principle that all code positions shall have specific character assignments. In developing specific assignments, the following factors were weighed.

a. The relative need for extra graphics vs. extra control functions in the unassigned area of ASCII.

b. The need for a dual-case alphabet.

c. The implications of placing second-case letters in the unassigned area vs. placing them in the shift-out set with all 7 bits identical for the upper and lower case of each letter.

d. The merits of having upper-case vs. lower-case letters in the unassigned area.

e. The extensions required to accommodate foreign alphabets having more than 26 letters.

f. Methods of extending ASCII to provide many additional graphics and many additional controls.

g. Related subsets and supersets of ASCII for specialized uses, such as industry-wide choices.

h. Related subsets and supersets of ASCII for data processors, keyboards, printing devices and local or remote stations.

4.2 *Control versus Graphic Assignment*

4.2.1 The conclusion was reached by the Subcommittee that the previously unassigned area should receive specific assignments, either the other case of the alphabet or additional controls. The Subcommittee then considered the following arguments.

4.2.2 *Arguments for the Placement of Graphics in the Unassigned Area*

4.2.2.1 In order to provide for a keyboard shift operation between upper and lower case characters, as in office typewriter keyboards, without introducing delays to keyboard operation or requiring character storage, the unassigned area should be used for the second case of the alphabet.

4.2.2.2 Upper and lower case letters will be more widely used than special controls. One of the principal users will be the graphic arts industry which requires a dual-case alphabet.

4.2.2.3 Assuming the rule that substitute graphics in the basic mode (ASCII) and alternate graphics in the shift-out mode are to be restricted to the graphic positions in ASCII, a greater number of substitute and alternate graphics are available if graphics are assigned to the unassigned areas.

4.2.2.4 Discretely assigned dual-case letters are easier to represent directly in fixed-field documents such as punched cards, since shift characters would not be required.

4.2.2.5 For information retrieval, text manipulation, mathematics and other specialized fields more than 64 graphics are generally considered necessary.

4.2.3 *Arguments for the Placement of Controls in the Unassigned Area*

4.2.3.1 Many users of information interchange need additional controls that are not presently assigned positions in ASCII. If additional single character controls are not provided in the standard, the use of non-standard controls will proliferate.

4.2.3.2 The present code does not adequately consider the requirements of advanced communication systems.

4.2.4 *Argument Against the Placement of Controls in the Unassigned Area*

4.2.4.1 A number of the proposed controls may require additional protection against false generation. Such controls should be generated by a sequence of characters and preferably by a sequence of characters now assigned in ASCII rather than a repetition of a single character control occupying a new code position.

4.2.4.2 Since the use of coded characters as controls varies, it is difficult to know which of the many proposed characters need assignment in ASCII and where they should be placed.

4.2.5 *Arguments Against the Placement of Graphics in the Unassigned Area* (Arguments 4.2.5.2 and 4.2.5.3 were presented but have been wholly or partially negated by subsequent developments.)

4.2.5.1 The alphabetic sequence is already contained in ASCII. It is inefficient to have two nearly identical alphabetic sequences in the same basic code set. Shift characters could be used to shift to a second alphabet.

4.2.5.2 There is not really adequate room for the lower case alphabet in the unassigned area if it is desired to have an internationally acceptable code—the three Scandinavian letters following "z" will interfere with the controls in column 7.

4.2.5.3 If graphics were placed in the unassigned area and specifically in column 7, a mixture of controls (Acknowledge, Escape and Delete) and graphics within a column would exist and consequently would

