

# Optimizing Compression Schemes for Sparse Tensor Algebra

**Helen Xu**

**Lawrence Berkeley National Laboratory**

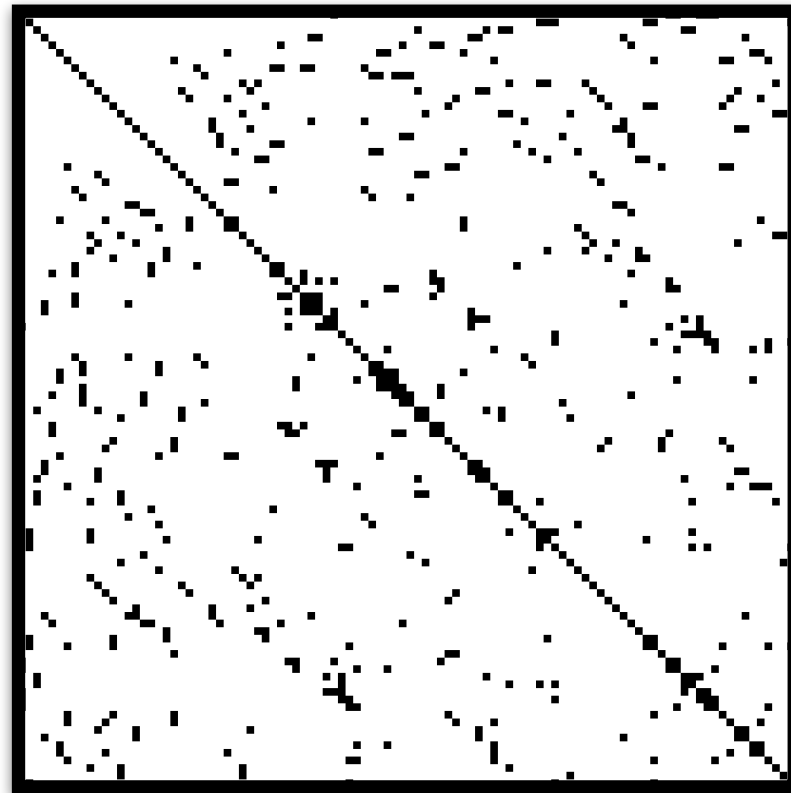
**Joint work with Tao B. Schardl, Michael Pellauer, and Joel S. Emer**

