Spline-Interpolation-Based Error-Bounded Lossy Compression for Scientific Data on GPUs

Introduction: Data-Heavy Scientific Applications

- High-quality for accurate post-analysis
- Data reduction rate at 10x in need
- High-throughput processing to ease I/O and communication pressure

Big-Data Scientific Application
- Data Reduction
  - HACC: 20 PB per year (challenge in I/O)
  - CESM: 50% of all the change in 2017
  - APS-U: 100x PB

Data management is a real-world problem to address when we advance in scientific exploration.

SZ [1, 2] Lossy Compression Essence

To lower bit randomness-prediction-based SZ, while guaranteeing the error-boundness, performs bit-level randomness elimination to increase the compression.

Interpolative Prediction Data-Access Pattern

We use 2D interpolation for conceptual demonstration, the prediction direction alters each stage.

An interpolation iteration contains nxd stages. Stages feature altering interpolation direction. When an iteration finishes, the interpolation distances shrinks by a factor of 2.

Starting points are distant from each other and beyond GPU thread blocks can handle, they become anchor points and are saved directly.

Error-Quantization Distribution

Compressing Ratio Throughout 3,000 RTM Snapshots

Compressing Ratio & Quality Across (GPU, CPU)-Spline and Default (Lorenzo)

Preliminary Evaluation (Seismic Data)
- GB/s compression
- A100: 253.5 (0.99x)
- V100: 232.7 (1.00x)
- decompression
- A100: 335.8 (1.39x)
- V100: 242.3 (1.00x)

Need further in-depth optimization.

Dataset: RTM Simulation #600 to #3000.

Impact on Error-Quantization

The larger the value, the more sparse.

Interpolating 32x18x8 data chunk. With anchor points, the error quantizations are smaller in amplitude throughout different error bounds. Anchor points are used for performance concern: fitting subproblem size to GPU hardware architecture.

Reference