IGCSE COMPUTER SCIENCE • EXERCISE SHEET

1.3 Data storage and compression

Name: ____________ Class: ________ Date: __________ **Total: 29 marks**

KEY WORDS compression 压缩 • lossless 无损 • lossless compression 无损压缩 • lossy 有损 • lossy compression 有损压缩

Objective

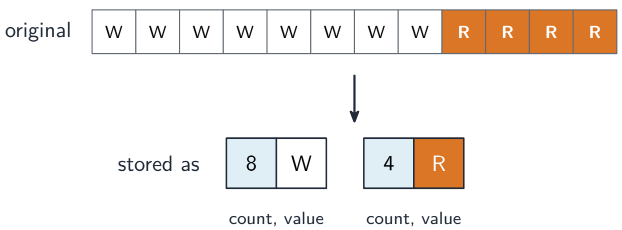
Build the skills to answer exam questions on **data storage and compression** 数据存储与压缩 — storage units, **lossless** 无损 and **lossy** 有损 compression, and **run-length encoding** 行程编码.

You must be able to:

- use the storage units (each ×1024)
- compare lossless and lossy compression
- apply run-length encoding (RLE)

1 Worked examples

■ Compression



RLE stores each run once as a count and a value (lossless)

- **lossless:** no data lost, original rebuilt exactly (text, programs); e.g. RLE.
- **lossy:** permanently removes data → much smaller (photos, music, video).
- units: byte → KiB → MiB → GiB, each ×1024.

■ Compression: what is thrown away, and what is not

Lossless compression 无损压缩 keeps every bit of the original: the file is rebuilt exactly. **Lossy compression** 有损压缩 permanently removes data the user is unlikely to notice, so the file cannot be rebuilt exactly but is much smaller.

Run-length encoding (RLE) is lossless. It replaces a run of identical values with a pair: how many, then which value. The pixel row `w w w w b b w w` becomes `4 w 2 b 2 w`. It works well on an image with large areas of one colour and badly on a photograph, where neighbouring pixels differ, and it can even make such a file larger.

Worked example. Compress `0 0 0 0 0 1 1 0 0 0` with RLE, then say how many bits are saved if each value and each count needs 4 bits.

1. The runs are five 0s, two 1s, three 0s.
2. The encoding is `5 0 2 1 3 0`: six values instead of ten.
3. Original: $10 \times 4 = 40$ bits. Compressed: $6 \times 4 = 24$ bits. Saved: 16 bits.

Where each is used. Lossless: text and program files, where a single wrong character matters, and image formats such as PNG. Lossy: MP3 audio (frequencies outside the range people hear well are removed, and quieter sounds masked by louder ones are dropped) and JPEG images (fine colour detail the eye is poor at seeing is discarded).

Why compress at all: the file takes less storage space, transmits in less time and uses less bandwidth, and a streamed file buffers less.

2 Practice

2.1 State one benefit of compressing a file. [1]

2.2 State how many bytes are in 1 KiB. [1]

2.3 Name the type of compression used for a text file, and give a reason. [2]

2.4 Identify two reasons for compressing a file before sending it over the internet. [2]

2.5 Give the run-length encoding of `b b b w w b b b b`. [2]

3 Exam-style questions

3.1 Which compression type permanently removes data? [1]

- **A** lossless
- **B** lossy
- **C** run-length encoding
- **D** both are reversible

3.2 A file must be compressed.

(a) State whether a text file should use lossless or lossy compression, and why. [2]

(b) Apply run-length encoding to the run `AAAAABBB`. [2]

3.3 Explain why RLE works well for an image with large areas of one colour. [2]

3.4 A photographer stores images and sound recordings.

(a) Explain the difference between lossy and lossless compression. [2]

(b) A black-and-white image row is `w w w w w b b b w w`. Give the run-length encoding of this row. [2]

(c) Each colour and each run length is stored in 4 bits. Calculate the number of bits saved by the encoding in part (b). [3]

(d) Explain why run-length encoding would be a poor choice for a detailed photograph. [2]

(e) The photographer compresses a sound recording with a lossy method. Describe two things the compression removes, and identify one situation in which she must not use lossy compression. [3]

(f) Give two benefits of compressing the files before uploading them to a website. [2]