



INTERNATIONAL TELECOMMUNICATION UNION

CCITT

THE INTERNATIONAL
TELEGRAPH AND TELEPHONE
CONSULTATIVE COMMITTEE

T.51

(09/92)

TERMINAL EQUIPMENT AND PROTOCOLS FOR TELEMATIC SERVICES

LATIN BASED CODED CHARACTER SETS FOR TELEMATIC SERVICES



Recommendation T.51

FOREWORD

The CCITT (the International Telegraph and Telephone Consultative Committee) is a permanent organ of the International Telecommunication Union (ITU). CCITT is responsible for studying technical, operating and tariff questions and issuing Recommendations on them with a view to standardizing telecommunications on a worldwide basis.

The Plenary Assembly of CCITT which meets every four years, establishes the topics for study and approves Recommendations prepared by its Study Groups. The approval of Recommendations by the members of CCITT between Plenary Assemblies is covered by the procedure laid down in CCITT Resolution No. 2 (Melbourne, 1988).

Recommendation revised T.51 was revised by Study Group VIII and was approved under the Resolution No. 2 procedure on the 18 September 1992.

CCITT NOTES

- 1) In this Recommendation, the expression "Administration" is used for conciseness to indicate both a telecommunication administration and a recognized private operating agency.
- 2) A list of abbreviations used in this Recommendation can be found in Annex C.

© ITU 1993

All rights reserved. No part of this publication may be reproduced or utilized in any form or by any means, electronic or mechanical, including photocopying and microfilm, without permission in writing from the ITU.

Recommendation T.51**LATIN BASED CODED CHARACTER SETS FOR TELEMATIC SERVICES**

*(Malaga-Torremolinos, 1984: Amended at Melbourne 1988;
revised 1992)*

1 Scope**1.1 The CCITT,**

considering

(a) the increasing interdependence of the various CCITT character sets and coding schemes in various telematic services;

(b) the introduction of new facilities such as code conversion and interworking between various telematic services;

(c) the advantage of having a single unified repertoire and coding of Latin based character set in a Recommendation to act as a reference for the telematic services;

(d) that Recommendations T.60/T.61 and T.100/T.101 define the character coding systems for teletex and videotex;

(e) that Recommendation T.50 specifies the International Reference Version (IRV) of the 7-bit coded character set,

provides the following Recommendation as a reference document

towards which the Latin based portion of the coded character sets of telematic services should migrate and from which coded character subsets and elements of code extension mechanisms can be derived for individual telematic services.

1.2 This Recommendation specifies a primary set and a 96-character supplementary set of graphic characters. When various telematic services restrict their primary and supplementary sets to be respective subsets of those given in this Recommendation, it will be ensured that no code position in any of the specified code tables is assigned more than one meaning within different telematic services. 94-character subsets of the supplementary code table can be found in Recommendations of specific telematic services,

1.3 This Recommendation gives the escape sequences for designating the primary and supplementary sets of graphic characters, to be used according to the code extension techniques specified.

1.4 Non-Latin based character sets are to be dealt with in Recommendation T.52.

1.5 This Recommendation describes those code extension mechanisms that are relevant to existing telematic services. Additional mechanisms will be included in this Recommendation as the need for such is identified for one or more telematic services. The purpose of this Recommendation is to include an up-to-coding systems in various telematic services.

1.6 In this Recommendation 7-bit code tables are described which can be used either in a 7-bit or in an 8-bit environment, with applicable code extension mechanisms that are given in other Recommendations specific to given telematic services.

1.7 This Recommendation gives a unified superset of the repertoire of Latin based alphanumeric characters (see Annex A).



Documents

Document Sharing Hub

Recommendation T.51 (09/92)

1

1.8 This Recommendation gives a table of character and control sets used in CCITT telematic services (see Annex B).

1.9 There is no conformance clause in this Recommendation specifying the mandatory and optional subsets of code extension mechanisms and coded character sets. Conformance requirements will be the subject of other CCITT Recommendations specific to particular telematic services.