**DCC 2023**

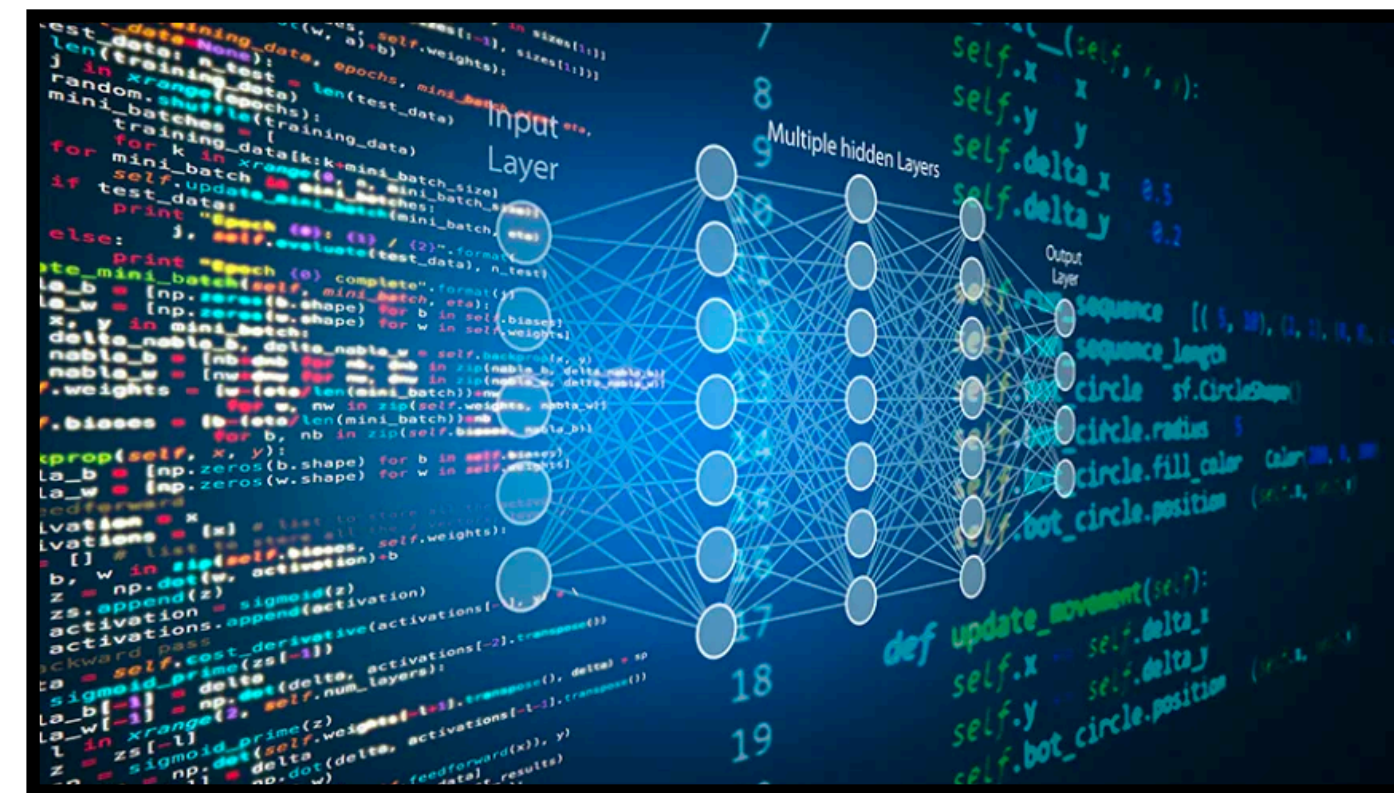


# Sparse Matrices and Tensors are Ubiquitous

**Sparse matrices and tensors**, which have many fewer elements than the total possible number of elements, underlie most real-world applications.



Scientific computing [Saad03]



Data science [JoupYoPa17]



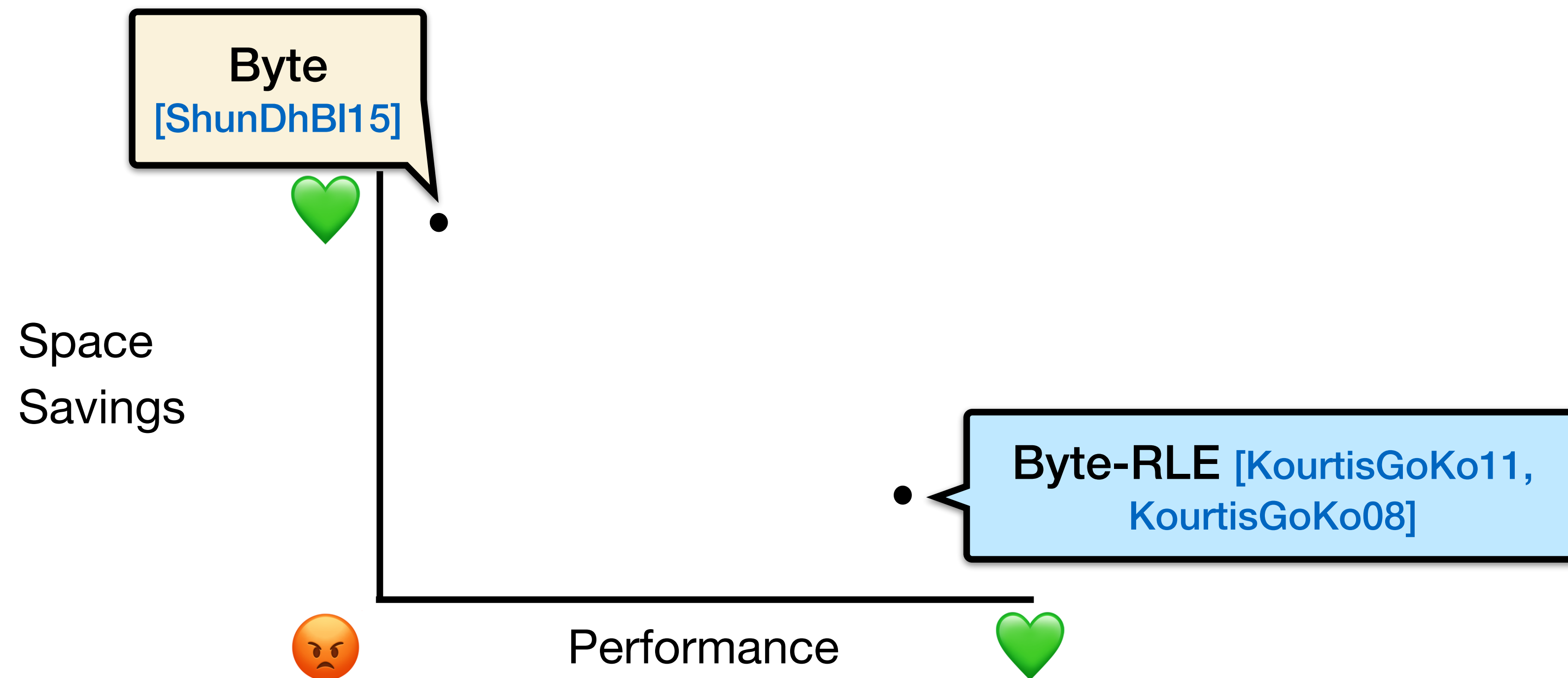
Graph analytics [MattBaBe13]  
...and others!

