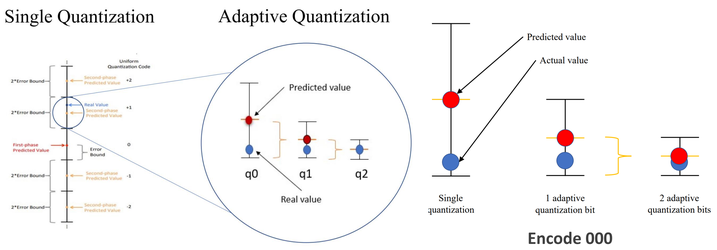
ADAPTIVE ERROR QUANTIZATION FOR LOSSY COMPRESSION OF SCIENTIFIC DATASETS

Student: Samuel Quan **Advisor:** Dr. Jon C. Calhoun

INTRODUCTION

- Scientific research uses prediction-based error bounded lossy compressors due to larger compression ratios (CRs) in relation to lossless compressors [1]
- Unpredictable data points fall outside the error bound of the prediction (hard-to-compress), reducing advantage of lossy compression over lossless methods
- To address hard-to-compress data points, this work introduces adaptive error quantization, which further quantizes the error of hard-to-compress values as needed to achieve higher data fidelity at similar or better CRs



EXPERIMENTAL SETUP

Compressors

- SZ3 [2] – version 3.1.5.4
  - Append SQ & Pack Adaptive
  - Append SQ & Huffman Encode Adaptive
  - Append Adaptive to SQ
  - Pack Adaptive
  - Huffman Encode Adaptive

Datasets

- HACC (1×1073726487)
  - 6 fields
  - Particle simulation dataset

Test Metrics

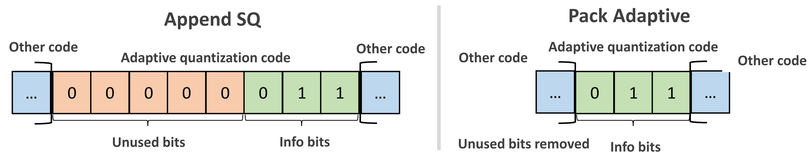
- Peak Signal to Noise Ratio (PSNR) for data fidelity
- Compression Ratio (CR)

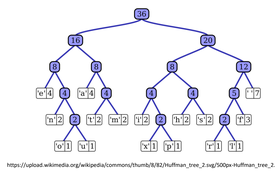
$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(e_{abs})^2}$$

$$PSNR = 20 \cdot \log_{10}\left(\frac{R_{max}}{RMSE}\right)$$

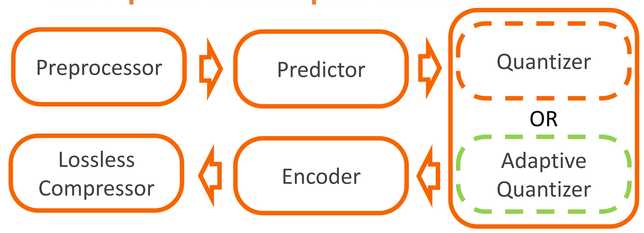
$$CR = \frac{\text{Original Data Size}}{\text{Compressed Data Size}}$$

Adaptive Quantization Storage Methods:



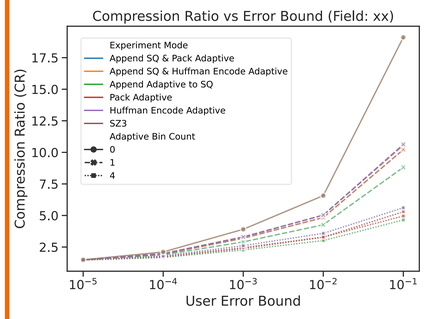


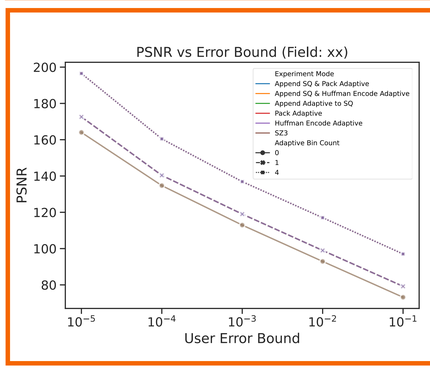
SZ3 Pipeline with Adaptive Quantization:

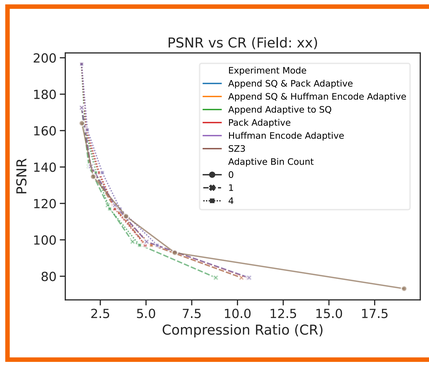


RESULTS:

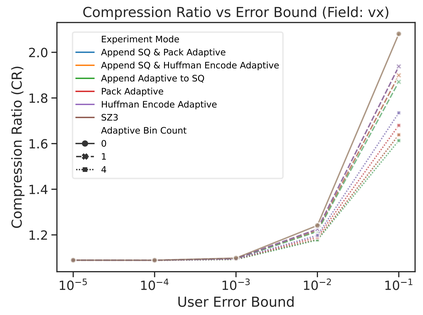
Position

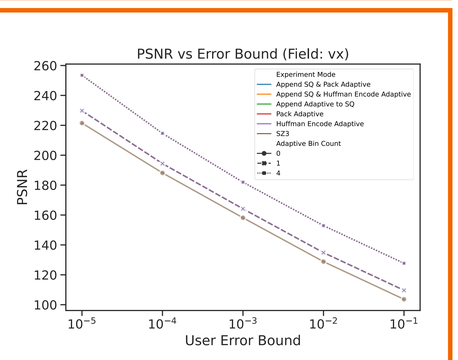


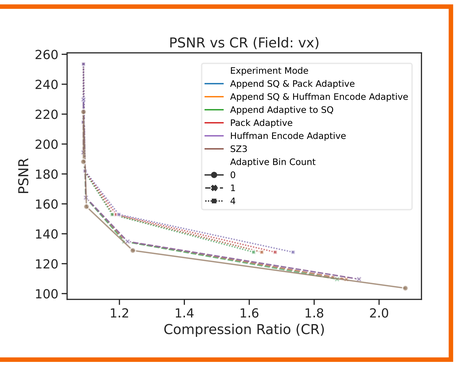




Velocity







CR vs Error Bound

- HACC fields are either velocity or position, *xx* and *vx* fields chosen as representative samples
- Velocity is harder to compress field than position, CR tradeoff of adaptive quantization is observably lower
- Huffman Encoding most efficient storage

PSNR vs Error Bound

- Adaptive quantization achieves higher PSNR than baseline SZ3 with very little CR tradeoff at lower error bounds
- CR tradeoff is most competitive for the harder to compress velocity field
- No PSNR change between storage method (expected)

PSNR vs CR

- For harder to compress Velocity field, adaptive quantization can provide both higher compression ratios *and* higher PSNR on the same dataset than SZ3
- For more predictable position fields, adaptive quantization shows advantage over SZ3 when desired CR is lower

CONCLUSIONS

- **Adaptive quantization methods applied to hard-to-compress data can increase both compression ratio and data quality over a baseline SZ3 prediction-based method**
- **Flexible choice of adaptive bin count allows user control over how much storage to sacrifice for higher compression quality**
- **Refining storage method for adaptive quantization codes reduces the necessary compression ratio tradeoff to achieve higher data fidelity, extending applicability**
- **Potential for future work to further refine storage methods**

CONTACT

Email: quansamuelh@gmail.com
Github: github.com/shquan04

Clemson University is acknowledged for generous allotment of compute time on the Palmetto cluster.

This material is based upon work supported by the National Science Foundation under Grant No. SHF-1943114.

[1] Kai Zhao, Sheng Di, Maxim Dmitriev, Thierry-Laurent D. Tonellot, Zizhong Chen, and Franck Cappello. "Optimizing Error-Bounded Lossy Compression for Scientific Data by Dynamic Spline Interpolation" , Proceeding of the 37th IEEE International Conference on Data Engineering (ICDE 21), Chania, Crete, Greece, Apr 19 - 22, 2021.
[2] Xin Liang, Kai Zhao, Sheng Di, Sihuan Li, Robert Underwood, Ali M Gok, Jiannan Tian, Junjing Deng, Jon C Calhoun, Dingwen Tao, Zizhong Chen, and Franck Cappello. "SZ3: A modular framework for composing prediction-based error-bounded lossy compressors", IEEE Transactions on Big Data, 2022.