1.10 The T51String is defined in Annex D.



2 Graphic character sets

2.1 Primary set

2.1.1 The primary set of graphic characters specified in Figure 1/T.51 is identical with the set of graphic characters of the International Reference Version (IRV) of the 7-bit coded character set of Recommendation T.50.

2.1.2 The primary set is designated as G0 by the sequence ESC 2/8 4/2. It can also be alternatively designated as G1, G2 or G3 by the sequences ESC 2/9 4/2, ESC 2/10 4/2 or ESC 2/11 4/2 respectively. See § 3 for details on code extension techniques.

Terminals used for telematic services which make reference to the 1988 version of Recommendation T.51 use, for the designation of the primary set as G0, the sequence ESC 2/8 4/0 and alternatively as G1, G2 and G3, the sequences ESC 2/9 4/0, ESC 2/10 4/0, ESC 2/11 4/0.

2.2 Supplementary set

2.2.1 The supplementary set of graphic characters is specified in Figure 2/T.51.

2.2.2 Unallocated code positions are subject to future standardization and will be allocated when a need for such is identified.

2.2.3 The supplementary set is designated as G2 by the sequence ESC 2/14 5/2. It can be alternatively designated as G1 or G3 by the sequences ESC 2/13 5/2 or ESC 2/15 5/2 respectively.

Terminals used for telematic services which make reference to the 1988 version of Recommendation T.51 use, for the designation of the supplementary set as G2, the sequence ESC 2/10 6/2 and alternatively as G0, G1 and G3, the sequences ESC 2/8 6/2, ESC 2/9 6/2, ESC 2/11 6/2.

2.2.4 Notes on the primary and supplementary sets of graphic characters for Figures 1/T.51 and 2/T.51

In the figures the number of the note being referred to is encircled.

Note 1 – All the characters in column 4 of the supplementary set are non-spacing characters. They are all diacritical marks.

Note 2 – Cross-shaded code positions are reserved for future standardization by the CCITT.

Note 3 – Terminals used for current ITU-T defined telematic services may send and receive the codes 2/6 and 2/4 of the supplementary set for the NUMBER SIGN and DOLLAR SIGN, respectively. When receiving codes 2/3 and 2/4 from the primary set of graphic characters, terminals may interpret them as # and ¤ respectively. Future applications in telematic services should code the NUMBER SIGN, CURRENCY SIGN and DOLLAR SIGN in accordance with Figures 1/T.51 and 2/T.51.



Note 4 – Terminals used for ITU-T defined telematic services should send only the codes 4/1 of the supplementary set followed by SPACE for a stand-alone grave accent, 4/3 of the supplementary set followed by SPACE for a stand-alone circumflex accent, and 4/4 of the supplementary set followed by SPACE for a stand-alone tilde. Whenever a telematic terminal is capable of receiving and interpreting codes 6/0, 5/14 and 7/14 from the primary set of graphic characters, terminals shall interpret them as GRAVE, CIRCUMFLEX and TILDE respectively.



Note 5 – This code position is reserved and shall not be used.

Note 6 – Current telematic services may interpret this as the non-spacing underline. The non-spacing underline character is never used individually but always in combination with some other graphic character to represent the graphic rendition “underlined” for the associated character. The non-spacing underline character can be used in combination with any graphic character of the repertoire, including an accented letter or an umlaut, or space. It is recommended to implement the “underline” function by means of the control function SGR(4) instead of the “non-spacing underline” graphic character. However, both must be correctly interpreted when received.

					b ₇	0	0	0	0	1	1	1	1
					b ₆	0	0	1	1	0	0	1	1
					b ₅	0	1	0	1	0	1	0	1
						0	1	2	3	4	5	6	7
b ₄	b ₃	b ₂	b ₁										
0	0	0	0	0				SP	0	@	P	④	p
0	0	0	1	1				!	1	A	Q	a	q
0	0	1	0	2				"	2	B	R	b	r
0	0	1	1	3				# _③	3	C	S	c	s
0	1	0	0	4				\$ _③	4	D	T	d	t
0	1	0	1	5				%	5	E	U	e	u
0	1	1	0	6				&	6	F	V	f	v
0	1	1	1	7				'	7	G	W	g	w
1	0	0	0	8				(	8	H	X	h	x
1	0	0	1	9				)	9	I	Y	i	y
1	0	1	0	10				*	:	J	Z	j	z
1	0	1	1	11				+	;	K	[	k	{
1	1	0	0	12				,	<	L	\	l	
1	1	0	1	13				—	=	M	]	m	}
1	1	1	0	14				.	>	N	④	n	④
1	1	1	1	15				/	?	O	—	o	

T0811620-93

Note – Notes to this figure are contained in § 2.2.4.

