UDC 003.62:681.325.3.048

JIS

This standard was revised in 6, 1995

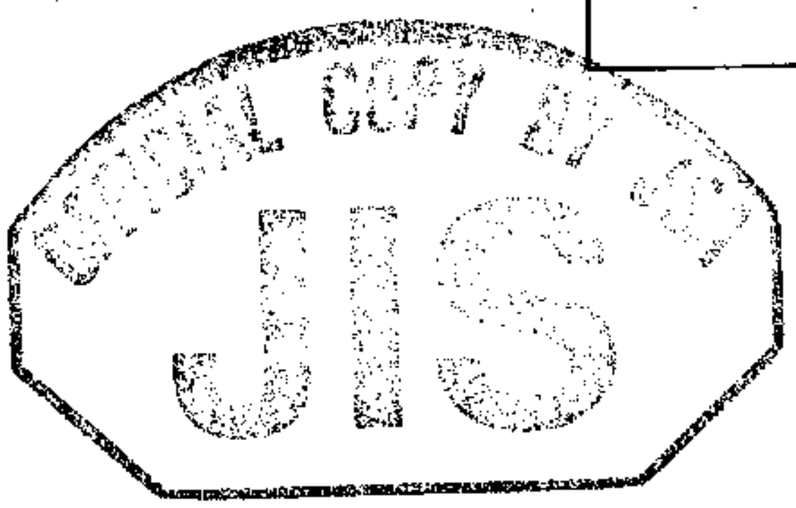
JAPANESE INDUSTRIAL STANDARD

Code Extension Techniques for Use with the Code for Information Interchange

JIS ~~C 6228~~ —1984

The code number of this standard JIS C 6228 changed to the following on 1 March 1987.

JIS X 0202-1984



This copy is an English version of Japanese Industrial Standard prepared by Japanese Standards Association, the copyright holder. Any further reproduction without permission is prohibited.

Translated and Published

by

Japanese Standards Association

Printed

20 S

Documents
Document Sharing Hub

Full document details: JIS X 0202-1984 (1995) JIS C 6228-1984

Translation without guarantee
In the event of any doubt arising, the original
standard in Japanese is to be evidence

Full document details: JIS X 0202-1984 (1995) JIS C 6228-1984

UDC 003.62:681.325.3.048

JAPANESE INDUSTRIAL STANDARD

J I S

Code Extension Techniques for Use with the Code for Information Interchange

X 0202-1984

1. Scope

1.1 Code Extension Techniques This Japanese Industrial Standard specifies method of extending the code for information interchange in a 7-bit environment and in an 8-bit environment. These code extension techniques are described in four inter-related clauses dealing respectively with:

(1) The extension of the 7-bit code remaining in a 7-bit environment,

(2) The structure of a family of 8-bit codes,

(3) The extension of an 8-bit code remaining in an 8-bit environment,

(4) The relation between the 7-bit code and an 8-bit code.

1.2 Extension Principle In this Japanese Industrial Standard code extension techniques are classified and the structures are specified.

The principles established in this Japanese Industrial Standard may be utilized to form code extension facilities other than those specified in this standard.

Reference: ISO 6429 has followed such extension principle to formulate some parameterized control function. Assignments of bit combinations associated with the designation of the classes are made in accordance with ISO 2375. Specific assignments of bit combinations to relate individual codes with their invocation or designation are also to be made in accordance with ISO 2375.

1.3 Sequential Process The code extension techniques are designed to be used for data to be processed serially in a forward direction. Use of these techniques in strings of data which are processed other than serially in a forward direction or included in data formatted for fixed-length record processing may have undesirable results or may require additional special treatment to ensure correct interpretation.

Applicable Standards, Corresponding International Standard and Reference Standards: See page 34.

Documents
Document Sharing Hub

Full document details: JIS X 0202-1984 (1995) JIS C 6228-1984

1.4 Conformance This Japanese Industrial Standard addresses whole classes of provisions, and it is not intended that they are all implemented in all instances.

To comply with this standard, the following is required:

(1) When code extension techniques are used, the applicable parts of this standard shall be followed:

(2) When two systems with different levels of implementation of code extension techniques are required to communicate with one another, they shall do so using the code extension techniques they have in common.

Remarks 1. Full conformance to a standard means that all its requirements are met. For such conformance to be unique the standard must contain no options. This is typically the case for hardware standards. This standard is of a different nature and as a result, it is only practicable to envisage limited conformance to it, as defined above.

2. The class means each code extension technique when the code extension techniques specified in this standard are classified, and it is identified by the first intermediate character and the second intermediate character.

