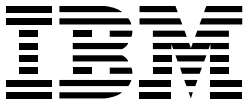


IBM Simplified Chinese Graphic Character Set, GB 18030 code

National Standard and DBCS-Host



IBM Simplified Chinese Graphic Character Set, GB 18030 code

National Standard and DBCS-Host

**IBM Simplified Chinese Graphic Character Set, GB 18030 code**

National Standard and DBCS-Host

Contents

Introduction	1	6.9 Area 5 (Chinese Character Set)	22
1. Scope	1	6.10 Area 6a/6b (Chinese Character Set)	22
1.1 Abstract	1	6.10.1 ISO 10646-1 CJK Unified Ideographs	23
1.2 Objective	1	6.10.2 ISO 10646-1 CJK Compatibility Ideographs	23
1.3 Application	1	6.10.3 Additional Chinese Characters and Radicals/Components	23
1.4 Effective Date	1	6.11 Area 7 (Extended Non-Chinese Character Set)	25
2. Document Administration	1	6.12 Area 9a/9b (CJK Unified Ideograph Extension-A)	28
2.1 Originating Area and Responsibility	1	7. Code Translation Correspondence	29
2.2 Authorization	2	7.1 Code Translation Correspondence of GB 2312-80	
2.3 Compliance	2	Characters in GB 18030 Code	29
2.4 Property Statement	2	7.1.1 Between GB 18030 National Standard and	
3. Related Documents	2	GB-PC/DBCS-EUC	29
3.1 Superseded Documents	2	7.1.2 Between GB 18030 DBCS-Host and GB-Host	29
3.2 Referenced External Documents	2	7.1.3 Between GB 18030 National Standard and GB 18030	
3.3 Referenced IBM Documents	2	DBCS-Host	29
3.4 Copyright Permission	2	7.2 Code Translation Correspondence of non-GB 2312-80	
Requirements	3	Characters in GB 18030 Code	29
4. External Standards	3	7.2.1 Between GB 18030 National Standard and	
4.1 PRC National Standard GB 2312-80	3	GB-PC/GB-Host/DBCS-EUC	29
4.2 ISO 10646-1	3	7.2.2 Between GB 18030 DBCS-Host and	
4.3 GB 13000	3	GB-PC/GB-Host/DBCS-EUC	29
4.4 GB 18030	3	7.2.3 Between GB 18030 National Standard and GB 18030	
4.4.1 GB 18030 two-byte code set	5	DBCS-Host	29
4.4.2 GB 18030 four-byte code set	5	7.3 Code Translation Correspondence of User Definable	
4.4.3 Implementation Phase 1 and 2	5	Characters in GB 18030 Code	30
4.4.4 Relationship of GB18030 DBCS-Host with GB18030	5	7.3.1 Code Translation Correspondence between GB 18030	
5. Specification of Chinese Repertoire	6	National Standard and GB 18030	30
5.1 Identification of Graphic Characters	6	7.3.2 Code Translation Correspondence between GB 18030	
6. Character Sets and Code Pages	6	National Standard and DBCS-EUC	30
6.1 Registration of GCSGID and CPGID	6	7.3.3 Code Translation Correspondence between GB 18030	
6.2 GB 18030 National Standard Specification	9	DBCS-Host and GB-Host	31
6.2.1 GB 18030 two-byte code set	9	7.3.4 Code Translation Correspondence between GB 18030	
6.2.2 Area 1 (Non-Chinese Character Set)	11	DBCS-Host and GB-PC	31
6.2.2.1 GB 2312-80 Set	11	7.3.5 Code Translation Correspondence between GB 18030	
6.2.2.2 Non GB 2312-80 Set	11	DBCS-Host and DBCS-EUC	31
6.2.3 Area 2 (Chinese Character Set)	12	7.4 The Code Correspondence for IBM Selected Characters	32
6.2.4 Area 3 (Chinese Character Set)	12	Supplementary Information	34
6.2.5 Area 4 (Chinese Character Set)	12	8. Definition of Terms	34
6.2.5.1 ISO 10646-1 CJK Unified Ideographs	12	8.1 Byte	34
6.2.5.2 ISO 10646-1 CJK Compatibility Ideographs	12	8.2 Code	34
6.2.5.3 Additional Chinese Characters and		8.3 Code Page	34
Radicals/Components	12	8.4 Code Page Global Identifier (CPGID)	34
6.2.6 Area 5 (Non-Chinese Character Set)	15	8.5 Code Point	34
6.3 GB 18030 four-byte code set Specification	19	8.6 Coded Graphic Character	34
6.4 GB 18030 DBCS-Host Specification	20	8.7 Coded Graphic Character Set	34
6.5 Area 1 (Non-Chinese Character Set)	22	8.8 Coded Graphic Character Set Global Identifier (CGCSGID)	34
6.5.1 Non-Chinese Character Set	22	8.9 DBCS	34
6.5.2 Extended Non-Chinese Character Set	22	8.10 DBCS-EUC	34
6.6 Area 2 (GB 18030 Extension: Uyгур)	22	8.11 DBCS-Host	34
6.7 Area 3 (GB 2312-80 Chinese Character Set)	22	8.12 DBCS-PC	34
6.8 Area 4 (GB 18030 Extension: Tibetan and others)	22	8.13 EBCDIC	34
		8.14 GBK	34
		8.15 Graphic Character	34
		8.16 Graphic Character Global Identifier (GCGID)	35
		8.17 Graphic Character Set	35
		8.18 Graphic Character Set Global Identifier (GCSGID)	35



IBM Simplified Chinese Graphic Character Set, GB 18030 code

National Standard and DBCS-Host

8.19 PC Code 35
8.20 Product 35
8.21 Registered Code Page 35
8.22 Registered Coded Graphic Character Set 35
8.23 Unrestricted Code Page 35
8.24 User Definable Character 35
8.25 Ward 35
9. History of CH-S Coded Character Set Specification 36
9.1 DBCS-PC(GB National Standard) code 36
9.1.1 Change from GB to GBK (C-H 3-3220-130 1993-11 to C-H 3-3220-020 1997-02) 36
9.1.1.1 Overall 36
9.1.1.2 Character Coverage 36
9.1.1.3 Code Translation Correspondence from C-H 3-3220-130 to C-H 3-3220-020) 36
9.1.2 Change from GBK to GB18030 (C-H 3-3220-020 1997-02 to C-H 3-3220-020 2001-03) 37
9.1.2.1 Overall 37
9.1.2.2 Character Coverage 37
9.1.2.3 Code Translation Correspondence from C-H 3-3220-020 1997-02 to C-H 3-3220-020 2001-03) 37
9.2 DBCS-Host code 38
9.2.1 Change from DBCS-Host to GBK DBCS-Host (C-H 3-3220-130 1993-11 to C-H 3-3220-020 1997-02) 38
9.2.1.1 Overall 38
9.2.1.2 Character Coverage 38
9.2.1.3 Code Translation Correspondence from C-H 3-3220-130 to C-H 3-3220-020) 38
9.2.2 Change from GBK DBCS-Host to GB18030 DBCS-Host (C-H 3-3220-020 1997-02 to 2001-03) 39
9.2.2.1 Overall 40
9.2.2.2 Character Coverage 40
9.2.2.3 Code Translation Correspondence from C-H 3-3220-020 1997-02 to C-H 3-3220-020 2001-03) 40
10. Graphic Character Representation 41

Tables

1. Character Coverage of GB 18030 two-byte code set 5
2. Character Coverage of GB 18030 four-byte code set 5
3. Registration of GCSGID and CPGID for GB 18030 Code 6
4. Specification of GB 18030 two-byte code set 9
5. Non-GB 2312 Symbol, Area 1, GB 18030 National Standard 11
6. ISO 10646-1 CJK Compatibility Ideographs, Area 4, GB 18030 National Standard 12
7. Additional Chinese characters and radicals/components, Area 4, GB 18030 National Standard 12
8. Non-Chinese Character Set, Area 5, GB 18030 National Standard 15
9. Code Assignment of GB 18030 four-byte code set 19
10. Specification of GB 18030 DBCS-Host 20
11. Non-GB 2312 Symbol, Area 1, GB 18030 DBCS-Host 22
12. ISO 10646-1 CJK Compatibility Ideographs, Area 6b, GB 18030 DBCS-Host 23
13. Additional Chinese characters and radicals/components, Area 6b, GB 18030 DBCS-Host 23
14. Non-Chinese Character Set, Area 7, GB 18030 DBCS-Host 25
15. The code correspondence for IBM selected characters 32



IBM Simplified Chinese Graphic Character Set, GB 18030 code

National Standard and DBCS-Host

Introduction

1. Scope

1.1 Abstract

This document is a registry of the IBM Simplified Chinese *Graphic Character Set*, GB 18030 Code, National Standard and DBCS-Host, in which the following are set forth:

- Repertoires of the IBM Simplified Chinese *Graphic Character Set*, GB 18030 Code, National Standard and DBCS-Host, are specified.
- Each *graphic character* of the repertoire is identified by a *Graphic Character Global Identifier (GCGID)* or *Graphic Character UCS Identifier (GCUID)*.
- *graphic character sets* are identified by *Graphic Character Set Global Identifiers (GCSGIDs)*.
- three *code pages*: GB 18030 Code, National Standard and GB 18030 Code, DBCS-Host for the *graphic character sets* are specified and identified by *Code Page Global Identifiers (CPGIDs)*.
- Code translation correspondence between GB 18030 Code, National Standard and GB 18030 Code, DBCS-Host is specified.
- Code translation correspondence between GB 18030 Code, National Standard or GB 18030 Code, DBCS-Host and the following code pages are specified.
 - *DBCS-PC*
CPGID 01380 in C-S 3-3220-130
 - *DBCS-Host*
CPGID 00837 in C-S 3-3220-130
 - *DBCS-EUC*
CPGID 01382 in C-S 3-3220-132
 - GB 2312-80

1.2 Objective

The objectives of this document are:

- 1) To make available to all developers supporting CH-S the specification of the CH-S *graphic character sets* and *code pages* for GB 18030 Code.
- 2) To ensure consistent implementation of CH-S *graphic character sets* and *code pages* for GB 18030 Code.

Note: This document is one of a set of documents defined in C-S 3-3220-019, *Coded Character Sets, Implementation*. The reader should be familiar with the requirements specified in C-S 3-3220-019 which describe the relationship of the documents. Note also that the requirements for compliance with this specification are described in C-S 3-3220-019.

Terms in italics (except in headings) are defined in the Supplementary Information section.

“Simplified Chinese” will be abbreviated to “CH-S” hereinafter.

X'nn', X'nnnn' or X'nnnnnnnn' represents a hexadecimal notation.

“GB 18030 National Standard” will be synonymously used for “GB 18030 Code, National Standard.,” and “GB 18030 DBCS-Host” will be synonymously used for “GB 18030 Code, DBCS-Host.”

1.3 Application

This specification is applicable to all IBM *products* which implement the GB 18030 Code as specified in Section 1.1.

1.4 Effective Date

This document takes effect with its publication.

2. Document Administration

2.1 Originating Area and Responsibility

This document is maintained by the Standards Project Authority (SPA) for:

SIRS Code: 029 (Double-Byte Character Set)



IBM Simplified Chinese Graphic Character Set, GB 18030 code

National Standard and DBCS-Host

Location Code: 93K (Yamato)

Division: NHD

SCRL Group: None

Page Count: 41

(UCS), Part One: Systematic Structure and Basic Multi-Language Plane (idt ISO /IEC 10646.1: 1993)

2.2 Authorization

Publication of this document was approved by the Standards Authorities of the affected operating units.

2.3 Compliance

Compliance with the requirements of this document is recommended unless it is called out as a requirement in a standard or equivalent document. It then assumes the same compliance requirements as that document.

2.4 Property Statement

This document is the property of IBM. Its use outside IBM is authorized only for responding to a request for quotation or for the performance of work for IBM. All supplier/vendor questions must be referred to the IBM purchasing department.

3. Related Documents

3.1 Superseded Documents

C-H 3-3220-020 dated 1997-02

IBM Simplified Chinese Graphic Character Set, GBK Code

Note: Changes in this edition are indicated by the character “|” (vertical line) in the revision column.

3.2 Referenced External Documents

GB 2312-80

PRC National Standard, Code of Chinese Graphic Character Set for Information Interchange, Primary Set

GB 11383-1989

PRC National Standard, Information Processing - Structure and Encoding Rules for 8-bit Code for Information Interchange (ISO 4873: 1986)

GB 13000.1-1993

*PRC National Standard, Information Technology * Universal Multiple Octet Coded Character Set*

GB 18030-2000

Information technology - Chinese ideograms coded character set for information interchange - Extension for the basic set

ISO 646

ISO Standard for 7-Bit Coded Character Set for Information Interchange

ISO 10646-1

ISO Information technology - Universal Multiple-Octet Coded Character Set (UCS), Part 1: Architecture and Basic Multilingual Plane

3.3 Referenced IBM Documents

C-S 3-3220-002

Extended BCD Interchange Code (EBCDIC)

C-S 3-3220-019

Coded Character Sets, Implementation

C-H 3-3220-050

Registry, Graphic Character Sets and Code Pages, Online Database

C-H 3-3220-055

Graphic Character Identification System, Graphic Character Global Identifier (GCGID) Structure

C-S 3-3220-102

Double-Byte Character Set (DBCS), Terminology and Code Scheme

C-S 3-3220-130

IBM Simplified Chinese Graphic Character Set, DBCS-Host and DBCS-PC

C-S 3-3220-132

IBM Simplified Chinese Graphic Character Set for Extended UNIX Code (EUC)

3.4 Copyright Permission

Not applicable.

IBM Simplified Chinese Graphic Character Set, GB 18030 code

National Standard and DBCS-Host

Requirements

4. External Standards

4.1 PRC National Standard GB 2312-80

GB 2312-80, PRC National Standard, Code of Chinese Graphics Characters Set for Information Interchange, Primary Set, published in March 1981, specifies a primary set of graphics characters with their binary-coded representation for Chinese information interchange. It applies to Chinese information-processing systems, communication systems and so on. It covers 682 non-Chinese characters and 6763 Chinese characters, 7445 graphic characters in total.

The non-Chinese characters include general characters, ordinal numbers, numerical characters, Latin alphabet, Japanese Kana, Greek alphabet, Russian alphabet, Chinese phonetic symbols and Chinese phonetic-annotated letters.

The Chinese Characters are divided into 2 levels, 3755 of them are included in Level 1 and 3008 in Level 2, 6763 Chinese characters in total.

4.2 ISO 10646-1

ISO 10646 ("ISO 10646-1" hereinafter) is international standard for the universal character encoding scheme for written characters and text. It defines a consistent way of encoding multilingual text that enables the exchange of text data internationally and creates the foundation for global software. It contains the characters from major international and national standards as well as prominent industry character sets. It contains 20902 CJK unified ideographs and additional 6582 ideographs.

4.3 GB 13000

GB13000, PRC National Standard, published in December 1994 is equivalent to ISO 10646-1(:1993).

4.4 GB 18030

GB 18030, PRC National Standard, contains all characters defined in ISO 10646-1, but they have totally different code assignment. In GB 18030, one-byte, two-byte and four-byte encoding systems are adopted. The total capability is over 1.5 millions of code positions. Currently, GB 18030 contains more than 27000 Chinese characters which have been defined in the latest version of ISO 10646-1. This standard provides utter solutions for the urgent needs of Chinese characters used in names and addresses.

