

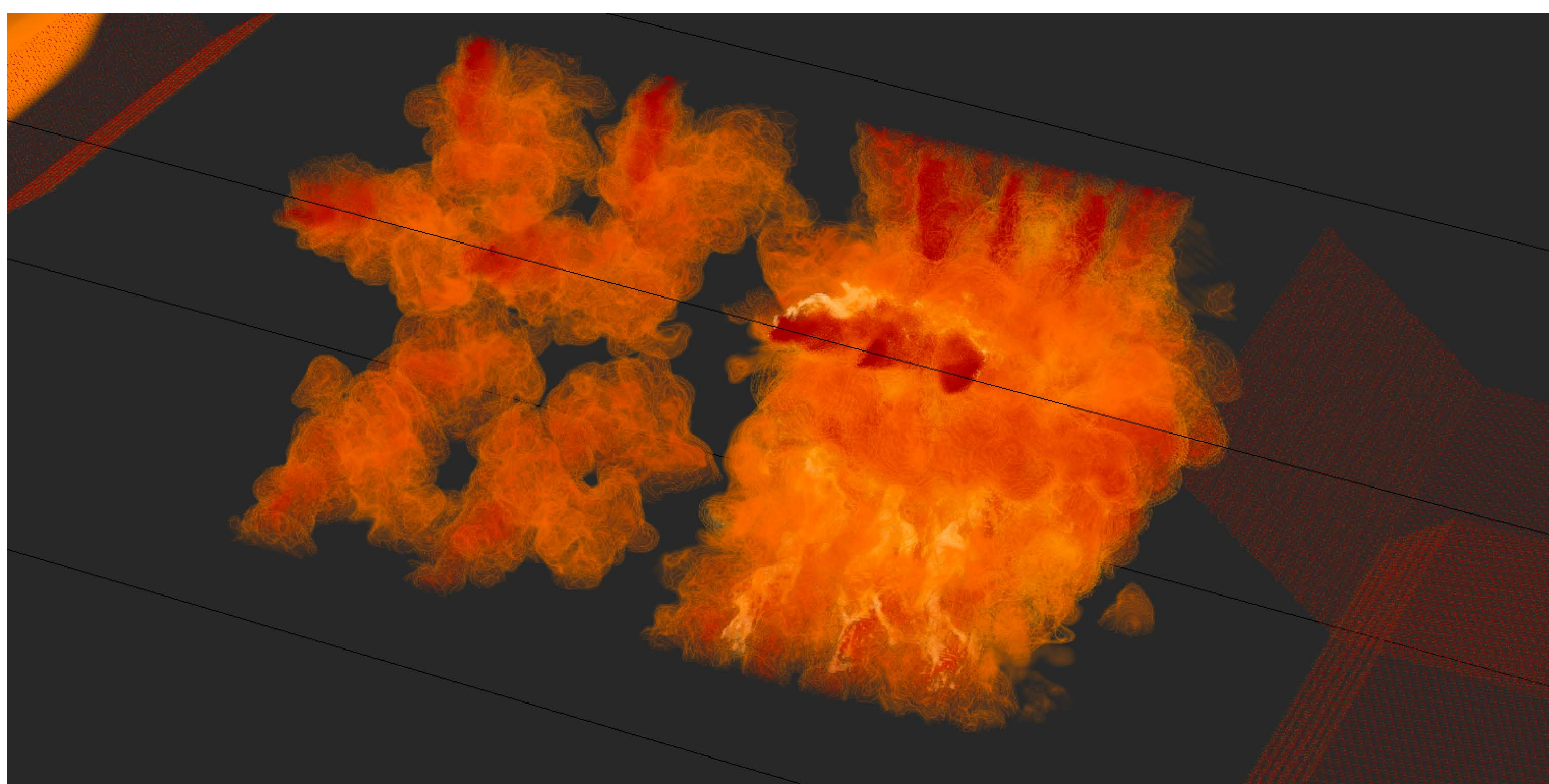
ViSUS

Data streaming techniques for progressive processing and visualization of large scientific datasets.

Exploit the coupling between time-critical algorithms and progressive multi-resolution data-structures to realize an end-to-end optimized flow of data from the original source, such as remote storage or large scientific simulation, to the rendering hardware.

