Exercise walk-through

Data storage and compression ▪ 1.3

eXercise

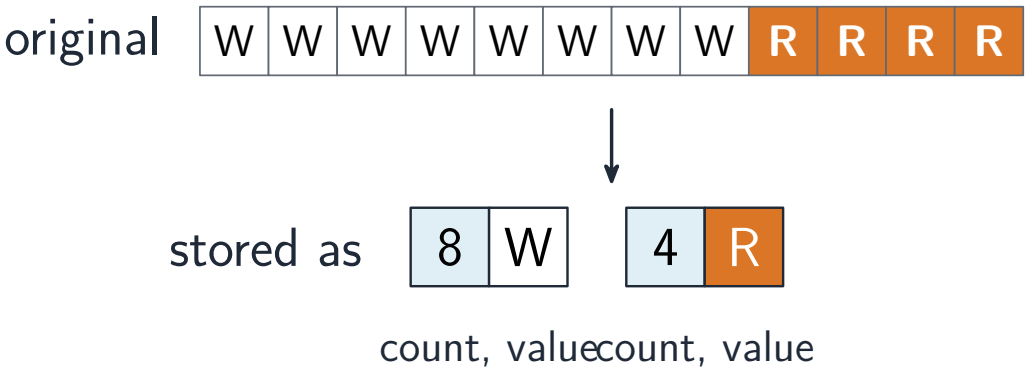
You must be able to

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

Method — Compression

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

Method — Compression



Run-length encoding: runs stored as count + value

Method — 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.

Method — Compression: what is thrown away, and what is not

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.

Practice 2.1 ▪ 1 mark

State one benefit of compressing a file.

Solution 2.1

Worked steps

any of: uses less storage; needs less bandwidth; shorter transmission time B1.

Practice 2.2 ▪ 1 mark

State how many bytes are in 1 KiB.

Solution 2.2

Worked steps

1024 bytes B1.

Practice 2.3 ▪ 2 marks

Name the type of compression used for a text file, and give a reason.

Solution 2.3

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

Worked steps

Lossless B1; every character must be reconstructed exactly, because losing even one would change the meaning of the text B1.

Practice 2.4 ▪ 2 marks

Identify two reasons for compressing a file before sending it over the internet.

Solution 2.4

Worked steps

Any two of: the file transmits faster; it uses less bandwidth; it takes less storage space at the other end; a streamed file buffers less B2.

Practice 2.5 ▪ 2 marks

Give the run-length encoding of `b b b w w b b b b`.

Solution 2.5

Worked steps

`3 b 2 w 4 b` B2 *one mark for the correct runs, one for the correct pairing.*

Exam-style 3.1 ▪ 1 mark

Which compression type permanently removes data?

A lossless

B lossy

C run-length encoding

D both are reversible

Solution 3.1

A lossless

B lossy ✓

C run-length encoding

D both are reversible

Worked steps

B. Lossy compression permanently removes data.

Exam-style 3.2 ▪ 2 marks

A file must be compressed.
(a) State whether a text file should use lossless or lossy compression, and why.
(b) Apply run-length encoding to the run `AAAAABBB`.

Solution 3.2

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

Worked steps

(a) lossless **M1**; because every character must be kept exactly (a text file cannot lose data) **A1**.
(b) `5A 3B` (5 A's then 3 B's) **M1** **A1**.

Exam-style 3.3 ▪ 2 marks

Explain why RLE works well for an image with large areas of one colour.

Solution 3.3

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

Worked steps

(a) Lossy compression permanently removes data the user is unlikely to notice, so the original cannot be restored exactly **B1**; lossless compression keeps all the original data, so the file can be rebuilt exactly **B1**.
(b) `5 w 3 b 2 w` **B2** *one mark for the runs, one for correct pairs.*
(c) Original: $10 \times 4 = 40$ bits **B1**; compressed: $6 \times 4 = 24$ bits **B1**; saved: $40 - 24 = 16$ bits **A1**.
(d) In a photograph, neighbouring pixels are rarely identical, so the runs are very short **B1**; each run still needs a count and a value, so the encoded file may be no smaller, or

Solution 3.3

even larger B1.
(e) Any two of: frequencies outside the range people hear well are removed; quieter sounds masked by louder ones are dropped; the sampling rate or resolution is reduced B2. She must not use lossy compression when the original quality must be preserved, for example a master recording that will be edited further B1.
(f) Any two of: the pages load faster for visitors; less storage is used on the server; less bandwidth is used, so costs are lower B2.

Exam-style 3.3 ▪ 2 marks

A photographer stores images and sound recordings.
(a) Explain the difference between lossy and lossless compression.
(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.
(c) Each colour and each run length is stored in 4 bits. Calculate the number of bits saved by the encoding in part (b).
(d) Explain why run-length encoding would be a poor choice for a detailed photograph.
(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.
(f) Give two benefits of compressing the files before uploading them to a website.

Solution 3.3

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

Worked steps

(a) Lossy compression permanently removes data the user is unlikely to notice, so the original cannot be restored exactly **B1**; lossless compression keeps all the original data, so the file can be rebuilt exactly **B1**.
(b) `5 w 3 b 2 w` **B2** *one mark for the runs, one for correct pairs.*
(c) Original: $10 \times 4 = 40$ bits **B1**; compressed: $6 \times 4 = 24$ bits **B1**; saved: $40 - 24 = 16$ bits **A1**.
(d) In a photograph, neighbouring pixels are rarely identical, so the runs are very short **B1**; each run still needs a count and a value, so the encoded file may be no smaller, or

Solution 3.3

even larger B1.
(e) Any two of: frequencies outside the range people hear well are removed; quieter sounds masked by louder ones are dropped; the sampling rate or resolution is reduced B2. She must not use lossy compression when the original quality must be preserved, for example a master recording that will be edited further B1.
(f) Any two of: the pages load faster for visitors; less storage is used on the server; less bandwidth is used, so costs are lower B2.