Figure 1 on page 4 shows the code range of GB18030. In the one-byte portion, encoding structure and rules set in GB 11383 are adopted, using codes from X'00' to X'80'. In the two-byte portion, two 8-bit binary strings are used to represent one character, whose first byte includes codes from X'81' to X'FE', and the end byte includes codes from respectively X'40' to X'7E' and X'80' to X'FE'. In the four-byte portion, codes from X'30' to X'39' are used as suffix of extended two-byte codes. This extended four-byte codes range from X'81308130' to X'FE39FE39'. The codes of four-byte characters start from the fourth byte, the code-position is from X'30' to X'39', then the third byte, its code-position is from X'81' to X'FE', and the the second byte, its code-position is from X'30' to X'39', finally the first byte, its code-position is from X'81' to X'FE'. Figure 2 on page 4 shows code structure of four-byte characters.

IBM Simplified Chinese Graphic Character Set, GB 18030 code

National Standard and DBCS-Host

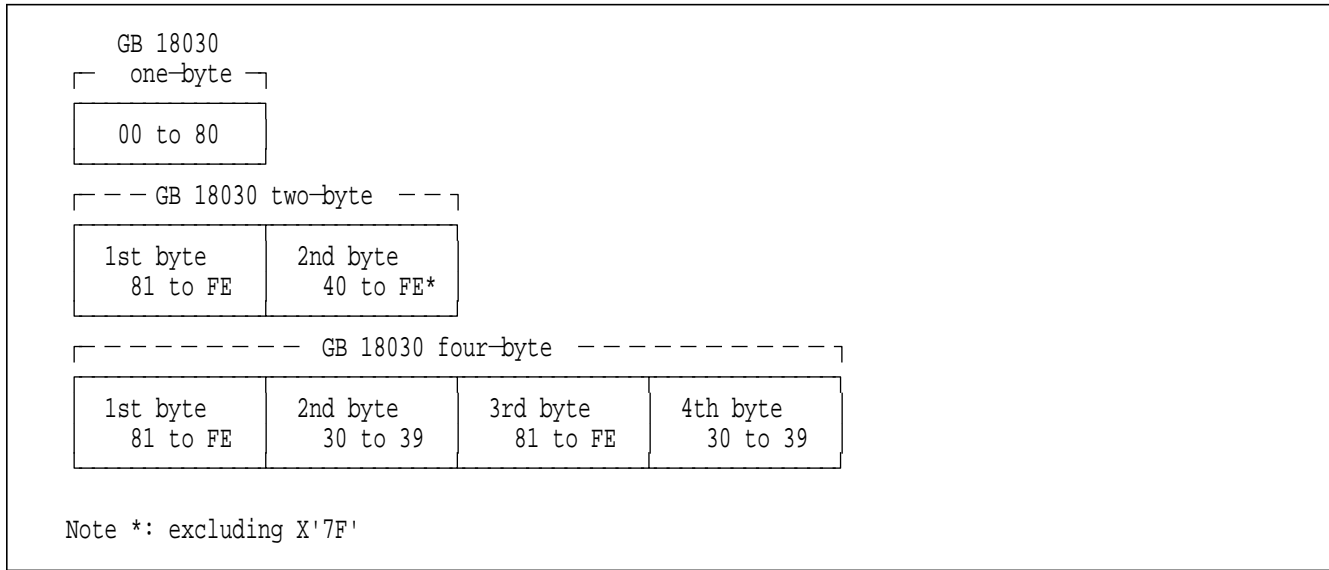


Figure 1. Code range of GB 18030

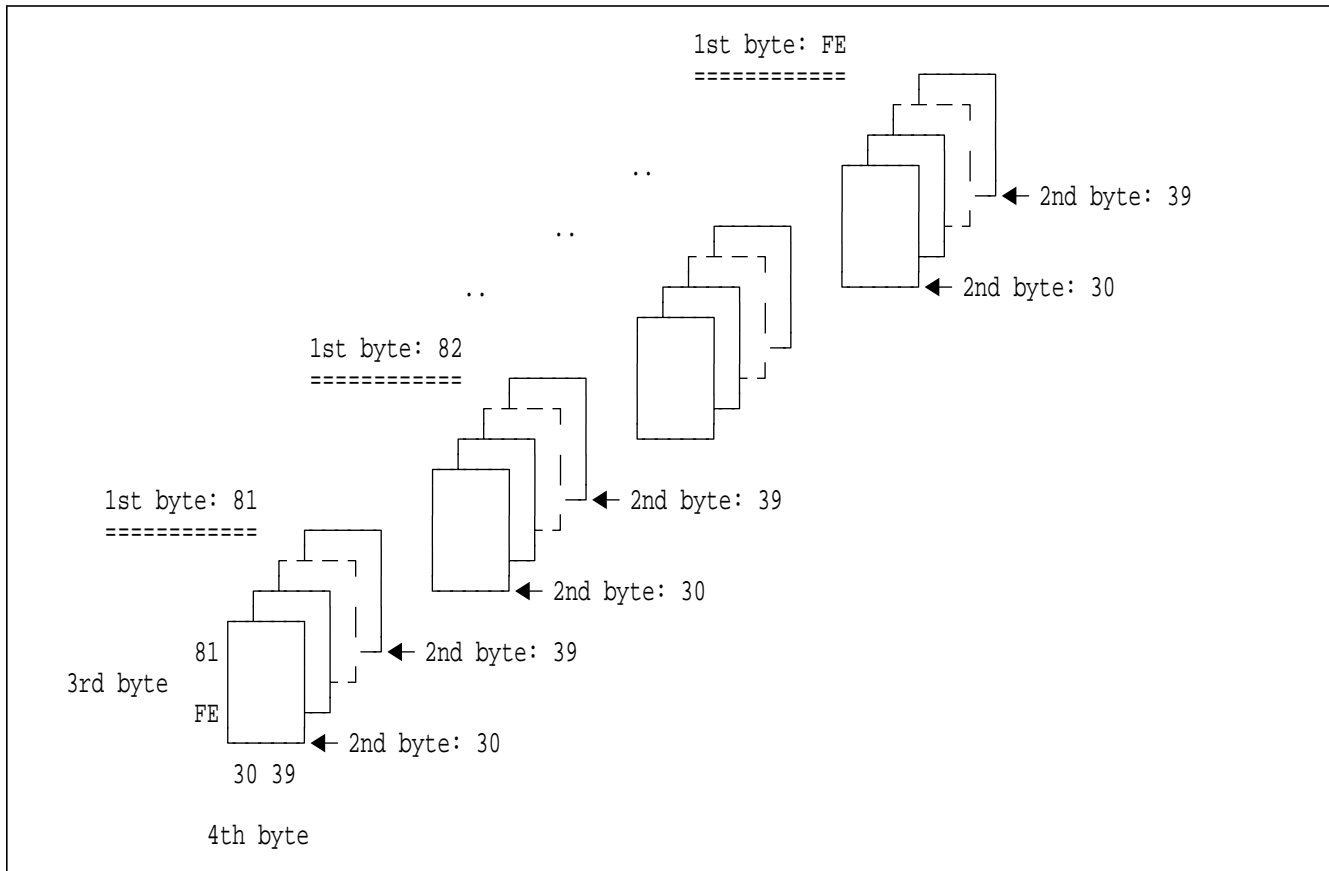


Figure 2. Code structure of GB 18030 four-byte code set

IBM Simplified Chinese Graphic Character Set, GB 18030 code

National Standard and DBCS-Host

4.4.1 GB 18030 two-byte code set. GB 18030 two-byte code set is the super set of GB 2312-80 (PRC National Standard) and totally compatible with that. GB 18030 two-byte code set expands GB 2312-80 character set to 20902 characters in all of ISO 10646-1 CJK unified ideographs and contains additional *DBCS* symbols defined in Big-5 code (a de-facto industry standard for traditional Chinese used in PC of Taiwan, R.O.C.) and Euro currency sign, as shown in Table 1.

Table 1. Character Coverage of GB 18030 two-byte code set

Content	Number of Characters
Non-Chinese Character Set <ul style="list-style-type: none"> • GB 2312-80 set (682) <ul style="list-style-type: none"> – Latin Alphabet – Greek – Russian – Japanese Katakana/Hiragana – Numeric – Special Symbols • non-Chinese characters (187)*1 <ul style="list-style-type: none"> – Vertical Forms – Box Drawing – Other Special Symbols – Euro Currency Sign • Chinese Radicals (28)*2 • Ideographic Structure Symbols (13)*3 • PinYin*4(2)*5 	912
Chinese Character Set <ul style="list-style-type: none"> • ISO 10646-1 CJK unified ideographs, including GB 2312-80 set (20 902) • ISO 10646-1 CJK compatibility ideographs (21) • ISO 10646-1 CJK unified ideographs Extension-A (52)*6 	20975
User Definable Character Set	1 894
Notes: The number in parentheses indicates the number of characters assigned. <ol style="list-style-type: none"> 1) 34 characters with no asterisk (GCGID column) in Table 5 on page 11 and 153 characters with no asterisk (GCGID column) in Table 8 on page 15. 2) identified by double asterisk in Table 7 on page 12. 3) identified by double asterisk in Table 8 on page 15. 4) used for Latin Transcription of Mandarin Chinese. 5) identified by double asterisk in Table 5 on page 11. 6) characters with no asterisk (GCGID column) in Table 7 on page 12. 	

4.4.2 GB 18030 four-byte code set. All character set in ISO 10646 except the above two-byte characters are specified in GB 18030 four-byte code set. Table 2 shows the character coverage of four-byte code set.

Table 2. Character Coverage of GB 18030 four-byte code set

Content	Number of Characters
Non-Chinese Character Set <ul style="list-style-type: none"> • Symbols (9 245), including <ul style="list-style-type: none"> – Uygur (Arabic in ISO 10646-1) (206) – Tibetan (193) – Mongolian (155) – Yi Syllables (1 165) – Yi Radical (50) • Hangul Syllables (11 172) 	20 417
Chinese Character Set <ul style="list-style-type: none"> • ISO 10646-1 CJK unified ideographs Extension-A (6 530)*1 • ISO 10646-1 CJK compatibility ideographs (281) 	6 811
Notes: 1) 52 characters out of Extension-A 6 582 are contained in GB 18030 two-byte code set.	

4.4.3 Implementation Phase 1 and 2. Two implementation phases were defined by Chinese government. CJK Unified Ideograph Extension-A and some minor scripts in China which are Uygur, Tibetan, Mongolian and Yi syllables/radicals need to be supported in phase 1. The others need to be supported in phase 2. Note that Uygur scripts (used in northern west of China) are equivalent to Arabic scripts with extensions involving both BiDi and character shape handling for proper rendering.

4.4.4 Relationship of GB18030 DBCS-Host with GB18030. GB18030 DBCS-Host in this specification has equivalent character coverage to GB18030 phase 1 in Section 4.4.3.



IBM Simplified Chinese Graphic Character Set, GB 18030 code

National Standard and DBCS-Host

5. Specification of Chinese Repertoire

5.1 Identification of Graphic Characters

This document specifies the repertoire of the CH-S *graphic character set* for GB 18030 Code.

Identification of Graphic Character refers to C-H 3-3220-055. It is identified using Graphic Character UCS Identifier (GCUID) or Graphic Character Global Identifier (GCGID).

The GCUID format is for defining additional characters and sets of characters that (mostly) exist in UCS-4 and need to be used on IBM code pages. The format allows all current and future characters from UCS-2 and at least the first sixteen additional planes from UCS-4 to be described.

The GCGID identifies graphic characters defined by IBM. The GCGID definition uniformly associates arbitrary graphic character shape with an eight character identifier GCGID.

6. Character Sets and Code Pages

6.1 Registration of GCSGID and CPGID

This specification registers the growing character set with the minimum repertoires and the subsets of the IBM CH-S *Graphic Character Set* for GB 18030 Code, and identifies them by *Graphic Character Set Global Identifiers (GCSGIDs)*.

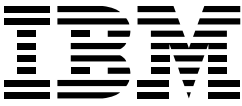
It also specifies a relevant code page for its *graphic character sets* and identifies it by *Code Page Global Identifier (CPGID)*.

Table 3 specifies the *GCSGIDs* and the *CPGIDs* of the *registered coded graphic character set* for GB 18030 Code.

The registration category “U” means “Unrestricted,” which implies that *graphic character sets* and *code pages* in this category have potential broad implementation in information processing *products*.

Note: See C-H 3-3220-050 for the detail of registration categories of *graphic character sets* and *code pages*.

Table 3 (Page 1 of 3). Registration of GCSGID and CPGID for GB 18030 Code				
CGCSGID GCSGID CPGID	Name	Set Size	References	Category of Registration
65535 01385	GB 18030 two-byte code set including 7 445 GB 2312-80 characters, 35 Non-Chinese characters on GB 2312-80 area, 194 Non-Chinese characters, 14 212 Chinese characters, 1 894 UDCs, Euro sign, and other growing characters Wards 81 to FE	23 781 (Mini - mum)	Area 1, 2, 3, 4, 5, 6, 7 and 8 of Table 4 on page 9	U

**IBM Simplified Chinese Graphic Character Set, GB 18030 code**

National Standard and DBCS-Host

Table 3 (Page 2 of 3). Registration of GCSGID and CPGID for GB 18030 Code

CGCSGID GCSGID CPGID	Name	Set Size	References	Category of Regis- tration
02084 01385	GB 18030 two-byte code set including 7 445 GB 2312-80 characters, 35 Non-Chinese characters on GB 2312-80 area, 194 Non-Chinese characters, 14 212 Chinese characters, 1 894 UDCs, and Euro sign Wards 81 to FE	23 781	Area 1, 2, 3, 4, 5, 6, 7 and 8 of Table 4 on page 9	R
01084 01385	GB 18030 two-byte code set excluding Euro sign without UDCs	21 886	Area 1, 2, 3, 4 and 5 of Table 4 on page 9	R
65535 01391	GB 18030 four-byte code set for all characters in ISO 10646 except characters in GB 18030 two-byte code set and other growing characters	27 228 (Mini - mum)	Area 1 of Table 9 on page 19	U
02085 01391	GB 18030 four-byte code set, for all characters in ISO 10646 except characters in GB 18030 two-byte code set	27 228	Area 1 of Table 9 on page 19	R
65535 00837	GB 18030 DBCS-Host including 7 445 GB 2312-80 characters, 232 Non-Chinese characters, 14 212 Chinese characters, 1 894 UDCs, Euro sign, 206 Uygur characters, 193 Tibetan characters, 155 Mongolian characters, 1 165 Yi Syllables, 50 Yi Radical, 6 530 CJK Unified Ideograph Extension-A, and other growing characters Wards 81 to FE	32 083 (Mini - mum)	Area 0, 1, 2, 3, 4, 5, 6a, 6b, 7, 8, 9a and 9b of Table 10 on page 20	U

**IBM Simplified Chinese Graphic Character Set, GB 18030 code**

National Standard and DBCS-Host

Table 3 (Page 3 of 3). Registration of GCSGID and CPGID for GB 18030 Code

CGCSGID GCSGID CPGID		Name	Set Size	References	Category of Regis- tration
02086	00837	GB 18030 DBCS-Host including 7 445 GB 2312-80 characters, 232 Non-Chinese characters, 14 212 Chinese characters, 1 894 UDCs, Euro sign, 206 Uygur characters, 193 Tibetan characters, 155 Mongolian characters, 1 165 Yi Syllables, 50 Yi Radical, 6 530 CJK Unified Ideograph Extension-A, Wards 81 to FE	32 083	Area 0, 1, 2, 3, 4, 5, 6a, 6b, 7, 8, 9a and 9b of Table 10 on page 20	R
01085	00837	GB 18030 DBCS-Host excluding Euro sign, 206 Uygur characters, 193 Tibetan characters, 155 Mongolian characters, 1 165 Yi Syllables, 50 Yi Radical, 6 530 CJK Unified Ideograph Extension-A, Wards 81 to FE without UDCs	21 889	Area 0, 1, 3, 5, 6a, 6b and 7 of Table 10 on page 20	R