Three major visualization modalities with ViSUS :

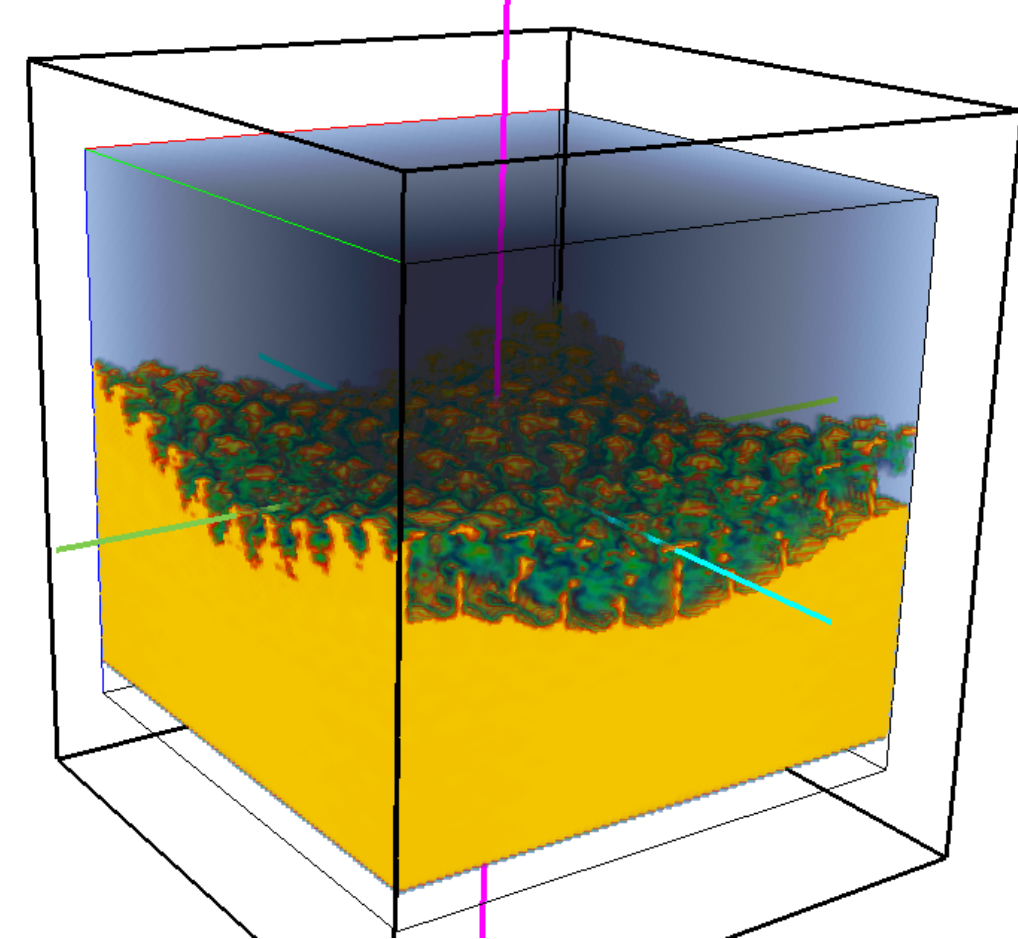
- Interactive visualization on high resolution power-walls.
- Interactive visualization on desktop workstations of large datasets that cannot be stored locally.
- Immediate monitoring of remote simulations from a desktop workstation.



Origins: Kraken NSF machine and the U. of Utah's ember cluster.