2. Field of Application

The 7-bit code of JIS C 6220 allows the representation of up to 128 characters. Additionally, that standard allows the representation of other graphic characters by the combination of two graphic characters with the control character BS. In some application fields, the code of JIS C 6220 lacks sufficient control functions or graphic characters to satisfy the needs of the application.

Such lack may be supplemented by means of the code extension techniques specified in this standard.

This standard presents a review of the salient structure of the 7-bit code and then builds up that structure to describe various means of extending the control function and graphic sets of the code. It also describes structures and techniques to construct and formalize codes related to the 7-bit code. These related codes are structured so as to allow application-dependent usage without preventing the interchangeability of data employing them. It describes:

The structure of the 7-bit code;

The extension of the 7-bit code, remaining in a 7-bit environment and making use of code extension techniques;

The structure of a family 8-bit codes, remaining compatible with the 7-bit structure; and

The extension of an 8-bit code, remaining in an 8-bit environment, and making use of code extension techniques.

Documents
Document Sharing Hub

Full document details: JIS X 0202-1984 (1995) JIS C 6228-1984

In order to use identical extension techniques in each of the above cases, and to facilitate conversion between them, standard rules for code extension are necessary. This has the advantage of:

Reducing the risk of conflict between systems required to inter-operate;

Permitting provision for code extension in the design of systems;

Providing standardized methods of calling into use agreed sets of characters;

Allowing the interchange of data between 7-bit environment and 8-bit environment.

This standard also describes the structure of families of codes which are related to the code of JIS C 6220 by their structure.

3. Definitions

For the purpose of this standard, in addition to the definitions specified in JIS C 6230, the following definitions apply:

(1) bit combination An ordered set of bits used for the representation of characters.

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

Reference: Characters are classified into the graphic character of (13) and the control character of (7). A character is represented by a combination of the bits of (1).

(4) coded character set or code A sit of unambiguous rules that establishes a character set and the one-to-one relationship between the characters of the set and their bit combinations.

(5) code extension The technique for encoding of characters that are not included in the character set of a given code.

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

(7) control character a control function the coded representation of which consists of a single bit combination.

(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.

(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.

Documents
Document Sharing Hub

Full document details: JIS X 0202-1984 (1995) JIS C 6228-1984

(10) environment The characteristic that identifies the number of bits used to represent a character in a data processing or data communication system.

(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 ESC.

(12) final character The character the bit combination of which terminates and escape sequence.

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

(14) intermediate character A character the bit combination of which occurs between that of the ESC character and that of the final character in an escape sequence consisting of more than two bit combinations.

(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.

(16) position That part of a code table identified by its column and row coordinates.

(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.

(18) version of the 7-bit code A code table in which the positions corresponding to #, $, @, [¥], ^, _, {‾} and ‾ among the 7-bit code of Roman letters specified in JIS C 6220 are optional and the same characters as those of JIS C 6220 or other characters are allocated to such optional positions, or such positions are declared unused.

(19) version of the 8-bit code A code table in which the positions corresponding to #, $, @, [¥], ^, _, {‾}, ‾ and the columns of 10 to 15 (positions of 10/0 and 15/15 are excluded) are optional, and the same characters as those of JIS C 6220 or other characters are allocated to such optional positions, or such positions are declared unused.

Documents
Document Sharing Hub

Full document details: JIS X 0202-1984 (1995) JIS C 6228-1984

4. Notation

In this standard the following notations are used:

For bits of a 7-bit combination	-	b_7	b_6	b_5	b_4	b_3	b_2	b_1
For bits of an 8-bit combination	b_8	b_7	b_6	b_5	b_4	b_3	b_2	b_1
Bit weights for column and row reference	2^3	2^2	2^1	2^0	2^3	2^2	2^1	2^0
	Column				Row			

A bit combination is sometimes referred to by the column and row numbers of its position in the code table. The column number is the decimal equivalent of bits b_7 to b_5 (or b_8 to b_5) and the row number is the decimal equivalent of bits b_4 to b_1, giving to those bits the weights shown above. The column and row numbers are separated by a solidus.

In representing the decimal equivalents, the convention is to append a leading zero to the column number for 8-bit columns 00 to 09. As an example, the position of SP in the 7-bit code table is 2/0, the position of the same character in an 8-bit code table in 02/0.

5. Extension of the 7-bit Code Remaining in a 7-bit Environment

5.1 Introduction

5.1.1 Structure of the 7-bit Code The 7-bit code table used in this standard is the basis of code extension techniques for use with the 7-bit coded character set of JIS C 6220, consists of areas for an ordered set of control characters and graphic characters and grouped as given below. The grouping is illustrated in Figure 1.

