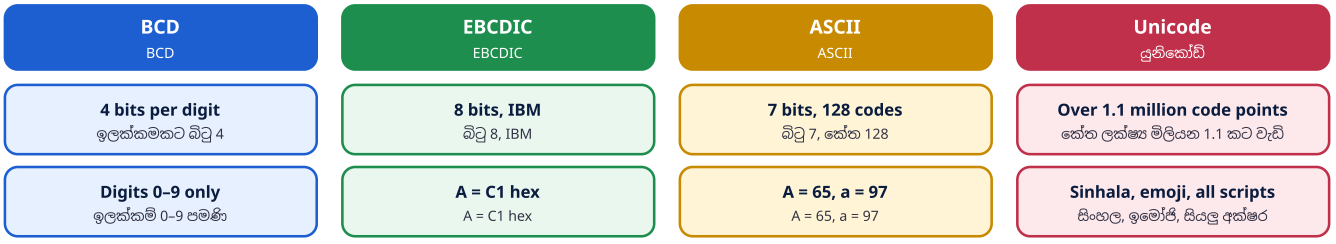
Grade 12 A/L ICT · Lesson 3

3.2 Character Representation: BCD, EBCDIC, ASCII and Unicode

Syllabus 3.2. When you press **A** on the keyboard, the computer does not store a picture of the letter. It stores a **number code**. In this part you learn the four main coding schemes – **BCD, EBCDIC, ASCII and Unicode** – how to **convert symbols to codes and back**, and the **advantages and disadvantages** of each.

1. What is a character code?

A **character** is any letter, digit, symbol or space, such as A, 7, ?, අ or 😀. A **character set (coding scheme)** is an agreed table that gives every character a **unique number**. Because both computers use the same table, text typed on one machine shows correctly on another.



The four coding schemes in this lesson

2. BCD – Binary Coded Decimal

In **BCD** each **decimal digit** is changed to its own **4-bit** code (0 = 0000 … 9 = 1001). The codes 1010 to 1111 are **not used**. BCD is used where numbers must be **shown as digits** quickly and exactly: **calculators, digital clocks, meters and banking**.

> 🧮 **Worked example: Write 2026 in BCD.**
> Take each digit on its own: 2 → 0010, 0 → 0000, 2 → 0010, 6 → 0110. So 2026 = **0010 0000 0010 0110** (16 bits).

Digit	2	0	2	6
BCD	0010	0000	0010	0110

> 🧮 **Worked example: Compare 59 in BCD and in pure binary.**
> BCD: 5 → 0101, 9 → 1001, so **0101 1001** (8 bits). Pure binary: 59 = 32 + 16 + 8 + 2 + 1 = **111011** (6 bits). BCD needs **more bits**, but each digit is easy to read.

© 2026 schoolict.net (Ruwan Mangala Suraweera) · All rights reserved

> 🧮 **Worked example: Which decimal number is the BCD code 0111 0000 0100?**
> Split into groups of 4: 0111 = 7, 0000 = 0, 0100 = 4. The number is **704**. (Do not convert the whole thing as one binary number!)

3. EBCDIC

EBCDIC (Extended Binary Coded Decimal Interchange Code) is an **8-bit** code made by **IBM** in the 1960s for its **mainframe** computers. It has 256 codes. The letters are **not in one continuous run**: A–I are C1–C9, J–R are D1–D9 and S–Z are E2–E9 (hex). Digits 0–9 are F0–F9. It is still used in **banks and big IBM mainframes** today.

> 🧮 **Worked example: Write the word ICT in EBCDIC (hex and binary).**
> I = C9, C = C3, T = E3, so **C9 C3 E3** (hex) = 11001001 11000011 11100011.

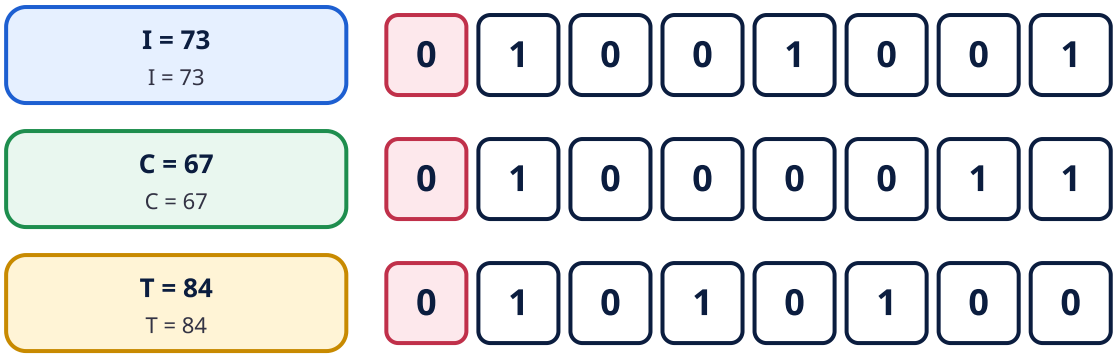
4. ASCII

ASCII (American Standard Code for Information Interchange, 1963) is a **7-bit** code with **128 codes (0–127)**. It is stored in one byte. **Extended ASCII** uses all 8 bits for 256 codes (extra symbols such as é and £). ASCII is the base of almost all modern codes.