IBM Simplified Chinese Graphic Character Set, GB 18030 code

National Standard and DBCS-Host

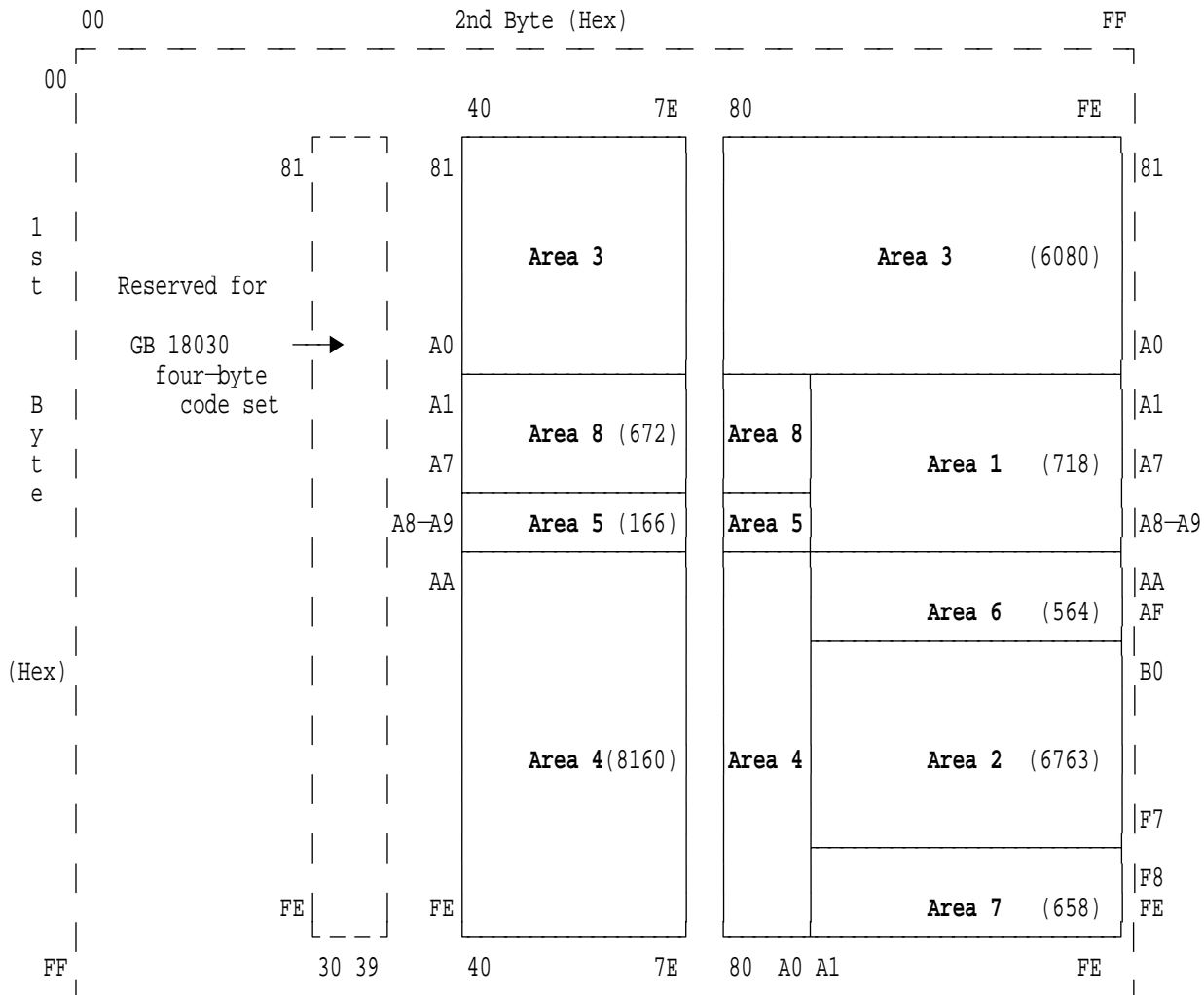
6.2 GB 18030 National Standard Specification | **6.2.1 GB 18030 two-byte code set** Table 4 summarizes
 | In this section, the specification of GB 18030 National | the code specification of GB 18030 two-byte code set
 | Standard *coded graphic character sets* is provided. | and Figure 3 on page 10 shows the structure of it.

Table 4. Specification of GB 18030 two-byte code set

Area	First Byte	Second Byte	Contents	Number of Characters	Specification
1	A1 to A9	A1 to FE	GB 2312-80 Non-Chinese Character Set	682	Equivalent to CH-S DBCS-PC (CPGID 01380) Refer to C-H 3-3220-130
			Extended Non-Chinese Character Set (Non GB 2312-80)	36	Table 5 on page 11
2	B0 to F7	A1 to FE	GB 2312-80 Chinese Character Set	6763	Equivalent to CH-S DBCS-PC (CPGID 01380) Refer to C-H 3-3220-130
3	81 to A0	40 to FE(*1)	ISO 10646-1 CJK unified ideographs	6080	ISO 10646-1 CJK unified ideographs are sequentially assigned except GB 2312-80 Chinese Character Set
4	AA to FE	40 to A0(*1)	ISO 10646-1 CJK unified ideographs	8059	
			ISO 10646-1 CJK compatibility ideographs	21	
			Additional Chinese characters, radicals/components	80	Table 7 on page 12
5	A8 to A9	40 to A0(*1)	Extended Non-Chinese Character Set	166	Table 8 on page 15
6	AA to AF	A1 to FE	User Definable Character (Part-1)	564	UDCs sequentially assigned
7	F8 to FE	A1 to FE	User Definable Character (Part-2)	658	
8	A1 to A7	40 to A0(*1)	User Definable Character (Part-3)	672	
Note:					
1) X'7F' is excluded.					

IBM Simplified Chinese Graphic Character Set, GB 18030 code

National Standard and DBCS-Host



- Notes:
- 1) The number in parentheses indicates the number of characters assigned.
 - 2) The code point of DBCS Space is X'A1A1'.
 - 3) Each area includes the following character set:
 - Area 1, 5 ... Non-Chinese Character Set
(Area 1 ... including GB2312-80 Non-Chinese character set)
 - Area 2, 3, 4 ... Chinese Character Set
(Area 2 ... GB2312-80 Chinese character set)
 - Area 6, 7, 8 ... User Definable Character Area
 - 4) 128 code points in Area 1 are reserved for future assignment.
 - 5) 5 code points in Area 2 are reserved for future assignment.
 - 6) 26 code points in Area 5 are reserved for future assignment.

Figure 3. Code Structure for GB 18030 two-byte code set

IBM Simplified Chinese Graphic Character Set, GB 18030 code

National Standard and DBCS-Host

6.2.2 Area 1 (Non-Chinese Character Set)

6.2.2.1 GB 2312-80 Set. The specification of GB 2312-80 Non-Chinese Character Set (682 characters) in GB 18030 National Standard is equivalent to *DBCS-PC Code* of Non-Chinese Character Set in C-H 3-3220-130.

6.2.2.2 Non GB 2312-80 Set

Line Number	Graphic	GB 18030 National Standard	GB 18030 DBCS-Host	GCGID
1	i	A2A1	41B1***	NR010080
2	ii	A2A2	41B2***	NR020080
3	iii	A2A3	41B3***	NR030080
4	iv	A2A4	41B4***	NR040080
5	v	A2A5	41B5***	NR050080
6	vi	A2A6	41B6***	NR060080
7	vii	A2A7	41B7***	NR070080
8	viii	A2A8	41B8***	NR080080
9	ix	A2A9	41B9***	NR090080
10	x	A2AA	41BA***	NR100080
11	€	A2E3	45EB	SC200080
12	˘	A6E0	4741	SP240080
13	˙	A6E1	4742	SP250080
14	˚	A6E2	4743	SP240081
15	ˇ	A6E3	4744	SP250081
16	ˆ	A6E4	4745	SP240082
17	˘	A6E5	4746	SP250082
18	˜	A6E6	4747	SP240083
19	˝	A6E7	4748	SP250083

Line Number	Graphic	GB 18030 National Standard	GB 18030 DBCS-Host	GCGID
20	˘	A6E8	4749	SP360080
21	˙	A6E9	474A	SP370080
22	˚	A6EA	474B	SP360081
23	˛	A6EB	474C	SP370081
24	˜	A6EE	474D	SP240084
25	˝	A6EF	477E	SP250084
26	˘	A6F0	474F	SP340080
27	˙	A6F1	4750	SP350080
28		A6F2	4751	SV070080
29		A6F4	4752	SV530080
30	{	A6F5	4753	SV450080
31	ɑ	A8BB	465B	LA010081
32	ɑ	A8BC*	465C	LM110080**
33	ɑ	A8BD	465D	LN110080
34	ɑ	A8BE	465E	LN210080
35	ɑ	A8BF	465F	LN130080**
36	g	A8C0	4660	LG010081

Note:

- * X'A8BC' is not contained in ISO 10646-1.
- ** They are PinYin characters used for Latin transcription of Mandarin Chinese.
- ***They are equivalent to DBCS-Host (C-H 3-3220-130).

IBM Simplified Chinese Graphic Character Set, GB 18030 code

National Standard and DBCS-Host

6.2.3 Area 2 (Chinese Character Set). The specification of Area 2 is equivalent to *DBCS-PC Code* of Chinese Character Sets (Level 1 and Level 2) in C-H 3-3220-130.

6.2.4 Area 3 (Chinese Character Set). On Area 3, ISO 10646-1 CJK unified ideographs are sequentially assigned except the characters corresponding to Chinese Character in Area 2.

6.2.5 Area 4 (Chinese Character Set)

6.2.5.1 ISO 10646-1 CJK Unified Ideographs. On Area 4, ISO 10646-1 CJK unified ideographs are sequentially assigned continuously from Area 3, except the characters corresponding to Chinese Character in Area 2.

6.2.5.2 ISO 10646-1 CJK Compatibility Ideographs

Table 6. ISO 10646-1 CJK Compatibility Ideographs, Area 4, GB 18030 National Standard

Line Number	Graphic	GB 18030 National Standard	GB 18030 DBCS-Host	GCGID
14	臈	FE48	CE4E	U000FA1F
15	蘊	FE49	CE4F	U000FA20
16	蚌	FE4A	CE50	U000FA21
17	赳	FE4B	CE51	U000FA23
18	返	FE4C	CE52	U000FA24
19	鐸	FE4D	CE53	U000FA27
20	鐸	FE4E	CE54	U000FA28
21	隄	FE4F	CE55	U000FA29

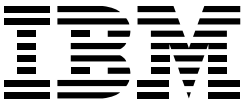
Table 6. ISO 10646-1 CJK Compatibility Ideographs, Area 4, GB 18030 National Standard

Line Number	Graphic	GB 18030 National Standard	GB 18030 DBCS-Host	GCGID
1	郎	FD9C	CE41	U000F92C
2	涼	FD9D	CE42	U000F979
3	季	FD9E	CE43	U000F995
4	裏	FD9F	CE44	U000F9E7
5	隣	FDA0	CE45	U000F9F1
6	兀	FE40	CE46	U000FA0C
7	設	FE41	CE47	U000FA0D
8	夔	FE42	CE48	U000FA0E
9	塔	FE43	CE49	U000FA0F
10	崎	FE44	CE4A	U000FA11
11	禡	FE45	CE4B	U000FA13
12	樺	FE46	CE4C	U000FA14
13	礼	FE47	CE4D	U000FA18

6.2.5.3 Additional Chinese Characters and Radicals/Components

Table 7 (Page 1 of 3). Additional Chinese characters and radicals/components, Area 4, GB 18030 National Standard

Line Number	Graphic	GB 18030 National Standard	GB 18030 DBCS-Host	GCGID
1	𠂇	FE50*	CE56	EL1Y0080**
2	𠂈	FE51	CE57	EL1Z0080**
3	𠂉	FE52	CE58	EL200080**
4	𠂊	FE53	CE59	EL210080**
5	𠂋	FE54*	CE5A	EL220080**
6	儻	FE55*	CE5B	EL230080
7	儻	FE56*	CE5C	EL240080
8	𠂌	FE57*	CE5D	EL250080**
9	𠂍	FE58*	CE5E	EL260080**
10	マ	FE59	CE5F	EL270080**



IBM Simplified Chinese Graphic Character Set, GB 18030 code

National Standard and DBCS-Host

Table 7 (Page 2 of 3). Additional Chinese characters and radicals/components, Area 4, GB 18030 National Standard				
Line Number	Graphic	GB 18030 National Standard	GB 18030 DBCS-Host	GCGID
11	唎	FE5A*	CE60	EL280080
12	唎	FE5B*	CE61	EL290080
13	唎	FE5C*	CE62	EL2A0080
14	灬	FE5D*	CE63	EL2B0080**
15	小	FE5E*	CE64	EL2C0080**
16	愒	FE5F*	CE65	EL2D0080
17	怡	FE60*	CE66	EL2E0080
18	手	FE61	CE67	EL2F0080**
19	扌	FE62*	CE68	EL2G0080
20	扌	FE63*	CE69	EL2H0080
21	扌	FE64*	CE6A	EL2I0080
22	扌	FE65*	CE6B	EL2J0080
23	丰	FE66	CE6C	EL2K0080**
24	丰	FE67	CE6D	EL2L0080**
25	柄	FE68*	CE6E	EL2M0080
26	殒	FE69*	CE6F	EL2N0080
27	灑	FE6A*	CE70	EL2O0080
28	生	FE6B*	CE71	EL2P0080**
29	夂	FE6C	CE72	EL2Q0080**
30	山	FE6D	CE73	EL2R0080**
31	正	FE6E*	CE74	EL2S0080**
32	睪	FE6F*	CE75	EL2T0080
33	穆	FE70*	CE76	EL2U0080
34	夂	FE71*	CE77	EL2V0080**

Table 7 (Page 2 of 3). Additional Chinese characters and radicals/components, Area 4, GB 18030 National Standard				
Line Number	Graphic	GB 18030 National Standard	GB 18030 DBCS-Host	GCGID
35	紬	FE72*	CE78	EL2W0080
36	卩	FE73*	CE79	EL2X0080**
37	𠂇	FE74*	CE7A	EL2Y0080**
38	𠂇	FE75*	CE7B	EL2Z0080**
39	𠂇	FE76	CE7C	EL300080**
40	糴	FE77*	CE7D	EL310080
41	糴	FE78*	CE7E	EL320080
42	丰	FE79**	CE7F	EL330080**
43	膊	FE7A*	CE81	EL350080
44	勞	FE7B*	CE82	EL360080
45	禛	FE7C*	CE83	EL370080
46	禛	FE7D*	CE84	EL380080
47	𠂇	FE7E	CE85	EL390080**
48	訢	FE80*	CE86	EL3A0080
49	讖	FE81*	CE87	EL3B0080
50	賻	FE82*	CE88	EL3C0080
51	賻	FE83*	CE89	EL3D0080
52	足	FE84*	CE8A	EL3E0080**
53	鎬	FE85*	CE8B	EL3F0080
54	𠂇	FE86*	CE8C	EL3G0080
55	𠂇	FE87*	CE8D	EL3H0080
56	鎬	FE88*	CE8E	EL3I0080
57	鎬	FE89*	CE8F	EL3J0080
58	鎬	FE8A*	CE90	EL3K0080



IBM Simplified Chinese Graphic Character Set, GB 18030 code

National Standard and DBCS-Host

Table 7 (Page 3 of 3). Additional Chinese characters and radicals/components, Area 4, GB 18030 National Standard

Line Number	Graphic	GB 18030 National Standard	GB 18030 DBCS-Host	GCGID
59	纓	FE8B*	CE91	EL3L0080
60	闕	FE8C*	CE92	EL3M0080
61	闌	FE8D*	CE93	EL3N0080
62	闈	FE8E*	CE94	EL3O0080
63	阡	FE8F*	CE95	EL3P0080
64	草	FE90	CE96	EL3Q0080**
65	𦉳	FE91	CE97	EL3R0080**
66	廡	FE92*	CE98	EL3S0080
67	𩚑	FE93*	CE99	EL3T0080
68	𩚒	FE94*	CE9A	EL3U0080
69	𩚓	FE95*	CE9B	EL3V0080
70	廩	FE96*	CE9C	EL3W0080
71	𩚕	FE97*	CE9D	EL3X0080
72	𩚖	FE98*	CE9E	EL3Y0080
73	𩚗	FE99*	CE9F	EL3Z0080
74	𩚘	FE9A*	CEA0	EL400080
75	𩚙	FE9B*	CEA1	EL410080
76	𩚚	FE9C*	CEA2	EL420080
77	𩚛	FE9D*	CEA3	EL430080
78	𩚜	FE9E*	CEA4	EL440080
79	𩚝	FE9F*	CEA5	EL450080

Table 7 (Page 3 of 3). Additional Chinese characters and radicals/components, Area 4, GB 18030 National Standard

Line Number	Graphic	GB 18030 National Standard	GB 18030 DBCS-Host	GCGID
80	𩚞	FEA0	CEA6	EL460080**

Note:

* These characters are contained in either CJK Unified Ideographic Extension-A or CJK Radicals Supplement of ISO10646-1.

