

INTERNATIONAL  
STANDARD

**ISO/IEC**  
**10646-1**

First edition  
1993-05-01

---

---

**Information technology — Universal  
Multiple-Octet Coded Character Set  
(UCS) —**

**Part 1:**

**Architecture and Basic Multilingual Plane**

<https://standards.iso.org/standards/catalog/standards/si/4580:116f-fa44-4a84-9100-f088f71d111e/iec-10646-1-1993>

*Technologies de l'information — Jeu universel de caractères codés à  
plusieurs octets —*

*Partie 1: Architecture et table multilingue*



Reference number  
ISO/IEC 10646-1:1993(E)

## Contents

	Page
1 Scope .....	1
2 Conformance .....	1
3 Normative references .....	2
4 Definitions .....	2
5 General structure of the UCS .....	3
6 Basic structure and nomenclature .....	4
7 Special features of the UCS .....	7
8 The Basic Multilingual Plane .....	7
9 Other planes .....	7
10 The Restricted Use zone .....	7
11 Private Use groups and planes .....	8
12 Revision and updating of the UCS .....	8
13 Subsets .....	8
14 Coded representation forms of the UCS .....	8
15 Implementation levels .....	9
16 Use of control functions with the UCS .....	9
17 Declaration of identification of features .....	9
18 Structure of the code tables and lists .....	11
19 Block names .....	11
20 Characters in bi-directional context .....	12
21 Special characters .....	12
22 Order of characters .....	12
23 Combining characters .....	13
24 Hangul syllable composition method .....	14
25 Code tables and lists of character names .....	14
26 CJK unified ideographs .....	262

## Annexes

A Collections of graphic characters for subsets .....	699
B List of combining characters .....	701
C Mirrored characters in Arabic bi-directional context .....	704
D Alternate format characters .....	706

© ISO/IEC 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 publisher.

ISO/IEC Copyright Office • Case Postale 56 • CH-1211 Genève 20 • Switzerland

Printed in Switzerland

<b>E</b>	Alphabetically sorted list of character names .....	709
<b>F</b>	The use of "signatures" to identify UCS .....	743
<b>G</b>	UCS transformation format (UTF-1) .....	744
<b>H</b>	Recommendation for combined receiving/originating devices with internal storage .....	746
<b>J</b>	Notations of octet value representations .....	747
<b>K</b>	Character naming guidelines .....	748
<b>L</b>	Sources of characters .....	750
<b>M</b>	External references to character repertoires .....	752
<b>N</b>	Scripts under consideration for future editions of ISO/IEC 10646 .....	754

## **iTeh STANDARD PREVIEW** **(standards.iteh.ai)**

[ISO/IEC 10646-1:1993](https://standards.iteh.ai/catalog/standards/sist/4580d16f-fa44-4a84-9100-fb288f3e47dd/iso-iec-10646-1-1993)

<https://standards.iteh.ai/catalog/standards/sist/4580d16f-fa44-4a84-9100-fb288f3e47dd/iso-iec-10646-1-1993>

## Foreword

ISO (the International Organization for Standardization) and IEC (the International Electrotechnical Commission) form the specialized system for worldwide standardization. National bodies that are members of ISO or IEC participate in the development of International Standards through technical committees established by the respective organization to deal with particular fields of technical activity. ISO and IEC technical committees collaborate in fields of mutual interest. Other international organizations, governmental and non-governmental, in liaison with ISO and IEC, also take part in the work.

In the field of information technology, ISO and IEC have established a joint technical committee, ISO/IEC JTC 1. Draft International Standards adopted by the joint technical committee are circulated to national bodies for voting. Publication as an International Standard requires approval by at least 75 % of the national bodies casting a vote.

International Standard ISO/IEC 10646-1 was prepared by Joint Technical Committee ISO/IEC JTC 1, *Information technology*, Sub-Committee SC 2, *Character sets and information coding*.

ISO/IEC 10646 consists of the following parts under the general title *Information technology — Universal Multiple-Octet Coded Character Set (UCS)*:

— *Part 1: Architecture and Basic Multilingual Plane*

Additional parts will specify other planes.

Annexes A and B form an integral part of this part of ISO/IEC 10646. Annexes C to N are for information only.

## Introduction

ISO/IEC 10646 specifies the Universal Multiple-Octet Coded Character Set (UCS). It is applicable to the representation, transmission, interchange, processing, storage, input and presentation of the written form of the languages (scripts) of the world as well as additional symbols.

