NBS TECHNICAL NOTE

478

Some Evolving Conventions and Standards For Character Information Coded in Six, Seven, and Eight Bits

U.S. DEPARTMENT OF COMMERCE
National Bureau of Standards

UNITED STATES DEPARTMENT OF COMMERCE
Maurice H. Stans, Secretary
NATIONAL BUREAU OF STANDARDS • A. V. Astin, Director

NBS TECHNICAL NOTE 478

ISSUED MAY 1969

Nat. Bur. Stand. (U.S.), Tech. Note 478, 27 pages (May 1969)
CODEN: NBTNA

Some Evolving Conventions and Standards for Character Information Coded in Six, Seven, and Eight Bits

John L. Little

Office of Information Processing Standards
Center for Computer Sciences and Technology
National Bureau of Standards
Washington, D.C. 20234

NBS Technical Notes are designed to supplement the Bureau's regular publications program. They provide a means for making available scientific data that are of transient or limited interest. Technical Notes may be listed or referred to in the open literature.

For sale by the Superintendent of Documents, Government Printing Office
Washington, D.C., 20402 - Price 35 cents

CONTENTS

ILLUSTRATIONS

Figures:

Some Evolving Conventions and Standards for
Character Information Coded in Six, Seven, and Eight Bits

John L. Little

ABSTRACT

This Technical Note describes some of the properties of the USA Standard Code for Information Interchange, widely known as ASCII or USASCII. It also relates this code to similar international codes and presents the national code for Russia as an example of the worldwide acceptance of this code, which is comprised of seven information bits representing 128 coded characters. Some of the conventions which are evolving to relate these seven bit codes to six and eight bit computer codes are given, and conventions for extending the code to represent an unlimited repertoire of concepts is also given. Two alternate arrangements of the code table are shown to facilitate an understanding of its structure and application.

Key words: ASCII; character codes; code extension; coded character sets; collating sequence; computer codes; data transmission codes; information code structure; international information code; Russian code standard; standard information codes; USASCII.

1. The Evolution of the USA Standard Code for Information Interchange

Coding of character-oriented information into binary bit patterns is as old as automatic telegraphy, or about a century. Perforated tape and punched cards have been in use since about the turn of the century. Magnetic wire, tape, cores, drums and disks have come into widespread use in conjunction with automatic recording of data and processing by digital computers. Independent development of such equipment resulted in a chaotic variety of conventions, and practices, for the physical techniques as well as the logical coding patterns.

Punched card codes, usually attributed to Herman Hollerith, represented originally numbers and signs, later letters and still later special symbols. The evolution always retained historical features of earlier structure, and these in turn influenced the coding of numbers, letters and symbols in many electronic computers developed much later than the card codes. Although the grouping of characters in card codes influenced the grouping of characters in computer codes, the card codes were never carried over and used internally in computers, since they are not well suited for that purpose. Minor variations in the special symbol codes used in punched cards is still creating much confusion, but standardization is being attempted.

Perforated tape and magnetic tape grew and proliferated in much greater variety of widths, recording densities and coding schemes. Standardization is attempting to bring order out of the chaos.

Magnetic disk packs are currently proliferating, and the chaos is somewhat in evidence, but relatively early attempts are being made to provide interchangeability of disk packs in order to increase their utility to users. This requires standards for physical dimensions, magnetic properties, track layout, recording techniques and density, and coding of character information as well as automatic labels for identification and data structure or format specification.

By 1960 the chaos of data representation had reached such proportions that a committee on computers and information processing, known as X3, was established by the American Standards Association. Among its goals was the establishment of a standard code for the representation of character-oriented information, and standard implementations of tnat code on the various media, such as magnetic tape, perforated tape, punched cards, and electrical communication.

The complexity of the problem and the many conflicting requirements facing the subcommittee X3.2, known as Codes and Input/Output, posed an awesome task, requiring such diverse considerations as collating sequence, keyboard conventions, provision of an adequate assortment of graphic symbols, format effectors, communication controls, device controls, and information separators, as well as appropriate subsets for use in smaller coded character sets, and provisions for expansion to encompass an unlimited repertoire of concepts not explicitly embedded in the code proper. These considerations and their resolution as reflected in the code are given in greater detail in the appendices to the standard (References 1 and 12).

The initials ASCII (pronounced ask'-ee) standing for "American Standard Code for Information Interchange" as well as USASCII (pronounced you-sass'-key) standing for "USA Standard Code for Information Interchange" have become familiar by now in many circles dealing with automated information in coded form. The code is the United States' national version of an international code adopted by several international standardizing bodies and manufacturing associations. Many countries have adopted their own national versions of this code and have given it a mandatory legal status. The United States Government has adopted it as a Federal Information Processing Standard (FIPS-1) and it will become a common requirement in many government-supported and sponsored systems: federal, state, and local.

2. Structure of Seven Bit ASCII

ASCII is a seven bit code having 128 coded character positions, as shown in Figure 1. The seven bits are designated b1 through b7 from low order to high order of significance. The eight columns are numbered 0 through 7 corresponding to the octal values of the three high order bits b5, b6, and b7. The sixteen rows are numbered 0 through 15 in accordance with the four low order bits b1, b2, b3, and b4. The code extends from seven "zero" bits, the Null character (NUL) to seven "one" bits, the Delete character (DEL). Columns 0 and 1 contain 32 control characters and columns 2 through 7 contain 95 graphic characters and the Delete character.

