illustrate the physical and probability - based models used in compression (lossless), highlighting their application

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

1. Physical Models
2. Probability Based Models

1. Physical Models

Physical models exploit the structural repitition of patterns in data. They remove spatial redundancy by replacing repeated sequence with shorter symbols.

Working Principle

- Detect repeated strings
- Replace them with pointer or dictionary reference
- Reconstruct original data during decoding

Techniques

a) Run length Encoding (RLE)

Stores repeated symbols as

Symbol + count

Example: AAAA → A4

Used in bitmap images and fax transmission

b) Lempel-Ziv (LZ77/LZ78)

Uses sliding window or dictionary

Repeated patterns are encoded as (offset, length)

Used in ZIP and GZIP

Advantages: Simple, fast

Limitation: Inefficient if repetition is low

2. Probability-Based Models

These models use statistical frequency of symbols. Frequently occuring symbols are assigned shorter codes, reducing statistical redundancy

Working Principle

- Calculate Symbol probability
- Assign variable length code
- Minimize average code length

Techniques

a) Huffman coding

- Construct binary tree using Frequencies
- Generate prefix-free codes
- Used in image and audio compression

b) Arithmetic coding

- Represent entire message as a fraction b/w 0 and 1
- Provides higher compression efficiency

Conclusion

Physical models remove redundancy using repeated patterns, while probability models use symbol frequency to assign optimal codes.