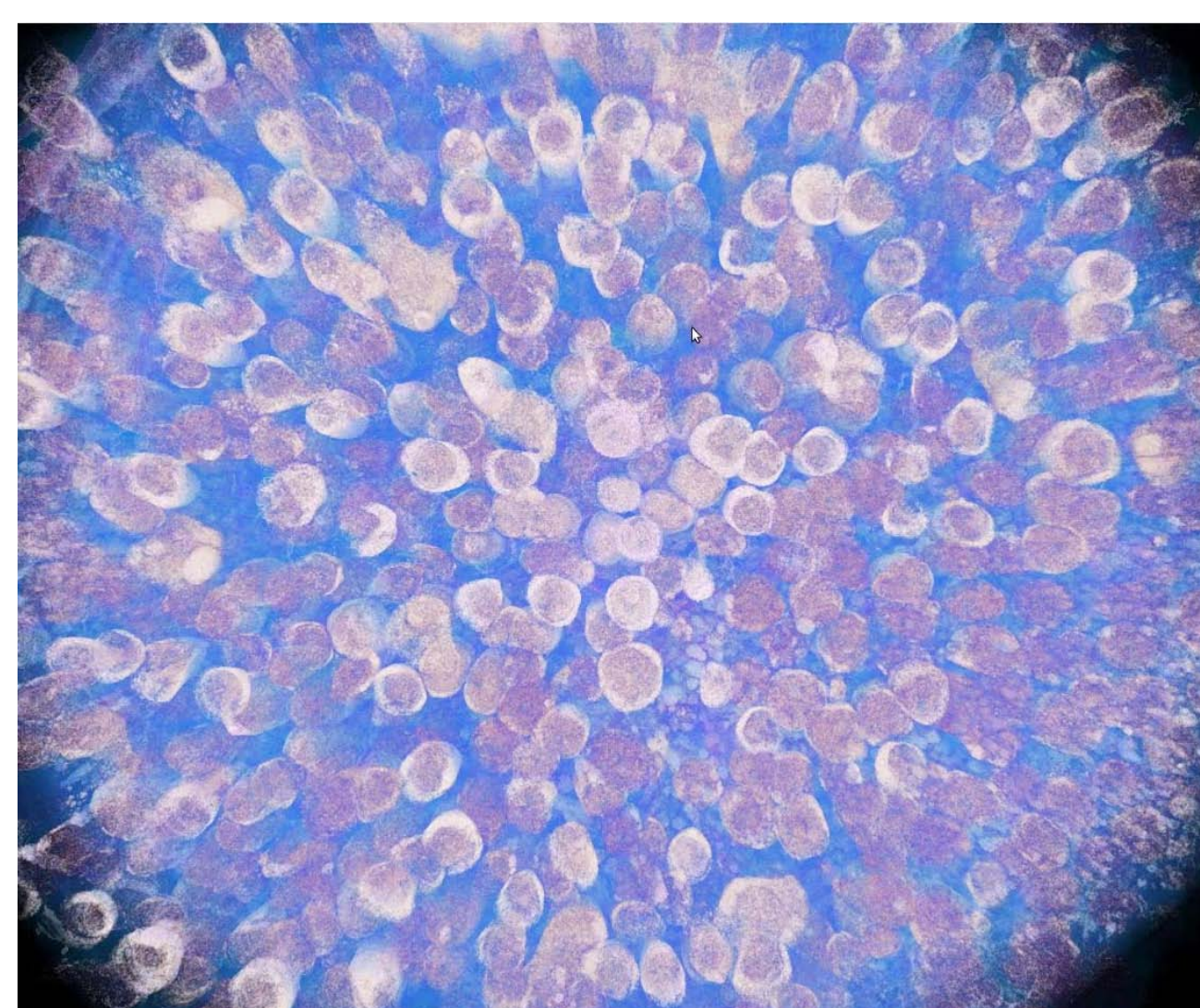
Scientific Importance: The scenario simulated here is one of a clean-coal technology, wherein the coal is fired with a mixture of CO₂ and O₂ rather than pure air. The combustion products are nearly pure CO₂. This stream is then easily "cleaned", ready for CO₂ capture, eliminating the source of greenhouse gas emissions from coal combustion process.

