Amended. JR. 17.9.64

Erratum to Memorandum No. 3.

Table 5: Monitor typewriter character/code set should be replaced with the corrected table overleaf.

TABLE 5 : Monitor typewriter character/code set.

B / A	00		10		20		30		40		50		60	70	
	L	U	L	U	L	U	L	U	L	U	L	U		L	U
0			Q	Q			F	F	Y	Y	;	:	Space	J	J
1	B	B	C	C			N	N	L	L	E	E	Backspace	X	X
2	W	W	A	A			.	.	●	O	'	"	Tabulate	M	M
3			6	¢			3	#	0	)	7	&	New line	Z	Z
4	-	_	,	,			=	+	/	?	P	P	Upper Case	G	G
5			K	K			U	U	H	H	D	D		T	T
6			I	I			V	V	S	S	R	R	Lower Case	!	°
7	9	(	8	*			2	@	4	$	5	%		1	±

COMPUTING RESEARCH SECTION

Memorandum No.3

Character & Code Sets for the Basic Network

Contents :

CRS3/4/64

COMPUTING RESEARCH SECTION

Memorandum No.3

CHARACTER & CODE SETS FOR THE BASIC NETWORK

1. INTRODUCTION

Considerable detailed effort has been made by the C.S.I.R.O. Computing Research Section in the definition and adoption of suitable sets of characters, their associated codes and their implementation on the media to be handled by the network. This memorandum deals mainly with the actual 'hardware' character/code set, those which apply to media like magnetic and paper tapes, and their corresponding symbols produced via printing units.

A standard character/code set for transmission of data between parts of the network, members, users and other participants and standards for identification of the properties of character/code sets to which network members and users may be locally limited and may require for acceptance by the network, will be defined in due course.

The Section has taken the opportunity to choose character/code sets in view of the probable development and implementation of advanced programming procedures during the next few years. It has long been felt that the character/code sets commonly available with punched card data preparation equipment, and utilised by a number of very important source language schemes is inadequate and restrictive to their wealth of expression, rendering difficult both human reading and understanding, as well as the establishment of compiler systems using them.

The only readily available medium upon which a suitable wealth of characters can be provided at present is 7 and 8

channel paper tape. Even so, data preparation equipment currently available with this medium provides only for 90 different 'graphic characters' plus a number of 'control characters' which control the format of the data on a printed page, e.g. 'new line'. The 90 graphic characters occur in two 'cases', each pair corresponding to one key of the data preparation keyboard of 45 keys not including control keys.

Although the main character set chosen for the system is currently restricted to a two-case 90 graphic set fitted to a 6-bit code, one of the factors affecting the choice has been the belief that in due course, an even wider character set will be needed requiring an internal code occupying at least 8 bits. In the present scheme suitable choice of the graphics and control functions provides facility for some extension by the construction of composite graphics from the 90 graphic set adopted.

Two separate but related character sets are distinguished :

A. A set of 90 graphic characters assigned between 'upper' and 'lower' case and represented by odd-parity 7-bit codes adapted to paper tape.

B. A set of 63 graphic characters assigned to 6-bit codes, suited to line-printer and magnetic-tape recording.

No attempt has been made to differ from an existing punched card character/codes and no attempt made to achieve compatibility between any 5-bit code set as a subset of code A.

The only internal code considered is that currently available or readily adaptable to the CDC 6-bit internal code for which software exists. Although it is expected that full implementation of the ALGOL symbolism will occupy 118 different characters in an 8-bit internal code, such a code is not as yet implemented in software.

2. AIMS

The following have been the objectives :

i) To enable use of FORTRAN, COBOL and ALGOL programming languages specifically, with special regard to the future use and development of languages requiring extension to a wider character set.

ii) To enable programs and data to be accepted from punched paper tape and punched cards.

iii) To allow for listing programs on a line printer, and via punched paper tape on a tapewriter, in forms identical with or closely corresponding to the original source language notation.

iv) To simplify conversion to internal and existing standard codes such as those for punched cards and the 'external' BCD codes for magnetic tape recording.

v) To simplify conversion between internal codes and 'hardware' codes, i.e. those for tapewriters, monitor typewriters and line printers.

vi) To simplify any data conversion required to conform with expected code standards for data communication on a wide scale.

