Go Data Compression

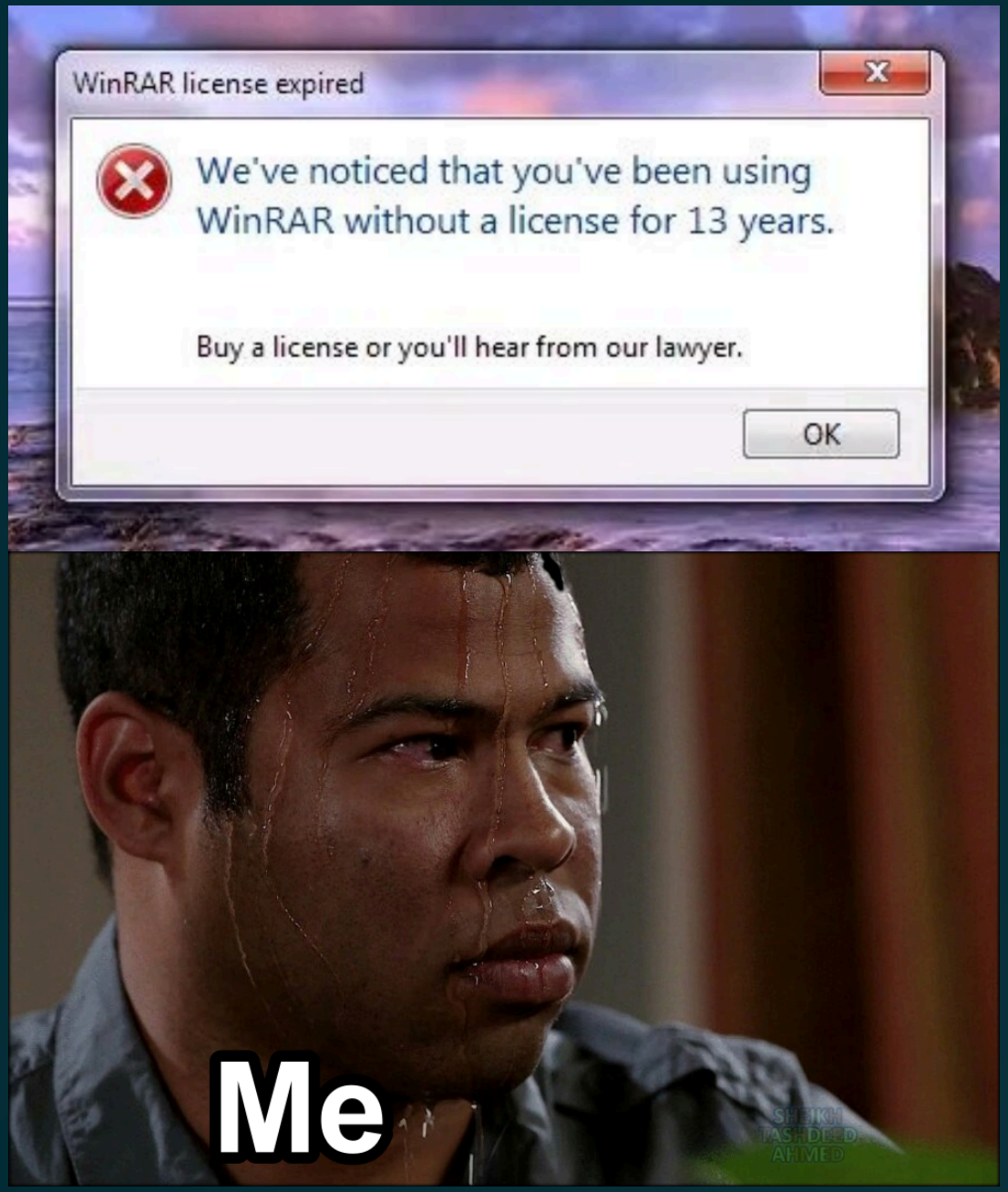
Klaus Post @ Golab 2026

𝕏: @sh0dan

🦋: @klauspost.com

MINIO

100% meat-generated (except stats)


WinRAR license expired
We've noticed that you've been using WinRAR without a license for 13 years.
Buy a license or you'll hear from our lawyer.
OK
Me

Data Compression Primer

Minimize transfer and storage

Always a CPU/size tradeoff

Sometimes Compatibility

Handwaving will happen 👋

Compressor Classes

- Generic Compression
- Domain Specific (+lossy)

moderate rock @lookoutitsbbear · May 27

i got 4.1



5 17 514 342K

i @usrbinishan · May 27

This converts back losslessly?

