

Illustrate the physical-based & probability based models used in lossless compression tech. How are they used to evaluate efficiency?

A Lossless compression reduces file size without losing information. The original data can be perfectly reconstructed. Based on modeling tech, it is classified into:

1. Physical Models.

2. Probability-based Models.

→ Physical Models.

don't exploit the structural repetition of patterns in our data. They remove statistical redundancy by replacing repeated sequence with shorter references.

Working Principle:

- Detect repeated strings
- Replace them with pointer / length or dict ref.
- Reconstruct original data during decoding.

Techniques

a) Run Length Encoding (RLE)

stores repeated symbols as:

Symbol + Count

Eg) AAAA → A4

Used in bitmaps, img & fax transmission

b) Lempel-Ziv (LZ77/LZ78)

Uses sliding window or dict.

Repeated patterns are encoded as (offset, length)

Used in ZIP & GZIP

advantages: simple, fast.

Limitations: sufficient if entropy is low

## 2. Probability-based models.

use statistical freq of symbols, freq occurring  
symbols core assignment shortest codes, reducing  
statistical redundancy.

Working Principle.

Calculate symbol probability  
Assign variable-length codes.  
minimize avg code length.

Tree

### a) Huffman coding

construct binary tree using freq

generate prefix-free codes

Used in LZ77 & audio compressors.

### b) Arithmetic coding

Represent entire message as a

provide higher compression

Better compression ratio.

Adv:

Limitation: More complex

Hybrid model remove redundancy using repeated patterns,  
while probability-based model use symbol freq  
to assign optimal codes.