The meanings of the two-or three-letter groups in the ASCII code table are given below. Definitions of these functions and names of the single-symbol graphic characters are given in the standard (Reference 1).

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 (punched card skip)	EM	End of Medium
LF	Line Feed	SUB	Substitute
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
		SP	Space
		DEL	Delete

ASCII is frequently referred to as "an eight level code." This is usually due to the addition of a parity bit in paper tape and a parity bit included in the character structure of communication systems. The parity bit is not included in the code standard (Reference 1) but its inclusion and sense (odd or even) are basic to certain other standards which specify the implementation of the code in media or in communication. There are also various implementations of the seven bit code in eight bit environments, and the method of embedment of the seven code bits within an eight bit environment has caused much debate, but a resolution appears to have been achieved, as will be discussed.

The 128 characters of ASCII are arranged in a collating sequence, with NUL low and DEL high in the sequence. The non-printing graphic character Space (SP) collates ahead of all other graphic characters. The collating sequence presents a natural arrangement for machine sorting and file structuring, and it is the principal consideration in the structure of ASCII.

The ten decimal digits are positioned in order in column 3 so that their value is represented by the binary values of the four low order bits b1 through b4. Upper case and lower case Latin alphabets are each located in 26 contiguous positions. The code for an upper case letter differs from its corresponding lower case letter in only one bit, b6. Other collating and character positioning considerations are given in section 10 and in the standard (Reference 1).

3. Six, Seven, and Eight Bit Considerations

Original efforts to develop a standard code for general information interchange, in the early 1960's, attempted to limit the code to six bits, in which there can be only 64 uniquely coded characters. This was soon deemed to be inadequate, although most computers of that era represented characters in only six bits. Seven bits was deemed to be the optimum size, with its attendant 128 unique coded bit patterns, and that size was selected for ASCII. Eight bits was deemed to be wasteful because of the need to transmit the extra bit in communication systems.

Because computer design favors eight-bit groups in preference to seven bits for several cogent technical reasons, an eight-bit form of ASCII may become even more prevalent than the seven-bit form currently used in several communication systems.

4. Structure of Eight Bit USASCII-8

Placement of the 128 ASCII characters within the framework of eight-bits, with its 256 positions, has been a subject of intense controversy in standardizing circles for several years, but it appears to have stabilized into the form shown in Figure 2, in which the ASCII characters are in the left-half of the eight-bit code table, where the high order bit (E8) is zero.

In ASCII the seven bits are designated b1 through b7, while in USASCII-8 the bits are designated E1 through E8. In these notations, "b" stands for "bit" and "E" is used for distinction, apparently having come from the first letter of the word "eight."

Several arbitrary conventions have been evolved for the eight-bit code table of Figure 2. The table has 16 rows designated 0 through 15, and 16 columns designated with dual decimal digits 00 through 15. Positions in the table are specified by the column and row number separated by a slant (slash) symbol. Thus the capital letter "T" occupies position 05/4 and the control "Carriage Return" (CR) occupies 00/13. The double digits 00 through 15 for column number distinguish this notation from the similar seven-bit ASCII notation which has only one digit for column number (0 through 7).

Specific standard character assignments for the right half of Figure 2, in which the high order bit E8=1, have not yet been made, except possibly for EO, the eight ones position, which is immutable. However, it is well agreed that columns 08 and 09 should contain additional control characters, designated for reference as KO through K31 and that the other six columns 10 through 15 should contain 95 additional graphic characters and EO. The graphics designated as NO through N63 map into the center four columns of ASCII. The graphics designated GO through G30 differ from the lower case alphabet positions only in bit E8, and hence are said to map from columns 14 and 15 into columns 6 and 7 if bit E8 is ignored or dropped.

5. Preferred 64-Character Subset

The center four columns (2,3,4,5) of ASCII have become recognized as a preferred 64-character subset, which can be represented in six bits by dropping bit b6, (or by dropping bits E6 and E8 in USASCII-8). This 64-character subset can be (and often is) represented in seven bits and eight bits.

6. Representation of ASCII in Six Bits

A six-bit code is constrained to 64 unique characters. Sometimes it is desired to represent the ASCII set of 128 characters or even more in six-bit systems. In the six-bit world of seven-track magnetic tape, especially in Europe, the six bits normally represent the 64 characters shown in columns 2,3,4, and 5 of Figure 1 with bit b6 being ignored. (In library work, though, the 6-bits usually represent characters in columns 2,3,6, and 7.) In this situation, the character "Underline" in position 5/15 serves as a flag character whose meaning is: the next six-bit character following the flag is from one of the outside four columns of ASCII, two positions to the left or right of the normal position from the 64 character set. This scheme allows for the representation of 64 additional concepts as two-character sequences. A "real Underline" is represented as two successive Underlines, and hence the character "Delete" in 7/15 cannot be represented by this particular convention.