This part of ISO/IEC 10646 specifies the overall architecture and the Basic Multilingual Plane (BMP) of the UCS.

## iTeh STANDARD PREVIEW (standards.iteh.ai)

[ISO/IEC 10646-1:1993](https://standards.iteh.ai/catalog/standards/sist/4580d16f-fa44-4a84-9100-fb288f3e47dd/iso-iec-10646-1-1993)

<https://standards.iteh.ai/catalog/standards/sist/4580d16f-fa44-4a84-9100-fb288f3e47dd/iso-iec-10646-1-1993>

**iTeh STANDARD PREVIEW**  
**(standards.iteh.ai)**

ISO/IEC 10646-1:1993

<https://standards.iteh.ai/catalog/standards/sist/4580d16f-fa44-4a84-9100-fb288f3e47dd/iso-iec-10646-1-1993>

# Information technology — Universal Multiple-Octet Coded Character Set (UCS) —

## Part 1:

### Architecture and Basic Multilingual Plane

#### 1 Scope

ISO/IEC 10646 specifies the Universal Multiple-Octet Coded Character Set (UCS). It is applicable to the representation, transmission, interchange, processing, storage, input and presentation of the written form of the languages of the world as well as additional symbols.

This part of ISO/IEC 10646 specifies the overall architecture, and

- defines terms used in ISO/IEC 10646;
- describes the general structure of the coded character set;
- specifies the Basic Multilingual Plane (BMP) of the UCS, and defines a set of graphic characters used in scripts and the written form of languages on a world-wide scale;
- specifies the names for the graphic characters of the BMP, and the coded representations;
- specifies the four-octet (32-bit) canonical form of the UCS: UCS-4;
- specifies a two-octet (16-bit) BMP form of the UCS: UCS-2;
- specifies the coded representations for control functions;
- specifies the management of future additions to this coded character set.

The UCS is a coding system different from that specified in ISO 2022. The method to designate UCS from ISO 2022 is specified in 17.2.

#### 2 Conformance

##### 2.1 General

Whenever Private Use characters are used as specified in ISO/IEC 10646, the characters themselves shall not be covered by these conformance requirements.

##### 2.2 Conformance of information interchange

A coded-character-data-element (CC-data-element) within coded information for interchange is in conformance with ISO/IEC 10646 if

- a) all the coded representations of graphic characters within that CC-data-element conform to clauses 6 and 7, to an identified form chosen from clause 14, and to an identified implementation level chosen from clause 15;
- b) all the graphic characters represented within that CC-data-element are taken from those within an identified subset (clause 13);
- c) all the coded representations of control functions within that CC-data-element conform to clause 16.

A claim of conformance shall identify the adopted form, the adopted implementation level and the adopted subset by means of a list of collections and/or characters.

##### 2.3 Conformance of devices

A device is in conformance with ISO/IEC 10646 if it conforms to the requirements of item a) below, and either or both of items b) , and c).

NOTE - The term *device* is defined (in 4.17) as a component of information processing equipment which can transmit and/or receive coded information within CC-data-elements. A device may be a conventional

input/output device, or a process such as an application program or gateway function.

A claim of conformance shall identify the document that contains the description specified in a) below, and shall identify the adopted form(s), the adopted implementation level, the adopted subset (by means of a list of collections and/or characters), and the selection of control functions adopted in accordance with clause 16.

a) Device description: A device that conforms to ISO/IEC 10646 shall be the subject of a description that identifies the means by which the user may supply characters to the device and/or may recognise them when they are made available to the user, as specified respectively, in subclauses b), and c) below.

b) Originating device: An originating device shall allow its user to supply any characters from an adopted subset, and be capable of transmitting their coded representations within a CC-data-element in accordance with the adopted form and implementation level.

c) Receiving device: A receiving device shall be capable of receiving and interpreting any coded representation of characters that are within a CC-data-element in accordance with the adopted form and implementation level, and shall make any corresponding characters from the adopted subset available to the user in such a way that the user can identify them.

Any corresponding characters that are not within the adopted subset shall be indicated to the user in a way which need not allow them to be distinguished from each other.

#### NOTES

1 An indication to the user may consist of making available the same character to represent all characters not in the adopted subset, or providing a distinctive audible or visible signal when appropriate to the type of user.

2 See also annex H for receiving devices with re-transmission capability.

### 3 Normative references

The following standards contain provisions which, through reference in this text, constitute provisions of this part of ISO/IEC 10646. At the time of publication, the editions indicated were valid. All standards are subject to revision, and parties to agreements based on this part of ISO/IEC 10646 are encouraged to investigate the possibility of applying the most recent

editions of the standards listed below. Members of IEC and ISO maintain registers of currently valid International Standards.