(1) The area for a set of 32 control characters allocated to column 0 and 1.

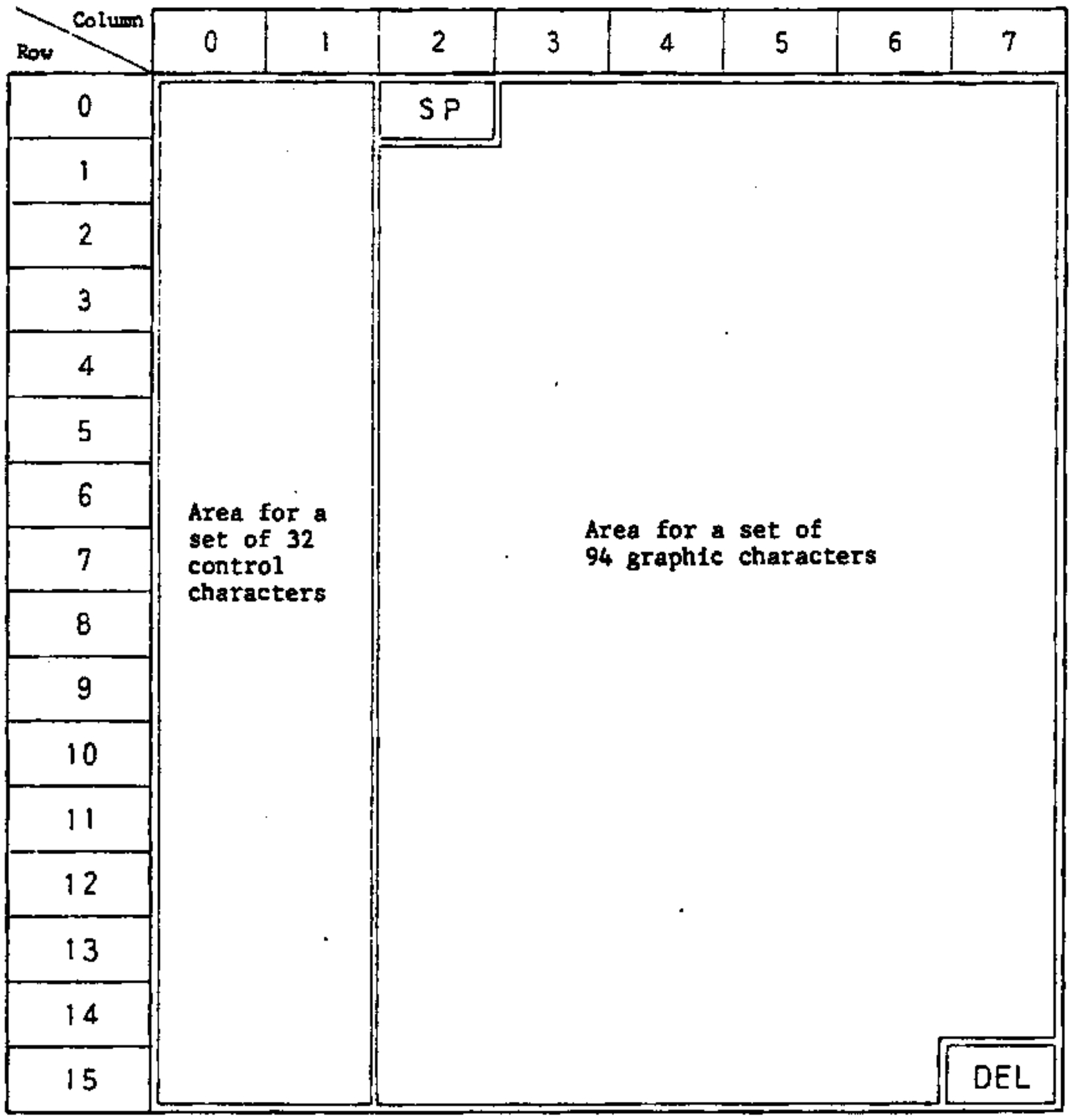
(2) The character SP in position 2/0 which may be regarded either as a control character or a graphic character.

(3) The area for a set of 94 graphic characters allocated to columns 2 to 7.

(4) The character DEL in position 7/15.

Full document details: JIS X 0202-1984 (1995) JIS C 6228-1984

Fig. 1. Structure of 7-bit Code



5.1.2 Extension by Substitution In many cases, the requirements of an application is satisfied by allocating necessary characters to the optional positions in the version of the 7-bit code. Other applications are satisfied by the use of a similarly structured code in which substituted characters are allocated to the optional position in the version of the 7-bit code and some other positions.