7. Representation of Concepts Beyond the Code

The 7-bit ASCII is constrained to the 128 characters shown in Figure 1 or in the left half of Figure 2. Pressure has naturally arisen from many user sources to allow for the representation of concepts not explicitly representable in ASCII. This pressure has been noted in 6, 7, and 8 bit environments. Several preferred solutions are becoming apparent.

If a system is constrained to exist in seven bits, then additional concepts can be provided with the aid of four of the ASCII control characters: "Shift Out" (SO) in 0/14, "Shift In" (SI) in 0/15, "Escape" (ESC) in 1/11, and "Data Link Escape" (DLE) in 1/0.

SO is used as a precedence character to alter the interpretation of succeeding graphic characters from columns 2 through 7 until the corresponding SI character appears and restores the normal ASCII interpretation to those graphic characters. In seven-bits, the meanings of the control characters in columns 0 and 1 remains unchanged by SO or SI.

The character "Data Link Escape" (DLE) in 1/0 is used exclusively to alter the interpretation of immediately succeeding ASCII communication control characters from columns 0 and 1 of Figure 1. There are ten of these (SOH, STX, ETX, EOT, ENQ, ACK, DLE itself, NAK, SYN, and ETB). All ten of these are from the top half of columns 0 or 1.

The character "Escape" (ESC) in 1/11 provides the greatest flexibility and unlimited capability of any. A convention appears to be evolving prescribing the use of ESC in seven bits, as follows: If the character

ESC is immediately followed by any graphic character from columns 3, 4, 5, 6, or 7, then that two-character sequence defines a new concept, which may be graphic or control in nature. If the character ESC is immediately followed by an appropriate control character from columns 0 or 1 (such as a "format effector," a "device control" or an "information separator" character) then that two-character sequence defines a new control concept which is closely related to the normal control function of the character following ESC. The expressions "locking" or "non-locking" are commonly used in an obvious manner to describe the duration of the newly defined concept. The "locking" actions can endure until an "unlocking" action occurs. "Locking" features can be cascaded one over the other, and a master "restore" convention has not been prescribed as a standard.

Unlimited possibilities occur if the character ESC is immediately followed by a graphic character from column 2 of Figure 1, still confining our attention to seven bits. Complete details are yet lacking, but in Europe a convention is evolving around the definition of three-character Escape sequences in which the first character is ESC, the second character is a graphic from column 2, and the third is a "final" character, which can be any graphic character from columns 3, 4, 5, 6, or 7, or any suitably selected character from columns 0 or 1. The class of a concept defined by a three-character sequence is determined by the middle character, from column 2, and the classification is as follows, where "(F)" represents a "final" character:

Sequence	to Represent
ESC ! (F)	Device controls
ESC " (F)	Format controls
ESC # (F)	Sets of controls "K" in 8 bits, columns 08 and 09
ESC $ (F)	Control sets in 7 bits
ESC % (F)	Shift Out sets in 7 bits
ESC & (F)	Sets (columns 0 to 7) in 7 bits
ESC ' (F)	Sets with more or less than 7 bits
ESC * (F)	National choice sets in 7 bits

Four-character Escape sequences can be generated using ESC followed by two characters from column 2 and a final character. All four-character Escape sequences are reserved for private use, not subject to standardization.

Having defined means for representing an unlimited repertoire of concepts in seven bits, we now come to the matter of representing these concepts in eight bits and of relating the seven and eight bit representations.

8. Mapping Between Seven and Eight Bit Codes

Figure 2 is an eight-bit code table with the left half fully defined as representing the 128 characters of ASCII. The right half is defined only as having 32 controls, 95 graphics and the Eight Ones (EO) character. It is unlikely that the right half will ever be defined as fully as the left half for general use. Instead, it is likely that a family of eight bit codes will evolve, each for a particular community of interest, representing such communities as libraries, typesetting, mathematics, organic chemistry, non-Latin alphabets, and many others.

Referring again to Figure 2, we can think of the whole table as representing an eight-bit code. Or if we choose, we can ignore bit E8 and think of the table as representing two seven bit code tables, the left one being ASCII, and the right one containing the Shift Out set of graphics.

We note that there are 95 graphic characters in ASCII, including Space (SP) in 02/0, and we note that there are 95 graphic characters of either the N-type or G-type on the right half. Each of these 190 graphic characters has a unique bit code in eight bits, and hence, in eight bits, they can be interspersed in a character stream in any order, without the use of the shift characters Shift Out (SO) or Shift In (SI). The information conveyed by SO and SI (in seven bits) is now inherent in the eighth bit E8.

In making a translation from a seven bit character stream into an eight bit stream, we can monitor for SO and SI, use these to set bit E8 for subsequent eight bit characters, and then either discard or retain the SO and SI characters.

In going from a stream of eight bit graphic characters into a seven bit stream, we can monitor bit E8, and when it changes, we can insert an SO or SI character into the seven bit stream, to retain the information conveyed by bit E8.