1 2 922 349K

moderate rock
@lookoutitsbbear

Follow

not only does it convert back losslessly it also has the noise removed

1:21 AM · May 27, 2024 · **4.3M** Views

Match Coding

String of Operations

Literals and back-references

Most common scheme

Example

Entropy Coding

Symbols (bytes) -> Bit Stream

- Huffman
- Range Coding
- ANS

Not order dependent

DEFLATE

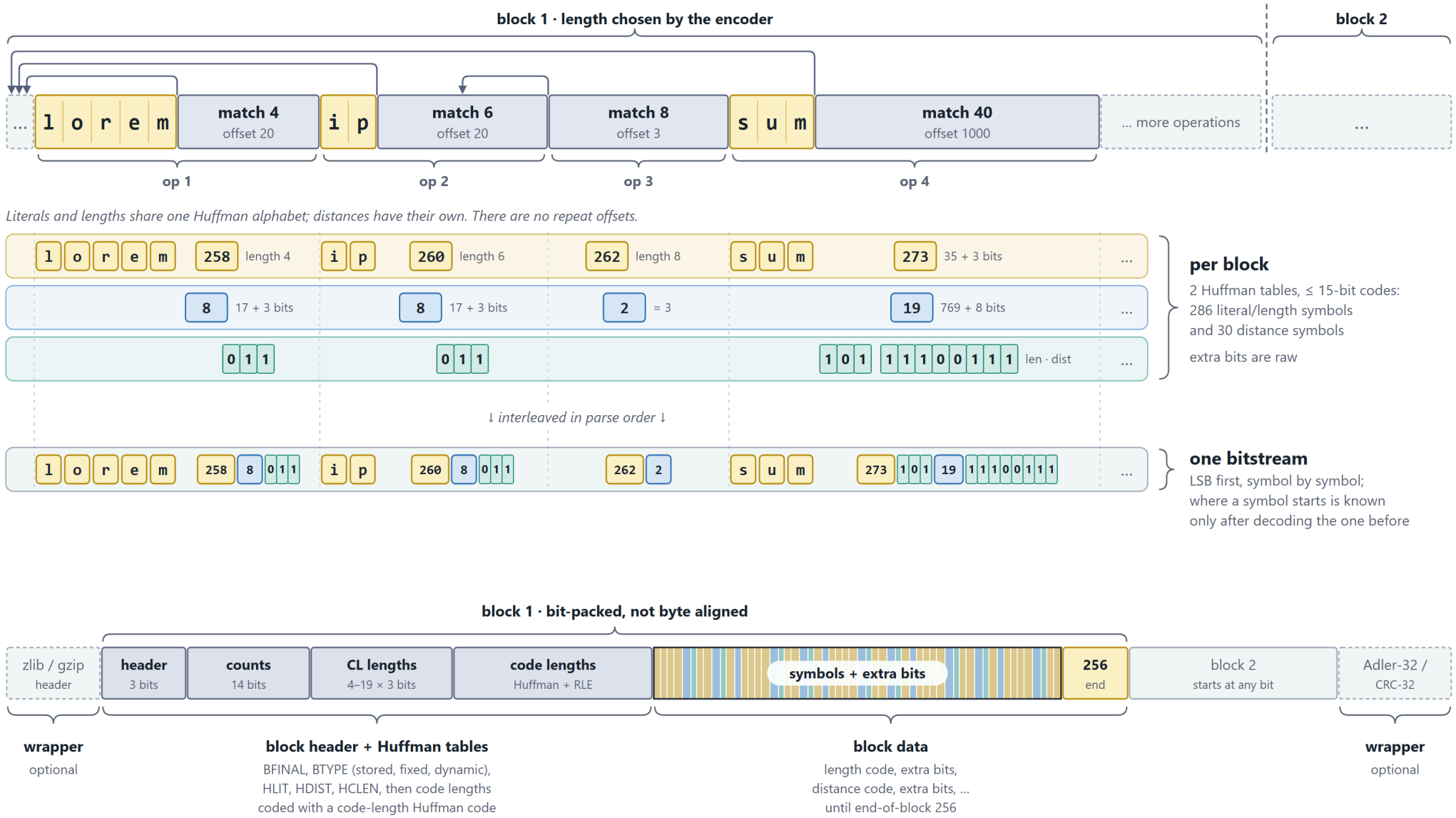
Create by Phil Katz for PKZIP



Huffman Stream with either a literal or a match.

Deflate: encoder buffers

Literals, lengths and distances are Huffman coded into a single bitstream, in parse order.



Everything is one LSB-first bitstream: each symbol starts where the previous one ends, so the decoder works through it serially.