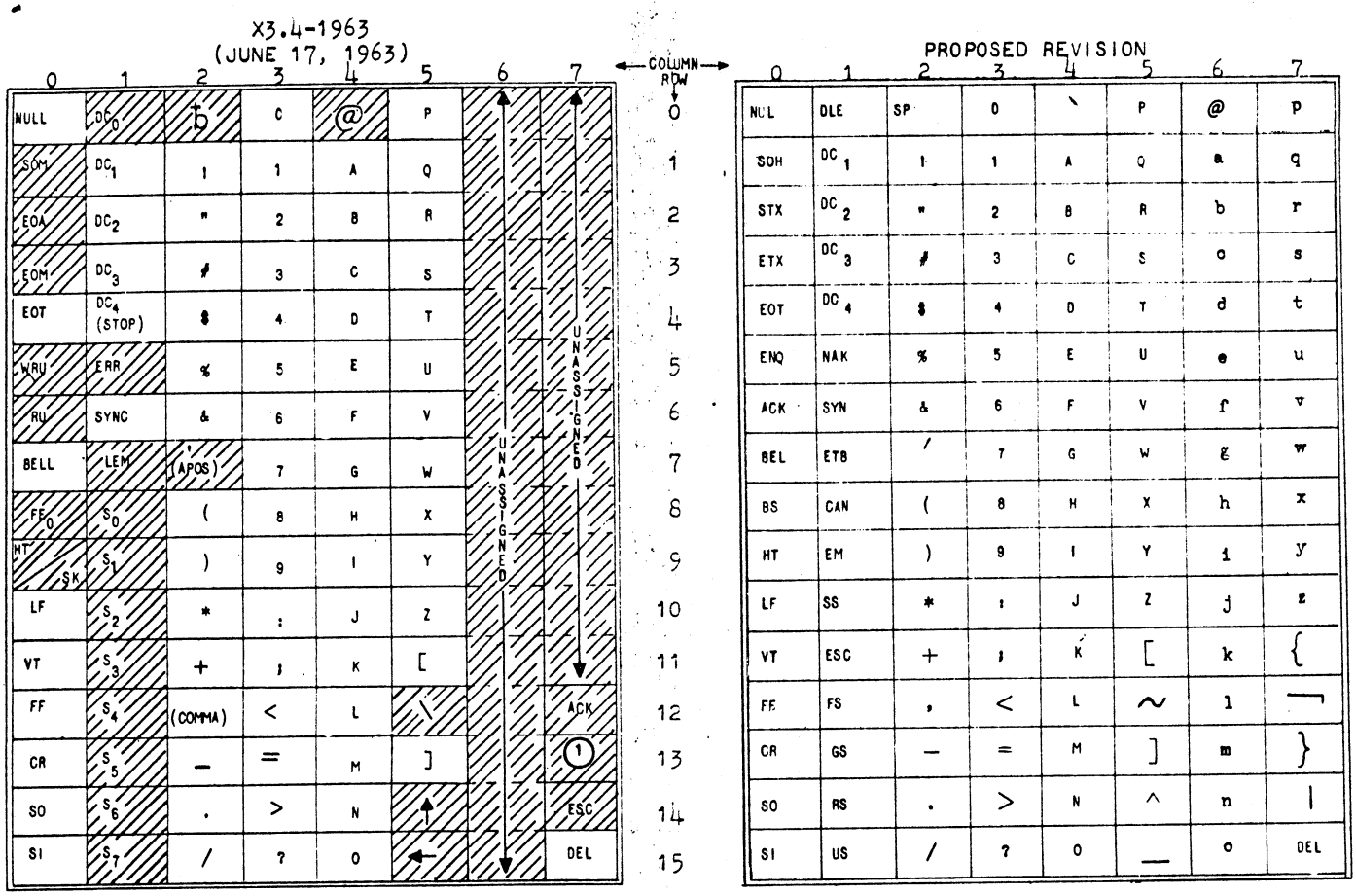

FIG. 1. ASCII table (X3.4-1963) and proposed revised code table

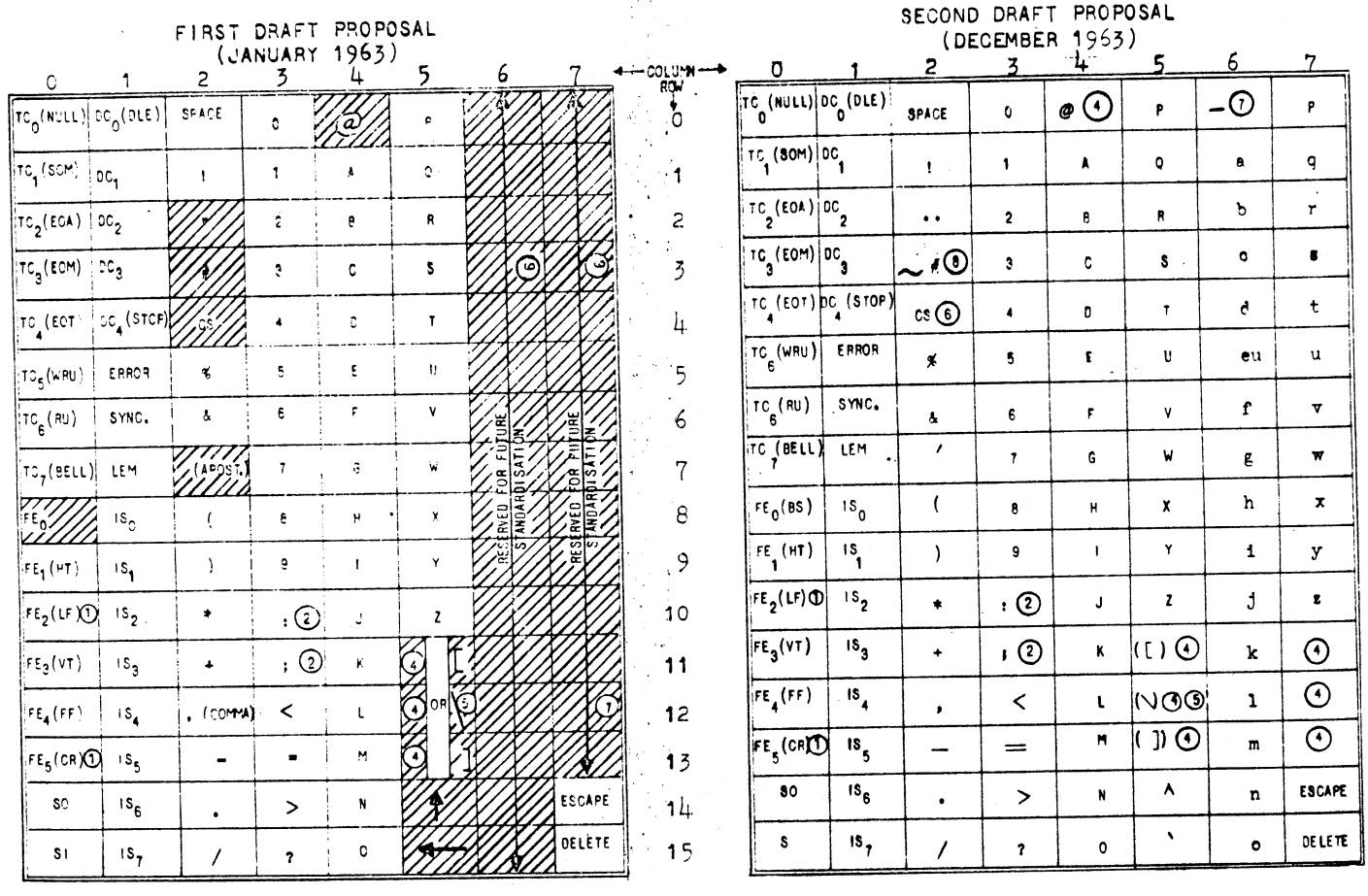

FIG. 2. ISO 7-bit code (first and second draft proposals)

SECOND DRAFT PROPOSAL
(DECEMBER 1963)

0	1	2	3	4	5	6	7	ROW
TC_0 (NULL)	DC_0 (DLE)	SPACE	0	(@) ④	P	— ⑦	p	0
TC_1 (SOM)	DC_1	!	1	A	Q	a	q	1
TC_2 (EOA)	DC_2	"	2	B	R	b	r	2
TC_3 (EOM)	DC_3	~ # ⑧	3	C	S	c	s	3
TC_4 (EOT)	DC_4 (STOP)	CS ⑥	4	D	T	d	t	4
TC_5 (WRU)	ERROR	%	5	E	U	e	u	5
TC_6 (RU)	SYNC	&	6	F	V	f	v	6
TC_7 (BELL)	LEM	'	7	G	W	g	w	7
FE_0 (BS)	IS_0	(	8	H	X	h	x	8
FE_1 (HT)	IS_1	)	9	I	Y	i	y	9
FE_2 (LF) ①	IS_2	*	: ②	J	Z	j	z	10
FE_3 (VT)	IS_3	+	; ②	K	([) ④	k	④	11
FE_4 (FF)	IS_4	,	<	L	(\) ④ ⑤	l	④	12
FE_5 (CR) ①	IS_5	—	=	M	(]) ④	m	④	13