Rayleigh-Taylor Instability Simulation



Data Dimension : 2048 x 2048 x 2048

Rabbit Retina Dataset



Data Dimension : 131072 x 131072 x 340

Gigapixel, multi-view, multi-resolution rendering of The Digital Michelangelo Project's David



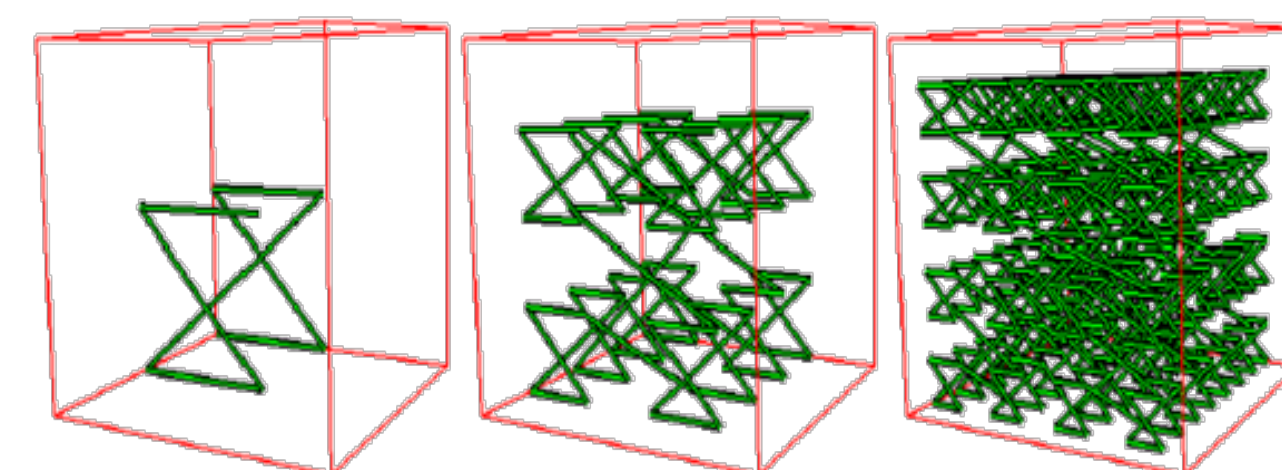
Recent Publication

Sidharth Kumar, Venkatram Vishwanath, Philip Carns, Joshua A. Levine, Robert Latham, Giorgio Scorzelli, Hemanth Kolla, Ray Grout, Robert Ross, Michael E. Papka, Jacqueline Chen, Valerio Pascucci "Efficient Data Restructuring and Aggregation for I/O Acceleration in PIDX," accepted at *Supercomputing 2012*

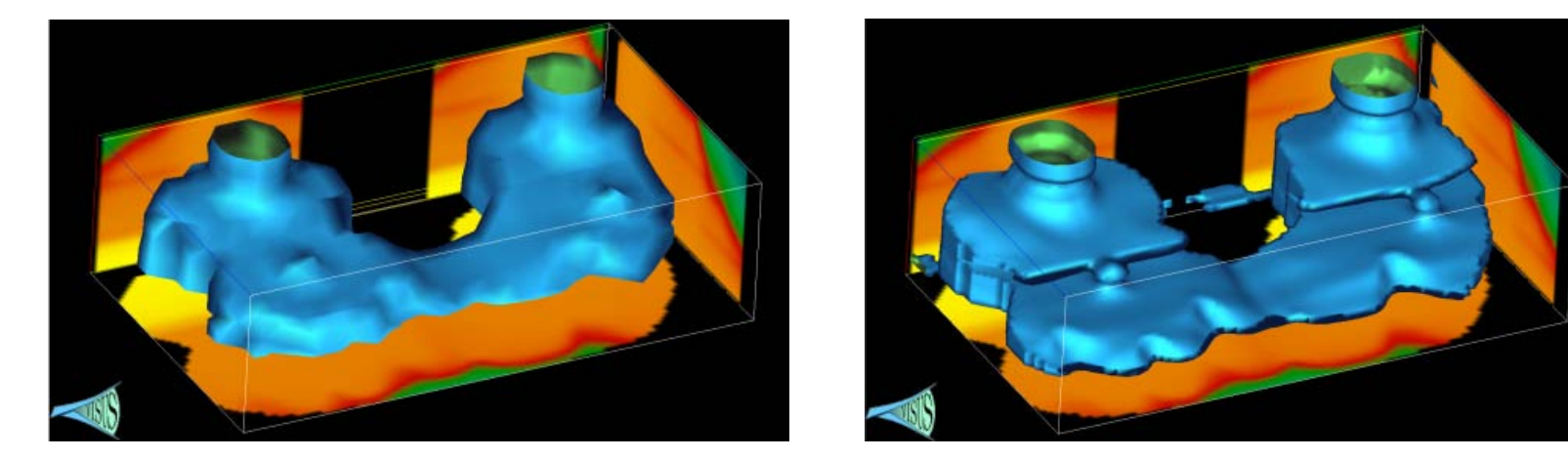
ViSUS - IDX File Format

IDX is a hierarchical multi-resolution data format that provides efficient, cache oblivious, and progressive access to large-scale, multi-dimensional scientific datasets by storing them in a hierarchical Z (HZ) order.