** These characters are radicals.

**IBM Simplified Chinese Graphic Character Set, GB 18030 code**

National Standard and DBCS-Host

6.2.6 Area 5 (Non-Chinese Character Set)

Line Number	Graphic	GB 18030 National Standard	GB 18030 DBCS-Host	GCGID
1	ˊ	A840	CD41	SB380080
2	ˋ	A841	CD42	SB400080
3	˙	A842	CD43	SD290080 SB410080
4	ˉ	A843	CD44	SS680080
5	—	A844	CD45	SM120080
6	¨	A845	447E*	SV430080
7	˘	A846	CD46	SD130080
8	‰	A847	CD47	SS640080
9	°F	A848	CD48	SM850080
10	↖	A849	CD49	SM970080
11	↗	A84A	CD4A	SM950080
12	↘	A84B	CD4B	SM990080
13	↙	A84C	CD4C	SM980080
14	/	A84D	CD4D	SP120081
15	└	A84E	CD4E	SA420080
16	┘	A84F	CD4F	U0002223
17	≡	A850	CD50	SA700081
18	≡	A851	CD51	SA520082
19	≡	A852	CD52	SA530082
20	∠	A853	CD53	SA850080
21	—	A854	CD54	SF430080
22		A855	CD55	SF240080
23	┌	A856	CD56	SF510080

Line Number	Graphic	GB 18030 National Standard	GB 18030 DBCS-Host	GCGID
24	┐	A857	CD57	SF520080
25	┌	A858	CD58	SF390080
26	└	A859	CD59	SF220080
27	┐	A85A	CD5A	SF210080
28	┐	A85B	CD5B	SF250080
29	└	A85C	CD5C	SF500080
30	└	A85D	CD5D	SF490080
31	└	A85E	CD5E	SF380080
32	┘	A85F	CD5F	SF280080
33	┘	A860	CD60	SF270080
34	┘	A861	CD61	SF260080
35	┘	A862	CD62	SF360080
36	┘	A863	CD63	SF370080
37	┘	A864	CD64	SF420080
38	┘	A865	CD65	SF190080
39	┘	A866	CD66	SF200080
40	┘	A867	CD67	SF230080
41	┘	A868	CD68	SF470080
42	┘	A869	CD69	SF480080
43	┘	A86A	CD6A	SF410080
44	┘	A86B	CD6B	SF450080
45	┘	A86C	CD6C	SF460080
46	┘	A86D	CD6D	SF400080
47	┘	A86E	CD6E	SF540080
48	┘	A86F	CD6F	SF530080



IBM Simplified Chinese Graphic Character Set, GB 18030 code

National Standard and DBCS-Host

Table 8 (Page 2 of 4). Non-Chinese Character Set, Area 5, GB 18030 National Standard

Line Number	Graphic	GB 18030 National Standard	GB 18030 DBCS-Host	GCGID
49	+	A870	CD70	SF440080
50	∩	A871	CD71	SS160080
51	∪	A872	CD72	SS180080
52	∩	A873	CD73	SS190080
53	∪	A874	CD74	SS170080
54	/	A875	CD75	SH020080
55	\	A876	CD76	SH030080
56	×	A877	CD77	SH040080
57	—	A878	CD78	SF700081
58	—	A879	CD79	SF710081
59	■	A87A	CD7A	SF720081
60	■	A87B	CD7B	SF570080 SF730081
61	■	A87C	CD7C	SF740081
62	■	A87D	CD7D	SF750081
63	■	A87E	CD7E	SF760081
64	■	A880	CD7F	SF610080
65	■	A881	CD81	SF830081
66	■	A882	CD82	SF820081
67	■	A883	CD83	SF810081
68	■	A884	CD84	SF800081
69	■	A885	CD85	SF790081
70	■	A886	CD86	SF780081
71	■	A887	CD87	SF770081
72	■	A888	CD88	SF160080

Table 8 (Page 2 of 4). Non-Chinese Character Set, Area 5, GB 18030 National Standard

Line Number	Graphic	GB 18030 National Standard	GB 18030 DBCS-Host	GCGID
73	—	A889	CD89	SF670080
74		A88A	CD8A	SF650080
75	▼	A88B	44EC*	SV040080
76	▽	A88C	44EB*	SM740080
77	▲	A88D	CD8B	SF840080
78	▶	A88E	CD8C	SF850080
79	▶	A88F	CD8D	SF870080
80	▼	A890	CD8E	SF860080
81	⊙	A891	CD8F	U0002609
82	⊕	A892	CD90	SA550080
83	〒	A893	446C*	SS730080
84	"	A894	CD91	SP210081
85	"	A895	CD92	SP220081
86		A940	CD93	NC010080
87		A941	CD94	NC020080
88		A942	CD95	NC030080
89	×	A943	CD96	NC040080
90	∩	A944	CD97	NC050080
91	∪	A945	CD98	NC060080
92	≡	A946	CD99	NC070080
93	≡	A947	CD9A	NC080080
94	⊗	A948	CD9B	NC090080
95	⊕	A949	CD9C	SS780080
96	mg	A94A	CD9D	SS850080
97	kg	A94B	CD9E	SS860080

**IBM Simplified Chinese Graphic Character Set, GB 18030 code**

National Standard and DBCS-Host

Line Number	Graphic	GB 18030 National Standard	GB 18030 DBCS-Host	GCGID
98	mm	A94C	CD9F	SS810080
99	cm	A94D	CDA0	SS820080
100	km	A94E	CDA1	SS830080
101	m ²	A94F	CDA2	SS840080
102	cc	A950	CDA3	SS870080
103	KM	A951	CDA4	SS830081
104	ln	A952	CDA5	SS890080
105	log	A953	CDA6	SS880080
106	mil	A954	CDA7	SS800080
107	:	A955	CDA8	SV550080
108	┌	A956	425F*	SM660080
109	┆	A957	426A*	SM650080
110	TEL	A959	446F*	SS710080
111	(株)	A95A	446D*	SS740080
112	-	A95C	445A*	SP320080
113	一	A960	4358*	JX700080
114	ˆ	A961	43BE*	JX710080
115	°	A962	43BF*	JX720080
116	˘	A963	43DC*	JQ750080
117	˙	A964	43DD*	JQ760080
118	˚	A965	445E*	SS720080
119	˛	A966	44DC*	RQ750080
120	˜	A967	44DD*	RQ760080
121	---	A968	CDA9	SV470080
122	---	A969	CDAA	SV490080

Line Number	Graphic	GB 18030 National Standard	GB 18030 DBCS-Host	GCGID
123	----	A96A	CDAB	SV500080
124	---	A96B	CDAC	SV510080
125	---	A96C	CDAD	SV120080
126	---	A96D	CDAE	SV480080
127	----	A96E	CDAF	SV460080
128	,	A96F	CDB0	SP081080
129	,	A970	CDB1	SP081084
130	.	A971	CDB2	SP111080
131	;	A972	CDB3	SP141080
132	:	A973	CDB4	SP131080
133	?	A974	CDB5	SP151080
134	!	A975	CDB6	SP021080
135	(A976	CDB7	SP061080
136)	A977	CDB8	SP071080
137	{	A978	CDB9	SM111080
138	}	A979	CDBA	SM141080
139	(A97A	CDBB	SP061081
140)	A97B	CDBC	SP071081
141	#	A97C	CDBD	SM011080
142	&	A97D	CDBE	SM031080
143	*	A97E	CDBF	SM04008A
144	+	A980	CDC0	SA011080
145	-	A981	CDC1	SA001080
146	<	A982	CDC2	SA031080
147	>	A983	CDC3	SA051080

**IBM Simplified Chinese Graphic Character Set, GB 18030 code**

National Standard and DBCS-Host

Table 8 (Page 4 of 4). Non-Chinese Character Set, Area 5, GB 18030 National Standard

Line Number	Graphic	GB 18030 National Standard	GB 18030 DBCS-Host	GCGID
148	=	A984	CDC4	SA041080
149	\	A985	CDC5	SM071080
150	\$	A986	CDC6	SC031080
151	%	A987	CDC7	SM021080
152	@	A988	CDC8	SM051080
153	☎	A989	CDC9	SI010080**
154	☏	A98A	CDCA	SI020080**
155	☐	A98B	CDCB	SI030080**
156	☒	A98C	CDCC	SI040080**
157	☓	A98D	CDCD	SI050080**
158	☑	A98E	CDCE	SI060080**
159	☔	A98F	CDCF	SI070080**
160	☕	A990	CDD0	SI080080**
161	☖	A991	CDD1	SI090080**
162	☗	A992	CDD2	SI100080**
163	☘	A993	CDD3	SI110080**
164	☙	A994	CDD4	SI120080**
165	☚	A995	CDD5	SI130080**
166	○	A996	445F*	ND100087

Note:

* They are equivalent to DBCS-Host (C-H 3-3220-130).

** They are Ideographic Structure Symbols.



IBM Simplified Chinese Graphic Character Set, GB 18030 code

National Standard and DBCS-Host

6.3 GB 18030 four-byte code set Specification

The GB 18030 four-byte code set is structured as shown in Figure 2 on page 4 before. On this structure, all characters in ISO 10646 BMP (Basic Multilingual Plane) except characters in GB 18030 two-byte code set are sequentially assigned in the area from 1st byte X'81' to X'84'.

Table 9. Code Assignment of GB 18030 four-byte code set

Area	Contents	1st byte code range	Num. of char/Num. of Code Point
1	UCS-2 BMP including <ul style="list-style-type: none"> • CJK Unified Ideograph Extension-A • CJK Compatibility Ideograph • Minor scripts in China <ul style="list-style-type: none"> – Uygur (Arabic in ISO 10646-1) – Tibetan – Mongolian – Yi Syllables – Yi Radical • Other scripts 	81 to 84	27 228 /50 400
	Reserved for Future Assignments	85 to FE	0 /1 537 200

**IBM Simplified Chinese Graphic Character Set, GB 18030 code**

National Standard and DBCS-Host

6.4 GB 18030 DBCS-Host Specification

In this section, the specification of GB 18030 DBCS-Host coded graphic character sets is provided. Table 10 shows the summary of specification on the structure of Figure 4 on page 21.

Table 10. Specification of GB 18030 DBCS-Host

Area (First Byte)	Contents	Number of Characters	Specification
0 (X'4040')	Space	1	Equivalent to CH-S DBCS-Host Refer to C-H 3-3220-130
1 (41 to 47)	Non-Chinese Character Set incl. GB 2312-80 Non-Chinese Chars	712	
		Extended Non-Chinese Character Set incl. Euro sign	26
2 (47 to 48)	GB 18030 Extension: Uygur	206	ISO 10646-1 characters are sequentially assigned.
3 (48 to 6C)	GB 2312-80 Chinese Character Set	6763	Equivalent to CH-S DBCS-Host Refer to C-H 3-3220-130
4 (6D to 75)	GB 18030 Extension: Tibetan	193	ISO 10646-1 characters are sequentially assigned.
	GB 18030 Extension: Mongolian	155	
	GB 18030 Extension: Yi Syllables	1165	
	GB 18030 Extension: Yi Radical	50	
5 (81 to A1)	ISO 10646-1 CJK unified ideographs	6080	ISO 10646-1 CJK unified ideographs are sequentially assigned except GB 2312-80 Chinese Character Set
6a (A1 to CC)	ISO 10646-1 CJK unified ideographs	8059	
6b (CE)	ISO 10646-1 CJK compatibility ideographs	21	Table 12 on page 23
	Additional Chinese characters, radicals/components	80	Table 13 on page 23
7 (CD)	Extended Non-Chinese Char Set	148	Table 14 on page 25
8 (76 to 80)	User Definable Characters	1894	UDCs sequentially assigned
9a, 9b (CF to EF, F6 to F7)	GB 18030 Extension: CJK Unified Ideograph Extension-A	6530	ISO 10646-1 CJK unified ideographs are sequentially assigned except Chinese Characters on Area 6b