ISO 2022:1986 *Information processing — ISO 7-bit and 8-bit coded character sets — Code extension techniques*.

ISO/IEC 6429:1992 *Information technology — Control functions for coded character sets*.

### 4 Definitions

For the purposes of ISO/IEC 10646, the following definitions apply :

**4.1 Basic Multilingual Plane (BMP)** : Plane 00 of Group 00.

**4.2 block** : A contiguous collection of characters that share common characteristics, such as script.

**4.3 canonical form** : The form with which characters of this coded character set are specified using four octets to represent each character.

**4.4 CC-data-element (Coded-Character-Data-Element)** : An element of interchanged information that is specified to consist of a sequence of coded representations of characters, in accordance with one or more identified standards for coded character sets.

**4.5 cell** : The place within a row at which an individual character may be allocated.

**4.6 character** : A member of a set of elements used for the organisation, control, or representation of data.

**4.7 character boundary** : Within a stream of octets the demarcation between the last octet of the coded representation of a character and the first octet of that of the next coded character.

**4.8 coded character** : A character together with its coded representation.

**4.9 coded character set** : A set of unambiguous rules that establishes a character set and the relationship between the characters of the set and their coded representation.

**4.10 code table** : A table showing the characters allocated to the octets in a code.

**4.11 combining character** : A member of an identified subset of the coded character set of ISO/IEC 10646 intended for combination with the preceding non-combining graphic character, or with a sequence of combining characters preceded by a non-combining character (see also 4.13).



NOTE - This part of ISO/IEC 10646 specifies several subset collections which include combining characters.

**4.12 compatibility character** : A graphic character included as a coded character of ISO/IEC 10646 primarily for compatibility with existing coded character sets.

**4.13 composite sequence** : A sequence of graphic characters consisting of a non-combining character followed by one or more combining characters (see also 4.11).

#### NOTES

1 A graphic symbol for a composite sequence generally consists of the combination of the graphic symbols of each character in the sequence.

2 A composite sequence is not a character and therefore is not a member of the repertoire of ISO/IEC 10646.

**4.14 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 octets.

**4.15 default state** : The state that is assumed when no state has been explicitly specified.

**4.16 detailed code table** : A code table showing the individual characters, and normally showing a partial row.

**4.17 device** : A component of information processing equipment which can transmit and/or receive coded information within CC-data-elements. (It may be an input/output device in the conventional sense, or a process such as an application program or gateway function.)

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

**4.19 graphic symbol** : The visual representation of a graphic character or of a composite sequence.

**4.20 group** : A subdivision of the coding space of this coded character set; of 256 x 256 x 256 cells.

**4.21 interchange** : The transfer of character coded data from one user to another, using telecommunication means or interchangeable media.

**4.22 interworking** : The process of permitting two or more systems, each employing different coded character sets, meaningfully to interchange character coded data; conversion between the two codes may be involved.

**4.23 octet** : An ordered sequence of eight bits considered as a unit.

**4.24 plane** : A subdivision of a group; of 256 x 256 cells.

**4.25 presentation; to present** : The process of writing, printing, or displaying a graphic symbol.

**4.26 presentation form** : In the presentation of some scripts, a form of a graphic symbol representing a character that depends on the position of the character relative to other characters.

**4.27 private use planes** : Planes within this coded character set the contents of which are not specified in ISO/IEC 10646 (see 10.1).

**4.28 repertoire** : A specified set of characters that are represented in a coded character set.

**4.29 row** : A subdivision of a plane; of 256 cells.

**4.30 script** : A set of graphic characters used for the written form of one or more languages.

**4.31 supplementary planes** : Planes that accommodate characters which have not been allocated to the Basic Multilingual Plane.

**4.32 user** : A person or other entity that invokes the service provided by a device. (This entity may be a process such as an application program if the "device" is a code converter or a gateway function, for example.)

**4.33 zone** : A sequence of cells of a code table, comprising one or more rows, either in whole or in part, containing characters of a particular class (see clause 8).

## 5 General structure of the UCS

The general structure of the Universal Multiple-Octet Coded Character Set (referred to hereafter as "this coded character set") is described in this explanatory clause, and is illustrated in figures 1 and 2. The normative specification of the structure is given in later clauses.