5.1.3 Extension by Increasing Total Number of Characters Procedure for adding further characters to the 128 characters provided by the 7-bit code shall be as follows:

(1) Additional single control functions

(2) Additional sets of 32 control characters

(3) Additional sets of 94 graphic characters

(4) Additional sets of more than 94 graphic characters, each represented by more than one byte.

Documents
Document Sharing Hub

Full document details: JIS X 0202-1984 (1995) JIS C 6228-1984

5.1.4 Elements of Code Extension Many applications require combinations of the code extension facilities of 5.1.3. The elements of code extension are shown in Fig. 2, where the names of elements are defined as follows:

(1) CO Set A set of 32 control characters (columns 0 and 1)

(2) C1 Set An additional set of 32 control characters

(3) GO Set A set of 94 graphic characters (columns 2 to 7)
A multiple-byte set may also function as a GO set.

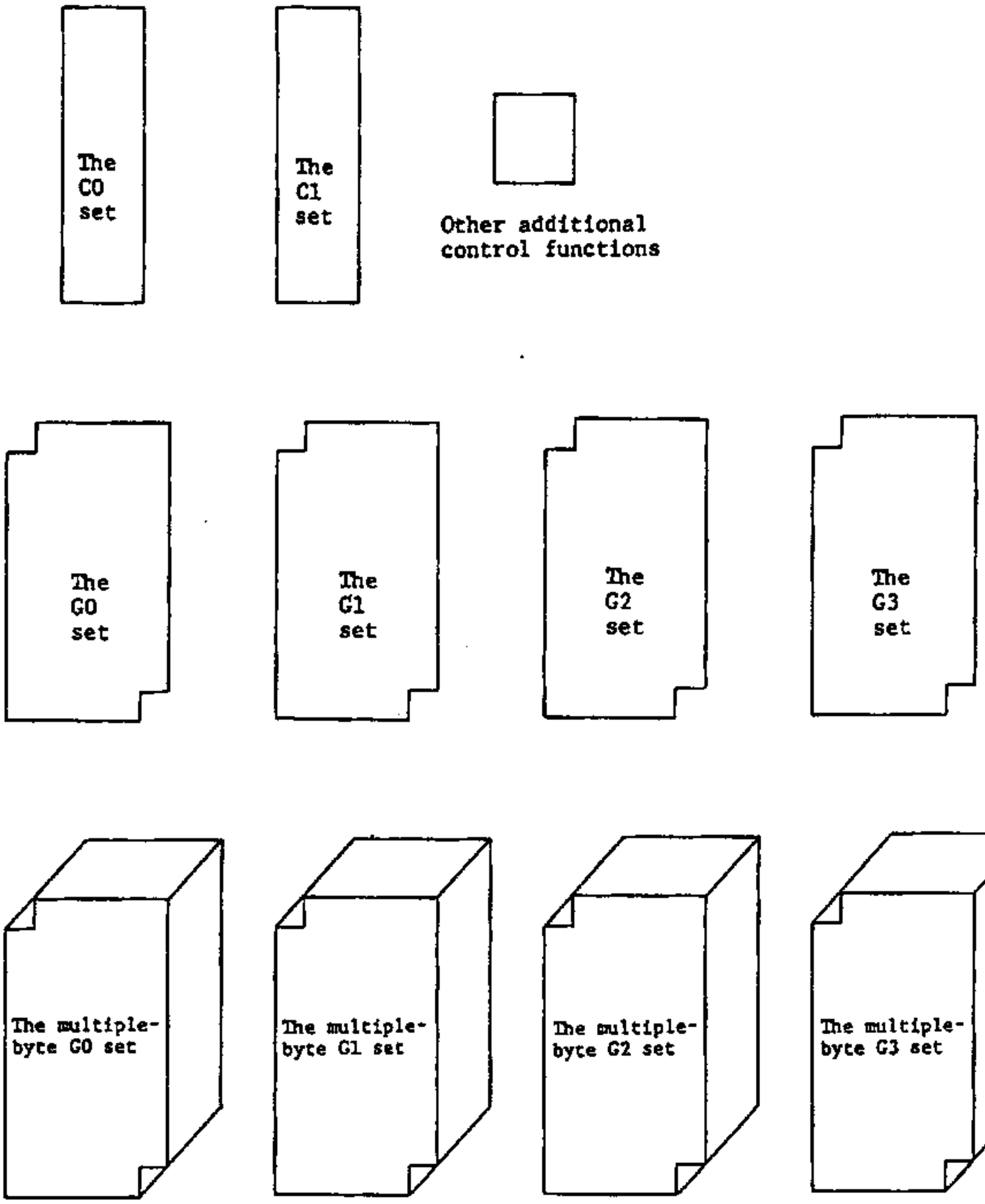
(4) G1, G2, G3 Set Additional sets of 94 graphic characters.
A multiple-byte set may also function as a G1, G2 or G3 set.