SO	IS_6	.	>	N	^	n	ESCAPE	14
SI	IS_7	/	?	O	`	o	DELETE	15

THIRD DRAFT PROPOSAL
(MAY 1964)

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

NOTE: SHADED AREAS DENOTE CHANGES

FIG. 3. ISO 7-bit code (second and third draft proposals)

ANTICIPATED FOURTH DRAFT PROPOSAL

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

ASCII
PROPOSED REVISION

0	1	2	3	4	5	6	7
NUL	DLE	SP	0	`	P	@	p
SOH	DC_1	!	1	A	Q	a	q
STX	DC_2	"	2	B	R	b	r
ETX	DC_3	#	3	C	S	c	s
EOT	DC_4	$	4	D	T	d	t
ENQ	NAK	%	5	E	U	e	u
ACK	SYN	&	6	F	V	f	v
BEL	ETB	'	7	G	W	g	w
BS	CAN	(	8	H	X	h	x
HT	EM	)	9	I	Y	i	y
LF	SS	*	:	J	Z	j	z
VT	ESC	+	;	K	[	k	{
FF	FS	,	<	L	~	l	¬
CR	GS	—	=	M	]	m	}
SO	RS	.	>	N	^	n	|
SI	US	/	?	O	_	o	DEL

① National Use Positions (other footnotes not shown).