A convention has been gaining favor relating the 32 "K" controls of columns 08 and 09 into two-character sequences in seven bits, having ESC as the first character and a corresponding graphic from columns 4 or 5 as the second character. In this transform, the six low-order bits of the graphic character (b1 through b6) match the six low-order bits of the K control (E1 through E6). In the reverse process, going from seven bits to eight bits, a two-character sequence comprised of ESC followed by a graphic character from columns 4 or 5, transforms into a single eight bit control character in columns 08 or 09, in a position corresponding to the position of the graphic character. Other multi-character Escape sequences remain the same in seven and eight bit streams.

9. Codes of More than Eight Bits

At the present time there does not appear to be any particular interest in conventions for representing characters in codes larger than eight bits, although devices exist, such as typesetting mechanisms, in which more than eight bits are used to represent each character, and coding schemes are used in computers in which two eight-bit groups are used to represent a character or an image of a character (Reference 6).

10. Collating Sequence of ASCII and USASCII-8

The 1963 version of the ASCII standard (Reference 3) disassociates ASCII from collating sequence with the sentences in section 5.1.

"The standard code does not include any redundancy or define techniques for error control. Further, it does not specify a standard collating sequence." However, that position is reversed in 1965 ASCII (Reference 2) and in the current ASCII (Reference 1) in section 6.3 which reads "The relative sequence of any two characters, when used as a basis for collation, is defined by their binary values."

Figure 2 defines a collating sequence for the 256 characters of the eight bit code, with the character NUL as the lowest in the sequence, SOH next, and so on, with EO (Eight Ones) being the highest. The graphic character Space (SP) collates lower than any other graphic character.

Complaints are sometimes heard that all of the upper case letters collate below all of the lower case letters, and that miscellaneous symbols collate between the upper and lower case alphabets. It is obvious from Figure 2 that these complaints are valid. However, the painstaking structure of ASCII, in which collating sequence is probably the dominant (though far from the only) feature allows for a variation such as that shown in Figure 3.

Figure 3 is a rearrangement of Figure 2 in which bit E6 has been moved to the lowest order position, and bits E1 through E5 are moved up one place.

In Figure 3, the alphabetic characters collate in the sequence A, a, B, b, C, c, etc., with no gaps in the sequence from A to z.

The character Space (SP) collates lower than all other characters, graphic as well as control, except for the character Null (NUL). The numerals 0 through 9 still collate in the proper sequence, but they are interspersed with control characters and other graphic characters intervene between the numerals and the letters, as in the ASCII collating sequence.

At first glance this appears to be a highly advantageous arrangement for purely alphabetic material, and such interspersed alphabets have accordingly been used in some computer codes. However, there are some highly subtle flaws in this collating sequence, as is vividly illustrated in Reference 10.

For example, if a sort were made on alphabetic material using the interleaved letter collating sequence of Figure 3, then all words beginning with lower case letter 'a" would follow all words beginning with capital "A", which is not the arrangement usually desired in listings of mixed upper and lower case material, such as those found in dictionaries, telephone directories, or library catalogs. The case distinction must be ignored, by ignoring bit b6 or E6, for sorting, but the case distinction must be retained for printing and for fine sorts. More detailed consideration is given in Reference 11.

Note that there is no bit reassignment whatever between Figure 2 and Figure 3. Only the order of presentation is different. But the interspersed alphabets of Figure 3 may suggest some of the flexibility inherent in ASCII. By suitable logical bit manipulation, one could obtain the interspersed alphabets of Figure 3 and the contiguous digits of Figure 2.

11. Octal and Hexadecimal Representation of the Coded Characters

The column/row notation of Figures 2 and 3 is readily convertible to hexadecimal (4bit) notation. Single decimal digits 0 through 9 are used for both column and row, while 10 through 15 are usually represented as single capital letters A through F for both column and row.

The arrangements of the code tables of Figures 1 or 2 make translation of the bit patterns into octal (3 bit) characters quite an awkward operation, since the bits are not grouped conveniently for that purpose, and the octal characters do not come out even for either seven or eight bits, although they would for six or nine bits, with no leftover bits.

When seven bit characters are represented in octal form, three octal digits are required to represent each character, and the high order octal digit is always either zero or one, as shown in the top half of Figure 4. In representing all 256 eight-bit characters, the high order octal digit can be 0, 1, 2, or 3. The other two octal digits can, of course, have any value from 0 to 7. The octal digits are shown explicitly in Figure 4 instead of column and row numbers. The code table arrangement, of Figure 4, was suggested by Crosby (Reference 7).

12. Three Evolutionary Forms of ASCII

The seven bit ASCII (American Standard Code for Information Interchange) has existed in its evolution in three distinct, but closely related versions. The first was known as American Standard X3.4 - 1963, approved by the American Standards Association on June 17, 1963 (Reference 3). It did not contain lower case letters, nor the characters Backspace, Underline, and several others which have been added subsequently.

The second version was an extensive revision approved by ASA as X3.4 - 1965 but never distributed to the public due to pending international modifications in the ISO 7-bit code. However, the 1965 version was published as a proposed revision (Reference 2) along with an exposition on the evolution of the ISO 7-bit code.