FIGURE 1/T.51

The primary set of graphic characters for telematic services
(coded representation when invoked in columns 2-7 of the code table)

1

					b ₇	0	0	0	0	1	1	1	1
					b ₆	0	0	1	1	0	0	1	1
					b ₅	0	1	0	1	0	1	0	1
						0	1	2	3	4	5	6	7
b ₄	b ₃	b ₂	b ₁										
0	0	0	0	0				NBSP	°	5	—	Ω	K
0	0	0	1	1				i	±	`	1	Æ	æ
0	0	1	0	2				¢	2	'	®	Ð	ð
0	0	1	1	3				£	3	^	©	ä	ä
0	1	0	0	4				©	x	~	™	Ë	h
0	1	0	1	5				¥	μ	-	♪	2	1
0	1	1	0	6				©	¶	˘	¬	IJ	ij
0	1	1	1	7				\$	.	.		L	l
1	0	0	0	8				□	÷	..	2	Ł	ł
1	0	0	1	9				'	'	5	2	Ø	ø
1	0	1	0	10				"	"	o	2	Œ	œ
1	0	1	1	11				«	»	„	2	Œ	œ
1	1	0	0	12				←	¼	6	⅛	Þ	þ
1	1	0	1	13				↑	½	"	¾	ƒ	ƒ
1	1	1	0	14				→	¾	„	⅝	N	η
1	1	1	1	15				↓	¿	˘	⅞	'n	SHY

T0811630-93

Note – Notes to this figure are contained in § 2.2.4.

FIGURE 2/T.51

The supplementary set of graphic characters for telematic services
(coded representation when invoked in columns 2-7 of the code table)

3 Code extension technique

3.1 General

3.1.1 Code extension techniques are required for the designation of various graphic or control character sets and their invocation in the 7-bit set or 8-bit set in use. Such techniques are derived from ISO Standard 2022.

3.1.2 This Recommendation describes only those code extension techniques currently specified for existing telematic services. Additional techniques will be further incorporated as they are identified for use in one or more telematic services.

3.2 Definitions

For the purpose of code extension techniques given in this Recommendation, the following definitions apply.

3.2.1 bit combination

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

3.2.2 byte

A bit string that is operated upon as a unit and the size of which is independent of redundancy or framing techniques.

3.2.3 character

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

3.2.4 coded character set; code

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

3.2.5 code extension

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

3.2.6 code table

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

3.2.7 control character

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

3.2.8 control function

An action that affects the recording, processing, transmission or interpretation of data and that has a coded representation consisting of one or more bit combinations.

3.2.9 to designate

To identify a set of characters that are to be represented, in some cases immediately and in others on the occurrence of a further control function, in a prescribed manner.

3.2.10 environment

The characteristic that identifies the number of bits used to represent a character in a data processing or data communication system or in part of such a system.



3.2.11 escape sequence

A bit string that is used for control purposes in code extension procedures and that consists of two or more bit combinations. The first of these bit combinations represents the character ESCAPE (1/11).

3.2.12 final character

The character the bit combination of which terminates an escape sequence.

3.2.13 graphic character

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

3.2.14 intermediate character

A character the bit combination of which occurs between that of the ESCAPE character and that of the Final character in an escape sequence consisting of more than two bit combinations.

3.2.15 to invoke

To cause a designated set of characters to be represented by the prescribed bit combinations whenever those bit combinations occur, until an appropriate code extension function occurs.

3.2.16 position

That part of a code table identified by its column and row coordinates.

3.2.17 to represent

- a) to use a prescribed bit combination with the meaning of a character in a set of characters that has been designated and invoked; or
- b) to use an escape sequence with the meaning of an additional control function.

