Understanding GPU-Based Lossy Compression for Extreme-Scale Cosmological Simulations

Sian Jin (The University of Alabama)
Pascal Grosset (Los Alamos National Laboratory)
Christopher M. Biwer (Los Alamos National Laboratory)
Jesus Pulido (Los Alamos National Laboratory)
Jiannan Tian (The University of Alabama)
Dingwen Tao (The University of Alabama)
James Ahrens (Los Alamos National Laboratory)

May 2020

Introduction

Understanding GPU-Based Lossy Compression for Extreme-Scale Cosmological Simulations

Why Compress/Lossy Compression?

- Huge amount of data from cosmological simulations.
  - Write speed.
  - Data storage.
- Much higher compression ratio compared to lossless compression.

Why Evaluate On Cosmological Simulations?

- Traditional distortion analysis are not sufficient.
- No prior work studying GPU-based lossy compression for large-scale cosmological simulations.

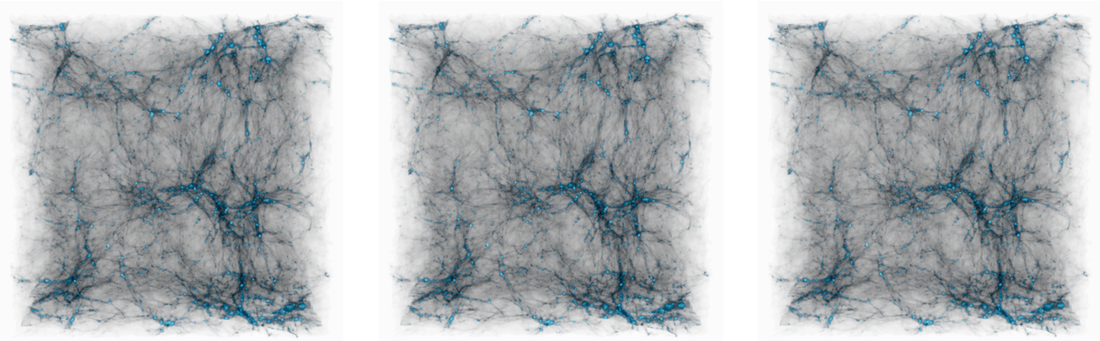
Why GPU?

- DoE supercomputers are moving towards GPU based architecture.
- Higher (de)compression throughput.
- Data is generated on GPU.

Introduction

What We Did

- Implement GPU-based lossy compressors into Foresight, our open-source compression benchmark and analysis framework.
- Comprehensively evaluate the practicality of using GPU-based lossy compressors with various compression configurations on two well-known cosmological simulation datasets.
- A general optimization guideline for domain scientists on how to determine the best-fit compression configurations for different GPU-based lossy compressors and cosmological simulations.



(a) Original (b) Reconstructed with PW_REL = 0.1 (c) Reconstructed with PW_REL = 0.25

Visualization of Nyx dataset compressed with lossy compressor with different configurations

Foresight is available at: https://github.com/lanl/VizAly-Foresight

Background

Cosmological Simulation: HPC code to simulate cosmological evolution of the universe in extreme time and particle scale.

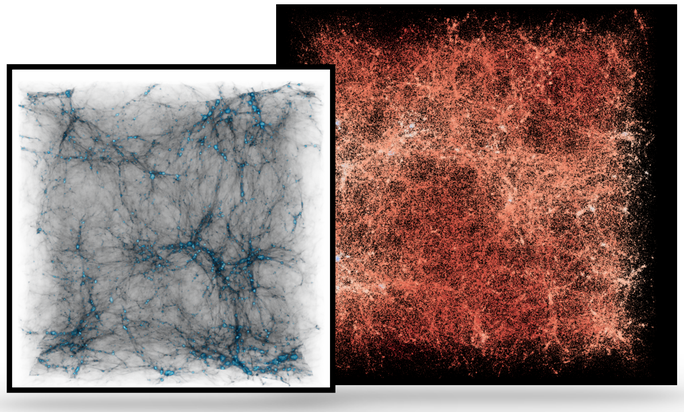
Lossy Compression: compress data with little information loss in the reconstructed data.

HACC

- Simulates the mass evolution of the universe for all available supercomputer architecture.
- Particle simulations, contains 1-D datasets.

Nyx

- Model astrophysical reacting flow on HPC systems.
- Field simulations, contains 3-D datasets.