FIG. 4. ISO 7-bit code (fourth draft proposal and proposed revision)

increase the complexity of systems design, since bits b_7 and b_8 could not be used to indicate whether a character was a control or graphic.

4.3 *Assignment of the Lower Case Alphabet to the Unassigned Area*

4.3.1 Consideration of the above arguments led the Subcommittee to the conclusion that graphics, specifically the lower case of the alphabet, should be assigned to the presently unassigned area.

5. Graphic Assignments—Other than Alphabet

5.1 *Diacritical Marks*

5.1.1 The anticipated ISO 7-bit code table includes 3 diacritical marks (*grave accent*, *tilde* and *circumflex*) in positions 4/0, 5/12 and 5/14 and specifies that the graphics in positions 2/2 and 2/7 have the significance of *quotation marks* and *apostrophe* respectively, but take on the significance of the diacritical marks *diaeresis* and *acute accent* when they follow BS (Backspace).

5.1.2 The ISO proposal to locate certain diacritical marks within the "international" area of the 7-bit code set appears to be a useful solution to the problem of multilingual international information interchange.

5.1.3 Subcommittee X3.2 accepted the principle of the use of *quotation marks* as *diaeresis* and *apostrophe* as *acute accent*. It was concluded that for increased utility the *opening single quotation mark* should be included in position 4/0; this mark may then serve as a *grave accent* when it appears above a letter, as may be done for all diacritical marks through the use of BS (Backspace). This action also compensated for the fact that an *apostrophe* which is shaped to allow its use as an *acute accent* can no longer be plausibly used as both opening and closing single quotation mark as has been the previous custom.

5.1.4 Provision has also been made for the use of the *comma* as the diacritical mark *cedilla* in a similar fashion.

5.2 *Keyboard Considerations*

5.2.1 Paragraph C2.17 of document X3.4-1963 gave as a criterion "that the codes for pairs of characters that normally appear on the same keytops on a typewriter shall differ only in a common single-bit position."

5.2.1.1 This was complied with in those cases where it did not conflict with basic principles of code design such as collating sequence, the translation requirements (in particular to and from numeric subsets), requirements for a dense graphic subset, etc.

5.2.2 The critical pairing on the same keytop of the underline with the hyphen, and the quotation marks with the apostrophe is characteristic of electric typewriters. The positions of these characters (and others) in the code table do not satisfy the one-bit criterion. The positioning of these four characters was reconsidered in Subcommittee X3.2 in the development of the present revision and it was reaffirmed that the structure of the code could not be altered to facilitate this pairing without violating basic code principles.

5.3 *Other Considerations*

5.3.1 *National Use Positions*

5.3.1.1 The anticipated ISO 7-bit code table, illustrated in Figure 4, shows eleven national use positions identified by note (3). Seven of the eleven national use positions have preferred graphic assignments as illustrated; these are included in the proposed revision of ASCII.

5.3.1.2 Subcommittee X3.2 made preferred assignments to the remaining four national use positions (7/11 through 7/14): the four graphics opening brace, overline, closing brace and vertical line.

5.3.1.3 The criteria for assignments to positions 7/11 through 7/14 were as follows:

a. Because the last two columns may collapse over the preceding two columns in some application: (1) symbols should have similar form or usage; and (2) symbols normally used for overprinting (e.g. those used with Backspace) should be assigned only to positions which collapse into other overprinting symbols.

b. Utility.

5.3.2 *Commercial at (@)*

5.3.2.1 In the proposed revision of ASCII the "commercial at" is relocated to position 6/0. The reason for the change is that the French represent the "at" with a single accented lower case, letter; therefore, the "@" should be in the lower case columns.

5.3.3 *Underline.* The addition of the control BS (Backspace) made practical the inclusion of the commonly used graphic *underline*. In the related codes used in some countries as many as 31 alphabetic characters will be required which have both upper- and lower-case versions. The location of *underline* in position 5/15 accommodates this requirement; it has no lower-case equivalent, and thus does not suffer from its pairing with DEL (Delete).