Codes	Characters	Remember
0–31, 127	Control codes (Enter = 13, Tab = 9)	Not printed
32	Space	32 = 0100000
48–57	Digits 0–9	“0” = 48
65–90	Capital A–Z	“A” = 65
97–122	Small a–z	“a” = 97

> 🧮 **Worked example: Write the word ICT in ASCII (decimal, hex and 8-bit binary). How many bytes does it need?**
> I = 73, C = 67, T = 84 → hex 49 43 54 → binary **01001001 01000011 01010100**. 3 characters × 1 byte = **3 bytes**.

© 2026 schoolict.net (Ruwan Mangala Suraweera) · All rights reserved



Each character takes one byte: 3 characters = 3 bytes

The word ICT stored in ASCII

🧱 **Worked example: Decode the ASCII codes 01000011 01000001 01010100.**
67 = C, 65 = A, 84 = T. The word is **CAT**.

🧱 **Worked example: The ASCII code of “E” is 69. Find the codes of “e” and “H” without a table.**
Small letters are **32 more** than capitals: “e” = 69 + 32 = **101**. Letters are in order: H is 3 after E, so “H” = 69 + 3 = **72**.

💡 The **character** “7” (code 55) is not the **number** 7 (binary 111). Phone numbers and NIC numbers are stored as text, so each digit uses one byte.

5. Unicode

ASCII has no space for Sinhala, Tamil, Chinese or emoji. **Unicode** gives every character in every language its own **code point**, written **U+** and a hex number. There is room for **1,114,112 code points** (U+0000 to U+10FFFF). The first 128 code points are the **same as ASCII**. Sinhala letters are in the block **U+0D80 to U+0DFF**.

UTF-8
1 to 4 bytes per character. ASCII letters use 1 byte. Used by almost all websites.

UTF-16
2 or 4 bytes per character. Used inside Windows, Java and JavaScript.

UTF-32
Always 4 bytes per character. Simple but uses the most space.

🧱 **Worked example: Find the Unicode code point and UTF-8 bytes of the Sinhala letter “අ”.**
“අ” is **U+0D85** (decimal 3461). In UTF-8 it takes **3 bytes: E0 B6 85**. In UTF-16 it takes 2 bytes.

© 2026 schoolict.net (Ruwan Mangala Suraweera) · All rights reserved



From a Sinhala letter to stored bits

🧱 **Worked example: How many bytes does the word “ලංකා” need in UTF-8 and in UTF-16? Compare with “ICT” in ASCII.**
“ලංකා” is 4 code points (ල, ං, ක, ා). UTF-8: 4 × 3 = **12 bytes**. UTF-16: 4 × 2 = **8 bytes**. “ICT” in ASCII is only 3 bytes. Unicode supports more languages but can need more space.

🧱 **Worked example: The emoji 😀 is code point U+1F600. How many bytes does it use in UTF-8?**
Code points above U+FFFF need **4 bytes** in UTF-8: **F0 9F 98 80**.

6. Advantages and disadvantages

Code	Bits	Advantages	Disadvantages
BCD	4 per digit	Easy to convert to and from decimal; exact for money; easy to show on displays	Digits only; wastes bits (1010–1111 unused); slower arithmetic
EBCDIC	8	256 codes; works with old IBM mainframe systems	Letters not continuous; not compatible with ASCII; English only
ASCII	7 (8 extended)	Simple; small (1 byte); letters in order; used everywhere	Only 128 (or 256) characters; no Sinhala, Tamil or emoji
Unicode	8 to 32 (UTF-8/16/32)	All languages and emoji; one world standard; UTF-8 matches ASCII	Needs more storage for non-English text; more complex

NEW 2026 **Unicode 18.0** was released on **16 September 2026** and has **172,808 characters** (Unicode 17.0 in 2025 had 159,801). Every new version adds scripts, symbols and a few new emoji, which reach phones in the following months.