Data stored in IDX format can be visualized in an interactive environment allowing for meaningful explorations with minimal resources.



Cache Friendly Layout

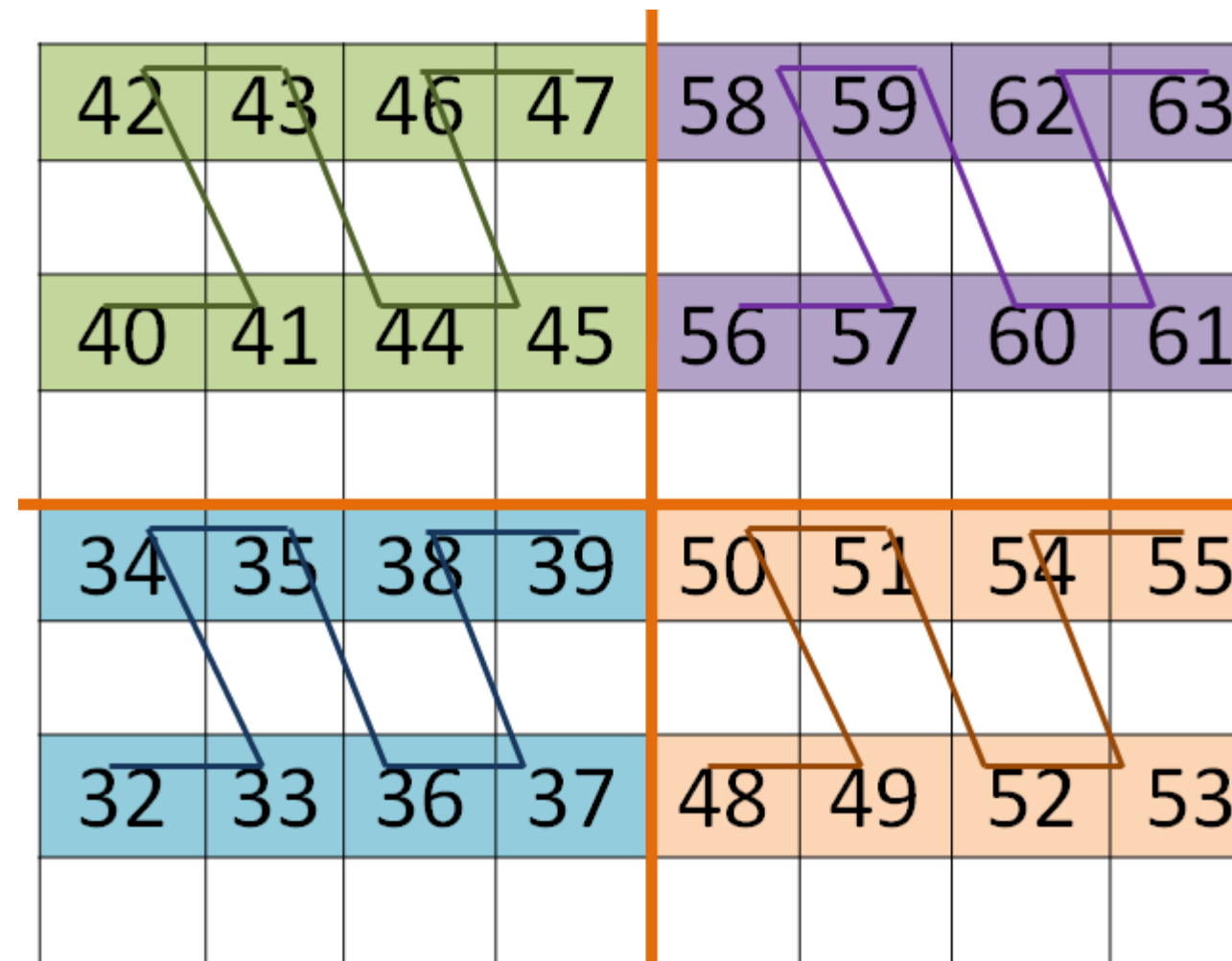


Progressive access of data

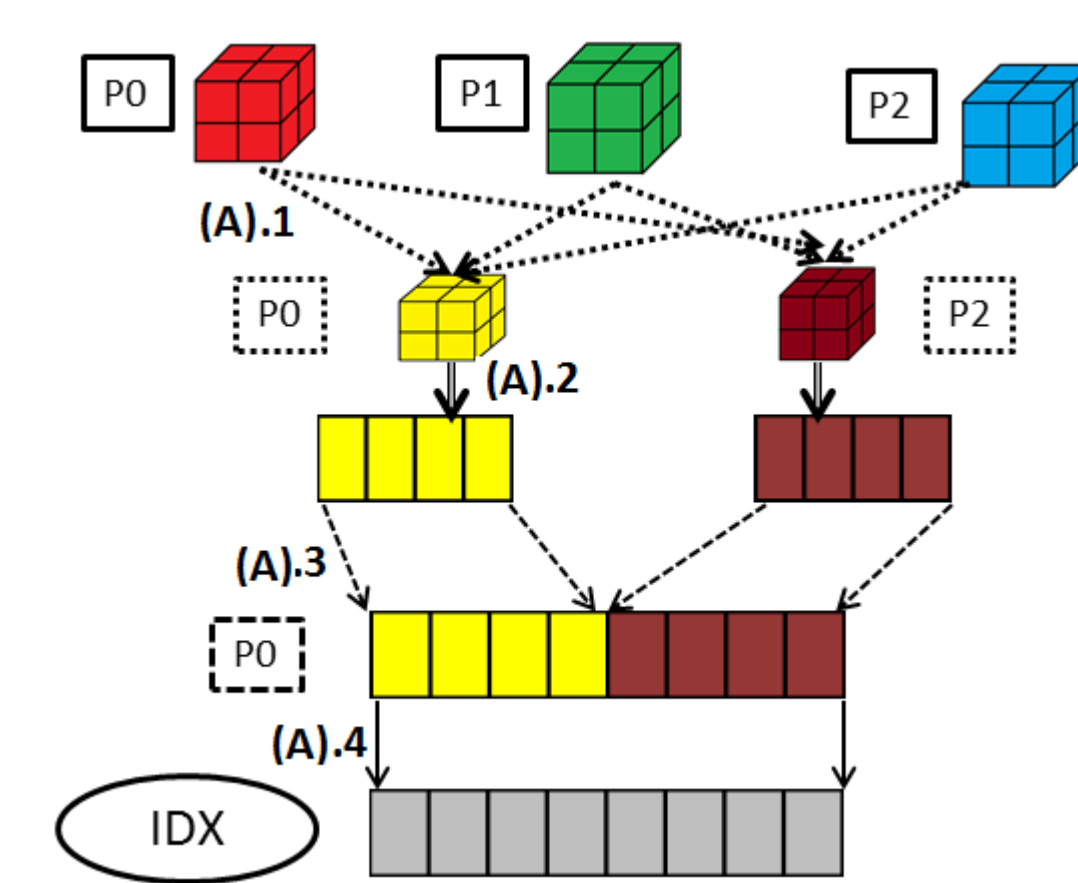
PIDX : Parallel IDX

PIDX is an I/O library that enables HPC applications to write distributed multi-dimensional data directly into a hierarchical multi-resolution data format (IDX Format) with minimal overhead.

Rate real-time remote visualization for large scale simulations.



Simple Parallel Implementation processing space filling curve.