3.2.18 Repertoire

A specified set of characters that are represented by the combination of a coded character set.

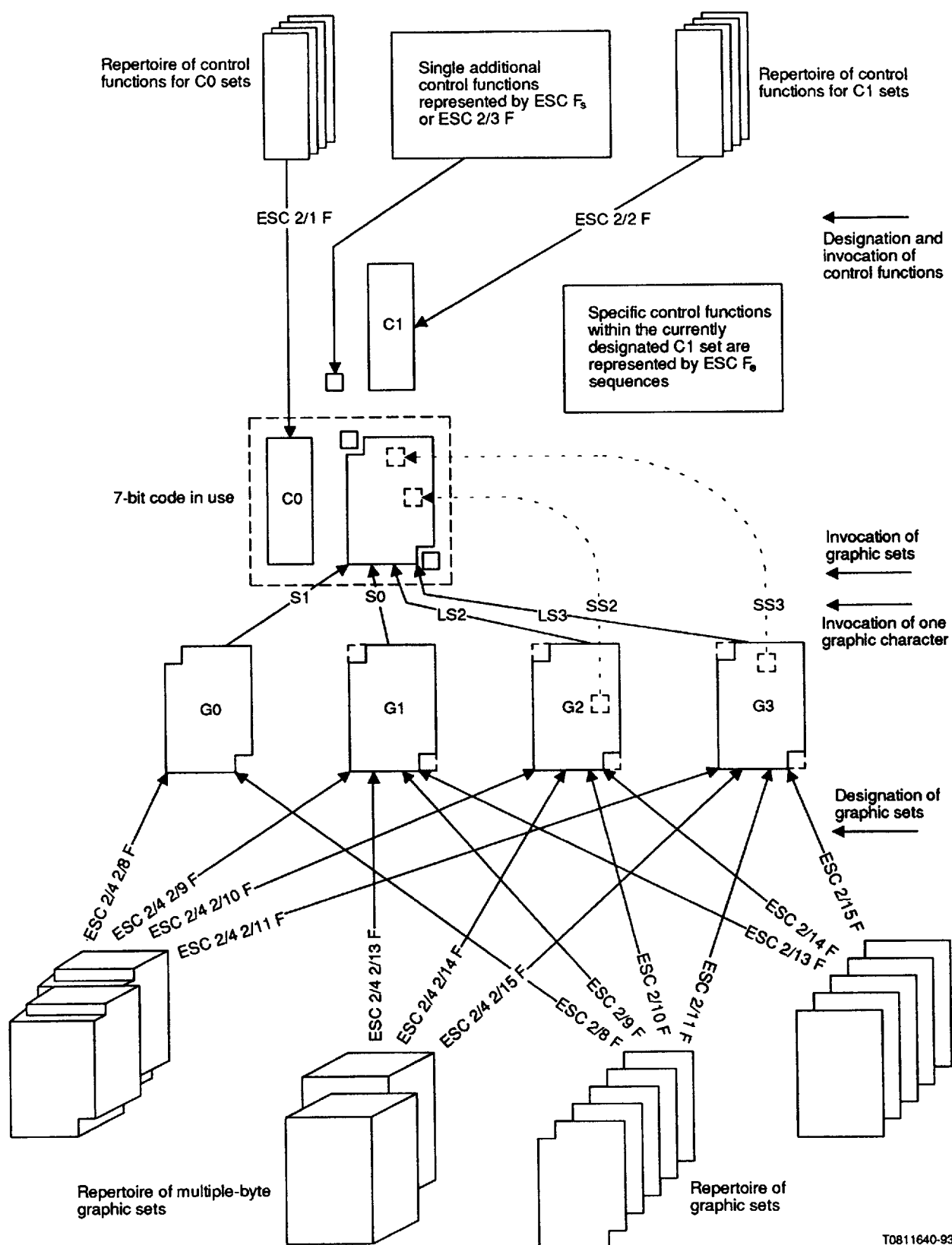
**3.3 Code extension facilities**

These are depicted in Figure 3/T.51 for the 7-bit environment and Figure 4/T.51 for the 8-bit environment. They include the following functions:

- a) designation and invocation of control sets C0 and C1 by means of the relevant escape sequences given in § 3.4;
- b) designation of a graphic character set G0 by means of the relevant escape sequence given in § 3.4;
- c) designation of up to three additional G-sets called G1, G2 and G3 by means of the relevant escape sequences given in § 3.4;
- d) invocation of the designated graphic sets, by means of locking and/or non-locking shift functions, given in § 3.5;
- e) designation and invocation of a complete code by means of the relevant escape sequence given in § 3.4.