3. CONDITIONS

The following were the main practical conditions.

i) The need to minimise when dealing with program source-language statements in FORTRAN and COBOL on paper tape, the number of case shifts. Both these languages use decimal digits, capital letters and certain punctuation

and special symbols amounting to 47 characters for FORTRAN and an additional 4 characters for COBOL. This requires that the symbols in excess of 45 be those of least frequent usage and available in upper case on paper tape. The 45th key is special.

ii) The maintenance of a preferred collation sequence which would provide codes in increasing order of value; format functions equipment functions, symbols for operators, punctuation and special delimiters, decimal digits 0 ...9 and letters A, a, B, b.... Z,z.

iii) For simplicity in conversion of decimal data to binary or binary coded decimal (BCD) form it is desirable that the codes for decimal digits be closely related to their equivalent binary values.

iv) The codes for letters should contain a five-bit array related to their collation sequence in the order of A 00001, B 00010, C 00011 and so on to Z 11010.

v) Since the network will be intimately concerned with data interchange between systems, a component of the chosen character/code scheme should conform closely to a widely accepted and promulgated standard.

4. EQUIPMENT CONSIDERATIONS

On-line peripheral devices are associated with special routines called "drivers" which control the functions of the associated devices and provide for code conversion from the internal data processing set to that appropriate to the unit. It is desirable that the storage and time demanded by such

devices be as small as possible, and this is attained if there is a close correspondence between the internal and external equipment codes.

Compatibility with other systems is possible through use of two particular media, namely 80-column punched cards and magnetic tapes to IBM standards. The latter is $\frac{1}{2}$ inch wide and is recorded at a density of 200, 556 or 800 frames per inch in a 7-track format and in either BCD (even parity) or binary (odd parity) mode. However, the varieties of existing punched paper tape and their codes are so numerous that some conversion will be inevitable. Thus a definite network character/code standard has been adopted for paper tape and associated data preparation units. This will serve future needs for a wide character set oriented towards program processing, suitably collatable and close to the Internal Standards Organisation (ISO) code and American Standard Code for Information Interchange (ASCII).

5. MAGNETIC TAPES

Six bit characters are recorded on IBM compatible magnetic tape in either of two modes; binary groups with additional odd parity and BCD groups with additional even parity. The code of the latter mode is referred to as the 'external' BCD code and is converted to the 'internal' code (as used by the computer for data processing) according to the following rules:

a) if the fifth bit is a '1', the sixth bit is complemented, thus an internal code 21_8 would be changed to an external code 61_8 and visa versa.

b) internal (core) 00_8 becomes external (tape) 12_8 and internal 12_8 is transferred unchanged. On return to core external 12_8 is converted to internal 00_8 (see appendix).

In the case of the FORTRAN character set, the internal data processing and magnetic tape external codes used by the Control Data 3600 and Control Data 3200's are the same as those used by IBM, as also are the physical track standards and code weightings. Thus there will be direct compatibility between the network and any such system as IBM 7090, 7040, 7044 etc. Additional characters have been provided on line printers and tape-writers etc. to occupy the remainder of the total of 64 code items.

It will be noted that the conversion between the internal and external BCD codes is by 'hardware' and not by program, thus simplifying the actual reading and writing routines.

Table 1 shows the relation between the full set of 64 graphic characters and their representations on magnetic tape i.e. in external BCD code.

6. PUNCHED CARDS

The punched card code (see table 2) is that provided by standard IBM Keyboard type 026H card punches, in which alphabetic and other non-numeric symbols are represented by multiple punching columns with an additional 12, 11, 0 or 8. The card codes are readily converted to internal BCD code used within the 3600 or 3200 by supplying suitable weightings to the rows 12, 11, 0 and 8. Table 2 also shows the connection between the card characters and the internal BCD codes.

The Control Data 405 card reader will be provided at the Centre and the Subsidiaries and is capable of reading cards in a column by column manner in either of two modes, in BCD or binary card image. The normal mode of reading converts punchings

in each column to an equivalent 6-bit character in the internal BCD code. In the binary mode each column read occupies a 12-bit byte as part of binary 'card image'.