The value of any octet is expressed in hexadecimal notation from 00 to FF in ISO/IEC 10646 (see annex J).

The canonical form of this coded character set — the way in which it is to be conceived — uses a four-dimensional coding space, regarded as a single entity, consisting of 128 three-dimensional groups.

NOTE - Thus, bit 8 of the most significant octet in the canonical form of a coded character can be used for internal processing purposes within a device as long as it is set to zero within a conforming CC-data-element.

Each group consists of 256 two-dimensional planes.

Each plane consists of 256 one-dimensional rows, each row containing 256 cells. A character is located and coded at a cell within this coding space or the cell is declared unused.

In the canonical form, four octets are used to represent each character, and they specify the group, plane, row and cell, respectively. The canonical form consists of four octets since two octets are not sufficient to cover all the characters in the world, and a 32-bit representation follows modern processor architectures.

The four-octet canonical form can be used as a four-octet coded character set in which case it is called UCS-4.

The first plane (Plane 00 of Group 00) is called the Basic Multilingual Plane. The Basic Multilingual Plane includes characters in general use in alphabetic, syllabic and ideographic scripts together with various symbols and digits. The BMP also has a Restricted Use (RU) zone in which the characters have special characteristics.

The subsequent planes are regarded as supplementary or private use planes, which will accommodate additional graphic characters.

The 32 planes with Plane-octet values E0 to FF of Group 00 are for Private Use. The 32 groups with Group-octet values 60 to 7F of this coded character set are also for Private Use. The contents of the cells in Private Use zones are not specified in ISO/IEC 10646.

Each character is located within the coded character set in terms of its Group-octet, Plane-octet, Row-octet, and Cell-octet.

In addition to the canonical form, a two-octet BMP form is specified. Thus, the Basic Multilingual Plane can be used as a two-octet coded character set identified as UCS-2.

Subsets of the coding space may be used in order to give a sub-repertoire of graphic characters.

A UCS Transformation Format (UTF-1) is specified in annex G which can be used to transmit text data through communication systems which are sensitive to octet values for control characters coded according to the structure of ISO 2022.

## 6 Basic structure and nomenclature

### 6.1 Structure

The Universal Multiple-Octet Coded Character Set as specified in ISO/IEC 10646 shall be regarded as

a single entity.

This entire coded character set shall be conceived of as comprising 128 groups of 256 planes. Each plane shall be regarded as containing 256 rows of characters, each row containing 256 cells. In a code table representing the contents of a plane (such as in figure 2), the horizontal axis shall represent the least significant octet, with its smaller value to the left; and the vertical axis shall represent the more significant octet, with its smaller value at the top.

Each axis of the coding space shall be coded by one octet. Within each octet the most significant bit shall be bit 8 and the least significant bit shall be bit 1.

Accordingly, the weight allocated to each bit shall be

bit 8	bit 7	bit 6	bit 5	bit 4	bit 3	bit 2	bit 1
128	64	32	16	8	4	2	1

### 6.2 Coding of characters

In the canonical form of the coded character set, each character within the entire coded character set shall be represented by a sequence of four octets. The most significant octet of this sequence shall be the group-octet. The least significant octet of this sequence shall be the cell-octet. Thus this sequence may be represented as

m.s.			l.s.
Group-octet	Plane-octet	Row-octet	Cell-octet

where m.s. means the most significant octet, and l.s. means the least significant octet.