The present version is known as the United States of America Standard Code for Information Interchange, X3.4 - 1967, approved by the United States of America Standards Institute (successor to ASA) on July 7, 1967. This version was adopted as a Federal Standard by President Lyndon B. Johnson in a memorandum for the Heads of Departments and Agencies, dated March 11, 1968. The latest version was published in 1968 as a combined ASCII and Federal Information Processing Standard, FIPS - 1 (Reference 12).

The evolution of ASCII, now also called USASCII, was influenced heavily by concurrent developments of international 7-bit codes, particularly the ISO 7-bit code (Reference 4) and the CCITT Alphabet No. 5, which is destined to become the successor to CCITT Alphabet No. 2, the international five-bit Baudot code, which is still in extensive but diminishing use.

The 1965 version of ASCII represented a major renaming and repositioning of the control functions from those originally appearing in the 1963 standard. The graphic characters "Up arrow" and "Left arrow" and the control character "RU" (Are you...?) were removed and replaced by other characters. The character "@" (Commercial At) was moved from 4/0 to 6/0, but because of many complaints, it was moved back to 4/0 in the 1967 ASCII. "Commercial At" was originally moved to make room for the character "Underline" in the preferred center four columns. In a CCITT meeting the Russian delegate requested that Underline be moved from 4/0 to 5/15 so as to pair with Delete, and make room for 31 Cyrillic alphabetic characters in both upper and lower case, displacing all of the characters from columns 4, 5, 6, and 7 except Underline and Delete, as shown in the right half of Figure 5.

13. National Versions of the Seven Bit Codes

The Seven Bit Code of the International Organization for Standardization (Reference 4) and the nearly identical international communications code of the International Consultative Committee on Telegraph and Telephone (CCITT) contain eleven code table positions which are reserved in varying degrees for "National Use." These eleven positions are 2/3, 4/0, 5/11, 5/12, 5/13, 5/14, 6/0, 7/11, 7/12, 7/13, and 7/14. Position 2/3 is known to contain only the number sign (#) and the symbol for "Pound Sterling." All of the other positions contain a variety of symbols. ASCII has assigned the United States' preferred symbols to these positions, and it is noted that many other nations have assigned the same symbols where there was no other need. Some of the ASCII national use symbols are shown as preferred assignments in the ISO code.

14. The Russian Code

In a USSR State Standard (Reference 8) the Russians have changed "Shift Out" to "Russian Register" and "Shift In" to "Latin Register." In the Russian register, they have placed uppercase Cyrillic characters in columns corresponding to lowercase Latin letters as shown in Figure 5, a seven bit code. Only 31 of the usual 33 Cyrillic letters are included. Omitted is the silent symbol which resembles "bl" and the "E" or "e" with the double-dot (diaeresis) diacritical overmark. The Cyrillic letters are arranged to match their Latin phonetic equivalents, rather than being placed in alphabetical order. This facilitates Latin-Cyrillic keyboard pairing, but complicates alphabetical sorting by computer of Cyrillic data.

The placement of uppercase Cyrillic letters in columns corresponding to lowercase Latin letters was done to facilitate a third seven-bit code arrangement, not illustrated, which has all lowercase letters removed, and both Latin and Cyrillic uppercase letters in their usual positions. This permits monocase operations with both Latin and Cyrillic alphabets without the use of Shift Out and Shift In characters.

Use of the Russian code is explained in more detail in Reference 9.

References

1. (1967 ASCII) "USA Standard Code for Information Interchange," X3.4-1967. Available for $2.00 from United States of America Standards Institute, 10 East 40th Street, New York, N. Y. 10016. Also approved as Federal Information Processing Std. (FIPS-1).

2. (1965 ASCII) "Proposed Revised American Standard Code for Information Interchange," Communications of the ACM, Vol. 8, No. 4, April 1965, pp. 207-214.

3. (1963 ASCII) "American Standard Code for Information Interchange," approved June 17, 1963 by the American Standards Association (now USASI, Ref. 1 above), no longer in print, but code table shown in Ref. 2.

4. (ISO 7 Bit Code) "6 and 7 Bit Coded Character Sets for Information Processing Interchange," ISO Recommendation R646, International Organization for Standardization, available for $6.00 from USA Standards Institute (see Ref. 1 above).

5. (USASCII-8) "Interchange Code Adds an Eighth Bit," by Eric H. Clamons, Control Engineering, July 1968, pp. 76-78.

6. (GPSDIC) "The Document Image Code System," National Bureau of Standards Technical News Bulletin, Vol. 52, No. 4, April 1968, pp. 86-88.

7. "Note on an ASCII-Octal Code Table" by Wm. S. Crosby, Communications of the ACM, Vol. 8, No. 10, October 1965, p. 616.

8. USSR State Standard GOST 13052-67, "Computers and Data Transmission Equipment Alpha-Numeric Codes," approved by the Committee of Standards, Measures, and Measuring Equipment of the Council of Ministers, USSR, July 10, 1967. Translator: Mrs. A. Werner, and Editor of Translation: D. A. Kerr, Bell Telephone Laboratories, Inc., Holmdel, New Jersey, June 25, 1968.

9. "Standard Data Transmission Code" by P. A. Yeskin and O. P. Kireyeo, Telecommunications (USSR), published in English as Telecommunications and Radio Engineering, Vol. 22, No. 2, February, 1968, pp. 24-28, Scripta Publishing Corp., Washington, D. C.