3.4 Types of character sets

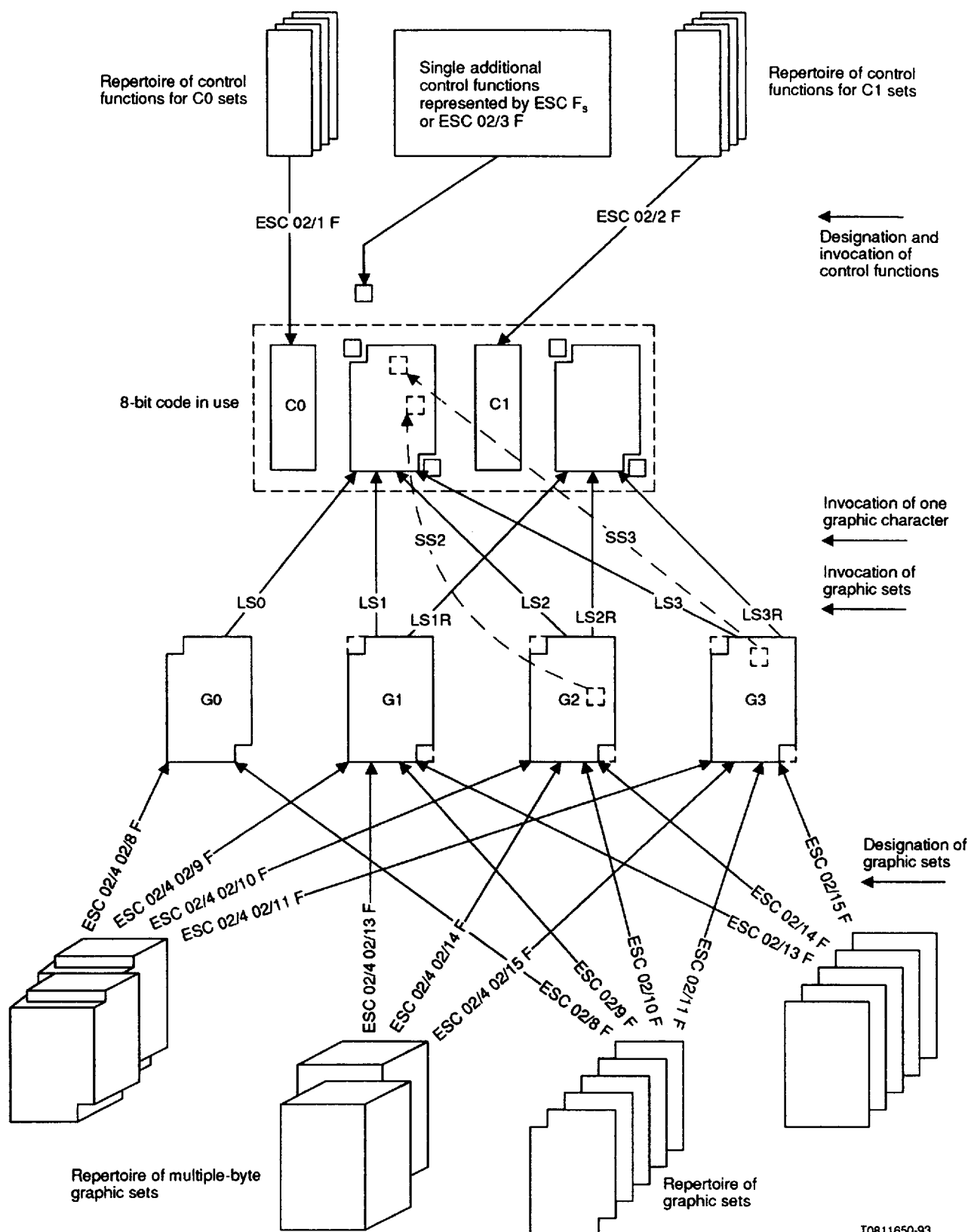
There are a number of different types of control and graphic character sets that can be designated and invoked for use in the 7-bit or 8-bit environment. These are listed in Table 3/T.51 and defined below. A given control or graphic character set can be designated by an escape sequence terminated by a Final character F from bit combination 4/0 to 7/14, specific to the character set to be designated. Final characters are allocated by ISO and registered in the ISO "International register of coded character sets to be used with escape sequences".