For brevity, the octets may be termed

m.s.			l.s.
G-octet	P-octet	R-octet	C-octet

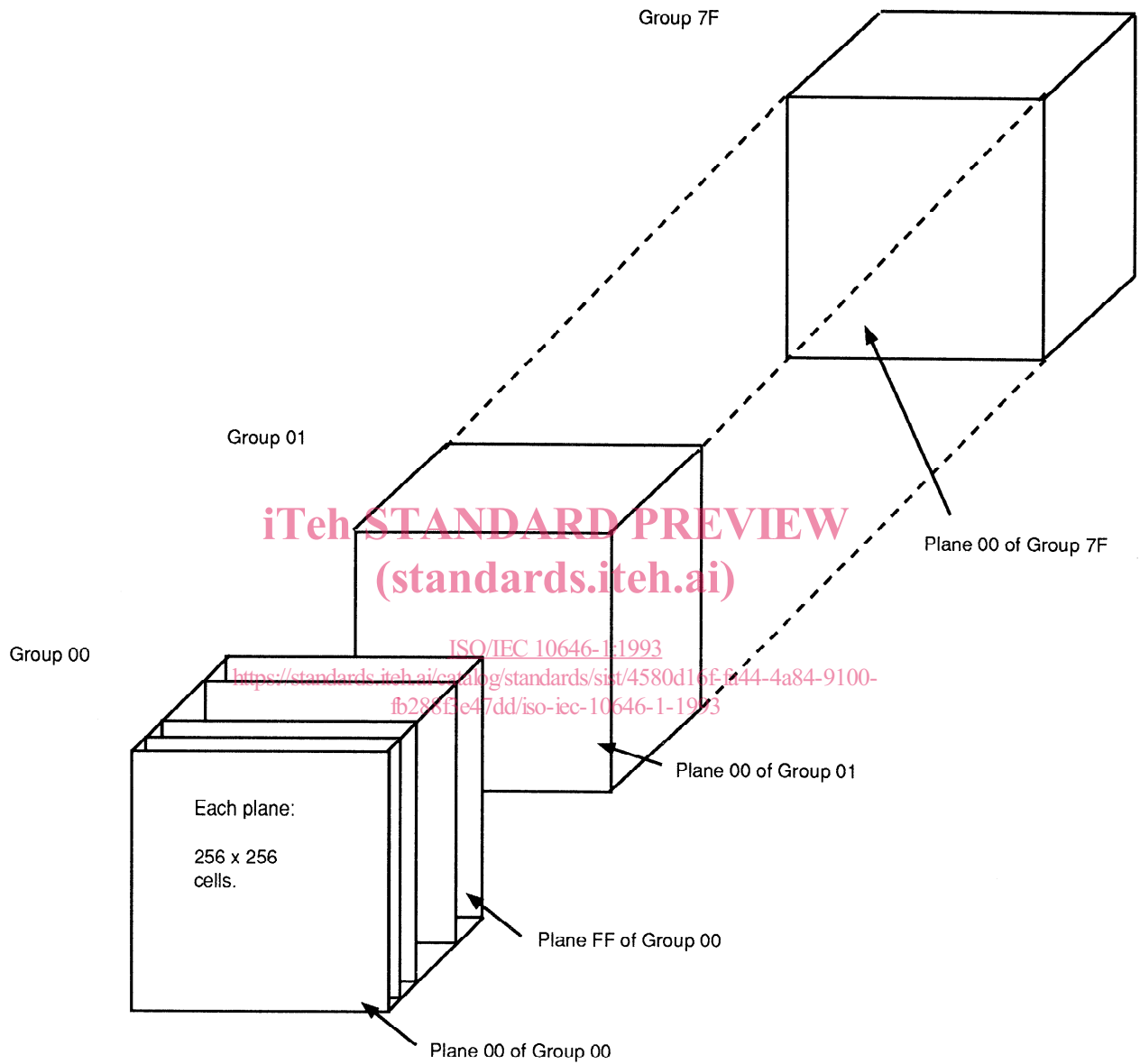
Where appropriate, these may be further abbreviated to G, P, R, and C.

The value of any octet shall be represented by two hexadecimal digits, for examples: 31 or FE. When a single character is to be identified in terms of the values of its group, plane, row and cell, this shall be represented such as:

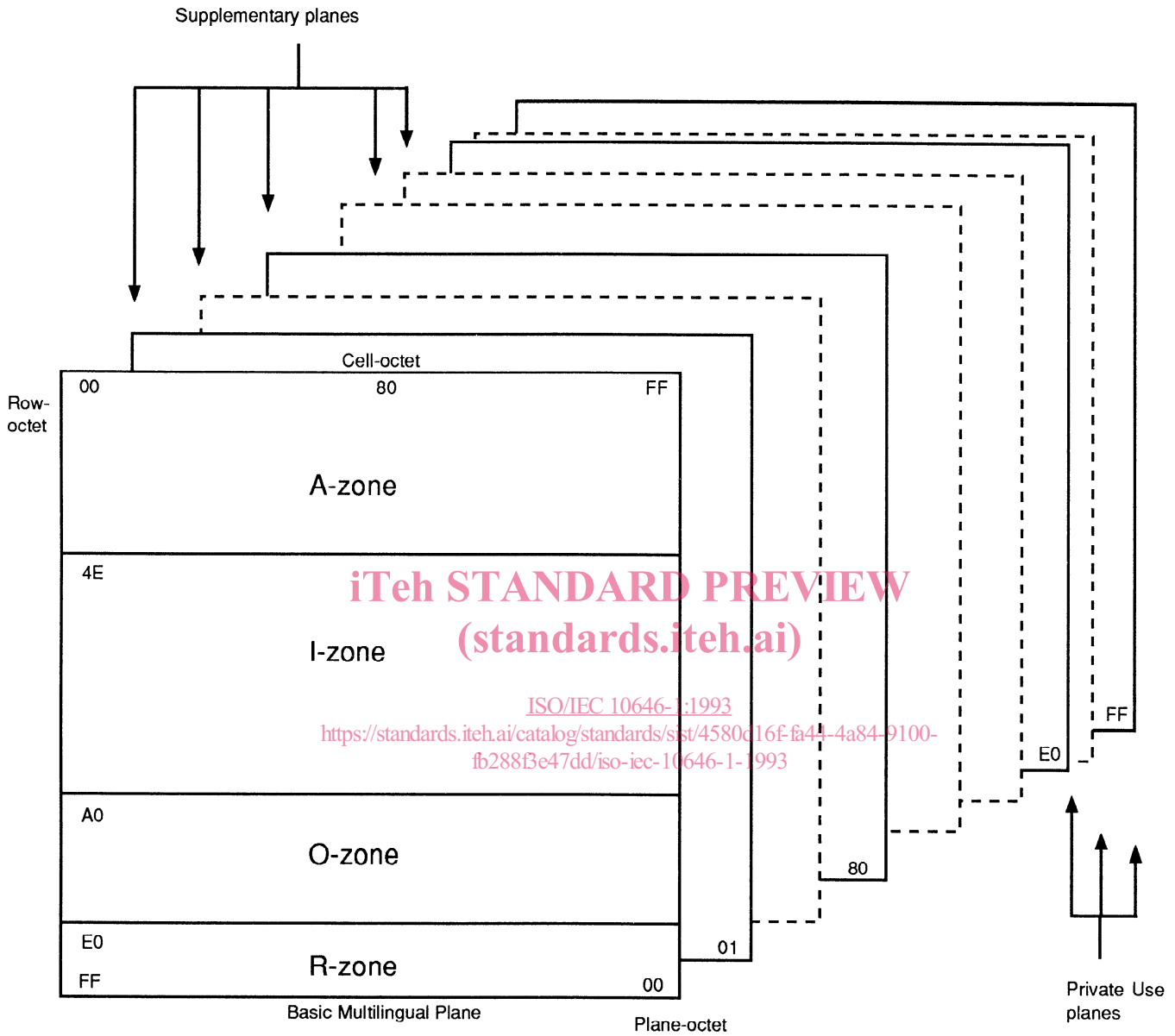
0000 0030 for DIGIT ZERO

0000 0041 for LATIN CAPITAL LETTER A

When referring to characters within a plane, the leading four zeros (for G-octet and P-octet) may be omitted. For example, 0030 may be used to refer to DIGIT ZERO.



**Figure 1 - Entire coding space of the Universal Multiple-Octet Coded Character Set**



Labels A-zone, I-zone, O-zone, and R-zone are specified in clause 8.

**Figure 2 - Group 00 of the Universal Multiple-Octet Coded Character Set**

### 6.3 Octet order

The sequence of the octets that represent a character, and the most significant and least significant ends of it, shall be maintained as shown above. When serialised as octets, a more significant octet shall precede less significant octets. When not serialised as octets, the order of octets may be specified by agreement between sender and recipient (see 17.1 and annex F).

## 7 Special features of the UCS

The following characteristics apply to the entire coded character set.

1. The values of P-, and R-, and C-octets used for representing graphic characters shall be in the range 00 to FF. The values of G-octets used for representation of graphic characters shall be in the range 00 to 7F. On any plane, code positions FFFE and FFFF shall not be used.

NOTE - Code position FFFE is reserved for "signature" (see annex F). Code position FFFF can be used for internal processing uses requiring a numeric value that is guaranteed not to be a coded character such as in terminating tables, or signaling end-of-text. Since it is the largest two-octet value, it may also be used as the final value in binary or sequential searching index.