It is possible to inhibit BCD mode and to read a binary card image by program control or by presence of a 7, 9 multiple punch on column 1 which will cause the remainder of a card to be read in binary mode. The latter feature is used extensively by the monitor systems (SCOPE 36 and SCOPE 32).

It will be noted that conversion to internal BCD is via hardware, thus simplifying reading routines for BCD data.

7. PUNCHED PAPER TAPE

The 'tapewriter' unit is intended primarily for the preparation of data on either 1" or $\frac{7}{8}$" wide paper tape (at choice of the user) via a 'typewriter' keyboard, and to provide simultaneously a typed copy of the data punched. It will automatically read pre-punched paper tape and optionally provide a page printed version of the data, and will produce a copy of a punched tape with or without simultaneous printing.

i) The Tapewriter Character Set

The tapewriter possesses a set of 46 keys laid out in the standard typewriter manner, 45 being associated with two graphic characters, thus providing a total set of 90 symbols. The 46th key remains unallocated Additional control functions are operated by depression of additional special keys placed on the keyboard, 'space', 'new line', 'shift to upper case', 'shift to lower case', 'backspace', 'tabulate'. Certain coded control functions are controlled by tabs placed above the keyboard which allow for switching tape readers

and tape punches into use and are used mainly for editing purposes. The set of 90 symbols used and their corresponding paper tape codes are given in Table 3. The character set has been chosen to provide easy preparation of both programs including ALGOL and FORTRAN and data. The character set of the ALGOL reference language can be constructed from the set provided by the use of the 'backspace' control function and over-printing of two or more characters. The use of backspace thus effectively provides for a much wider character set than the 90 characters provided on the keyboard. Typical of those in formal ALGOL would, for instance, be those provided by the use of the underline, vertical line, hyphen etc. Conventions concerning the use of superposed characters are given later in this memorandum (See Section 10). In accordance with common typewriter practice decimal digits are available in lower case, and for ease of recording programs in FORTRAN all characters commonly used in this language are available in lower case, all lower case letters being capitals of smaller size than corresponding letters in upper case. Only two FORTRAN characters namely '$' and '=' require insertion of an additional case change function immediately before and after either character. Further, although use of '=' cannot be avoided, use of '$' may be avoided by placing each statement on a separate line.

Two characters, namely % and ? in upper case, have been included for reasons of uniformity with respect to character/code sets being adopted by other data processing systems in Australia.

One further character requires comment, namely ◇ (paired with ↑). Although the character and its code 76 has no corresponding code in the internal BCD system of 'software' and the character does not appear in any other medium, it will be used as an 'escape' symbol. It is to be specially noted that the asterisk is available in lower case and it is intended to use this to represent the multiplication sign used in the ALGOL reference language.

ii) The Tapewriter Code Set

The restrictions imposed by the limited number of keys on the tapewriter keyboard and the use of case shift functions allow the corresponding code to be restricted to 7 channels with odd parity throughout.

Table 3 shows the relation between the character set and the code, from which it is seen that the collation sequence places the major control codes ahead of decimal digits, followed by the common symbols and letters.

The eighth channel available on 1" wide paper tape is unused by the tapewriter, and thus the 'delete', or 'erase' code corresponds to holes punched in all of the used channels.

Codes 03_8, 04_8 and 10_8 to 17_8 correspond to functions which affect the mode of operation of the tape-writer and will be used mainly for editing paper tapes but may be required as computer output codes for controlling printing formats. Code 75 is a special format code causing the tape to halt irrespective of the configuration of controls set by previous codes 03_8, 04_8, or 10_8 to 17_8.

8. LINE PRINTERS

Two types of line printer will be available. Two fast type 501's capable of printing 136-character lines at 1,000 lines per minute will be available at Canberra. Each subsidiary and Canberra will have two slower type 3253 line printers capable of printing 120-character lines at 300 lines per minute

Although the line printers are of two different types, it is intended that the character/code sets of all will be identical and will be as shown in Table 4, which relates the characters to their internal data processing codes.

Each printer contains a drum of type with sets of characters distributed around the periphery, one set for each print position. The positions of the characters around the drums are associated with a specially constructed decoding disc which associates each internal BCD code with the drum position of its corresponding character. Again, as with card readers and magnetic tape units, code conversion is carried out by hardware, namely by the appropriate code disc.