Nyx simulation (left) HACC simulation (right).

Compression Modes

- Absolute Error bound (ABS).
- Power Relative Error Bound (PW_REL).
- Fixed rate.

SZ

- Prediction Based.
- Suitable for ABS, PW_REL, etc.

ZFP

- Block transfer based.
- Suitable for Fixed rate.

Foresight Design

Understanding GPU-Based Lossy Compression for Extreme-Scale Cosmological Simulations

CBench

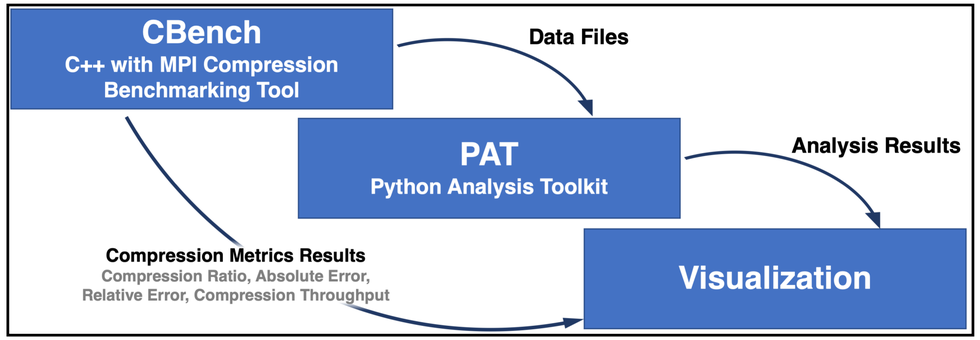
- A compressor benchmarking tool designed for scientific simulations.

PAT

- Python Analysis Toolkit, lightweight workflow submission Python package that contains a number of utilities for scheduling SLURM jobs.

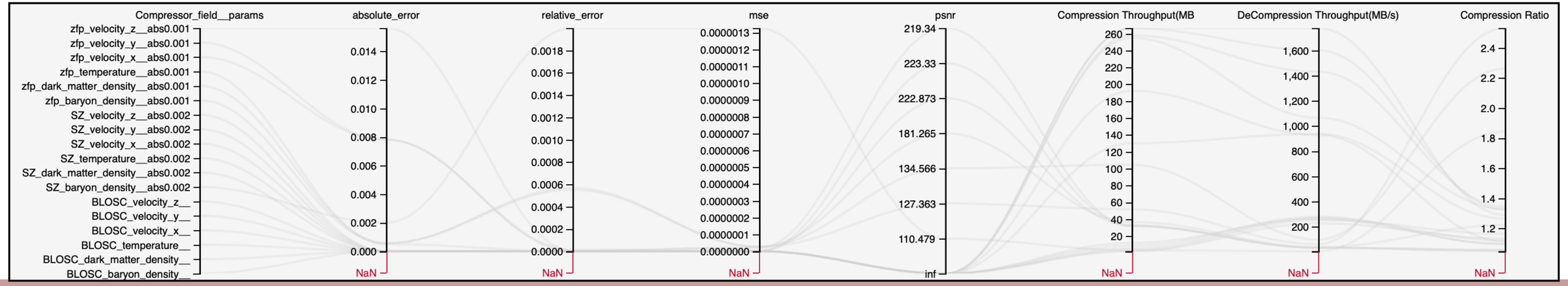
Visualization

- Takes metrics from CBench and analysis by PAT to generate parallel coordinate plots using the Cinema Framework.



↑Three components of foresight framework.

↓A visualization that demonstrate the result from CBench



Evaluation Methodology

Lossy Compressors

- SZ lossy compressor, GPU prototype.
- ZFP lossy compressor, GPU CUDA implementation.

Evaluation Datasets

- HACC dataset, particles generated with model M001 to cover a $(0.36\ \text{Gpc})^3$ volume and redshift value sets to be 0.
- Nyx dataset, single-level grid structure without adaptive mesh refinement (AMR).

Dataset	Dimension	Size	Field	Value Range
HACC	1,073,726,359	38 GB	Position (x, y, z)	$(0, 256)$
			Velocity (vx, vy, vz)	$(-10^4, 10^4)$
Nyx	512×512×512	6.6 GB	Baryon Density	$(0, 10^5)$
			Dark Matter Density	$(0, 10^4)$
			Temparature	$(10^2, 10^7)$
			Velocity	$(-10^8, 10^8)$

HACC and Nyx dataset details used in the experiments

Implementation Technique

- Dimension conversion for data dimension that is not yet supported with corresponding compressor.
- Logarithmic transformation for PW_REL compression mode.