**Compressed representations** such as Compressed Sparse Rows (CSR) store and compute only on the nonzero elements [TinneyWa67].

# Existing Compression Schemes

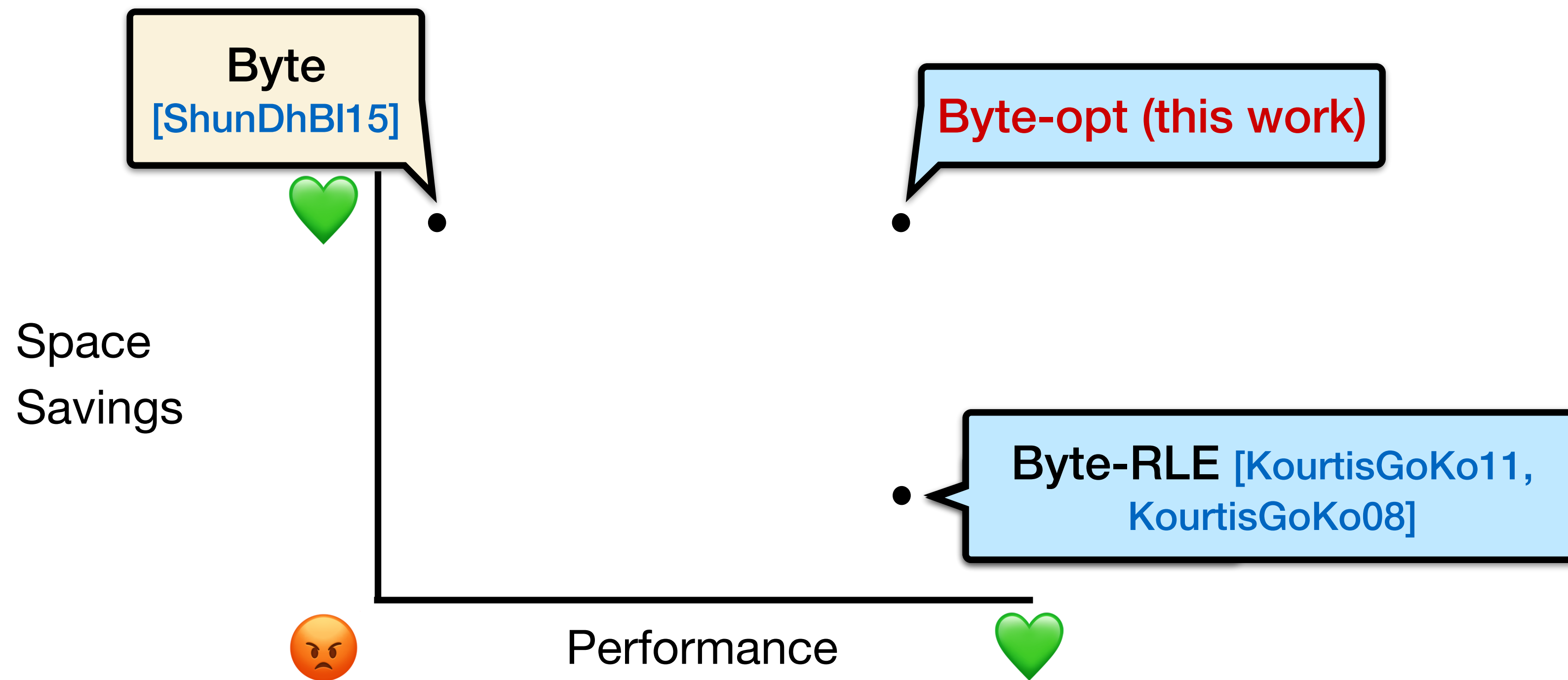
## Trade Space for Performance

More advanced compressed representations reduce memory usage but may become **latency-bound** depending on the structure of the input matrices [Goumas et al. '09].