IBM Simplified Chinese Graphic Character Set, GB 18030 code

National Standard and DBCS-Host

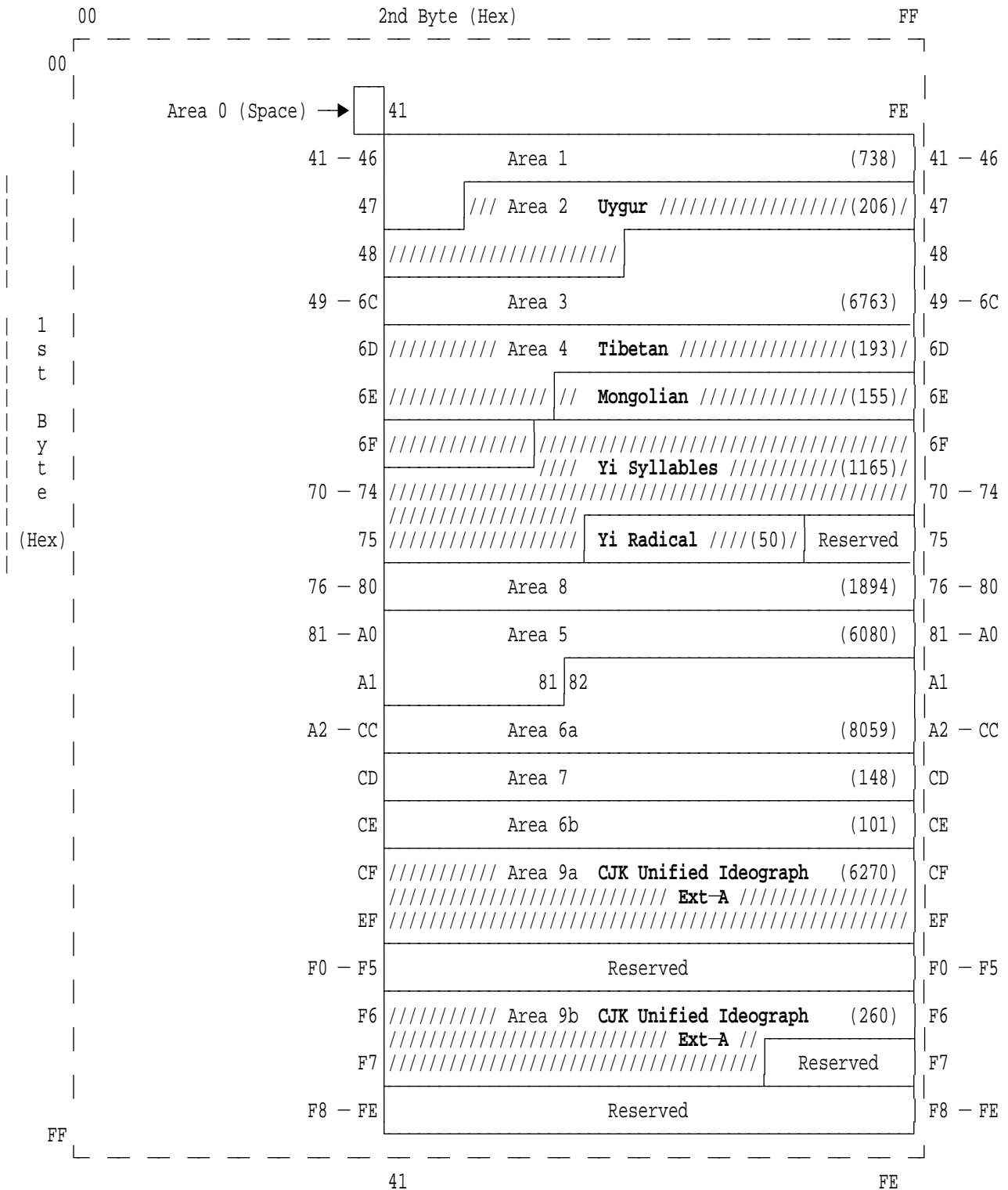


Figure 4. Code Structure for GB 18030 DBCS-Host

IBM Simplified Chinese Graphic Character Set, GB 18030 code

National Standard and DBCS-Host

6.5 Area 1 (Non-Chinese Character Set)

6.5.1 Non-Chinese Character Set. The specification of Area 1 Non-Chinese Character Set is equivalent to CH-S DBCS-Host (C-H 3-3220-130) Non-Chinese Character Set.

6.5.2 Extended Non-Chinese Character Set

Table 11. Non-GB 2312 Symbol, Area 1, GB 18030 DBCS-Host				
Line Number	Graphic	GB 18030 DBCS-Host	GB 18030 National Standard	GCGID
1	€	45EB	A2E3	SC200080
2	ɑ	465B	A8BB	LA010081
3	ř	465C	A8BC	LM110080
4	ń	465D	A8BD	LN110080
5	ň	465E	A8BE	LN210080
6	ñ	465F	A8BF	LN130080
7	g	4660	A8C0	LG010081
8	˘	4741	A6E0	SP240080
9	˙	4742	A6E1	SP250080
10	˚	4743	A6E2	SP240081
11	˛	4744	A6E3	SP250081
12	˜	4745	A6E4	SP240082
13	˝	4746	A6E5	SP250082
14	ˆ	4747	A6E6	SP240083
15	˜	4748	A6E7	SP250083
16	˘	4749	A6E8	SP360080
17	˙	474A	A6E9	SP370080
18	˚	474B	A6EA	SP360081
19	˛	474C	A6EB	SP370081

Table 11. Non-GB 2312 Symbol, Area 1, GB 18030 DBCS-Host				
Line Number	Graphic	GB 18030 DBCS-Host	GB 18030 National Standard	GCGID
20	˘	474D	A6EE	SP240084
21	˙	477E	A6EF	SP250084
22	˚	474F	A6F0	SP340080
23	˛	4750	A6F1	SP350080
24		4751	A6F2	SV070080
25		4752	A6F4	SV530080
26		4753	A6F5	SV450080

6.6 Area 2 (GB 18030 Extension: Uygur)

On Area 2, Uygur characters as Arabic on ISO 10646 U+0600 to U+06FF are sequentially assigned.

6.7 Area 3 (GB 2312-80 Chinese Character Set)

The specification of Area 3 (GB 2312-80 Chinese Character Set) is equivalent to *DBCS-Host Code* of Chinese Character Sets (Level 1 and Level 2) in C-H 3-3220-130.

6.8 Area 4 (GB 18030 Extension: Tibetan and others)

On Area 4, the following characters are sequentially assigned.

- Tibetan on ISO 10646 U+0F00 to U+0FFF
- Mongolian on ISO 10646 U+1800 to U+18FF
- Yi Syllables on ISO 10646 U+A000 to U+A48F
- Yi Radical on ISO 10646 U+A490 to U+A4CF

6.9 Area 5 (Chinese Character Set)

On Area 5 (Chinese Character Set), ISO 10646-1 CJK unified ideographs are sequentially assigned except the characters corresponding to Chinese Character in Area 3.

6.10 Area 6a/6b (Chinese Character Set)

IBM Simplified Chinese Graphic Character Set, GB 18030 code

National Standard and DBCS-Host

6.10.1 ISO 10646-1 CJK Unified Ideographs. On Area 6a (Chinese Character Set), ISO 10646-1 CJK unified ideographs are sequentially assigned continuously from Area 3, except the characters corresponding to Chinese Character in Area 2.

6.10.2 ISO 10646-1 CJK Compatibility Ideographs

Table 12. ISO 10646-1 CJK Compatibility Ideographs, Area 6b, GB 18030 DBCS-Host

Line Number	Graphic	GB 18030 DBCS-Host	GB 18030 National Standard	GCGID
1	郎	CE41	FD9C	U000F92C
2	凉	CE42	FD9D	U000F979
3	季	CE43	FD9E	U000F995
4	裏	CE44	FD9F	U000F9E7
5	隣	CE45	FDA0	U000F9F1
6	兀	CE46	FE40	U000FA0C
7	設	CE47	FE41	U000FA0D
8	雙	CE48	FE42	U000FA0E
9	塔	CE49	FE43	U000FA0F
10	崎	CE4A	FE44	U000FA11
11	柁	CE4B	FE45	U000FA13
12	榉	CE4C	FE46	U000FA14
13	礼	CE4D	FE47	U000FA18
14	臈	CE4E	FE48	U000FA1F
15	藕	CE4F	FE49	U000FA20
16	蚌	CE50	FE4A	U000FA21
17	赳	CE51	FE4B	U000FA23
18	返	CE52	FE4C	U000FA24
19	鏘	CE53	FE4D	U000FA27
20	鏘	CE54	FE4E	U000FA28

Table 12. ISO 10646-1 CJK Compatibility Ideographs, Area 6b, GB 18030 DBCS-Host

Line Number	Graphic	GB 18030 DBCS-Host	GB 18030 National Standard	GCGID
21	臨	CE55	FE4F	U000FA29

6.10.3 Additional Chinese Characters and Radicals/Components

Table 13 (Page 1 of 3). Additional Chinese characters and radicals/components, Area 6b, GB 18030 DBCS-Host

Line Number	Graphic	GB 18030 DBCS-Host	GB 18030 National Standard	GCGID
1	厂	CE56	FE50	EL1Y0080
2	ナ	CE57	FE51	EL1Z0080
3	ㄥ	CE58	FE52	EL200080
4	丿	CE59	FE53	EL210080
5	㇇	CE5A	FE54	EL220080
6	儻	CE5B	FE55	EL230080
7	佻	CE5C	FE56	EL240080
8	ㄎ	CE5D	FE57	EL250080
9	㇈	CE5E	FE58	EL260080
10	マ	CE5F	FE59	EL270080
11	囑	CE60	FE5A	EL280080
12	囑	CE61	FE5B	EL290080
13	囑	CE62	FE5C	EL2A0080
14	ㄣ	CE63	FE5D	EL2B0080
15	小	CE64	FE5E	EL2C0080
16	憐	CE65	FE5F	EL2D0080
17	怡	CE66	FE60	EL2E0080



IBM Simplified Chinese Graphic Character Set, GB 18030 code

National Standard and DBCS-Host

Table 13 (Page 2 of 3). Additional Chinese characters and radicals/components, Area 6b, GB 18030 DBCS-Host

Line Number	Graphic	GB 18030 DBCS-Host	GB 18030 National Standard	GCGID
18	𠂇	CE67	FE61	EL2F0080
19	𠂈	CE68	FE62	EL2G0080
20	𠂉	CE69	FE63	EL2H0080
21	𠂊	CE6A	FE64	EL2I0080
22	𠂋	CE6B	FE65	EL2J0080
23	𠂌	CE6C	FE66	EL2K0080
24	𠂍	CE6D	FE67	EL2L0080
25	𠂎	CE6E	FE68	EL2M0080
26	殯	CE6F	FE69	EL2N0080
27	𠂏	CE70	FE6A	EL2O0080
28	𠂐	CE71	FE6B	EL2P0080
29	𠂑	CE72	FE6C	EL2Q0080
30	𠂒	CE73	FE6D	EL2R0080
31	𠂓	CE74	FE6E	EL2S0080
32	𠂔	CE75	FE6F	EL2T0080
33	𠂕	CE76	FE70	EL2U0080
34	𠂖	CE77	FE71	EL2V0080
35	𠂗	CE78	FE72	EL2W0080
36	𠂘	CE79	FE73	EL2X0080
37	𠂙	CE7A	FE74	EL2Y0080
38	𠂚	CE7B	FE75	EL2Z0080
39	𠂛	CE7C	FE76	EL300080
40	𠂜	CE7D	FE77	EL310080
41	𠂝	CE7E	FE78	EL320080

Table 13 (Page 2 of 3). Additional Chinese characters and radicals/components, Area 6b, GB 18030 DBCS-Host

Line Number	Graphic	GB 18030 DBCS-Host	GB 18030 National Standard	GCGID
42	𠂞	CE7F	FE79	EL330080
43	𠂟	CE81	FE7A	EL350080
44	𠂠	CE82	FE7B	EL360080
45	𠂡	CE83	FE7C	EL370080
46	𠂢	CE84	FE7D	EL380080
47	𠂣	CE85	FE7E	EL390080
48	𠂤	CE86	FE80	EL3A0080
49	𠂥	CE87	FE81	EL3B0080
50	𠂦	CE88	FE82	EL3C0080
51	𠂧	CE89	FE83	EL3D0080
52	𠂨	CE8A	FE84	EL3E0080
53	𠂩	CE8B	FE85	EL3F0080
54	𠂪	CE8C	FE86	EL3G0080
55	𠂫	CE8D	FE87	EL3H0080
56	𠂬	CE8E	FE88	EL3I0080
57	𠂭	CE8F	FE89	EL3J0080
58	𠂮	CE90	FE8A	EL3K0080
59	𠂯	CE91	FE8B	EL3L0080
60	𠂰	CE92	FE8C	EL3M0080
61	𠂱	CE93	FE8D	EL3N0080
62	𠂲	CE94	FE8E	EL3O0080
63	𠂳	CE95	FE8F	EL3P0080
64	𠂴	CE96	FE90	EL3Q0080
65	𠂵	CE97	FE91	EL3R0080

IBM Simplified Chinese Graphic Character Set, GB 18030 code

National Standard and DBCS-Host

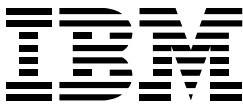
Table 13 (Page 3 of 3). Additional Chinese characters and radicals/components, Area 6b, GB 18030 DBCS-Host

Line Number	Graphic	GB 18030 DBCS-Host	GB 18030 National Standard	GCGID
66	廠	CE98	FE92	EL3S0080
67	鉀	CE99	FE93	EL3T0080
68	鎰	CE9A	FE94	EL3U0080
69	鎰	CE9B	FE95	EL3V0080
70	廠	CE9C	FE96	EL3W0080
71	騰	CE9D	FE97	EL3X0080
72	鸚	CE9E	FE98	EL3Y0080
73	鸚	CE9F	FE99	EL3Z0080
74	鸚	CEA0	FE9A	EL400080
75	鸚	CEA1	FE9B	EL410080
76	鸚	CEA2	FE9C	EL420080
77	鸚	CEA3	FE9D	EL430080
78	鸚	CEA4	FE9E	EL440080
79	葵	CEA5	FE9F	EL450080
80	繚	CEA6	FEA0	EL460080

6.11 Area 7 (Extended Non-Chinese Character Set)

Table 14 (Page 1 of 4). Non-Chinese Character Set, Area 7, GB 18030 DBCS-Host

Line Number	Graphic	GB 18030 DBCS-Host	GB 18030 National Standard	GCGID
1	ˊ	CD41	A840	SB380080
2	ˋ	CD42	A841	SB400080
3	˙	CD43	A842	SD290080 SB410080
4	ˉ	CD44	A843	SS680080
5	—	CD45	A844	SM120080
6	˘	CD46	A846	SD130080
7	%	CD47	A847	SS640080
8	°F	CD48	A848	SM850080
9	↖	CD49	A849	SM970080
10	↗	CD4A	A84A	SM950080
11	↘	CD4B	A84B	SM990080
12	↙	CD4C	A84C	SM980080
13	/	CD4D	A84D	SP120081
14	└	CD4E	A84E	SA420080
15		CD4F	A84F	U0002223
16	≡	CD50	A850	SA700081
17	≡	CD51	A851	SA520082
18	≡	CD52	A852	SA530082
19	∠	CD53	A853	SA850080
20	—	CD54	A854	SF430080
21		CD55	A855	SF240080
22	┐	CD56	A856	SF510080
23	┐	CD57	A857	SF520080



IBM Simplified Chinese Graphic Character Set, GB 18030 code

National Standard and DBCS-Host

Table 14 (Page 2 of 4). Non-Chinese Character Set, Area 7, GB 18030 DBCS-Host

Line Number	Graphic	GB 18030 DBCS-Host	GB 18030 National Standard	GCGID
24	┌	CD58	A858	SF390080
25	┐	CD59	A859	SF220080
26	└	CD5A	A85A	SF210080
27	┘	CD5B	A85B	SF250080
28	├	CD5C	A85C	SF500080
29	┤	CD5D	A85D	SF490080
30	└┘	CD5E	A85E	SF380080
31	┌┐	CD5F	A85F	SF280080
32	└┘	CD60	A860	SF270080
33	┌┐	CD61	A861	SF260080
34	├┤	CD62	A862	SF360080
35	├┤	CD63	A863	SF370080
36	├┤	CD64	A864	SF420080
37	├┤	CD65	A865	SF190080
38	├┤	CD66	A866	SF200080
39	├┤	CD67	A867	SF230080
40	├┤	CD68	A868	SF470080
41	├┤	CD69	A869	SF480080
42	├┤	CD6A	A86A	SF410080
43	├┤	CD6B	A86B	SF450080
44	├┤	CD6C	A86C	SF460080
45	├┤	CD6D	A86D	SF400080
46	├┤	CD6E	A86E	SF540080
47	├┤	CD6F	A86F	SF530080
48	├┤	CD70	A870	SF440080