Evaluation Results

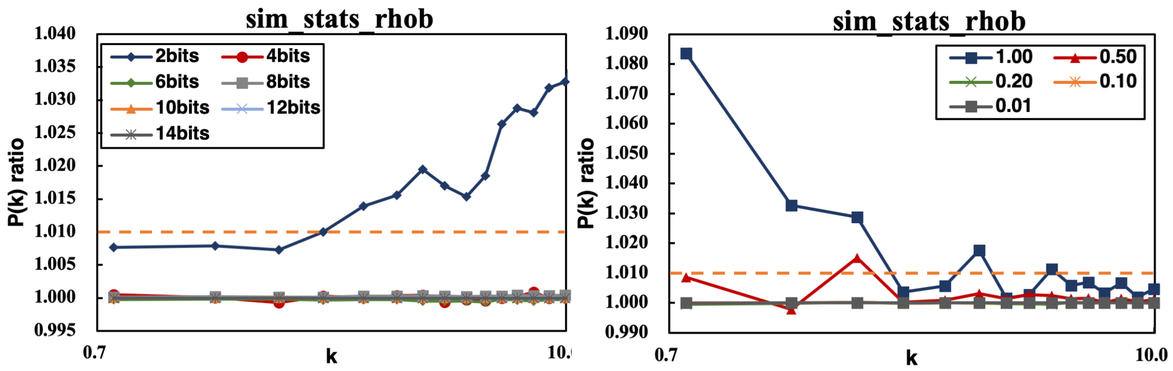
Understanding GPU-Based Lossy Compression for Extreme-Scale Cosmological Simulations

Rate-Distortion

- SZ provides lower rate-distortion than ZFP
- ABS mode has better performance than Fixed-rate mode on Nyx and HACC

Power Spectrum

- Maintain the pk ratio within ±1%.
- Overall compression ratio with cuZFP at 10.7x and GPU-SZ at 15.4x.

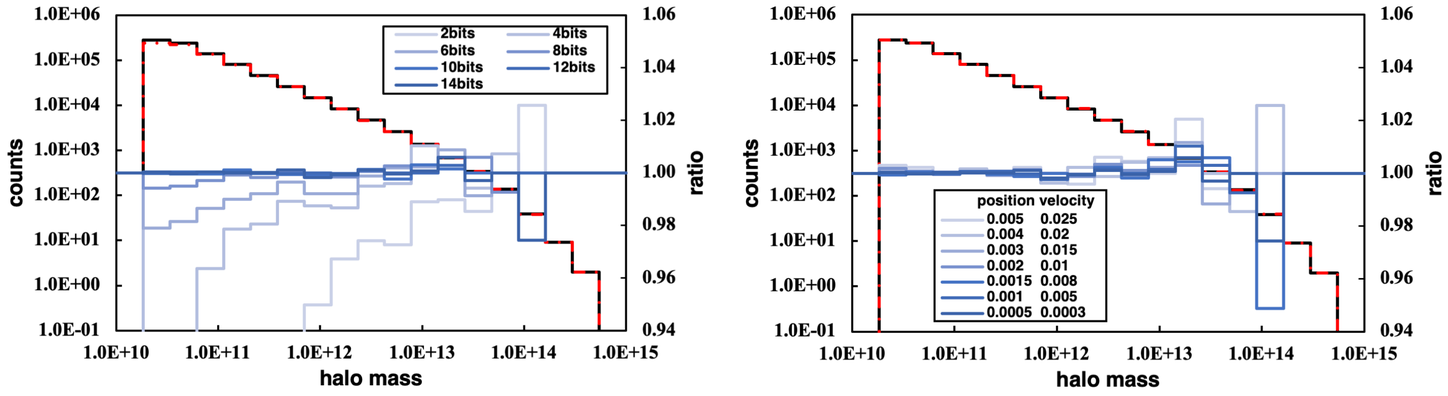


Power Spectrum of Nyx dataset with cuZFP (left) and GPU-SZ (right).

Halo Finder Analysis

- Similar results from original and reconstructed dataset.
- Overall compression ratio with cuZFP at 4.0x and GPU-SZ at 4.3x.