Remark: It is intended that, the control character set, graphic character set for Roman letters and graphic character set for KATAKANA specified in JIS C 6220 are assigned to the CO set, GO set and G1 set respectively if they are used.

Full document details: JIS X 0202-1984 (1995) JIS C 6228-1984

Documents
Document Sharing Hub

Fig. 2. Code Extension Elements



5.1.5 Compatibility For the purpose of interchange, various levels of compatibility which may be preserved when appling extension facilities, are identified. The following three such levels are distinguished in this standard:

(1) Version of the 7-bit code

Remark: The code utilizing only the graphic character sets for Roman letters specified in JIS C 6220 is considered as in compliance with this level.

Documents
Document Sharing Hub

Full document details: JIS X 0202-1984 (1995) JIS C 6228-1984

(2) A compatible variant, i.e. a set which complies with JIS C 6220 on all the following points:

(a) Columns 0 and 1 contain only control characters.

(b) Columns 2 to 7 are used for graphic characters only (apart from DEL).

(c) The ten transmission control characters ([1]) and NUL, SO, SI, CAN, SUB, ESC, SP and DEL remain unaltered in their meanings and in their positions in the code table.

Note ([1]) SOH, STX, ETX, EOT, ENQ, ACK, DLE, NAK, SYN and ETB

(d) The graphic characters allocated to the other positions than the optional positions in the version of the 7-bit code, are not moved to other positions (a non-latin alphabet containing graphic characters which are also included in the latin alphabet is not subject to this rule).

Remark: The code utilizing the graphic character sets for Roman letters and the graphic character sets for KATAKANA of JIS C 6220 are considered as in compliance with this level.

(3) Other sets structured as in 5.1.1. To be able to provide the facilities of code extension of this standard, the control characters ESC, SO and SI shall remain unaltered in their meanings and in their positions in the code table.

Remark: The code utilizing the Japanese graphic character sets of JIS C 6226 is considered as in compliance with this level.

5.1.6 Code Extension Characters of JIS C 6220 The following control characters specified in JIS C 6220 are used for code extension:

(1) ESC

(2) SO

(3) SI

(4) DLE

This standard does not specify the use of the DLE which is reserved for the provision of additional transmission control functions. The use of this character is specified in other Japanese Industrial Standard.

Documents
Document Sharing Hub

Full document details: JIS X 0202-1984 (1995) JIS C 6228-1984

5.1.7 Other Code Extension Characters For use within a 7-bit environment, this standard provides the following additional shift functions which are not included in JIS C 6220.

(1) LS2 locking-shift two

(2) LS3 locking-shift three

(3) SS2 single-shift two

(4) SS3 single-shift three

Reference: See the Reference in the end of text for the coding of these functions.

Locking-shift functions LS1R, LS2R, LS3R (see 7.2.2) are used in a 7-bit environment only when it is necessary to preserve their use for transformation between 7-bit and 8-bit environments(see 9.5).

5.1.8 Combination of Graphic Characters Some graphic sets may allow for the representation of additional graphic symbols such as accented characters by the combination of two or more graphic characters in the same character position, in the following manner:

(1) Use of graphic characters having implicit forward motion, hereinafter referred to as the "spacing characters" in conjunction with BS.

(2) Use of graphic characters having so implicit forward motion, hereinafter referred to as the "non-spacing characters" in combination with spacing graphic characters.

JIS C 6220 allows for the first of these methods to represent accented characters.

Reference: Sponsors of graphic sets applying for registration under the provision of ISO 2375 are expected to identify any characters in the set intended to be used in combination with other characters and to state the method by which this combination is to be achieved, including a clear indication of any characters in the set that are non-spacing.

Any restriction on the use of BS with any characters or the combination of spacing and non-spacing characters shall be clearly specified.

5.2 Extension of the Graphic Set by means of Shift Functions The shift functions for use in a 7-bit environment are as follows:

SO, SI, LS2, LS3, SS2 and SS3

5.2.1 Use of Locking-shift Functions In a 7-bit environment, the SO, SI, LS2 and LS3 shall be used exclusively for extension of graphic character set.

The sift functions SO, LS2 and LS3 shall each invoke an additional set of 94 graphic characters (^{2}) i.e. G1, G2 and G3. Such a set replaces other graphic character set previously invoked. Graphic characters need

Documents
Document Sharing Hub

Full document details: JIS X 0202-1984 (1995) JIS C 6228-1984