It will be seen that only 63 different printing characters are possible on either type of printer compared with the 90 available on the tapewriter (one of the 64 characters is 'space'). However, a simple correspondence between the two different sets is achieved allowing effective listing and easy reading, via line printing of program source statements etc. originally recorded on paper tape. No control codes exist and no case shifts are available on line printers, but it is possible to print a line with or without a line-feed function.

The effective cover of the tapewriter character set is obtained by allowing each line printer letter to substitute for both the corresponding upper and lower case tapewriter letters, thus leaving 38 positions available on the line printers for

allocation of the remaining 39 tape writer characters. The character which remains unallocated is the 'escape character ✧ which on tape refers to a component of the users program only.

Since it is possible to print a line without advancing the paper it is possible to overprint characters to form more complex characters; the organisation of overprinting being controlled by internal program. It will therefore be possible to generate all the standard ALGOL (and other) characters not directly available on the line printer/character set.

9. CONSOLE TYPEWRITERS

Console typewriters will be used for directives to and from the operator, consisting of only a small number of characters. In the interests of speed, longer messages to the operators such as diagnostic statements will be provided via the line printers. Thus it is not essential that the character/code set of the console typewriters correspond exactly to those of the tapewriters or line printers.

Table 5 shows the character/code set of the console typewriters for the 3600 and 3200 computers. It contains only upper case letters, decimal digit symbols and only those other symbols normally available on a standard typewriter. The code is that which actuates the printing hammers and conversion to it from internal code is arranged by program.

In particular all letters may be input via the keyboard in either upper or lower case setting, but the printer must be set to upper case for output of letters. All other characters will be printed according to the existing case setting and control

functions apply in either case. In the particular case of the 3200 monitor typewriter input to the processor occurs in upper case codes 13 and 04.

10. RELATION TO THE OTHER CHARACTER/CODE SETS

i) The ISO and ASCII Sets

The character set for the line printers (Table 3) is identical with the 6-bit code accepted by the ISO (International Standards Organisation) character/code set with options suited to scientific processing as shown in Table 6, and is a simple subset of the 7-bit code. The ASCII (American Standard Code for Information Interchange) character code set differs from the ISO set by the replacement of the characters ! ∨ ∧ , 0 ¬ _ by ! # & @ \ ← respectively (see 'The American Standard Code for Information Interchange', Datamation, August, 1963, September, 1963 and Communications of the A.C.M., August 1963, by R.W. Bemer). The tapewriter character/code set of table 3 bears a close relationship to the ISO set, the correspondence between the sectors of the codes being as follows:-

CODE GROUP IN ISO	CODE GROUP ON TAPEWRITER
00	00
01 to 07	upper case 21 to 27
10	lower case 73
11	lower case 40
12 to 17	lower case 12 to 17
20 to 31	lower case 20 to 31
32 to 73	upper case 32 to 73
74, 75	upper case 30, 31
76	no correspondence
77	upper case 20

The correspondence between the line printer character set and the ISO 6-bit set allows for easy use of the ISO 6-bit code internally in due course. Association between the code and printer characters will be arranged by substitution of the code discs by discs adapted to the ISO code.

ii) The ALGOL 60 Set

As shown in Sections 7 and 8 almost all the ALGOL 60 symbols in the reference language either exist on the tapewriter or line printer character sets, or may be constructed by simple juxtaposition and/or superposition of existing symbols. However, in three respects the tapewriter and line printers will differ from the reference notation, firstly lower case letters on the tapewriter are not of the usual kind but are capitals of small size, and only full size capitals are available on the line printers. Secondly, the formal multiplication sign in ALGOL is replaced by the asterisk. Thirdly, opening and closing quotes differ from the formal ALGOL equivalents; the opening quote “ is replaced by " and the closing quote ” replaced by ' on both line printers and tapewriters.

It is possible to distinguish delimiters represented formally by letter strings from identifiers by underlining the initial letter on both tapewriters and line printers and in the case of the latter equipment it will not be possible to distinguish between the cases of letters.

11. COMPOSITE CHARACTER BUILDING

Composite characters may be built up on both tapewriter and line printers by overprinting. In the case of paper tape preparation, the use which will most concern the user, the convention will be adopted that the first character placed on the tape will be the most commonly used in building composite characters. The order in which characters of a composite group are line printed is not so important but a similar rule is adopted in those samples below which do not contain letters.

