Characters: ASCII and Unicode

Computer systems (Paper 1) · revision notes

How text is stored: every character has a unique binary code in a character set. You need to know how the number of bits per character decides how many characters can be represented, and to compare ASCII and Unicode. Expect definitions, questions where you use a given code to work out others, and comparisons.

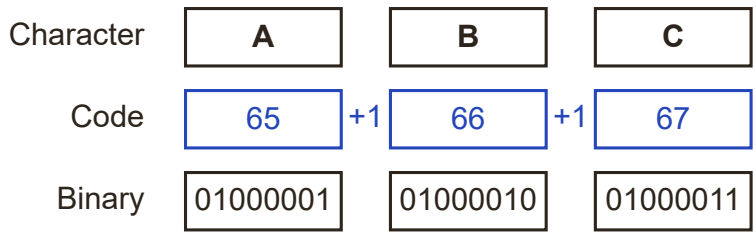
Key facts

- Character set: all the characters a computer can represent, each with a unique binary code
- ASCII: 7 bits, 128 characters (English letters, digits, punctuation, control codes)
- Unicode: more bits per character (e.g. 16 or 32), so many more characters: world languages, symbols, emoji
- The same text takes up more space in Unicode than in ASCII
- n bits per character → 2^n characters
- Extended ASCII: 8 bits, 256 characters
- Codes run in sequence: if 'A' = 65, then 'B' = 66, 'C' = 67

Notes

Character sets

- Every character (letter, digit, punctuation mark, space or symbol) is stored as a binary code.
- A **character set** is the set of all the characters a computer can represent, with the unique binary code for each one.
- Codes are grouped and run in sequence: letters in alphabetical order and digits in order. If 'A' = 65, then 'B' = 66, 'C' = 67 and so on. Uppercase and lowercase letters have different codes.



Codes run in order: if A = 65, then B = 66 and C = 67

- You do not have to memorise codes: an exam question gives you one and you count on or back from it.
- A digit typed as text, such as '7', is stored as its character code, not as the number 7.

Bits per character

- The number of bits used for each character decides how many different characters can be represented: n bits give 2^n different codes.
- 7 bits → 128 characters; 8 bits → 256 characters; 16 bits → 65,536 characters.
- Each extra bit doubles the number of possible characters.

ASCII and Unicode

- **ASCII** uses 7 bits per character, so it has 128 characters: the English upper- and lowercase letters, the digits 0 to 9, punctuation, and control codes such as a new line.
- **Extended ASCII** uses 8 bits, giving 256 characters, which adds some accented letters and extra symbols.
- ASCII does not have enough codes for languages such as Chinese, Arabic or Hindi, or for emoji.
- **Unicode** uses more bits per character than ASCII (e.g. 16 or 32 bits), so it can represent many more characters: the alphabets of most of the world's languages, plus symbols and emoji.
- The first 128 Unicode codes are the same as the ASCII codes.
- Drawback of Unicode: it uses more bits per character, so the same text takes up more storage and takes longer to send. Grade 7+ only

How to answer each type of question

Define a character set or state how many characters 1 to 2 marks **Grade 4**

1. For the definition, include 'each character has a unique binary code'.
2. For how many characters, work out 2 to the power of the number of bits.

Example. (a) State what is meant by a character set. [1]
(b) A character set uses 6 bits per character. State the maximum number of characters it can represent. [1]

Model answer.

(a) The set of characters that a computer can represent, each with its own unique binary code (1)

(b) $2^6 = 64$ (1)

Use a given character code

1 to 2 marks | Grade 5

1. Count on (or back) through the alphabet from the code you are given.
2. Convert to binary if asked, giving all 8 bits.

Example. In a character set, the code for 'H' is 72 in denary.
(a) State the code for 'K'. [1]
(b) Give the code for 'K' as an 8-bit binary number. [1]

Model answer.

(a) H, I, J, K, so 72 + 3 = 75 (1)
(b) 75 = 64 + 8 + 2 + 1, so 01001011 (1)

Explain why Unicode is used instead of ASCII

2 to 4 marks | Grade 7

1. State the bits and number of characters for ASCII.
2. Explain that Unicode uses more bits, so it covers far more characters, and link this to the scenario.
3. If asked for a drawback: more bits per character, so larger files.

Example. A company's website is used by customers in Japan, Greece and the UK. Explain why the website uses Unicode rather than ASCII, and give **one** drawback of this. [3]

Model answer. ASCII uses 7 (or 8) bits, so it can only represent 128 (or 256) characters, which is not enough for English, Greek and Japanese characters together (Japanese alone uses thousands of characters) (1). Unicode uses more bits per character, so it can represent the characters of many languages, including Japanese and Greek (1). Drawback: each character uses more bits, so the pages take up more storage / take longer to download (1).

Calculate the bits needed per character

2 marks | Grade 8

1. Count how many characters are needed in total.
2. Find the smallest power of 2 that is at least that number.
3. Show the power of 2 that is too small and the one that is big enough.

Example. A new character set must include 26 uppercase letters, 26 lowercase letters, 10 digits and 40 other symbols. Calculate the minimum number of bits needed for each character. Show your working. [2]

Model answer. 26 + 26 + 10 + 40 = 102 characters (1). $2^6 = 64$ is too few, but $2^7 = 128$ is enough, so 7 bits (1).

Don’t lose marks

- Saying 7-bit ASCII has 256 characters. 7 bits give 128; 8 bits (extended ASCII) give 256.

- Thinking 'a' and 'A' have the same code. Uppercase and lowercase letters have different codes.
- Saying Unicode is 'better' without saying why (more characters) or what it costs (larger files).
- Forgetting that codes start at 0, so 7 bits give the codes 0 to 127.

More tips

Memory tricks

- Powers of 2 worth knowing: $2^6 = 64$, $2^7 = 128$, $2^8 = 256$, $2^{16} = 65{,}536$.
- Given one code, count on through the alphabet on your fingers: H 72, I 73, J 74, K 75.
- ASCII began as an American standard for English; think of Unicode as 'universal code' for every language.

Exam technique

- When a question gives you a code (e.g. 'A' = 65), use it: you are not expected to have memorised the codes.
- In comparisons, cover both character sets: bits per character, number of characters, what they can represent and file size.
- Show the power of 2 you used when working out bits per character.

What each grade needs

What you need to be able to do, from the first marks up to the top grade.

Grade 3 — **Define a character set**
The set of characters a computer can represent, each with its own unique binary code.

Grade 4 — **State how many characters ASCII represents**
ASCII uses 7 bits (128 characters); extended ASCII uses 8 bits (256 characters).

Grade 5 — **Use a given code to find others**
Codes run in sequence, so if 'A' is 65 then 'D' is 68.

Grade 5 — **Explain why Unicode was introduced**
ASCII has too few codes for other alphabets and symbols; Unicode uses more bits so it can represent far more characters.

Grade 6 — **Link bits per character to characters represented**
n bits give 2^n characters, so each extra bit doubles the number of characters.

Grade 7 — **Compare ASCII and Unicode, including file size**
Unicode covers many more characters but uses more bits per character, so text files are larger.

Grade 8 — **Calculate the bits needed for a character set**
Find the smallest n for which 2^n is at least the number of characters needed.

Free notes, Full packs, practice papers and mock exams at chhetriacademy.com. Written for this site; not taken from past papers. Stuck? 1-to-1 tutoring with a free 30-minute trial lesson.