Table 14 (Page 2 of 4). Non-Chinese Character Set, Area 7, GB 18030 DBCS-Host

Line Number	Graphic	GB 18030 DBCS-Host	GB 18030 National Standard	GCGID
49	┐	CD71	A871	SS160080
50	┘	CD72	A872	SS180080
51	┘	CD73	A873	SS190080
52	┘	CD74	A874	SS170080
53	┘	CD75	A875	SH020080
54	┘	CD76	A876	SH030080
55	✕	CD77	A877	SH040080
56	—	CD78	A878	SF700081
57	—	CD79	A879	SF710081
58	—	CD7A	A87A	SF720081
59	—	CD7B	A87B	SF570080 SF730081
60	—	CD7C	A87C	SF740081
61	—	CD7D	A87D	SF750081
62	—	CD7E	A87E	SF760081
63	—	CD7F	A880	SF610080
64	—	CD81	A881	SF830081
65	—	CD82	A882	SF820081
66	—	CD83	A883	SF810081
67	—	CD84	A884	SF800081
68	—	CD85	A885	SF790081
69	—	CD86	A886	SF780081
70	—	CD87	A887	SF770081
71	—	CD88	A888	SF160080
72	—	CD89	A889	SF670080

**IBM Simplified Chinese Graphic Character Set, GB 18030 code**

National Standard and DBCS-Host

Line Number	Graphic	GB 18030 DBCS-Host	GB 18030 National Standard	GCGID
73		CD8A	A88A	SF650080
74	▲	CD8B	A88D	SF840080
75	▴	CD8C	A88E	SF850080
76	▵	CD8D	A88F	SF870080
77	▼	CD8E	A890	SF860080
78	⊙	CD8F	A891	U0002609
79	⊕	CD90	A892	SA550080
80	"	CD91	A894	SP210081
81	”	CD92	A895	SP220081
82		CD93	A940	NC010080
83		CD94	A941	NC020080
84		CD95	A942	NC030080
85	×	CD96	A943	NC040080
86	℄	CD97	A944	NC050080
87	⊥	CD98	A945	NC060080
88	≡	CD99	A946	NC070080
89	≡	CD9A	A947	NC080080
90	ㄨ	CD9B	A948	NC090080
91	⊕	CD9C	A949	SS780080
92	mg	CD9D	A94A	SS850080
93	kg	CD9E	A94B	SS860080
94	mm	CD9F	A94C	SS810080
95	cm	CDA0	A94D	SS820080
96	km	CDA1	A94E	SS830080
97	m ²	CDA2	A94F	SS840080

Line Number	Graphic	GB 18030 DBCS-Host	GB 18030 National Standard	GCGID
98	cc	CDA3	A950	SS870080
99	KM	CDA4	A951	SS830081
100	ln	CDA5	A952	SS890080
101	log	CDA6	A953	SS880080
102	mil	CDA7	A954	SS800080
103	:	CDA8	A955	SV550080
104	---	CDA9	A968	SV470080
105	---	CDAA	A969	SV490080
106	----	CDAB	A96A	SV500080
107	----	CDAC	A96B	SV510080
108	---	CDAD	A96C	SV120080
109	--	CDAE	A96D	SV480080
110	----	CDAF	A96E	SV460080
111	,	CDB0	A96F	SP081080
112	,	CDB1	A970	SP081084
113	.	CDB2	A971	SP111080
114	;	CDB3	A972	SP141080
115	:	CDB4	A973	SP131080
116	?	CDB5	A974	SP151080
117	!	CDB6	A975	SP021080
118	(CDB7	A976	SP061080
119)	CDB8	A977	SP071080
120	{	CDB9	A978	SM111080
121	}	CDBA	A979	SM141080
122	(CDBB	A97A	SP061081

IBM Simplified Chinese Graphic Character Set, GB 18030 code

National Standard and DBCS-Host

Table 14 (Page 4 of 4). Non-Chinese Character Set, Area 7, GB 18030 DBCS-Host

Line Number	Graphic	GB 18030 DBCS-Host	GB 18030 National Standard	GCGID
123)	CDBC	A97B	SP071081
124	#	CDBD	A97C	SM011080
125	&	CDBE	A97D	SM031080
126	*	CDBF	A97E	SM04008A
127	+	CDC0	A980	SA011080
128	-	CDC1	A981	SA001080
129	<	CDC2	A982	SA031080
130	>	CDC3	A983	SA051080
131	=	CDC4	A984	SA041080
132	\	CDC5	A985	SM071080
133	\$	CDC6	A986	SC031080
134	%	CDC7	A987	SM021080
135	@	CDC8	A988	SM051080
136	☐	CDC9	A989	SI010080
137	☐	CDCA	A98A	SI020080
138	☐	CDCB	A98B	SI030080
139	☐	CDCC	A98C	SI040080
140	☐	CDCD	A98D	SI050080
141	☐	CDCE	A98E	SI060080
142	☐	CDCF	A98F	SI070080
143	☐	CDD0	A990	SI080080
144	☐	CDD1	A991	SI090080
145	☐	CDD2	A992	SI100080
146	☐	CDD3	A993	SI110080
147	☐	CDD4	A994	SI120080

Table 14 (Page 4 of 4). Non-Chinese Character Set, Area 7, GB 18030 DBCS-Host

Line Number	Graphic	GB 18030 DBCS-Host	GB 18030 National Standard	GCGID
148	☐	CDD5	A995	SI130080

6.12 Area 9a/9b (CJK Unified Ideograph Extension-A)

On Area 9a and 9b, CJK Unified Ideograph Extension-A on ISO 10646 U+3400 to U+4DBF are sequentially assigned, except 52 characters corresponding to Chinese characters in Area 6b.

IBM Simplified Chinese Graphic Character Set, GB 18030 code

National Standard and DBCS-Host

7. Code Translation Correspondence

The *code* translation correspondence between GB 18030 National Standard and GB 18030 DBCS-Host is defined in this specification. And also the code translation correspondences between GB 18030 National Standard or GB 18030 DBCS-Host and the following *code pages* are defined:

- *DBCS-PC*
CPGID 01380 in C-S 3-3220-130
- *DBCS-Host*
CPGID 00837 in C-S 3-3220-130
- *DBCS-EUC*
CPGID 01382 in C-S 3-3220-132
- GB 2312-80

Note: Only in this chapter, “GB-Host” will be used for “*DBCS-Host*”, and “GB-PC” will be used for “*DBCS-PC (CPGID 01380)*”.

7.1 Code Translation Correspondence of GB 2312-80 Characters in GB 18030 Code

7.1.1 Between GB 18030 National Standard and GB-PC/DBCS-EUC. Since the code scheme of GB-PC, *DBCS-EUC* and GB 18030 National Standard are designed based on GB 2312-80, they have the same structure and *code point* for each character. For code translation correspondence among them, no conversion is required.

7.1.2 Between GB 18030 DBCS-Host and GB-Host. Since the specification of GB 2312-80 Characters in GB 18030 DBCS-Host is equivalent to the one for GB-Host, no conversion is required for code translation.

7.1.3 Between GB 18030 National Standard and GB 18030 DBCS-Host. The code correspondence between GB 18030 National Standard and GB 18030 DBCS-Host is equivalent to the one between GB-PC and GB-Host that can be found in C-H 3-3220-130.

7.2 Code Translation Correspondence of non-GB 2312-80 Characters in GB 18030 Code

7.2.1 Between GB 18030 National Standard and GB-PC/GB-Host/DBCS-EUC. None of GB-PC, GB-Host and *DBCS-EUC* contain non-GB 2312-80 characters in GB 18030 Code except the corresponding *code points* to 31 IBM selected characters. The code translation of 31 IBM selected characters are shown in Table 15 on page 32. Others are substituted.

7.2.2 Between GB 18030 DBCS-Host and GB-PC/GB-Host/DBCS-EUC. None of GB-PC, GB-Host and *DBCS-EUC* contain non-GB 2312-80 characters in GB 18030 Code except the corresponding *code points* to 31 IBM selected characters. The code translation of 31 IBM selected characters are shown in Table 15 on page 32. Others are substituted. Note that the specification of 31 IBM selected characters in GB 18030 DBCS-Host is equivalent to the one for GB-Host.

7.2.3 Between GB 18030 National Standard and GB 18030 DBCS-Host. The code translation correspondence of Non-Chinese Characters is shown in Table 5 on page 11 and Table 8 on page 15 that are ordered in GB 18030 National Standard code sequence, or Table 11 on page 22 and Table 14 on page 25 that are ordered in GB 18030 DBCS-Host code sequence.

Also, the code translation correspondence of the following non-Chinese characters is sequentially given from the code point of the source in both translation of GB 18030 National Standard to GB 18030 DBCS-Host and vice-verse.

Uyгур

Uyгур characters from X'81318132' to X'81319934' in fourth byte from X'30' to X'39' on GB 18030 National Standard are mapped to X'4769' to X'489D' in second byte from X'41' to X'FE' on GB 18030 DBCS-Host.

Tibetan

Tibetan characters from X'8132E834' to X'8132FD31' in fourth byte from X'30' to X'39' on GB 18030 National Standard are mapped to X'6D41' to X'6E52' in second byte from X'41' to X'FE' on GB 18030 DBCS-Host.

Mongolian

Mongolian characters from X'8134D238' to X'8134E337' in fourth byte from X'30' to X'39' on GB 18030 National Standard are mapped to

**IBM Simplified Chinese Graphic Character Set, GB 18030 code**

National Standard and DBCS-Host

X'6E83' to X'6F6E' in second byte from X'41' to X'FE' on GB 18030 DBCS-Host.

Yi Syllables

Yi Syllables from X'82359833' to X'82368E37' in third byte from X'81' to X'FE' and fourth byte from X'30' to X'39' on GB 18030 National Standard are mapped to X'6F75' to X'758D' in second byte from X'41' to X'FE' on GB 18030 DBCS-Host.

Yi Radical

Yi Radical from X'82368F31' to X'82369435' in fourth byte from X'30' to X'39' on GB 18030 National Standard are mapped to X'7591' to X'75C7' on GB 18030 DBCS-Host.

Out of Chinese Characters, the code translation correspondence of ISO 10646-1 CJK Unified Ideographs is sequentially given from the code point of the source in both translation of GB 18030 National Standard to GB 18030 DBCS-Host and vice versa. For example, the first code point of the GB 18030 National Standard Area 3 (X'8140') can be mapped to the first code point of GB 18030 DBCS-Host Area 5 (X'8141'), and the mapping sequentially continues.

Also, the code translation correspondence of ISO 10646-1 CJK Unified Ideographs Extension-A except 52 characters corresponding to Chinese characters in Area 6b is sequentially given from the code point of the source in both translation of GB 18030 National Standard to GB 18030 DBCS-Host and vice versa. The first code point of the GB 18030 National Standard CJK Unified Ideographs Extension-A area (X'8139EE39') is mapped to the first code point of GB 18030 DBCS-Host Area 9 (X'CF41'), and the mapping sequentially continues.

The code translation correspondence of other Chinese Characters and Chinese radicals including above 52 characters in Area 6b is shown in Table 6 on page 12 and Table 7 on page 12 that are ordered in GB 18030 National Standard code sequence, or Table 12 on page 23 and Table 13 on page 23 that are ordered in GB 18030 DBCS-Host code sequence.

7.3 Code Translation Correspondence of User Definable Characters in GB 18030 Code

GB 18030 National Standard and GB 18030 DBCS-Host has 1894 UDCs, while both of GB-PC and GB-Host have 1880 UDCs and *DBCS-EUC* has 1360 UDCs.

7.3.1 Code Translation Correspondence between GB 18030 National Standard and GB 18030 DBCS-Host/GB-PC/GB-Host.

The *code translation* correspondence between GB 18030 National Standard and GB 18030 DBCS-Host/GB-PC/GB-Host is sequentially given from the *code point* of the source in both translation of GB 18030 National Standard to others and vice versa.

For example, the first *code point* of the GB 18030 National Standard UDC area (X'AAA1') can be mapped to the first *code point* of GB-PC UDC area (X'8DA1') or to the first *code point* of GB 18030 DBCS-Host /GB-Host UDC area (X'7641'), and the mapping sequentially continues.

The user definable character areas are, however, divided into 3 segments on GB 18030 National Standard. These three UDC areas are used in the following ordering.

1st UDC

1st 564 UDCs, from X'AAA1' to X'AFFE' except second *byte* from X'40' to X'A0'.

2nd UDC

2nd 658 UDCs, from X'F8A1' to X'FEFE' except second *byte* from X'40' to X'A0'.

3rd UDC

3rd 672 UDCs, from X'A140' to X'A7A0' except second *byte* X'7F' and from X'A1' to X'FE'.

The last 14 UDCs are substituted on the code translation correspondence from GB 18030 National Standard to GB-PC/GB-Host.

7.3.2 Code Translation Correspondence between GB 18030 National Standard and DBCS-EUC.

Since 1360 UDCs are scattered on *DBCS-EUC*, a logical conversion is not applicable in the code translation correspondence between GB 18030 National Standard and *DBCS-EUC*. The code translation correspondence can be found in C-H 3-3220-132 via *code points* either of GB-PC or GB-Host obtained from the code trans-



IBM Simplified Chinese Graphic Character Set, GB 18030 codeNational Standard and DBCS-Host

lation in Section 7.3.1. The last 534 UDCs are substituted on the code translation correspondence from GB 18030 National Standard to DBCS-EUC.

7.3.3 Code Translation Correspondence between GB 18030 DBCS-Host and GB-Host. Since the specification of User Definable Characters in GB 18030 DBCS-Host is equivalent to the one for GB-Host, no conversion is required for code translation. The last 14 UDCs of GB 18030 DBCS-Host are substituted on the code translation correspondence from GB 18030 DBCS-Host to GB-Host.

7.3.4 Code Translation Correspondence between GB 18030 DBCS-Host and GB-PC. The *code translation* correspondence between GB 18030 DBCS-Host and GB-PC is logically given from the *code point* of the source in both translation of GB 18030 DBCS-Host to other and vice versa. The last 14 UDCs of GB 18030 DBCS-Host are substituted on the code translation correspondence from GB 18030 DBCS-Host to GB-PC.

7.3.5 Code Translation Correspondence between GB 18030 DBCS-Host and DBCS-EUC. The code translation correspondence between GB 18030 DBCS-Host and DBCS-EUC is equivalent to the one between GB-Host and DBCS-EUC in C-H 3-3220-132. The last 534 UDCs are substituted on the code translation correspondence from GB 18030 DBCS-Host to DBCS-EUC.

**IBM Simplified Chinese Graphic Character Set, GB 18030 code**

National Standard and DBCS-Host

7.4 The Code Correspondence for IBM Selected Characters

The following table shows the code correspondence for IBM selected characters between GB 18030 National Standard, GB 18030 DBCS-Host and other S-CH code pages; *DBCS-PC*, *DBCS-Host* and *DBCS-EUC*.