The achievement of superposition of two characters on a line-printer will require first printing a line with no paper feed followed by a line, with paper feed, containing only the superposed characters in their appropriate places. Thus:

SYMBOL	TAPEWRITER	LINE PRINTER
≠	uc: bs lc -	: (nl) - (1)
≤	uc_ bs <	_ (nl) < (1)
≥	uc_ bs >	_ (nl) > (1)
≠	uc / bs =	/ (nl) = (1)
:=	uc:=	:= (1)
≡	uc_ bs =	_ (nl) = (1)
⊐	uc_ bs ¬	_ (nl) ¬ (1)

The tapewriter is assumed, at the commencement of construction of the composite character, to be in lower case (as in producing decimal data and FORTRAN statements) and lc, uc and bs denote 'shift to lower case', 'shift to upper case', and 'backspace' respectively. For the line printer (nl) and (1) indicate 'no line feed' and 'line feed' respectively, associated with the preceding character.

Allowance for accepting and printing composite characters is a matter of future programming, extension of which will be made in due course to allow also for such additional characters as: -

SYMBOL	TAPEWRITER	LINE PRINTER
$\not>$	uc / bs >	/ (nl) > (1)
$\not<$	uc / bs <	/ (nl) < (1)
$>>$	uc >>	>>
$<<$	uc <<	<<
$\approx$	uc : bs =	: (nl) = (1)

Those symbols in ALGOL which consist of letter strings may be distinguished from identifiers by underlining the first letter, e.g.:

SYMBOL	TAPEWRITER	LINE PRINTER
INTEGER	uc_ bs INTEGER	INTEGER (nl) ^^^^^^^ (1)

where ^ denotes 'space'

Extension to more complex composite characters is possible.

12. PLOTTER & DISPLAY

On both plotters at all sites, and the display unit at Canberra, it will be possible to 'write' alphanumeric character strings. The character/code set for both types of device are compatible with that for the line printer.* Thus, the devices will be 'fed with internal BCD codes and will 'write' the corresponding characters from the set in Table 4.

* the character % is omitted in the case of the display, the code being allocated to a special function.

The display unit possesses an internal buffer store into which the BCD character string is placed and is automatically and continuously scanned without further control by the 3600. The plotters however, are controlled in detail by a set of 64 subroutines each corresponding to a particular character to be written.

T. Pearcey

23/4/64

TABLE 1 : Magnetic Tape External BCD Codes.

A \ B	00	10	20	30	40	50	60	70
0	:	8	Space	Y	-	Q	+	H
1	1	9	/	Z	J	R	A	I
2	2	0	S	]	K	V	B	<
3	3	=	T	,	L	$	C	.
4	4	'	U	(	M	*	D	)
5	5	¬	V	10	N	↑	E	?
6	6	%	W	_	O	"	F	|
7	7	[	X	∧	P	>	G	;

Note : on magnetic tapes this code is recorded with a 7th bit/character so as to maintain even parity.

TABLE 2 : Card punch character/codes and related internal BCD* codes.

Zone Punch / Digit Punch	null		12		11		0	
	ch	BCD	ch	BCD	ch	BCD	ch	BCD
null	Space	60	+	20	-	40	0	00
1	1	01	A	21	J	41	/	61
2	2	02	B	22	K	42	S	62
3	3	03	C	23	L	43	T	63
4	4	04	D	24	M	44	U	64
5	5	05	E	25	N	45	V	65
6	6	06	F	26	O	46	W	66
7	7	07	G	27	P	47	X	67
8	8	10	H	30	Q	50	Y	70
9	9	11	I	31	R	51	Z	71
8 and 3	=	13	.	33	$	53	,	73
8 and 4	'	14	)	34	*	54	(	74

* BCD codes are in octal notation and are obtained from the card code by adding the following weightings for the zone and multiple punchings :

Zone	weighting (octal)
null	zero
12	20
11	40
0	60

In case of triple punchings add weight 10 for digit 8.

TABLE 3 : Tapewriter character/code set.