T0811640-93

FIGURE 3/T.51
Code extension in a 7-bit environment
(showing all shift functions)

3.4.1 C0 set of 32 control characters (bit combinations 0/0 to 1/15). A C0 set is designated and invoked by the sequence ESC 2/1 F, where F identifies a registered C0 set.



T0811650-93

FIGURE 4/T.51
Code extension in an 8-bit environment
(showing all shift facilities)

3.4.2 C1 set of 32 control characters (bit combinations 8/0 to 9/15 in an 8-bit environment, or ESC 4/0 to ESC 5/15 in a 7-bit environment). A C1 set is designated and invoked by the sequence ESC 2/2 F, where F identifies a registered C1 set.

3.4.3 G0 set of 94 graphic characters (bit combinations 2/1 to 7/14).

3.4.4 G1, G2 and G3 sets of 94 graphic characters (bit combinations 2/1 to 7/14 or 10/1 to 15/14). A registered set of 94 graphic characters can be designated by the escape sequence ESC 2/8 F, ESC 2/9 F, ESC 2/10 F or ESC 2/11 F in order to be used as a G0, G1, G2 or G3 set respectively, where F identifies the designated set.

3.4.5 G1, G2 and G3 sets of 96 graphic characters (bit combinations 2/0 to 7/15 or 10/0 to 15/15). A registered set of 96 graphic characters can be designated by the escape sequence ESC 2/13 F, ESC 2/14 F or ESC 2/15 F in order to be used as a G1, G2 or G3 set respectively, where F identifies the designated set.

3.4.6 Multiple-byte G0 set of more than 94 graphic characters each represented by more than one bit combinations from 2/1 to 7/14. The designation sequence is ESC 2/4 2/8 F.

3.4.7 Multiple-byte G1, G2 and G3 sets of more than 94 graphic characters each represented by more than one bit combination from 2/1 to 7/14 or from 10/1 to 15/14. A registered multiple-byte set of more than 94 graphic characters can be designated by the escape sequence ESC 2/4 F, ESC 2/4 2/9 F, ESC 2/4 2/10 F or ESC 2/4 2/11 F in order to be used as a G0, G1, G2 or G3 set respectively, where F identifies the designated set.

3.4.8 Multiple-byte G1, G2 and G3 sets of more than 96 graphic characters each represented by more than one bit combination from 2/0 to 7/15 or from 10/0 to 15/15. A registered multiple-byte set of this type identified by a Final character F can be designated by the escape sequence ESC 2/4 2/13 F, ESC 2/4 2/14 F or ESC 2/4 2/15 F in order to be used as G1, G2 or G3 set respectively.

3.4.9 Complete code containing all bit combinations 0/0 to 7/15 in 7-bit environment, or 0/0 to 15/15 in 8-bit environment. A complete code identified by a Final character F, can be designated and invoked by the escape sequence ESC 2/5 F.

Note – Complete code is now referred to in ISO 2022 as “Coding System different from that of ISO 2022”.

3.5 *Invocation functions*

Following the designation of a G0, G1, G2 or G3 set as specified in § 3.4, any one of these sets will require invocation into the 7-bit or 8-bit in-use code table. This is performed by the use of either locking-shift functions or non-locking shift functions listed in Table 4/T.51. Coding for these functions is given in Table 5/T.51.