PIDX Three-Phase I/O: (A).1 Data restructuring (A).2 HZ encoding (A).3 Data aggregation (A).4 I/O writes

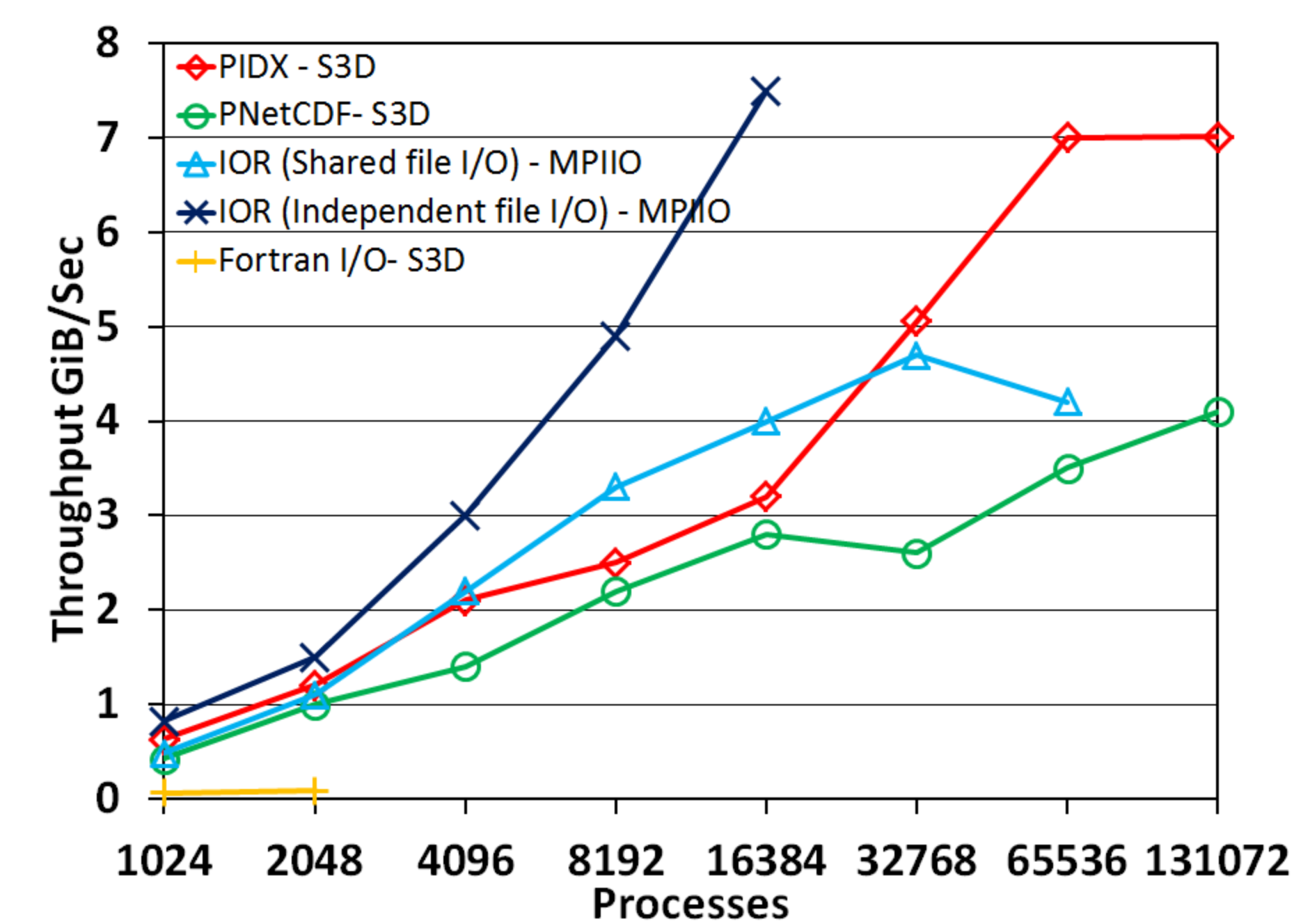


PIDX on Intrepid (BG/P)

PIDX Scaled up to 131072 Cores on Intrepid

Yields higher throughput than other file formats at almost all process counts

IOR (Independent file I/O) does not take into account the directory creation time.



PIDX on HOPPER (Cray)

PIDX Scaled up to 65536 Cores on Hopper

At 65536 core counts yields more than two fold improvement over other parallel file formats.