GPU-SZ provides a higher compression ratio than cuZFP



Halo Finder analysis on HACC dataset with cuZFP (left) and GPU-SZ (right).

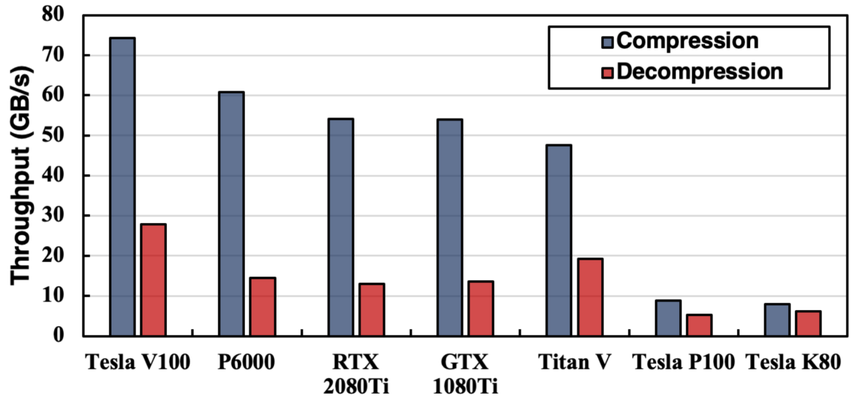
Evaluation Results

Understanding GPU-Based Lossy Compression for Extreme-Scale Cosmological Simulations

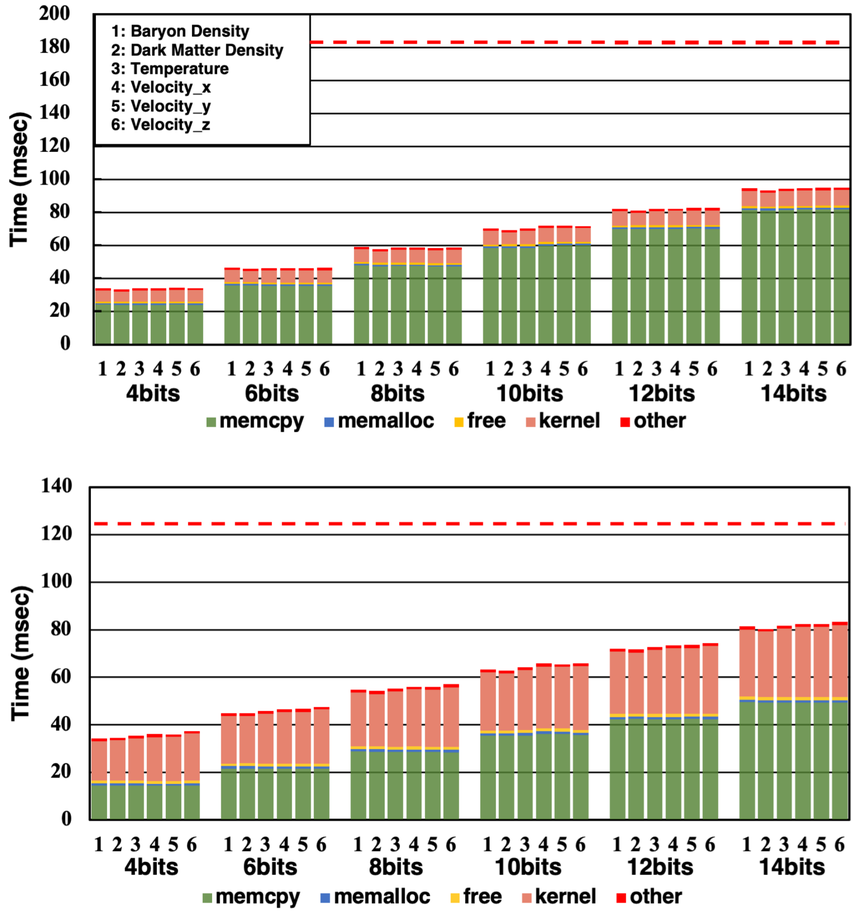
Throughput Evaluation

- High throughput with GPU-based lossy compressors.
- Overall transfer time still much lower than baseline.
- Kernel throughput increased by using a GPU with more shaders, higher pick performance and higher memory bandwidths.