Go 1.27 Improvements

- Custom level 1-6 encoders
- Huffman Encoding
- Level 7-9 tweaks

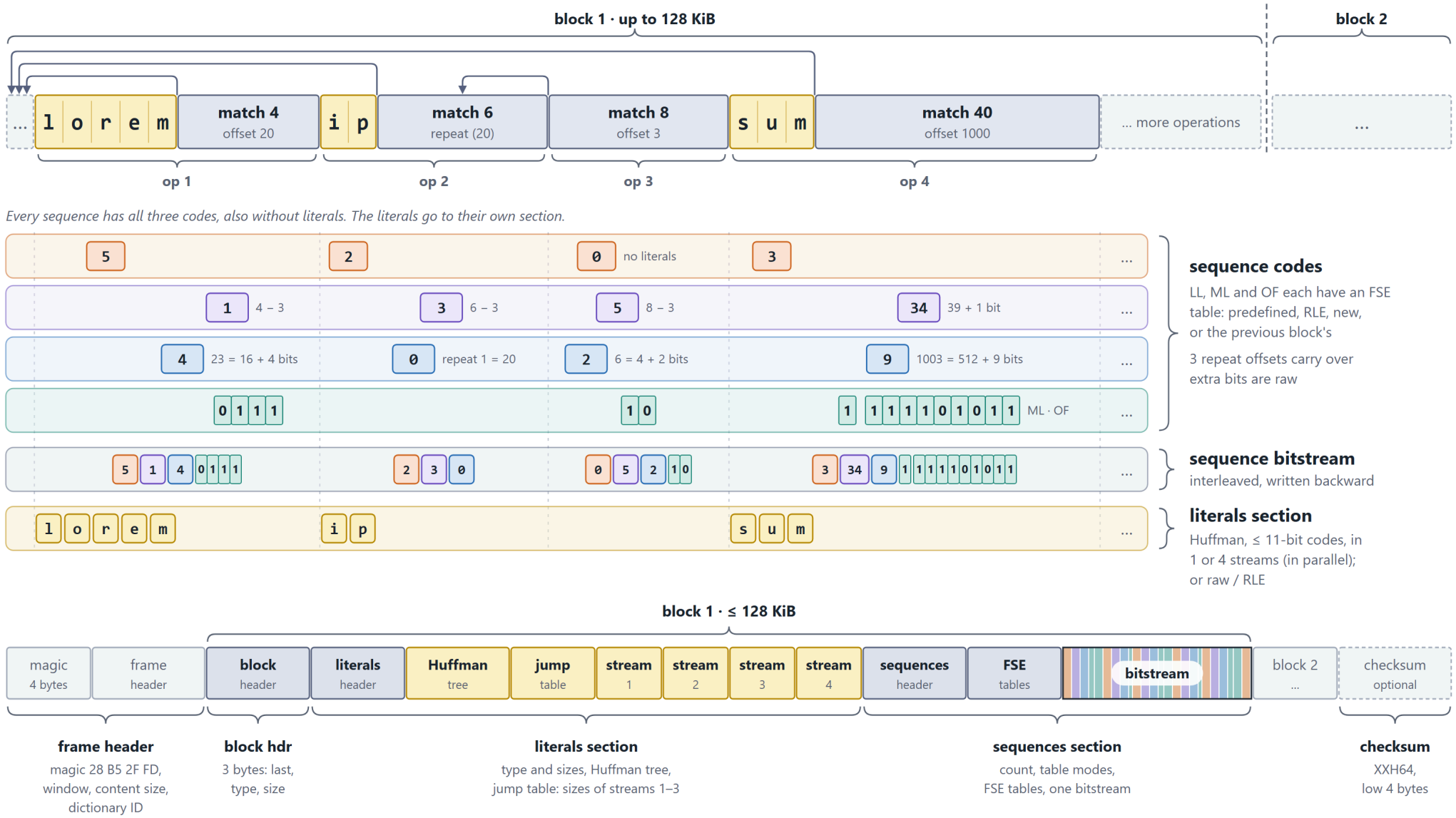
NUMBERS pls

Stdlib Zstandard

WIP - slow but steady.

Zstandard: encoder buffers

Literals are Huffman coded on their own; the sequences become one FSE bitstream with three interleaved states.



The 4 literal streams decode in parallel; the sequence bitstream is decoded one sequence at a time, from its end.

Transformers

(no, not those)

Usually doesn't compress by itself.