Table 15 (Page 1 of 2). The code correspondence for IBM selected characters

Line Number	Graphic	DBCS-PC	GB 18030 DBCS-Host and DBCS-Host	DBCS-EUC	GB 18030 National Standard	GCGID
1	i	8CE0	41B1	FEE0	A2A1	NR010080
2	ii	8CE1	41B2	FEE1	A2A2	NR020080
3	iii	8CE2	41B3	FEE2	A2A3	NR030080
4	iv	8CE3	41B4	FEE3	A2A4	NR040080
5	v	8CE4	41B5	FEE4	A2A5	NR050080
6	vi	8CE5	41B6	FEE5	A2A6	NR060080
7	vii	8CE6	41B7	FEE6	A2A7	NR070080
8	viii	8CE7	41B8	FEE7	A2A8	NR080080
9	ix	8CE8	41B9	FEE8	A2A9	NR090080
10	x	8CE9	41BA	FEE9	A2AA	NR100080
11	ㄟ	8CEA	425F	FEEA	A956	SM660080
12	丨	8CEB	426A	FEEB	A957	SM650080
13	ˇ	8CEC	427D	FEEC	A846(*1)	SP050080
14	一	8CED	4358	FEED	A960	JX700080
15	¥	8CEE	435B	EEEE	A3A4(*2)	SC050080
16	˘	8CEF	43BE	FEFF	A961	JX710080
17	°	8CF0	43BF	FEF0	A962	JX720080
18	˘	8CF1	43DC	FEF1	A963	JQ750080
19	˘	8CF2	43DD	FEF2	A964	JQ760080
20	—	8CF3	445A	FEF3	A95C	SP320080
21	仝	8CF4	445C	FEF4	D9DA(*2)	SS760080
22	↗	8CF5	445E	FEF5	A965	SS720080

IBM Simplified Chinese Graphic Character Set, GB 18030 code

National Standard and DBCS-Host

Table 15 (Page 2 of 2). The code correspondence for IBM selected characters						
Line Number	Graphic	DBCS-PC	GB 18030 DBCS-Host and DBCS-Host	DBCS-EUC	GB 18030 National Standard	GCGID
23	○	8CF6	445F	FEF6	A996	ND100087
24	〒	8CF7	446C	FEF7	A893	SS730080
25	栞	8CF8	446D	FEF8	A95A	SS740080
26	TBL	8CF9	446F	FEF9	A959	SS710080
27	••	8CFA	447E	FEFA	A845	SV430080
28	ゝ	8CFB	44DC	FEFB	A966	RQ750080
29	ゞ	8CFC	44DD	FEFC	A967	RQ760080
30	▽	8CFD	44EB	FEFD	A88C	SM740080
31	▼	8CFE	44EC	FEFE	A88B	SV040080

Notes:

- 1) There is no same glyph in GB 18030 National Standard. This is similar glyph.
- 2) GB 2312-80 has the corresponding character for this. Therefore, DBCS-PC, Host and EUC contain this character other than IBM selected character. Both of GB 2312-80 character and IBM selected character are mapped to this character. It is 2 to 1 code correspondence.



IBM Simplified Chinese Graphic Character Set, GB 18030 code

National Standard and DBCS-Host

Supplementary Information

8. Definition of Terms

The following terms are defined for the purposes of this document.

8.1 Byte

A unit of string consisting of 8 bits.

8.2 Code

A system of bit-patterns to which specific graphic or control meanings have been assigned.

8.3 Code Page

A specification of *code points* for each *graphic character* in a set, or in a collection of *graphic character sets*. Within a given *code page*, a *code point* can have only one specific meaning.

8.4 Code Page Global Identifier (CPGID)

A 5-digit decimal identifier assigned to a *code page*. See C-H 3-3220-050.

8.5 Code Point

One of the bit-patterns specified by a *code*.

8.6 Coded Graphic Character

A *graphic character* with its assigned *code point*.

8.7 Coded Graphic Character Set

A set of *graphic characters* with their assigned *code points*.

8.8 Coded Graphic Character Set Global Identifier (CGCSGID)

A 10-digit decimal identifier consisting of the concatenation of a *graphic character set global identifier (GCSGID)* and a *code page global identifier (CPGID)*. The *CGCSGID* is also represented as the concatenation of the hexadecimal representations of the *GCSGID* and *CPGID*. See C-H 3-3220-050.

8.9 DBCS

Abbreviation of a *Double-Byte Character Set*. It is a *graphic character set* in which each character is represented by two *bytes*. See C-S 3-3220-102 for detail.

8.10 DBCS-EUC

A *DBCS code page* which is mainly used with Extended UNIX Code (EUC) encoding scheme.

8.11 DBCS-Host

A *DBCS code page* which is mainly used in conjunction with *EBCDIC*. There are following DBCS-Host code pages in this specification. Note that, in this specification, *DBCS-Host* is specifically used only for DBCS code page on C-H 3-3220-130 1993-11.

DBCS-Host: DBCS-Host code page supporting GB2312-80 National Standard character set (C-H 3-3220-130 1993-11)

GBK DBCS-Host: DBCS-Host code page supporting GBK character set (C-H 3-3220-020 1997-02)

GB18030 DBCS-Host: DBCS-Host code page supporting GB18030 National Standard subset (C-H 3-3220-020 2001-03)

8.12 DBCS-PC

A *DBCS code page* which is mainly used in conjunction with *PC code*. There are following DBCS-PC code pages in this specification. Note that, in this specification, *DBCS-PC* is specifically used only for DBCS code page on C-H 3-3220-130 1993-11.

DBCS-PC: DBCS-PC code page supporting GB2312-80 National Standard character set (C-H 3-3220-130 1993-11)

GBK DBCS-PC: DBCS-PC code page supporting GBK character set (C-H 3-3220-020 1997-02)

8.13 EBCDIC

Abbreviation of Extended BCD Interchange Code which is defined in C-S 3-3220-002, *EBCDIC*.

8.14 GBK

GBK, Chinese Internal Code Specification is the super set of GB and totally compatible with GB 2312-80 (PRC National Standard). GBK expands its character set to 20902 characters in all of ISO 10646-1 CJK unified ideographs.

8.15 Graphic Character

A graphic symbol character, such as a numeric, alphabetic, special or Kanji.

IBM Simplified Chinese Graphic Character Set, GB 18030 codeNational Standard and DBCS-Host

8.16 Graphic Character Global Identifier (GCGID)

An up to eight-character identifier assigned to each *graphic character*. See C-H 3-3220-055.

8.17 Graphic Character Set

A set of *graphic characters*, treated as an entity.

8.18 Graphic Character Set Global Identifier (GCSGID)

A 5-digit decimal identifier assigned to a *graphic character set*. See C-H 3-3220-050.

8.19 PC Code

A *single-byte code page* which is primarily used in the personal computer environment. *PC Code* is based on ISO 646.

8.20 Product

A hardware or software unit that is individually packaged, that is marketable, and to which a type number/program number or equivalence has been assigned.

8.21 Registered Code Page

A *code page* that has been registered and assigned a *CPGID*.

8.22 Registered Coded Graphic Character Set

A *coded graphic character set*, whose *graphic character set* and *code page* have both been registered.

8.23 Unrestricted Code Page

A *registered code page* that has been approved by all affected operating units, and has been placed in the category of “*unrestricted*”.

8.24 User Definable Character

A *graphic character* whose graphic shape is definable by users.

8.25 Ward

A section of a *Double-Byte Character Set* where the first *bytes* of all of the *code points* belonging to it are the same. A *ward* has a unique numerical value which represents the first *byte* of the *code point* belonging to that *ward*.



IBM Simplified Chinese Graphic Character Set, GB 18030 code

National Standard and DBCS-Host

9. History of CH-S Coded Character Set Specification

9.1 DBCS-PC(GB National Standard) code

9.1.1 Change from GB to GBK (C-H 3-3220-130 1993-11 to C-H 3-3220-020 1997-02)

9.1.1.1 Overall. GBK DBCS-PC on C-H 3-3220-020 1997-02 has the superset of DBCS-PC on C-H 3-3220-130, except IBM selected characters and User definable characters, as shown in Figure 5.

9.1.1.3 Code Translation Correspondence from C-H 3-3220-130 to C-H 3-3220-020) For GB2312-80 characters, no conversion required. For IBM selected characters and User Definable characters, table conversion required.

9.1.1.2 Character Coverage. GBK DBCS-PC on C-H 3-3220-020 1997-02 extends the symbols on GBK/1 and GBK/5, Chinese characters on GBK/3 and GBK/4, then User definable characters on UDC-1, 2 and 3.

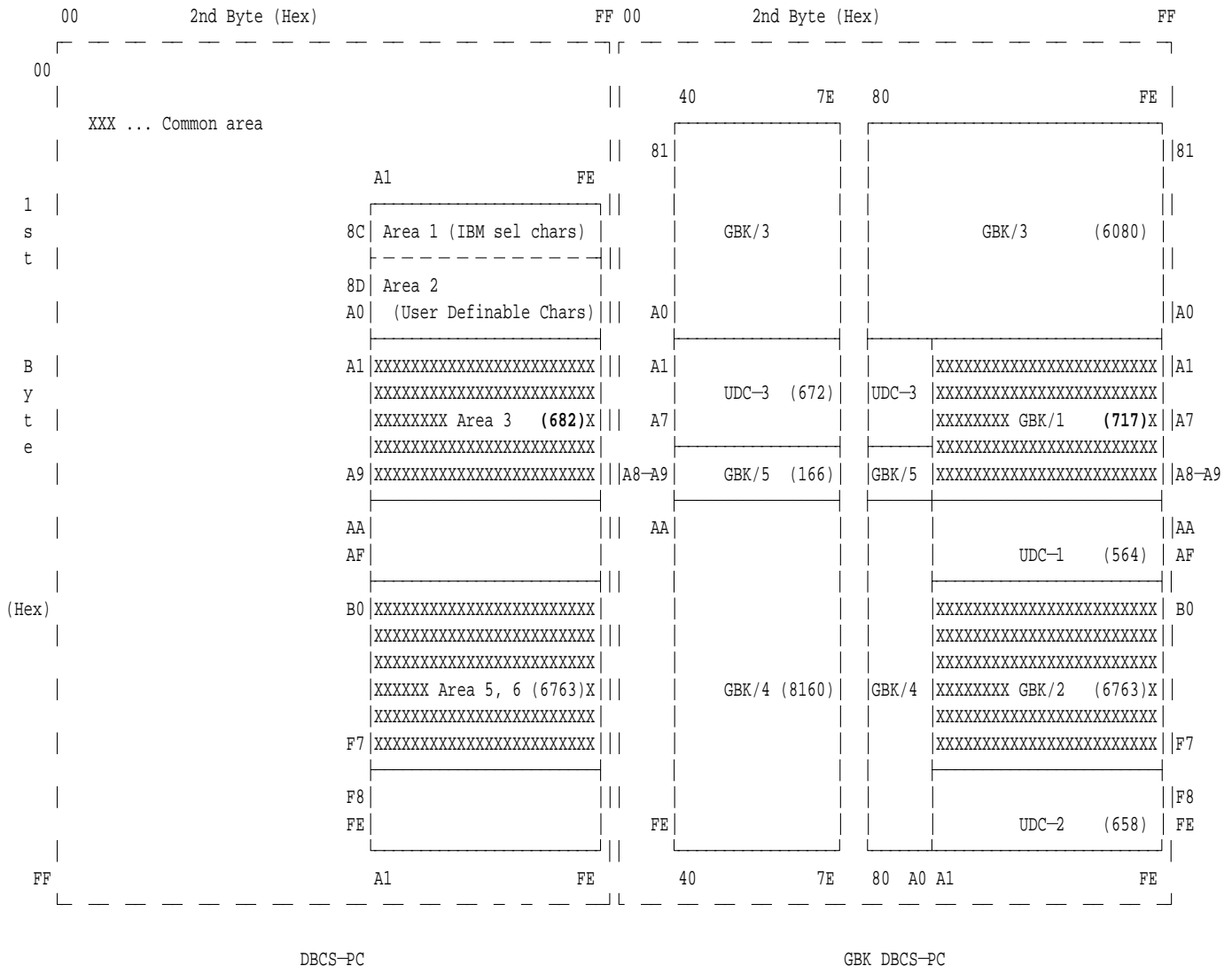


Figure 5. DBCS-PC and GBK DBCS-PC

IBM Simplified Chinese Graphic Character Set, GB 18030 code

National Standard and DBCS-Host

9.1.2 Change from GBK to GB18030 (C-H 3-3220-020 1997-02 to C-H 3-3220-020 2001-03)

9.1.2.1 Overall. GB18030 two-byte code set on C-H 3-3220-020 2001-03 has the superset of GBK DBCS-PC on C-H 3-3220-020 1997-02. GB18030 newly extends four-byte code set.

9.1.2.3 Code Translation Correspondence from C-H 3-3220-020 1997-02 to C-H 3-3220-020 2001-03) Since GB18030 two-byte code set on C-H 3-3220-020 2001-03 has the superset of GBK DBCS-PC on C-H 3-3220-020 1997-02, No conversion required.

9.1.2.2 Character Coverage. GB18030 two-byte code set on C-H 3-3220-020 2001-03 extends the Euro sign from GBK DBCS-PC and GB18030 four-byte code set is newly defined to support all characters on ISO 10646 except the ones on GB18030 two-byte characters.

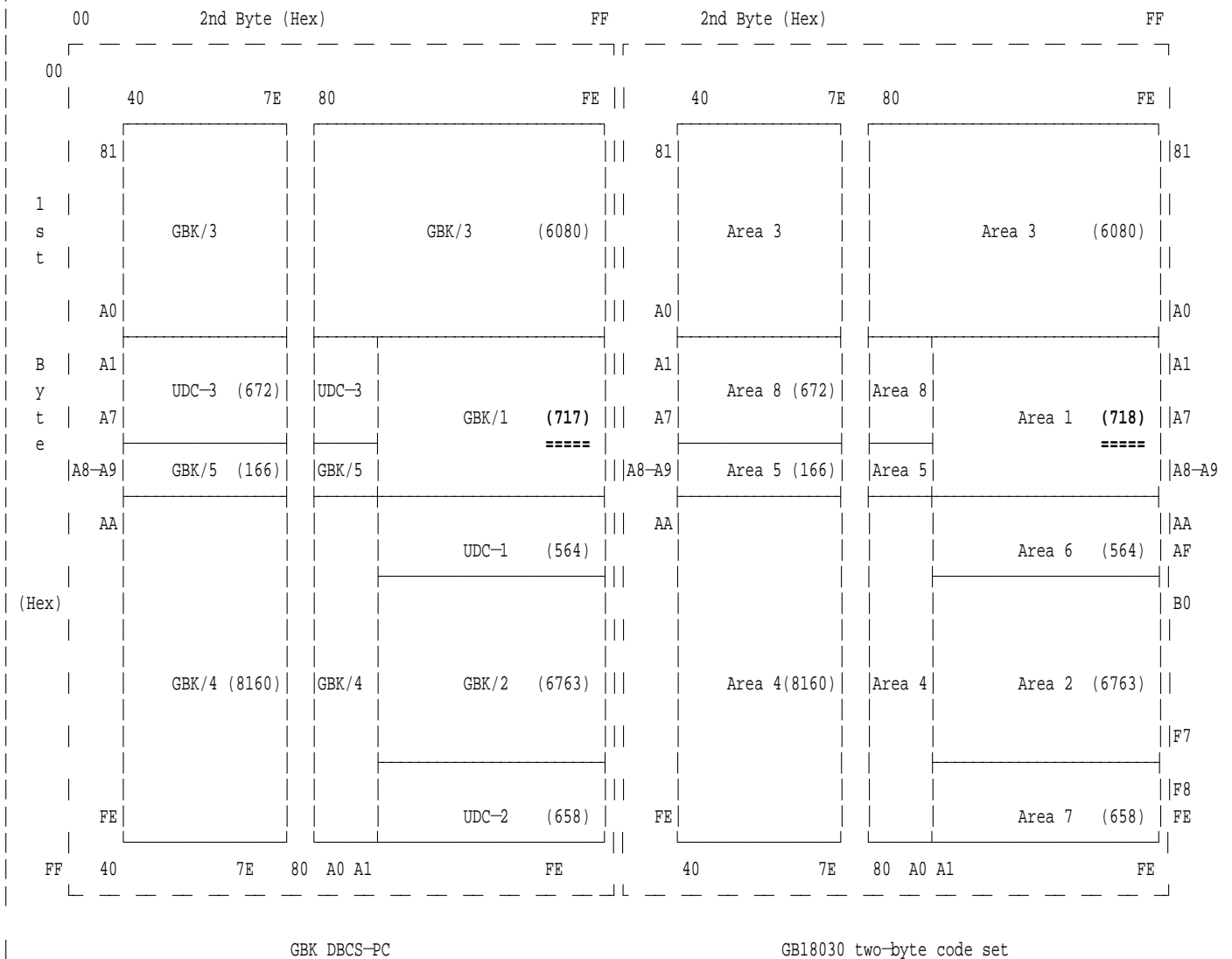


Figure 6. GBK DBCS-PC and GB18030 two-byte code set



IBM Simplified Chinese Graphic Character Set, GB 18030 code

National Standard and DBCS-Host

9.2 DBCS-Host code

9.2.1 Change from DBCS-Host to GBK DBCS-Host (C-H 3-3220-130 1993-11 to C-H 3-3220-020 1997-02)

9.2.1.1 Overall. GBK DBCS-Host on C-H 3-3220-020 1997-02 has the superset of DBCS-Host on C-H 3-3220-130, as shown in Figure 7.