🚀 Beyond the syllabus (2026 update) *Not required for the exam – read for interest.* **Beyond the syllabus – UTF-8 rules the web.** In 2026 about **98% of all websites** use UTF-8. Many emoji are really **several code points joined together**: a family emoji or a skin-tone emoji is built with a **Zero Width Joiner (U+200D)**, so one emoji on screen can take 10 or more bytes. Sinhala uses the same joiner: ශ්‍රී needs ශ + ් + ZWJ + ර + ී.

Key Words

Character – Any letter, digit, symbol or space that can be stored.

Character set – An agreed table that gives each character a unique number.

© 2026 schoolict.net (Ruwan Mangala Suraweera) · All rights reserved

BCD – A code that stores each decimal digit as its own 4 bits.	**EBCDIC** – An 8-bit IBM mainframe code with letters in separate runs.
ASCII – A 7-bit code with 128 characters, A = 65.	**Extended ASCII** – An 8-bit version of ASCII with 256 codes.
Unicode – A world standard that gives every character of every language a code point.	**Code point** – The number of a character in Unicode, written like U+0041.
UTF-8 – A Unicode encoding that uses 1 to 4 bytes and matches ASCII.	**UTF-16** – A Unicode encoding that uses 2 or 4 bytes per character.
Control code – A code that is not printed but controls the device, such as Enter.	**Encoding** – The rule for turning characters into bits.

Sample Q&A

1. Why do we need character codes?

> **Answer:** A computer can only store numbers in binary. A character code gives each character a number, and because everyone uses the same code, text can be shared between computers.

2. Write 2026 in BCD and say one place where BCD is used.

> **Answer:** 0010 0000 0010 0110. BCD is used in calculators and digital clocks.

3. Write “ICT” in 8-bit ASCII binary.

> **Answer:** 01001001 01000011 01010100

4. Give two differences between ASCII and Unicode.

> **Answer:** ASCII uses 7 bits and has 128 characters for English; Unicode has over a million code points for all languages. ASCII uses 1 byte; Unicode (UTF-8) uses 1 to 4 bytes.

5. Why was Unicode important for Sinhala users?

> **Answer:** Before Unicode, Sinhala text used special fonts that mapped letters onto ASCII codes, so text broke on other computers. Unicode gives Sinhala its own code points (U+0D80–U+0DFF), so Sinhala works on every phone, website and search engine.

6. Give one advantage and one disadvantage of EBCDIC.

> **Answer:** Advantage: works with existing IBM mainframe systems (256 codes). Disadvantage: letters are not continuous and it is not compatible with ASCII.

© 2026 schoolict.net (Ruwan Mangala Suraweera) · All rights reserved

MCQ

1. How many characters can standard 7-bit ASCII represent?

A. 128
B. 127
C. 256
D. 1,114,112
E. 64

2. What is the ASCII code of “A”?

A. 193
B. 65
C. 41
D. 64
E. 97

3. The ASCII code of “a” is 97. What is the code of “d”?

A. 100
B. 99
C. 104
D. 68
E. 101

4. What is 59 in BCD?

A. 0110 1001
B. 1001 0101
C. 0101 1001
D. 0101 1010
E. 0011 1011

5. Which of these is NOT a valid BCD digit?

A. 1001
B. 1100
C. 0000
D. 0111
E. 1000

6. The BCD code 0011 1000 stands for:

A. 56
B. 83
C. 11
D. 38
E. 3.8

© 2026 schoolict.net (Ruwan Mangala Suraweera) · All rights reserved

7. Which code was made by IBM for mainframes?

A. EBCDIC
B. UTF-16
C. BCD
D. UTF-8
E. ASCII

8. In EBCDIC the letter “A” is (hex):

A. 41
B. F1
C. 65
D. A1
E. C1

9. Which Unicode block holds the Sinhala script?

A. U+0D80 – U+0DFF
B. U+0900 – U+097F
C. U+0000 – U+007F
D. U+0B80 – U+0BFF
E. U+1F600 – U+1F64F

10. How many bytes does the letter “අ” (U+0D85) use in UTF-8?

A. 4
B. 3
C. 1
D. 2
E. 8

11. Which encoding is used by almost all websites in 2026?

A. EBCDIC
B. UTF-32
C. BCD
D. UTF-8
E. 7-bit ASCII only

12. The ASCII codes 01001001 01010100 spell:

A. it
B. TI
C. HT
D. IU
E. IT

© 2026 schoolict.net (Ruwan Mangala Suraweera) · All rights reserved

13. What is a main disadvantage of ASCII?

A. It is only used by IBM
B. It uses 32 bits per letter
C. Letters are not in order
D. It cannot store digits
E. It cannot represent Sinhala or Tamil