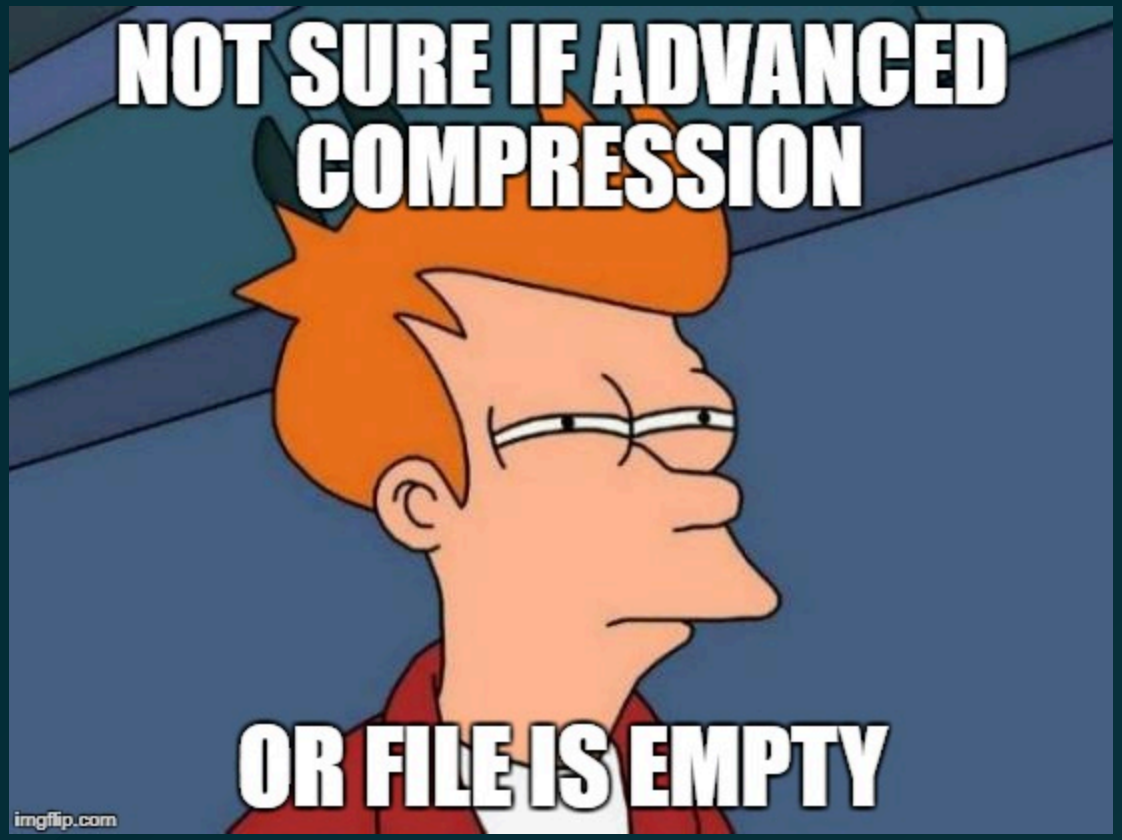
Example: Columnar data

Context Mixing / LLM

Select Encoder based on history.

Good. Exotic. (Too) Slow.

Benchmarking



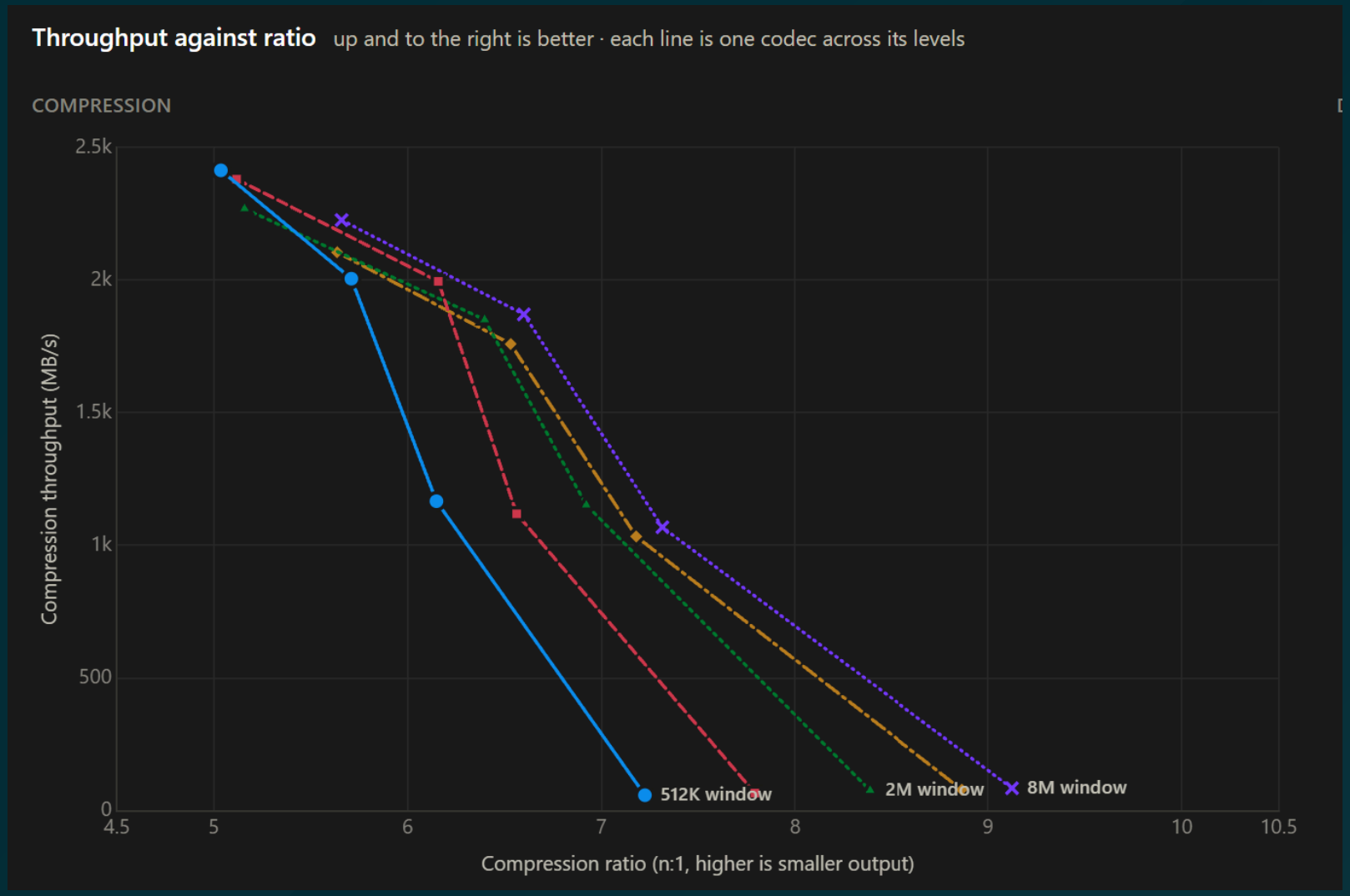
Benchmarking

Dimensions:

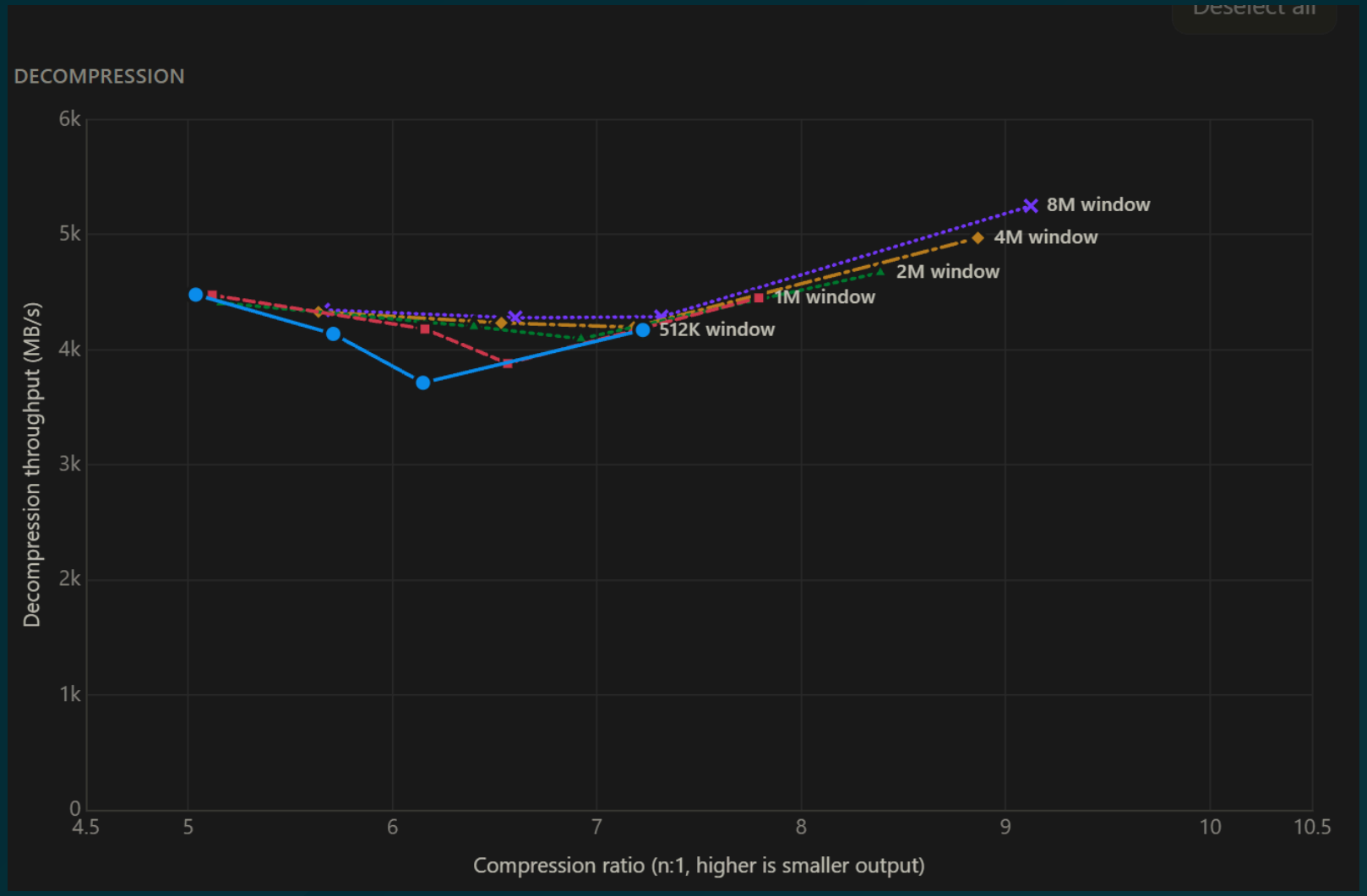
- Speed
- Compression Ratio
- Memory Usage
- Comp vs Decomp
- Representative test data
- Physical CPU/Memory