10. "ESC Facility in USASCII," a letter by R.W. Bemer, in Communications of the ACM, Vol. 10, No. 4, April 1967, p. 202.

11. "Computer Filing of Index, Bibliographic, and Catalog Entries," by T. C. Hines and J. L. Harris, Bro-Dart Foundation, Newark, N. J., 1966, L. C. Cat. Card No. 66-23484.

12. (1968 ASCII) "USA Standard Code for Information Interchange," published jointly as X3.4 - 1968 and FIPS-1. A very minor revision of Ref. 1. The revisions, allowing "New Line" as an option for "Line Feed" is not included in FIPS-1.

USA STANDARD CODE FOR INFORMATION INTERCHANGE
USAS X3.4-1967

Bits 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
b_4	b_3	b_2	b_1	Column → / Row ↓	0	1	2	3	4	5	6	7
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	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

FIGURE 1.

BITS E4	E3	E2	E1	COLUMN → / ROW ↓	00	01	02	03	04	05	06	07	08	09	10	11	12	13	14	15
E8 E7 E6 E5 →					0 0 0 0	0 0 0 1	0 0 1 0	0 0 1 1	0 1 0 0	0 1 0 1	0 1 1 0	0 1 1 1	1 0 0 0	1 0 0 1	1 0 1 0	1 0 1 1	1 1 0 0	1 1 0 1	1 1 1 0	1 1 1 1
0	0	0	0	0	NUL	DLE	SP	0	@	P	`	p	K0	K16	N0	N16	N32	N48	G0	G16
0	0	0	1	1	SOH	DC1	!	1	A	Q	a	q	K1	K17	N1	N17	N33	N49	G1	G17
0	0	1	0	2	STX	DC2	"	2	B	R	b	r	K2	K18	N2	N18	N34	N50	G2	G18
0	0	1	1	3	ETX	DC3	#	3	C	S	c	s	K3	K19	N3	N19	N35	N51	G3	G19
0	1	0	0	4	EOT	DC4	$	4	D	T	d	t	K4	K20	N4	N20	N36	N52	G4	G20
0	1	0	1	5	ENQ	NAK	%	5	E	U	e	u	K5	K21	N5	N21	N37	N53	G5	G21
0	1	1	0	6	ACK	SYN	&	6	F	V	f	v	K6	K22	N6	N22	N38	N54	G6	G22
0	1	1	1	7	BEL	ETB	'	7	G	W	g	w	K7	K23	N7	N23	N39	N55	G7	G23
1	0	0	0	8	BS	CAN	(	8	H	X	h	x	K8	K24	N8	N24	N40	N56	G8	G24
1	0	0	1	9	HT	EM	)	9	I	Y	i	y	K9	K25	N9	N25	N41	N57	G9	G25
1	0	1	0	10	LF	SUB	*	:	J	Z	j	z	K10	K26	N10	N26	N42	N58	G10	G26
1	0	1	1	11	VT	ESC	+	;	K	[	k	{	K11	K27	N11	N27	N43	N59	G11	G27
1	1	0	0	12	FF	FS	,	<	L	\	l	|	K12	K28	N12	N28	N44	N60	G12	G28
1	1	0	1	13	CR	GS	-	=	M	]	m	}	K13	K29	N13	N29	N45	N61	G13	G29
1	1	1	0	14	SO	RS	.	>	N	^	n	~	K14	K30	N14	N30	N46	N62	G14	G30
1	1	1	1	15	SI	US	/	?	O	_	o	DEL	K15	K31	N15	N31	N47	N63	G15	EO

ON E8=1 SIDE K,N,G ARE ARBITRARY DESIGNATIONS
K=CONTROL, N and G ARE GRAPHIC CHARACTERS
EO = EIGHT ONES

FIGURE 2. USASCII-8

E8 E7 E5 E4 →					0000	0001	0010	0011	0100	0101	0110	0111	1000	1001	1010	1011	1100	1101	1110	1111
BITS E3	E2	E1	E6	COLUMN → / ROW ↓	00	01	02	03	04	05	06	07	08	09	10	11	12	13	14	15
0	0	0	0	0	NUL	BS	DLE	CAN	@	H	P	X	K0	K8	K16	K24	N32	N40	N48	N56
0	0	0	1	1	SP	(	0	8	`	h	p	x	N0	N8	N16	N24	G0	G8	G16	G24
0	0	1	0	2	SOH	HT	DC1	EM	A	I	Q	Y	K1	K9	K17	K25	N33	N41	N49	N57
0	0	1	1	3	!	)	1	9	a	i	q	y	N1	N9	N17	N25	G1	G9	G17	G25
0	1	0	0	4	STX	LF	DC2	SUB	B	J	R	Z	K2	K10	K18	K26	N34	N42	N50	N58
0	1	0	1	5	"	*	2	:	b	j	r	z	N2	N10	N18	N26	G2	G10	G18	G26
0	1	1	0	6	ETX	VT	DC3	ESC	C	K	S	[	K3	K11	K19	K27	N35	N43	N51	N59
0	1	1	1	7	#	+	3	;	c	k	s	{	N3	N11	N19	N27	G3	G11	G19	G27
1	0	0	0	8	EOT	FF	DC4	FS	D	L	T	\	K4	K12	K20	K28	N36	N44	N52	N60
1	0	0	1	9	$	,	4	<	d	l	t	¦	N4	N12	N20	N28	G4	G12	G20	G28
1	0	1	0	10	ENQ	CR	NAK	GS	E	M	U	]	K5	K13	K21	K29	N37	N45	N53	N61
1	0	1	1	11	%	-	5	=	e	m	u	}	N5	N13	N21	N29	G5	G13	G21	G29
1	1	0	0	12	ACK	SO	SYN	RS	F	N	V	^	K6	K14	K22	K30	N38	N46	N54	N62
1	1	0	1	13	&	•	6	>	f	n	v	~	N6	N14	N22	N30	G6	G14	G22	G30
1	1	1	0	14	BEL	SI	ETB	US	G	O	W	_	K7	K15	K23	K31	N39	N47	N55	N63
1	1	1	1	15	'	/	7	?	g	o	w	DEL	N7	N15	N23	N31	G7	G15	G23	E0