# This Work: Overcoming the Space-Time Tradeoff

We introduce a novel-algorithm called **byte-opt** that saves space without sacrificing performance.



# Space and Performance Evaluation

We implemented the compression schemes on top of codes generated by `taco`, a sparse tensor algebra compiler [\[Kjolstad et al. 17\]](#).

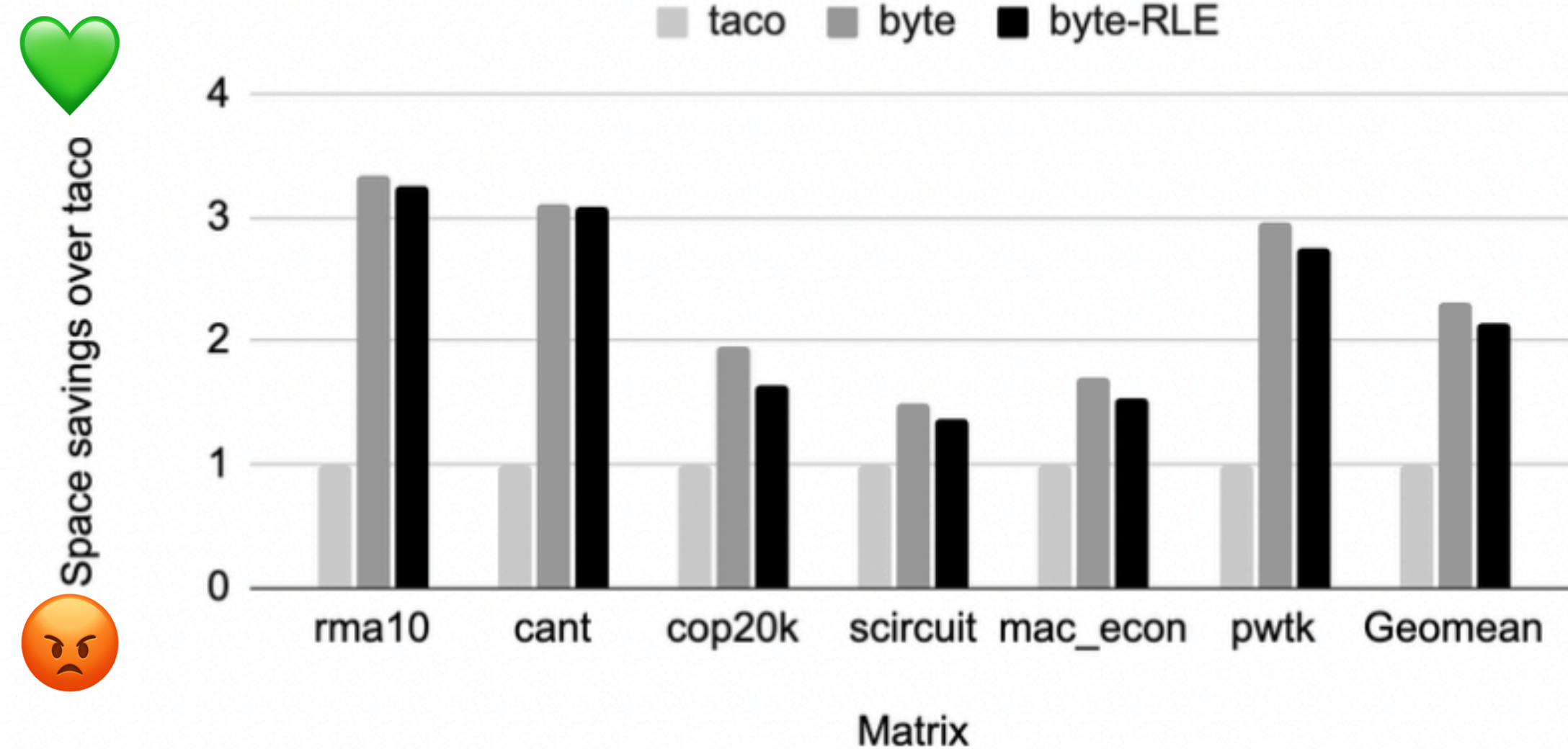


Figure 1: Space savings over original `taco` CSR.