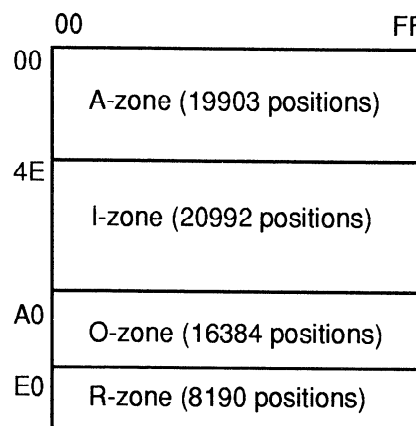
2. Code positions to which a character is not allocated, except for the positions reserved for Private Use characters, are reserved for future standardisation and shall not be used for any other purpose. Future editions of ISO/IEC 10646 will not allocate any characters to code positions reserved for Private Use characters.
3. The same graphic character shall not be allocated to more than one code position. There are graphic characters with similar shapes in the coded character set; they are used for different purposes and have different character names.
4. Compatibility characters are included in ISO/IEC 10646 primarily for compatibility with existing coded character sets to allow two-way code conversion without loss of information.

## 8 The Basic Multilingual Plane

Plane 00 of Group 00 shall be the Basic Multilingual Plane (BMP). The BMP can be used as a two-octet coded character set in which case it shall be called UCS-2 (see 14.1).

The Basic Multilingual Plane shall be divided into four zones:

A-zone: code positions 0000 to 4DFF  
 I-zone: code positions 4E00 to 9FFF  
 O-zone: code positions A000 to DFFF  
 R-zone: code positions E000 to FFFD



Code positions 0000 to 001F in the BMP are reserved for control characters, and code position 007F is reserved for the character DELETE (see clause 16). Code positions 0080 to 009F are reserved.

In the Basic Multilingual Plane, the A-zone is used for alphabetic and syllabic scripts together with various symbols. The I-zone is used for Chinese/Japanese/Korean (CJK) unified ideographs (unified East Asian ideographs). The O-zone is reserved for future standardisation. The R-zone shall be used for the Restricted Use zone in the BMP which contains Private Use characters, presentation forms, and compatibility characters (see clause 10).

## 9 Other planes

Planes 01 to DF in Group 00 and planes 00 to FF in Groups 01 to 5F are reserved for future standardisation, and thus those code positions shall not be used for any other purpose.

## 10 The Restricted Use zone

Sets of graphic characters that are used in particular ways are provided in the Restricted Use zone. These sets include:

- a) Private Use characters,
- b) Presentation forms of characters,
- c) Compatibility characters (see item 4 in clause 7).



## 10.1 Private Use characters

Private Use characters are not restrained in any way by ISO/IEC 10646. Private Use characters can be used to provide user-defined characters. For example, this is a common requirement for users of ideographic scripts.

NOTE 1 - For meaningful interchange of Private Use characters, an agreement, independent of ISO/IEC 10646, is necessary between sender and recipient.

Private Use characters can be used for dynamically-redefinable characters (DRCS) applications.

NOTE 2 - For meaningful interchange of DRCS, an agreement, independent of ISO/IEC 10646 is necessary between sender and recipient. ISO/IEC 10646 does not specify the techniques for defining or setting up dynamically-redefinable characters.

## 10.2 Presentation forms of characters

Each presentation form of character provides an alternative form, for use in a particular context, to the nominal form of the character or sequence of characters from the other zones of graphic characters. The transformation from the nominal form to the presentation forms may involve substitution, superimposition, or combination.

The rules for the superimposition, choice of differently shaped characters, or combination into ligatures, or conjuncts — which are often of extreme complexity — are not specified in ISO/IEC 10646.

In general, presentation forms are not intended to be used as a substitute for the nominal forms of the graphic characters specified elsewhere within this coded character set. However, specific applications may encode these presentation forms instead of the nominal forms for specific reasons among which is compatibility with existing devices. The rules for searching, sorting and other processing operations on presentation forms are outside the scope of ISO/IEC 10646.

## 11 Private Use groups and planes

The code positions of 32 planes from Plane E0 to Plane FF of Group 00 shall be for Private Use.

The code positions of the 32 groups from Group 60 to Group 7F shall be for Private Use.

The contents of these code positions are not specified in ISO/IEC 10646 (see 10.1).

## 12 Revision and updating of the UCS

The revision and updating of this coded character set will be carried out by ISO/IEC JTC1/SC2.

NOTE - It is intended that in future editions of ISO/IEC 10646, the names and allocation of the characters in this edition will remain unchanged.

## 13 Subsets

ISO/IEC 10646 provides the specification of subsets of coded graphic characters for use in interchange, by originating devices and by receiving devices.

There are two alternatives for the specification of subsets; limited subset and selected subset. An adopted subset may comprise either of them, or a combination of the two.

### 13.1 Limited subset

A limited subset consists of a list of graphic characters in the specified subset. This specification allows applications and devices that were developed using other codes to interwork with this coded character set.