USASCII-8 REARRANGED WITH BIT E6 IN THE LOWEST ORDER POSITION
FOR USASCII-7 OMIT E8 AND RIGHT HALF

FIGURE 3.

E8	E7	E6	E5	E4	E3→ E2→ E1→	0 0 0	0 0 1	0 1 0	0 1 1	1 0 0	1 0 1	1 1 0	1 1 1
↓	↓	↓	↓	↓	N8	0	1	2	3	4	5	6	7
0	0	0	0	0	00	NUL	SOH	STX	ETX	EOT	ENQ	ACK	BEL
0	0	0	0	1	01	BS	HT	LF	VT	FF	CR	SO	SI
0	0	0	1	0	02	DLE	DC1	DC2	DC3	DC4	NAK	SYN	ETB
0	0	0	1	1	03	CAN	EM	SUB	ESC	FS	GS	RS	US
0	0	1	0	0	04	SP	!	"	#	$	%	&	'
0	0	1	0	1	05	(	)	*	+	,	-	.	/
0	0	1	1	0	06	0	1	2	3	4	5	6	7
0	0	1	1	1	07	8	9	:	;	<	=	>	?
0	1	0	0	0	10	@	A	B	C	D	E	F	G
0	1	0	0	1	11	H	I	J	K	L	M	N	O
0	1	0	1	0	12	P	Q	R	S	T	U	V	W
0	1	0	1	1	13	X	Y	Z	[	\	]	^	_
0	1	1	0	0	14	`	a	b	c	d	e	f	g
0	1	1	0	1	15	h	i	j	k	l	m	n	o
0	1	1	1	0	16	p	q	r	s	t	u	v	w
0	1	1	1	1	17	x	y	z	{	¦	}	~	DEL
1	0	0	0	0	20								
1	0	0	0	1	21								
1	0	0	1	0	22								
1	0	0	1	1	23								
1	0	1	0	0	24								
1	0	1	0	1	25								
1	0	1	1	0	26								
1	0	1	1	1	27								
1	1	0	0	0	30								
1	1	0	0	1	31								
1	1	0	1	0	32								
1	1	0	1	1	33								
1	1	1	0	0	34								
1	1	1	0	1	35								
1	1	1	1	0	36								
1	1	1	1	1	37								EO

USASCII-8
Rearranged
to aid octal users

For USASCII-7
omit E8 and
the bottom half

From CROSBY,
Reference 7

FIGURE 4

RUSSIAN CODE TABLE, From Reference 8

					LATIN REGISTER [SHIFT]								RUSSIAN REGISTER [SHIFT]							
BITS $E_7E_6E_5$ →					000	001	010	011	100	101	110	111	000	001	010	011	100	101	110	111
E_4	E_3	E_2	E_1	COLUMN → / ROW ↓	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
0	0	0	0	0	NUL	DLE	SP	0		P	`	p	NUL	DLE	SP	0	ю	п	Ю	П
0	0	0	1	1	SOH	DC1	!	1	A	Q	a	q	SOH	DC1	!	1	а	я	А	Я
0	0	1	0	2	STX	DC2	"	2	B	R	b	r	STX	DC2	"	2	б	р	Б	Р
0	0	1	1	3	ETX	DC3	#	3	C	S	c	s	ETX	DC3	#	3	ц	с	Ц	С
0	1	0	0	4	EOT	DC4	¤	4	D	T	d	t	EOT	DC4	¤	4	д	т	Д	Т
0	1	0	1	5	ENQ	NAK	%	5	E	U	e	u	ENQ	NAK	%	5	е	у	Е	У
0	1	1	0	6	ACK	SYN	&	6	F	V	f	v	ACK	SYN	&	6	ф	ж	Ф	Ж
0	1	1	1	7	BEL	ETB	'	7	G	W	g	w	BEL	ETB	'	7	г	в	Г	В
1	0	0	0	8	BS	CAN	(	8	H	X	h	x	BS	CAN	(	8	х	ь	Х	Ь
1	0	0	1	9	HT	EM	)	9	I	Y	i	y	HT	EM	)	9	и	ы	И	Ы
1	0	1	0	10	LF	SUB	*	:	J	Z	j	z	LF	SUB	*	:	й	з	Й	З
1	0	1	1	11	VT	ESC	÷	;	K	[	k		VT	ESC	+	;	к	ш	К	Ш
1	1	0	0	12	FF	IS4	,	<	L	ˇ	l		FF	IS4	,	<	л	э	Л	Э
1	1	0	1	13	CR	IS3	-	=	M	]	m		CR	IS3	-	=	м	щ	М	Щ
1	1	1	0	14	RUS	IS2	.	>	N	^	n	‾	RUS	IS2	.	>	н	ч	Н	Ч
1	1	1	1	15	LAT	IS1	/	?	O	_	o	DEL	LAT	IS1	/	?	о	_	О	DEL