9.2.1.3 Code Translation Correspondence from C-H 3-3220-130 to C-H 3-3220-020) Since GBK DBCS-Host on C-H 3-3220-020 1997-02 has the superset of DBCS-Host on C-H 3-3220-130, no conversion required.

9.2.1.2 Character Coverage. DBCS-Host on C-H 3-3220-020 1997-02 extends the symbols on Area 1 and 5, Chinese characters on Area 3, 4a and 4b, then User definable characters on Area 6.

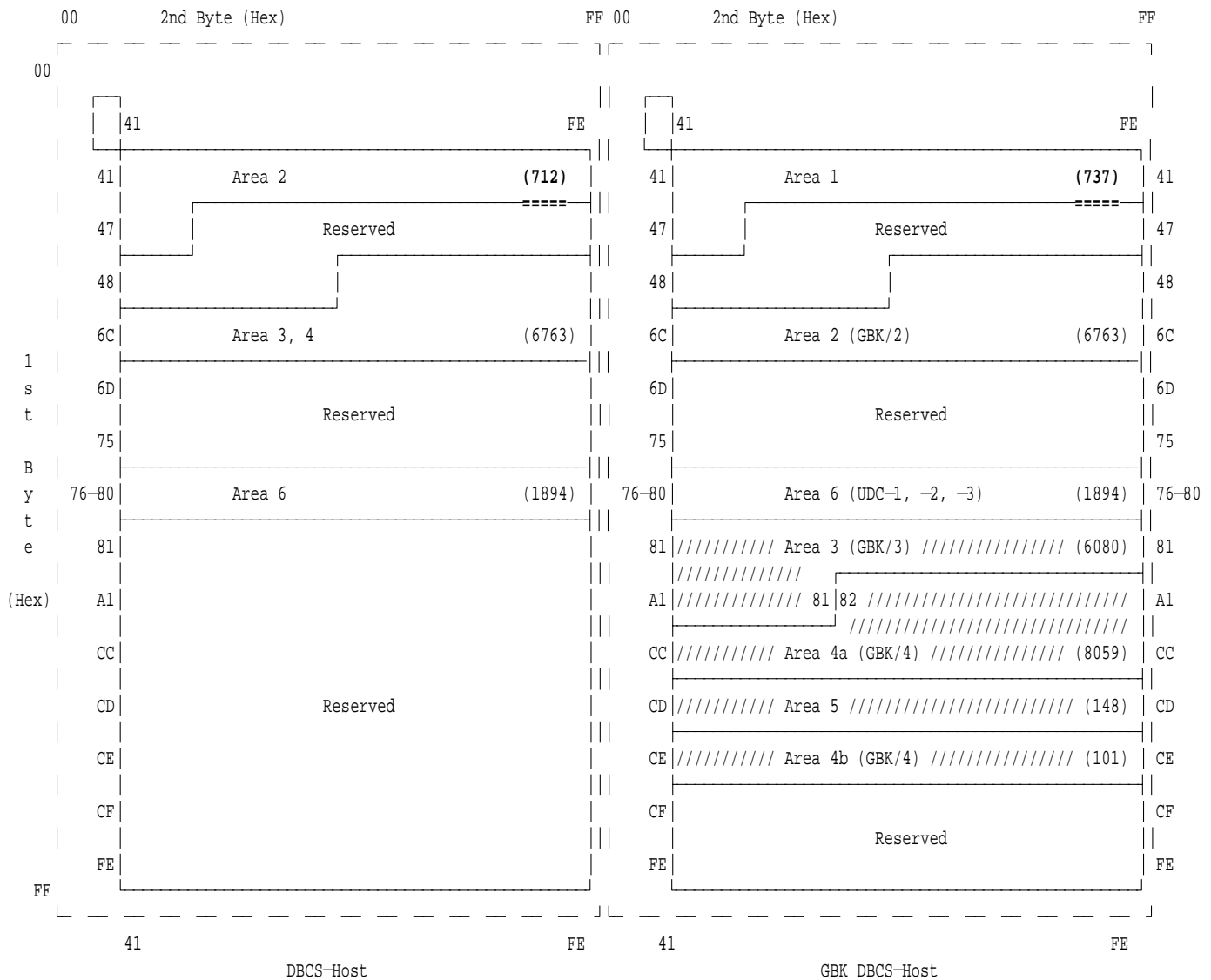


Figure 7. DBCS-Host and GBK DBCS-Host

IBM Simplified Chinese Graphic Character Set, GB 18030 code

National Standard and DBCS-Host

9.2.2 Change from GBK DBCS-Host to GB18030 DBCS-Host (C-H 3-3220-020 1997-02 to 2001-03)

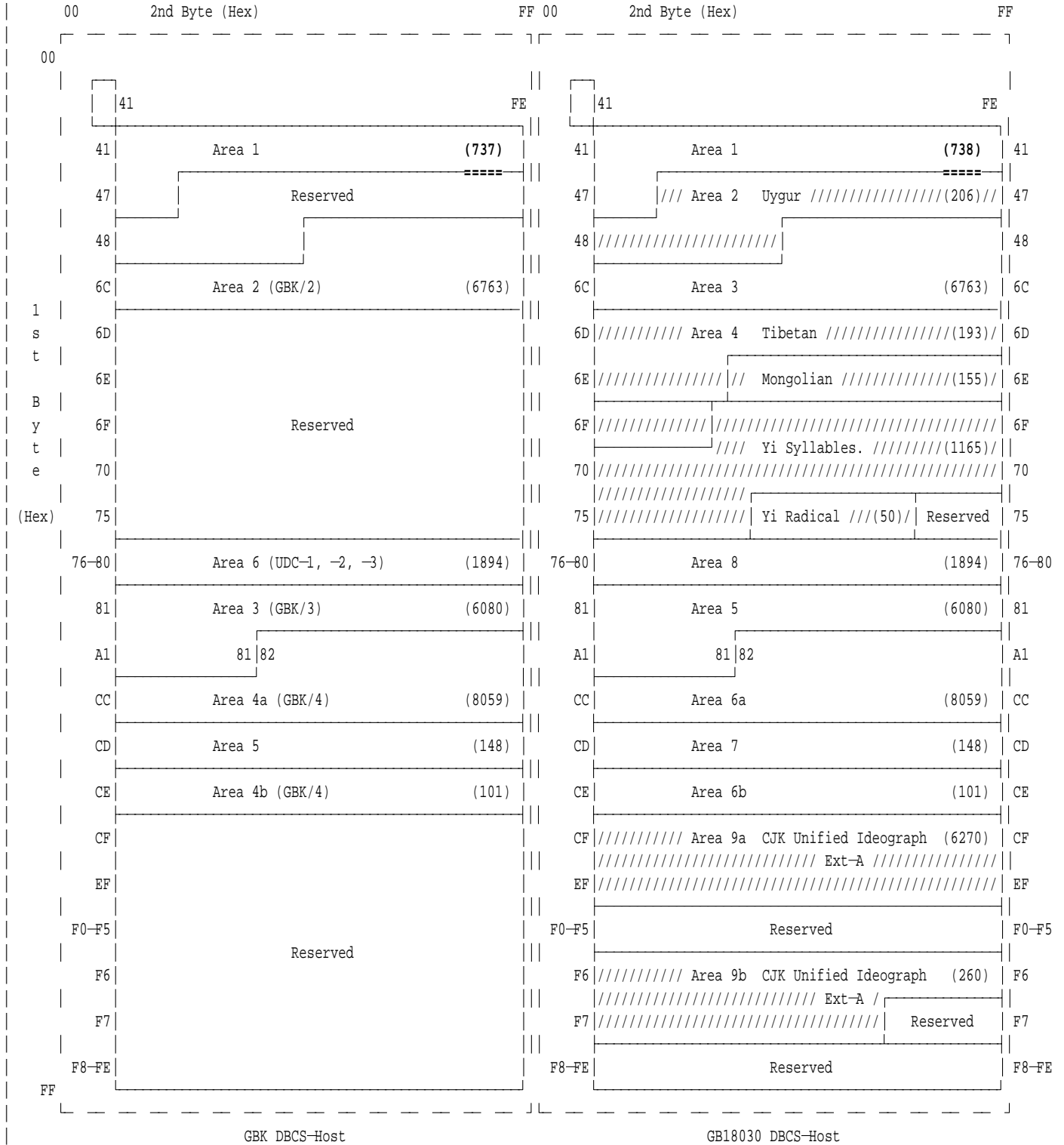
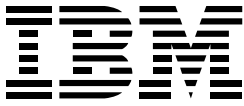


Figure 8. GBK DBCS-Host and GB18030 DBCS-Host



IBM Simplified Chinese Graphic Character Set, GB 18030 code

National Standard and DBCS-Host

| **9.2.2.1 Overall.** GB18030 DBCS-Host on C-H |
| 3-3220-020 2001-03 has the superset of GBK |
| DBCS-Host on C-H 3-3220-020 1997-02, as shown in |
| Figure 8.

| **9.2.2.2 Character Coverage.** GB18030 DBCS-Host |
| on C-H 3-3220-020 2001-03 extends the Euro sign on |
| Area 1, some scripts on Area 2 and 4, Chinese charac- |
| ters on Area 9.

| **9.2.2.3 Code Translation Correspondence from C-H** |
| **3-3220-020 1997-02 to C-H 3-3220-020 2001-03)** Since |
| GB18030 DBCS-Host on C-H 3-3220-020 2001-03 has |
| the superset of GBK DBCS-Host on C-H 3-3220-020 |
| 1997-02, no conversion required.



IBM Simplified Chinese Graphic Character Set, GB 18030 code

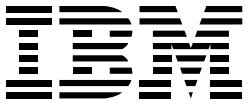
National Standard and DBCS-Host

10. Graphic Character Representation

Although character graphics are inserted as sample graphics of the *coded graphic characters* in the specification tables a *graphic character* is independent of actual size or shape in which it may appear.

This specification does not restrict the style of font

design. However, it is recommended that the font design of *graphic characters* for *products* follow the graphic representation in this document as close as possible to avoid confusion in identification of *graphic characters*. On questions about the font design to be implemented, the Standards Project Authority for DBCS (SIRS 029) may be consulted.



© Copyright IBM Corp. 2001

International Business Machines Corporation
IBM Standards Program
Old Orchard Road
Armonk, New York 10504, USA

Table Definitions

<u>id</u>	<u>File</u>	<u>Page</u>	<u>References</u>
CHARCV1	GB18030 SCRIPT	5	5, 5, 5
CHARCV2	GB18030 SCRIPT	5	5, 5
GID	GB18030 SCRIPT	6	6
SPCHDR	GB18030 SCRIPT	9	9
SPCBDY	GB18030 SCRIPT	9	9
TBL1	GB18030 SCRIPT	11	11, 11, 12, 12, 15, 22, 23, 23, 25
TBL2	GB18030 SCRIPT	11	11, 12, 12, 15, 22, 23, 23, 25
TBL3	GB18030 SCRIPT	11	11
TBL4	GB18030 SCRIPT	11	
STR11	GB18030 SCRIPT	19	19, 19
STR22	GB18030 SCRIPT	19	19
SPCHDR2	GB18030 SCRIPT	20	20
SPCBDY2	GB18030 SCRIPT	20	20
TBL5	GB18030 SCRIPT	32	32, 32
TBL6	GB18030 SCRIPT	32	32

Figures

<u>id</u>	<u>File</u>	<u>Page</u>	<u>References</u>
STR43	GB18030 SCRIPT	4	1 3
STR41	GB18030 SCRIPT	4	2 3, 19
STRUCT2	GB18030 SCRIPT	10	3 9
STRUCT4	GB18030 SCRIPT	21	4 20
GBGBKPC	GB18030 SCRIPT	36	5 36
GBKGBY2	GB18030 SCRIPT	37	6
GBGBKH	GB18030 SCRIPT	38	7 38
GBKGBYH	GB18030 SCRIPT	39	8 40

Headings

<u>id</u>	<u>File</u>	<u>Page</u>	<u>References</u>
ABST	GB18030 SCRIPT	1	1.1, Abstract 1
IBMDOC	GB18030 SCRIPT	2	3.3, Referenced IBM Documents
IMPL	GB18030 SCRIPT	5	4.4.3, Implementation Phase 1 and 2 5
UDCCONV	GB18030 SCRIPT	30	7.3.1, Code Translation Correspondence between GB 18030 National Standard and GB 18030 DBCS-Host/GB-PC/GB-Host 31

Processing Options

Runtime values:

Document fileid	GB18030 SCRIPT
Document type	STDDOC
Document style	IBMS1PL
Profile	EDFPRF40
Service Level	0031
SCRIPT/VS Release	4.0.0
Date	01.04.03
Time	13:57:09
Device	3820A
Number of Passes	3
Index	YES
SYSVAR A	NO
SYSVAR D	YES
SYSVAR S	OFFSET
SYSVAR X	YES

Formatting values used:

Annotation	NO
Cross reference listing	YES
Cross reference head prefix only	YES
Dialog	LABEL
Duplex	YES
DVCF conditions file	(none)
DVCF value 1	(none)
DVCF value 2	(none)
DVCF value 3	(none)
DVCF value 4	(none)
DVCF value 5	(none)
DVCF value 6	(none)
DVCF value 7	(none)
DVCF value 8	(none)
DVCF value 9	(none)
Explode	NO
Figure list on new page	NO
Figure/table number separation	YES
Folio-by-chapter	NO
Head 0 body text	(none)
Head 1 body text	(none)
Head 1 appendix text	Appendix
Hyphenation	YES
Justification	YES
Language	ENGL
Keyboard	395
Layout	OFF
Leader dots	YES
Master index	(none)
Partial TOC (maximum level)	4
Partial TOC (new page after)	INLINE
Print example id's	NO
Print cross reference page numbers	YES
Process value	(none)
Punctuation move characters
Read cross-reference file	(none)
Running heading/footer rule	NONE
Show index entries	NO
Table of Contents (maximum level)	5
Table list on new page	NO
Title page (draft) alignment	RIGHT
Write cross-reference file	(none)

Imbed Trace

Page 11	CUSCBC1A
Page 12	CUSCBC2A
Page 12	CUSCBC3A
Page 15	CUSCBC4A
Page 22	CUSCBC1B
Page 23	CUSCBC2B
Page 23	CUSCBC3B
Page 25	CUSCBC4B
Page 32	CUSCBC5A