A claim of conformance referring to a limited subset shall list the graphic characters in the subset by the names of graphic characters or code positions as defined in ISO/IEC 10646.

### 13.2 Selected subset

A selected subset consists of a list of collections of graphic characters as defined in ISO/IEC 10646. The collections from which the selection may be made are listed in annex A of each part of ISO/IEC 10646. A selected subset shall always automatically include the Cells 20 to 7E of Row 00 of Plane 00 of Group 00.

A claim of conformance referring to a selected subset shall list the collections chosen as defined in ISO/IEC 10646.

## 14 Coded representation forms of the UCS

ISO/IEC 10646 provides two alternative forms of coded representation of characters.

NOTE - The characters from the ISO 646 IRV repertoire are coded by simple zero extensions to their coded representations in ISO 646 IRV. Therefore, their coded representations have the same integer values when represented as 8-bit, 16-bit, or 32-bit integers. For implementations sensitive to a zero valued octet (e.g. for use as a string terminator), use of 8-bit based array data

type should be avoided as any zero valued octet may be interpreted incorrectly. Use of data types at least 16-bits wide is more suitable for UCS 2, and use of data types at least 32-bits wide is more suitable for UCS-4.

#### 14.1 Two-octet BMP form

This coded representation form permits the use of characters from the Basic Multilingual Plane with each character represented by two octets.

Within a CC-data-element conforming to the two-octet BMP form, a character from the Basic Multilingual Plane shall be represented by two octets comprising the R-octet and the C-octet as specified in 6.2.

NOTE - A coded graphic character using the two-octet BMP form may be implemented by a 16-bit integer for processing.

#### 14.2 Four-octet canonical form

The canonical form permits the use of all the characters of ISO/IEC 10646, with each character represented by four octets.

Within a CC-data-element conforming to the four-octet canonical form, every character shall be represented by four octets comprising the G-octet, the P-octet, the R-octet and the C-octet as specified in 6.2.

NOTE - A coded graphic character using the four-octet canonical form may be implemented by a 32-bit integer for processing.

### 15 Implementation levels

ISO/IEC 10646 specifies three levels of implementation. Combining characters are described in clause 23 and listed in annex B.

#### 15.1 Implementation level 1

When implementation level 1 is used, a CC-data-element shall not contain coded representations of combining characters (see clause B.1) nor of characters from HANGUL JAMO block (see clause 24).

#### 15.2 Implementation level 2

When implementation level 2 is used, a CC-data-element shall not contain coded representations of characters listed in clause B.2.

#### 15.3 Implementation level 3

When implementation level 3 is used, a CC-data-element may contain coded representations of any characters.

### 16 Use of control functions with the UCS

This coded character set provides for use of control functions encoded according to ISO 2022, ISO/IEC 6429 or similarly structured standards for control functions, and standards derived from these. A set or subset of such coded control functions may be used in conjunction with this coded character set. These standards encode a control function as a sequence of one or more octets.

When a C0 control character of ISO/IEC 6429 is used with this coded character set, its coded representation as specified in ISO/IEC 6429 shall be padded to correspond with the number of octets in the adopted form (see clause 14). Thus, the least significant octet shall be the bit combination specified in ISO/IEC 6429, and the more significant octet(s) shall be zeros.

For example, the control character FORM FEED is represented by "000C" in the two-octet form, and "0000 000C" in the four-octet form.

For escape sequences, control sequences, and control strings (see ISO/IEC 6429) consisting of a coded control character followed by additional bit combinations in the range 20 to 7F, each bit combination shall be padded by octet(s) with value 00.

For example, the escape sequence "ESC 02/00 04/00" is represented by "001B 0020 0040" in the two-octet form, and "0000 001B 0000 0020 0000 0040" in the four-octet form.

When using a C1 control character of ISO/IEC 6429 with this coded character set, it shall be coded as ESC Fe sequence (see ISO/IEC 6429) padded as specified above.

For example, the control character PARTIAL LINE BACKWARD – PLU (08/12 in ISO/IEC 6429 representation) is represented by "001B 004C" in the two-octet form, and "0000 001B 0000 004C" in the four-octet form.

Code extension control functions for the ISO 2022 code extension techniques (such as designation escape sequence, single shift and locking shift) shall not be used with this coded character set.

### 17 Declaration of identification of features

#### 17.1 Purpose and context of identification

CC-data-elements conforming to ISO/IEC 10646 are