[Translation note: Only the international specific control character designations are shown in this table. The original code table uses the Russian designations. For these designations, see Table 2. RUS (Russian register) and LAT (Latin register) correspond to SO (Shift Out) and SI (Shift In), respectively.]

Remarks:
1. Symbols of locations 0/0 - 3/15 are repeated in locations 8/0 - 11/15.
2. Locations 4/0 and 7/11 - 7/13 are reserve locations.

FIGURE 5.

☆ U. S. GOVERNMENT PRINTING OFFICE : 1969—341-027/189

NBS TECHNICAL PUBLICATIONS

PERIODICALS

JOURNAL OF RESEARCH reports National Bureau of Standards research and development in physics, mathematics, chemistry, and engineering. Comprehensive scientific papers give complete details of the work, including laboratory data, experimental procedures, and theoretical and mathematical analyses. Illustrated with photographs, drawings, and charts.

Published in three sections, available separately:

- **Physics and Chemistry**

Papers of interest primarily to scientists working in these fields. This section covers a broad range of physical and chemical research, with major emphasis on standards of physical measurement, fundamental constants, and properties of matter. Issued six times a year. Annual subscription: Domestic, $6.00; foreign, $7.25*.

- **Mathematical Sciences**

Studies and compilations designed mainly for the mathematician and theoretical physicist. Topics in mathematical statistics, theory of experiment design, numerical analysis, theoretical physics and chemistry, logical design and programming of computers and computer systems. Short numerical tables. Issued quarterly. Annual subscription: Domestic, $2.25; foreign, $2.75*.

- **Engineering and Instrumentation**

Reporting results of interest chiefly to the engineer and the applied scientist. This section includes many of the new developments in instrumentation resulting from the Bureau's work in physical measurement, data processing, and development of test methods. It will also cover some of the work in acoustics, applied mechanics, building research, and cryogenic engineering. Issued quarterly. Annual subscription: Domestic, $2.75; foreign, $3.50*.

TECHNICAL NEWS BULLETIN

The best single source of information concerning the Bureau's research, developmental, cooperative and publication activities, this monthly publication is designed for the industry-oriented individual whose daily work involves intimate contact with science and technology—*for engineers, chemists, physicists, research managers, product-development managers, and company executives.* Annual subscription: Domestic, $3.00; foreign, $4.00*.

*Difference in price is due to extra cost of foreign mailing.

NONPERIODICALS

Applied Mathematics Series. Mathematical tables, manuals, and studies.

Building Science Series. Research results, test methods, and performance criteria of building materials, components, systems, and structures.

Handbooks. Recommended codes of engineering and industrial practice (including safety codes) developed in cooperation with interested industries, professional organizations, and regulatory bodies.

Special Publications. Proceedings of NBS conferences, bibliographies, annual reports, wall charts, pamphlets, etc.

Monographs. Major contributions to the technical literature on various subjects related to the Bureau's scientific and technical activities.

National Standard Reference Data Series. NSRDS provides quantitative data on the physical and chemical properties of materials, compiled from the world's literature and critically evaluated.

Product Standards. Provide requirements for sizes, types, quality and methods for testing various industrial products. These standards are developed cooperatively with interested Government and industry groups and provide the basis for common understanding of product characteristics for both buyers and sellers. Their use is voluntary.

Technical Notes. This series consists of communications and reports (covering both other agency and NBS-sponsored work) of limited or transitory interest.

Federal Information Processing Standards Publications. This series is the official publication within the Federal Government for information on standards adopted and promulgated under the Public Law 89–306, and Bureau of the Budget Circular A–86 entitled, Standardization of Data Elements and Codes in Data Systems.

CLEARINGHOUSE

The Clearinghouse for Federal Scientific and Technical Information, operated by NBS, supplies unclassified information related to Government-generated science and technology in defense, space, atomic energy, and other national programs. For further information on Clearinghouse services, write:

Clearinghouse
U.S. Department of Commerce
Springfield, Virginia 22151

Order NBS publications from: Superintendent of Documents
Government Printing Office
Washington, D.C. 20402

U.S. DEPARTMENT OF COMMERCE
WASHINGTON, D.C. 20230

OFFICIAL BUSINESS

POSTAGE AND FEES PAID
U.S. DEPARTMENT OF COMMERCE