Let's see some Big Files

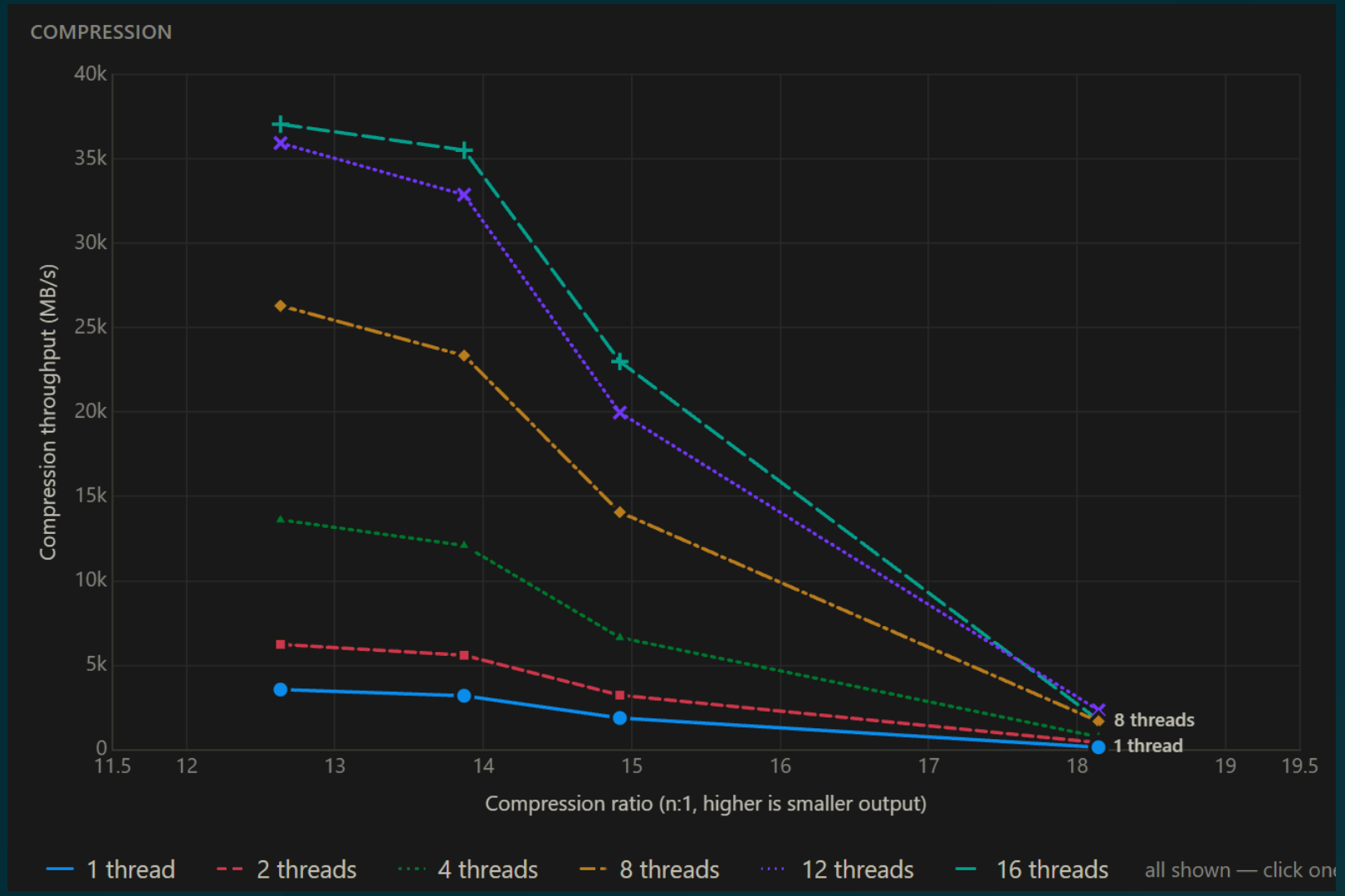
... and Small Files



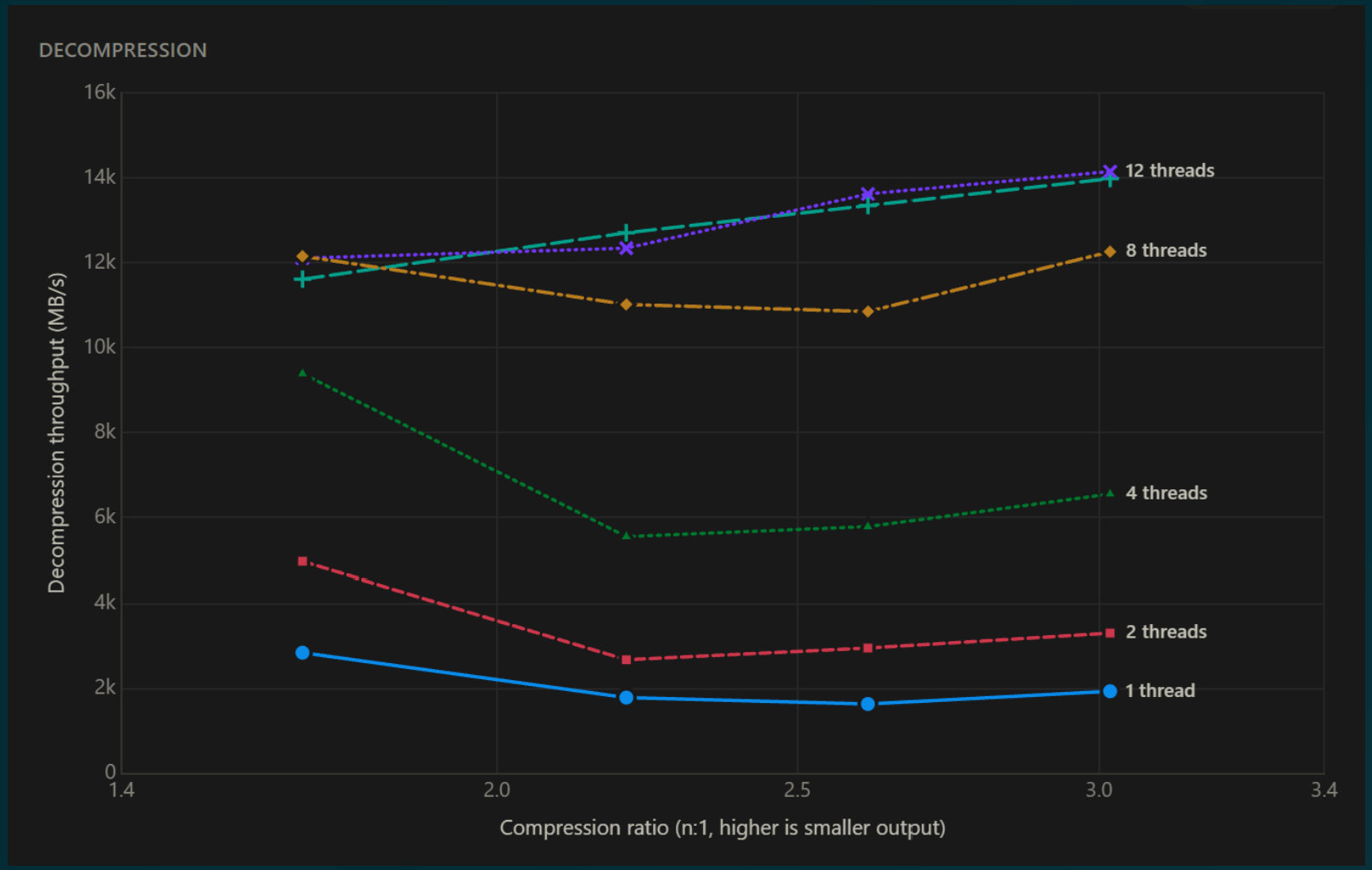
Window Size Compression



Window Size Decompression



Threads Compression



Threads Decompression

Third Party - Byte Aligned

Compressor	Main Upsides	Main Downside
Snappy	+Compatible, OK Speed	Low Comp
LZ4	+Compatible, +Speed, Simple	Low Comp
MinLZ	+Speed, Backcompat	Medium Comp

Bit Aligned

Compressor	Main Upsides	Main Downside
deflate (gz)	+Compatible	Slow Decomp
bzip2	Compatible	Very Slow, incl Decomp
zstd	Compatible, Versatile	Hard to multithread
brotli	HTML, Size	Rarely the best
lzma (xz, 7z)	Size, Flexible	Slow Comp/Decomp

New Tech?