B / A	00	10	20		30		40		50		60		70	
			L	U	L	U	L	U	L	U	L	U	L	U
0	Space	Switch Readers	0	_	8	¬	)	10	H	H	P	P	x	X
1	Tabulate	Punch On	1	|	9	]	A	A	I	I	Q	Q	y	Y
2	New line	Restore Print	2	"	*	:	B	B	J	J	R	R	z	Z
3	Code Control On	Skip On	3	∨	+	;	C	C	K	K	S	S	(	[
4	Code Control Off	Reader Off	4	$	,	<	D	D	L	L	T	T	Special Skip Off	
5	Backspace	Punch Off	5	%	−	=	E	E	M	M	U	U	Master Stop	
6	Upper Case	Suppress Print	6	∧	.	>	F	F	N	N	V	V	↑	Escape ✧
7	Lower Case	Skip Off	7	'	/	?	G	G	O	O	W	W	Delete	

Notes 1) Octal code = B + A
2) This code is recorded on paper tape with a 7^{th} bit/character so as to maintain odd parity.
3) "L" and "U" mean "lower" and "upper" case respectively.

Tape format

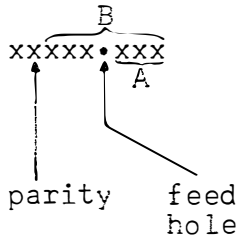

TABLE 4 : Line printer characters and internal BCD codes.

A \ B	00	10	20	30	40	50	60	70
0	0	8	+	H	−	Q	Space	Y
1	1	9	A	I	J	R	/	Z
2	2	:	B	<	K	∨	S	]
3	3	=	C	.	L	$	T	,
4	4	′	D	)	M	*	U	(
5	5	¬	E	?	N	↑	V	10
6	6	%	F	|	O	′	W	_
7	7	[	G	;	P	>	X	∧

TABLE 5 : Monitor typewriter character/code set.

B / A	00		10		20		30		40		50		60	70	
	L	U	L	U	L	U	L	U	L	U	L	U		L	U
0			Q	Q			E	E	Y	Y	;	:	Space	J	J
1	B	B	C	C			N	N	L	L	E	E	Backspace	X	X
2	W	W	A	A			.	.	O	O	'	"	Tabulate	M	M
3			6	¢			3	#	0	)	7	&	New line	Z	Z
4	-	—	,	,			=	+	/	?	P	P	Upper Case	G	G
5			K	K			U	U	H	H	D	D		T	T
6			I	I			V	V	S	S	R	R	Lower Case	!	°
7	9	(		*				@	4	$		%		1	±

TABLE 6 : The 6-bit ISO character/code set.

A \ B	00	10	20	30	40	50	60	70
0	Space	(	0	8	10	H	P	X
1	|	)	1	9	A	I	Q	Y
2	"	*	2	:	B	J	R	Z
3	V	+	3	;	C	K	S	[
4	$	,	4	<	D	L	T	¬
5	%	-	5	=	E	M	U	]
6	∧	.	6	>	F	N	V	↑
7	'	/	7	?	G	O	W	_

APPENDIX

Internal 00_8 and External 12_8

The only exceptions to the rule for conversion from internal (core) BCD codes to external BCD codes, (that if the 5 bit of a 6-bit BCD code transferred is unity, then the 6th bit is complemented) are the codes 00_8 and 12_8. The conversions are as follows :-

1) Magnetic Tape

   i) Internal (in core) 00_8 becomes external 12_8 (on tape)

   ii) Internal 12_8 becomes external 12_8 (i.e. unchanged).

   This ensures that no <u>even parity</u> code 00_8 occurs on the tape since such a code would appear as an unclocked gap on the tape and pass the reading head undetected.

   External 12_8 (on tape) becomes internal 00_8 (in core) so that a code originally appearing in core as 12_8 becomes 00_8 in core after transfer <u>to and from</u> tape.

2) Other Peripherals

   Peripherals other than magnetic tape units and paper tape and card punches convert from internal to external BCD with the exception that internal 00_8 becomes external 12_8 and internal 12_8 becomes external 00_8, the former for instance printing a 'zero' and the latter a 'colon'.
   When it is required to preserve <u>all</u> codes on transfer to and from magnetic tape, they must be transferred in the binary mode which occurs with odd parity only on tape.