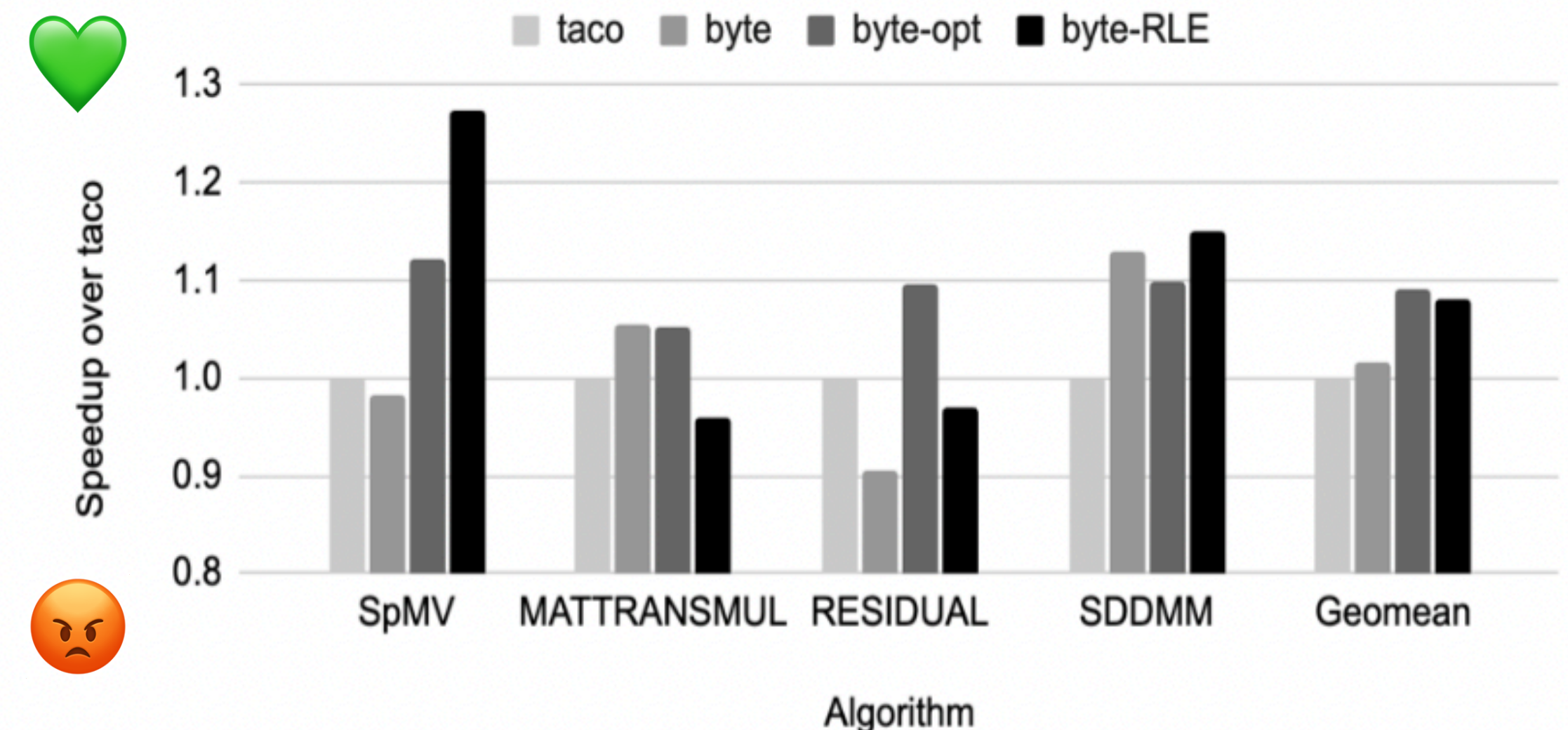


Figure 2: Speedup over original `taco` CSR when run on 48 hyperthreads.