Design a compressor purely around SIMD operations

`pivco-huffman` simd entropy

Goal: Entropy coding with minimal extra CPU

MinLZ SIMD Blocks

Achieved: Entropy coding with 2x-3x faster decompression

`amd64 avx2, avx512 - arm64 neon`

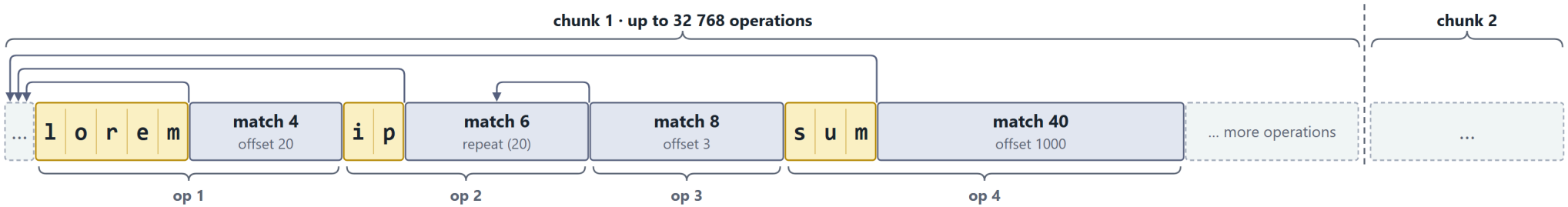
sure.. numbers pls

MinLZ SIMD blocks: encoder buffers

The LZ parse is split into one buffer per field, so the decoder can Huffman-decode and run the operations with SIMD.

1 Parse

input block (≤ 8 MiB)
→ literal run + match



2 Split

Each operation appends only the fields it has, so every buffer is packed.

Tokens
length code · L · R, 1 per op

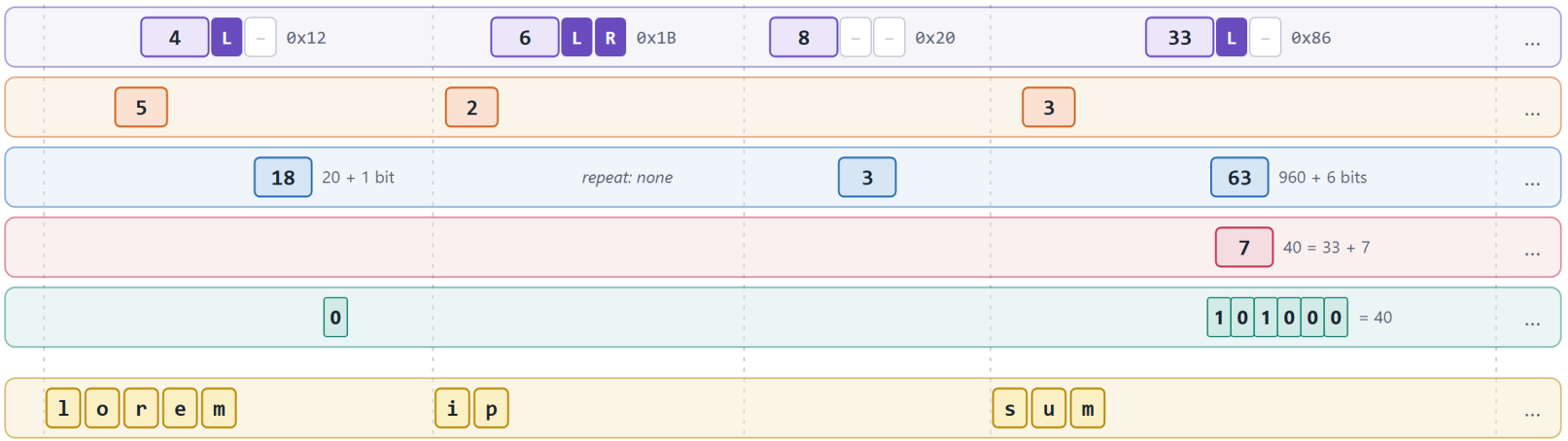
Literal lengths
1–33, one per literal run