6. Control Assignments

6.1 Following the agreement to assign the lower case of the alphabet to the last two columns, the first two columns were reviewed to establish whether or not the present controls were those most generally required in the general interchange of information.

6.2 After a review of the control character assignments, which included liaison with Subcommittee X3.3, Data Transmission, a total of 32 were chosen. These were as follows:

NUL	(Null)	DLE	(Data Link Escape)
SOH	(Start of Heading)	DC1	(Device Control 1)
STX	(Start of Text)	DC2	(Device Control 2)
ETX	(End of Text)	DC3	(Device Control 3)
EOT	(End of Transmission)	DC4	(Device Control 4 (stop))
ENQ	(Enquiry)	NAK	(Negative Acknowledge)
ACK	(Acknowledge)	SYN	(Synchronous Idle)
BEL	(Bell (audible or attention signal))	ETB	(End of Transmission Block)
BS	(Backspace)	CAN	(Cancel)
HT	(Horizontal Tabulation)	EM	(End of Medium)
LF	(Line Feed)	SS	(Start of Special Sequence)
VT	(Vertical Tabulation)	ESC	(Escape)
FF	(Form Feed)	FS	(File Separator)
CR	(Carriage Return)	GS	(Group Separator)
SO	(Shift Out)	RS	(Record Separator)
SI	(Shift In)	US	(Unit Separator)

6.3 The following were involved in the selection of the above 32 controls.

6.3.1 The three transmission controls SOM (Start of Message), EOA (End of Address) and EOM (End of Message) were designated SOH (Start of Heading), STX (Start of Text) and ETX (End of Text) respectively to eliminate the confusion associated with the terms "Address" and "Message."

6.3.2 The control WRU (Who Are You?) was designated ENQ (Enquiry) to reflect the more general definition of this control.

6.3.3 In subsequent deliberations within Subcommittees X3.2 and X3.3 the relative importance of the function RU (Are You . . . ?) was reduced and consequently RU was eliminated from the proposed revised code table to make room for a higher priority control function.

6.3.4 The control ACK (Acknowledge) was moved from column 7 to allow for the extension of the lower case alphabet to 29 characters to satisfy the Scandinavian countries. It was agreed that this was a desirable control and should be placed in one of the first two columns. Acknowledge was placed in position 0/6 so that it lies within the communication control portion of the code, and to create the maximum practical bit difference between NAK (Negative Acknowledge) and ACK.

6.3.5 With the addition of the diacriticals and underline, the BS (Backspace) character was mandatory and was placed in position 0/8.

6.3.6 The device control DC_0 was reserved for DLE (Data Link Escape) in the X3.4-1963 standard. Additional studies in Task Group X3.3.4 substantiated the fact that a unique control character was required for use in data communications to provide a means of code extension for obtaining additional data communication link control functions.

6.3.7 The X3.4-1963 standard contains a control character designated ERR (Error) which was not defined in terms of its direction, or use. It was considered necessary to specify the direction and use, and as a result NAK (Negative Acknowledge) was designated as a backward acting communication control and CAN (Cancel) was designated as a forward acting end-to-end control.

6.3.8 The control ETB (End of Transmission Block) was considered during the development of ASCII but was given a lower priority. Subsequent deliberations, both within Subcommittee X3.2 and X3.3, have raised the relative importance of this control and it is, therefore, included in the proposed revised code table.

6.3.9 The control LEM (Logical End of Medium) was designated EM (End of Medium) to reflect the more general definition. It was moved down to position 1/9 so that all communication controls could be concentrated in the upper half of columns 0 and 1.

6.3.10 The control SS (Start of Special Sequence) was considered necessary as a single character control to indicate the start of variable length sequences of characters which have special significance or which are to receive special handling.

6.3.11 Escape was moved to position 1/11 so that all of the control characters would be in columns 0 and 1, thus b_7 and b_8 could be used to distinguish controls from graphics.

6.3.12 In concurrence with Subcommittee X3.4, Programming Languages, the number of information separators was reduced from eight to four. This was judged to be ample for general information interchange, especially since SP (Space), which is contiguous with the information separators, may also be used as an information separator. If any additions are required, it will be possible to obtain them by means of code extension. The separators were identified as FS (File Separator), GS (Group Separator), RS (Record Separator), and US (Unit Separator).

7. Bibliography

A complete bibliography is available on request from Secretary X3, Business Equipment Manufacturers Association, 235 East 42 Street, New York, New York 10017.