14. How many bytes does UTF-32 use for every character?

A. 1
B. 4
C. 1 to 4
D. 3
E. 2

15. The first 128 Unicode code points are the same as:

A. Hexadecimal
B. EBCDIC
C. Extended EBCDIC
D. ASCII
E. BCD

16. Which term means: “Any letter, digit, symbol or space that can be stored”?

A. BCD
B. Code point
C. Character
D. Control code
E. Extended ASCII

17. Which term means: “An agreed table that gives each character a unique number”?

A. EBCDIC
B. Unicode
C. UTF-8
D. Character set
E. Encoding

18. Which term means: “A code that stores each decimal digit as its own 4 bits”?

A. UTF-16
B. ASCII
C. BCD
D. Character
E. Code point

© 2026 schoolict.net (Ruwan Mangala Suraweera) · All rights reserved

19. Which term means: “An 8-bit IBM mainframe code with letters in separate runs”?

A. Character set
B. Extended ASCII
C. EBCDIC
D. UTF-8
E. Control code

20. Which term means: “A 7-bit code with 128 characters, A = 65”?

A. BCD
B. Encoding
C. Unicode
D. UTF-16
E. ASCII

Structured Questions

1. (a) Write 59 in BCD. (b) Write it in pure binary. (c) Which needs more bits and why? *(3 marks)*

> (a) 0101 1001; (b) 111011; (c) BCD needs 8 bits against 6, because each digit uses 4 bits and codes 1010–1111 are wasted.

2. (a) Write “ICT” in ASCII decimal and binary. (b) How many bits are needed to store it in 8-bit ASCII? *(3 marks)*

> (a) 73, 67, 84; 01001001 01000011 01010100; (b) 3 × 8 = 24 bits.

3. Decode the ASCII binary 01000011 01000001 01010100 and write the same word in EBCDIC hex. *(4 marks)*

> 67 = C; 65 = A; 84 = T, word CAT; EBCDIC C3 C1 E3.

4. (a) What is a Unicode code point? (b) Give the code point and UTF-8 bytes of “ඇ”. (c) Give two advantages of Unicode. *(4 marks)*

> (a) the unique number of a character, written U+hex; (b) U+0D85, E0 B6 85; (c) supports all languages and emoji; one world standard compatible with ASCII.

5. Compare BCD, EBCDIC, ASCII and Unicode by number of bits and one use of each. *(4 marks)*

> BCD 4 bits per digit, calculators; EBCDIC 8 bits, IBM mainframes; ASCII 7 bits, English text and keyboards; Unicode 8 to 32 bits, websites and phones in all languages.

Summary

- A character code gives every character a unique number so text can be stored and shared.

© 2026 schoolict.net (Ruwan Mangala Suraweera) · All rights reserved

- BCD: each decimal digit in 4 bits (0000–1001); used in calculators and clocks.
- EBCDIC: 8-bit IBM mainframe code; A = C1 hex; letters not continuous.
- ASCII: 7 bits, 128 codes; A = 65, a = 97, 0 = 48; small = capital + 32.
- Unicode: a code point for every character in every language; UTF-8 uses 1–4 bytes.
- Sinhala is U+0D80–U+0DFF and takes 3 bytes per letter in UTF-8.

MCQ answer key – 3.2

1. **A** – $2^7 = 128$.
2. **B** – A = 65 (hex 41).
3. **A** – 97 + 3 = 100.
4. **C** – 5 = 0101, 9 = 1001.
5. **B** – Only 0000–1001 are used.
6. **D** – 0011 = 3, 1000 = 8.
7. **A** – EBCDIC is IBM's 8-bit code.
8. **E** – A = C1 in EBCDIC.
9. **A** – Sinhala is U+0D80–U+0DFF.
10. **B** – U+0800–U+FFFF need 3 bytes.
11. **D** – About 98% of sites use UTF-8.
12. **E** – 73 = I, 84 = T.
13. **E** – Only 128 characters, English only.
14. **B** – 32 bits = 4 bytes, always.
15. **D** – Unicode keeps ASCII for compatibility.
16. **C** – Character – Any letter, digit, symbol or space that can be stored.
17. **D** – Character set – An agreed table that gives each character a unique number.
18. **C** – BCD – A code that stores each decimal digit as its own 4 bits.
19. **C** – EBCDIC – An 8-bit IBM mainframe code with letters in separate runs.
20. **E** – ASCII – A 7-bit code with 128 characters, A = 65.

© 2026 schoolict.net (Ruwan Mangala Suraweera). All rights reserved.