Offset symbols
1–175, new offsets only

Escapes
length − 33, for 33 and up

Offset bits
raw low bits of new offsets

Literals
every literal byte



chunk 1

5 buffers, ≤ 32 768 operations

each stored raw or Huffman coded (pivco); a new table carries over to later chunks

offset bits are always raw

literal buffer

whole block, across chunks, cut into literal blocks of ≤ 65 535, raw or Huffman coded

3 Serialize

encoded block sizes at the end



header
indicator (not in streams),
varint: type 1 + decoded size,
flags: chunk limit

LZ section
chunk record: flags + 5 entries,
entry: size · 8 + mode (+ new table)
R: size of the chunk records

literal section
one record per literal block,
entry: size · 8 + mode (+ new table)
R: size of the literal records

L
size of LZ section

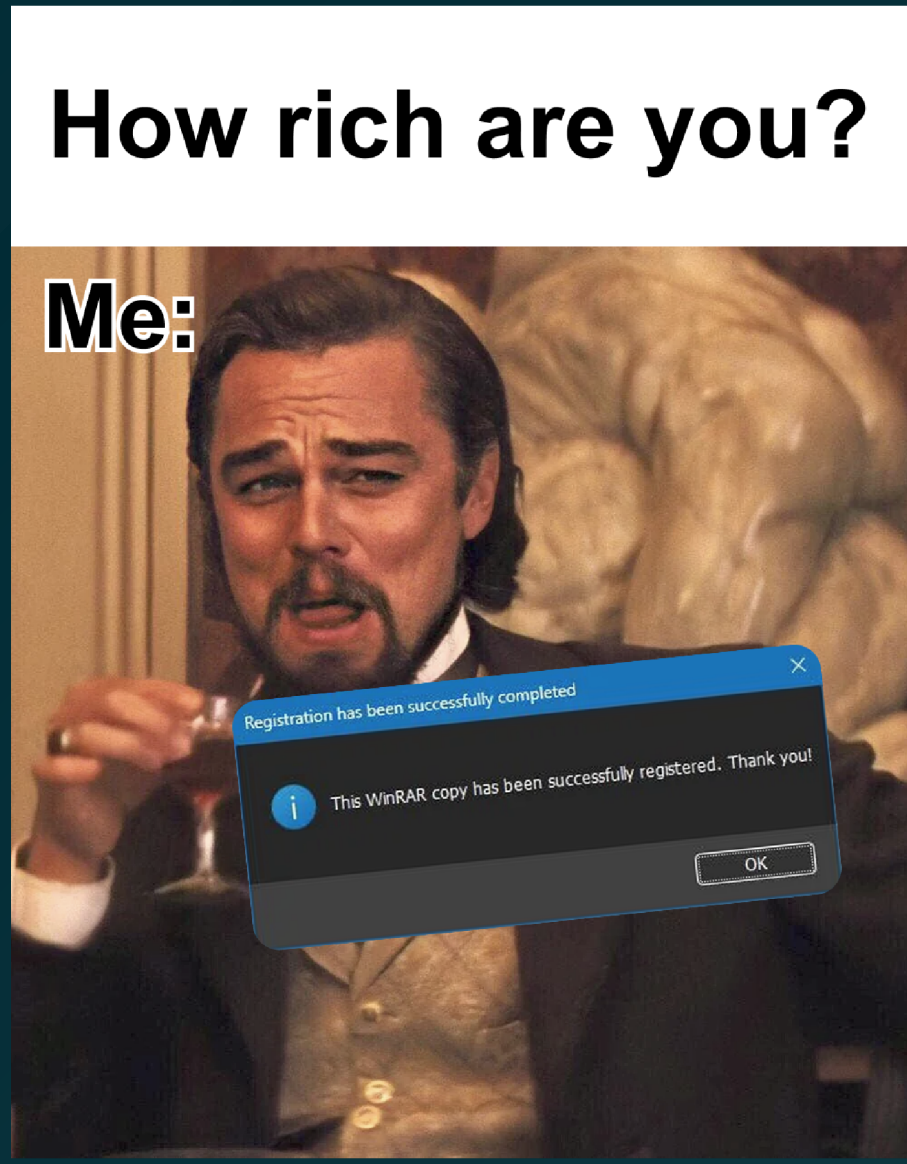
R and L are reverse varints stored after the data they describe: a decoder reads the block from its end.

MinLZ

- Blocks/Streams
- Fully Seekable
- Bloom Filter Search
- Multithreaded
- Snappy Backcompat
- Neat CLI
- Complete Spec
- `github.com/minio/minlz`

Q & A


????


How rich are you?
Me:
Registration has been successfully completed
This WinRAR copy has been successfully registered. Thank you!
OK

Thank You!

Klaus Post @ GoLab 2026

𝕏: @sh0dan

🦋: @klauspost.com