3.5.1 *Use of locking-shift functions*

There are seven locking-shift functions as given in Tables 4/T.51 and 5/T.51. A locking-shift function invokes the relevant G0, G1, G2 or G3 set into columns 2 to 7 or into columns 10 to 15 (8-bit code only) in order to replace the previously invoked G-set. The occurrence of a locking-shift function shall not affect those bit combinations that are included in any escape sequence or those that follow single-shift function SS2 or SS3. If a particular set has already been invoked, use of the corresponding locking-shift function has no effect.

3.5.2 *Use of single-shift functions*

The single-shift function SS2 shall invoke one character from the last designated G2 set. Similarly SS3 shall invoke one character from the last designated G3 set. In the case of using single-byte graphic sets, these invocations alter the meaning of the immediately following bit combination only and ascribe to it the meaning of the corresponding bit combination of the G2 or G3 set. The bit combination permitted to follow SS2 or SS3 is limited to one from columns 2 to 7. If a single-shift function is used to invoke a character from a multiple-byte set, the shift function will affect two or more bit-combinations to represent that character. The use of a single-shift function does not affect the current shift status established by a locking-shift function.

TABLE 3/T.51

Types of character sets

Description	Character set	Designation sequence	Bit combinations	
			7-bit environment	8-bit environment
Sets of 32 control characters (Note 1)	C0 C1	ESC 2/1 F ESC 2/2 F	0/0 to 1/15 ESC 4/1 to ESC 5/15	00/00 to 01/15 08/00 to 09/15
Sets of 94 graphic characters	G0	ESC 2/8 F	2/1 to 7/14	02/01 to 07/14
	G1 G2 G3	ESC 2/9 F ESC 2/10 F ESC 2/11	2/1 to 7/14	02/01 to 07/14 or 10/01 to 15/14
	G1 G2 G3	ESC 2/13 F ESC 2/14 F ESC 2/15 F	2/0 to 7/15	02/00 to 07/15 or 10/00 to 15/15
Sets of more than 94 graphic characters each represented by more than one byte	G0	ESC 2/4 F (Note 2) ESC 2/4 2/8 F	2/1 to 7/14	02/01 to 07/14
	G1 G2 G3	ESC 2/4 2/9 F ESC 2/4 2/10 F ESC 2/4 2/11 F	2/1 to 7/14	02/01 to 07/14 or 10/01 to 15/14
	G1 G2 G3	ESC 2/4 2/13 F ESC 2/4 2/14 F ESC 2/4 2/15 F	2/0 to 7/15	02/00 to 07/15 or 10/00 to 15/15
Complete code		ESC 2/5 F	0/0 to 7/15	00/00 to 15/15

Note 1 – Control sets C0 and C1, and complete codes are simultaneously designated and invoked by the relevant escape sequences indicated.

Note 2 – Multi-byte character sets are to be designated by four-character escape sequences. However, the following three-character escape sequences have in the past been registered and their use is valid for the registered multi-byte character sets:

ESC 2/4 4/0;
ESC 2/4 4/1;
ESC 2/4 4/2.

TABLE 4/T.51

Allocation of shift functions to the graphic character sets to be invoked

A set	Locking-shift functions		Non-locking shift functions
	Columns 2 to 7 of 7-bit or 8-bit code	Columns 10 to 15 of 8-bit code	Columns 2 to 7 of 7-bit or 8bit code
G0	SI(7-bit), LS0(8-bit)	—	—
G1	SO(7-bit), LS1(8-bit)	LS1R	—
G2	LS2	LS2R	SS2
G3	LS3	LS3R	SS3

TABLE 5/T.51

Coding for shift functions

Shift functions		Coding
Single-shift two	SS2	1/9
Single-shift three	SS3	1/13
Shift in SI(7-bit), locking-shift zero	LS0(8-bit)	0/15
Shift out SO(7-bit), locking-shift one	LS1(8-bit)	0/14
Locking-shift one right	LS1R	ESC 7/14
Locking-shift two	LS2	ESC 6/14
Locking-shift two right	LS2R	ESC 7/13
Locking-shift three	LS3	ESC 6/15
Locking-shift three right	LS3R	ESC 7/12