cuZFP provide higher throughput than GPU-SZ



Comparison of kernel compression and decompression throughput with cuZFP on different GPUs



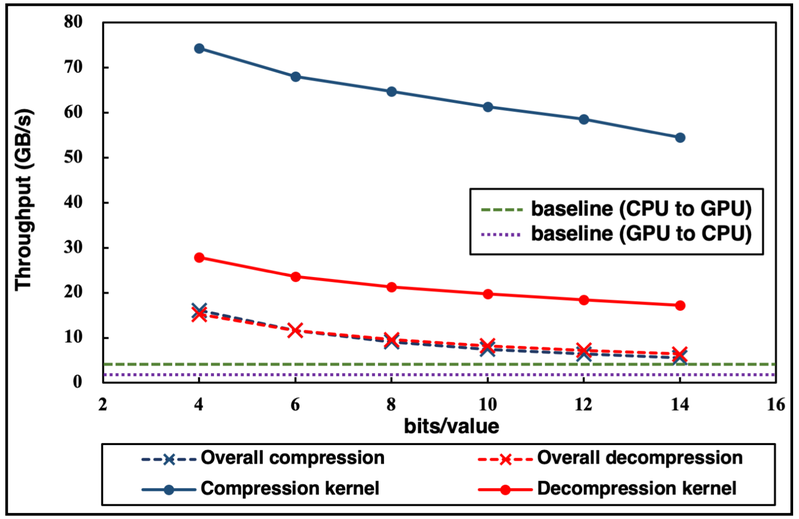
Breakdown of compression (top) and decompression (bottom) time with cuZFP on Nyx dataset. Red dashed line is baseline

Evaluation Results

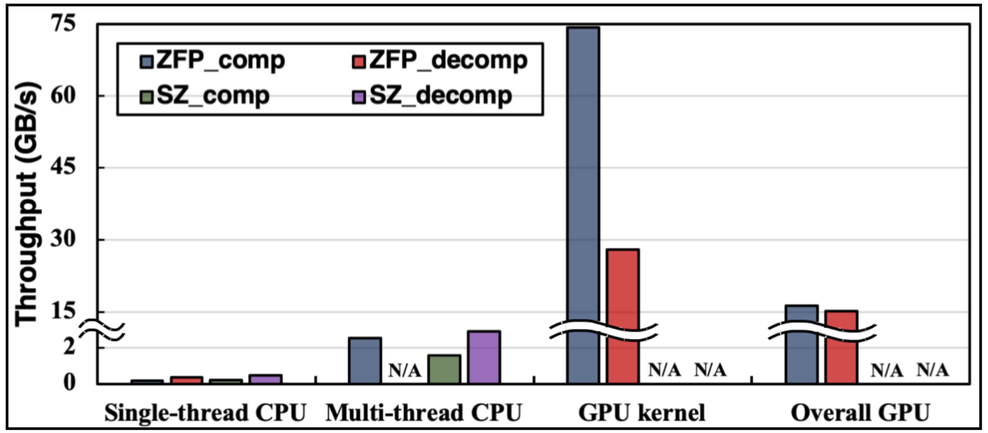
Understanding GPU-Based Lossy Compression for Extreme-Scale Cosmological Simulations

Guidelines

- Use our Foresight framework to benchmark different GPU- based lossy compressors with various configurations targeting cosmological simulation datasets.
- Identify a set of configurations to produce acceptable reconstructed data using power spectrum and halo finder analysis.
- Choose the optimal configuration with the highest compression ratio as the best-fit setting.



Compression and decompression throughput along with bit rate (compression ratio).



Comparison of compression and decompression throughput with SZ and ZFP on CPU and GPU.

Conclusion & Future Work

Understanding Impact of Lossy Compression On Exa-Scale HPC Applications And Developing In Situ Capability

Conclusion

- Implemented GPU-based lossy compressors into our open-source compression benchmark and analysis tool Foresight.
- Conduct a thorough empirical evaluation for two leading GPU-based error-bounded lossy compressors on the real-world extreme-scale cosmological simulation datasets HACC and Nyx.
- Evaluated a different compression configurations and their affection on general compression quality and post-analysis quality.
- Provided general optimization guidelines for cosmology scientists on how to determine the best-fit configurations for different GPU-based lossy compressors and extreme-scale cosmological simulations.

If you have further questions, fell free to contact Dingwen Tao: dingwen.tao@ieee.org