

A Combined Eulerian-Lagrangian Data Representation for Large-scale Applications

Franz Sauer, Jinrong Xie, and Kwan-Liu Ma University of California, Davis

VISUALIZATION & INTERFACE DESIGN INNOVATION

Background and Motivation



Scientific Simulation Data

Field-based (Eulerian) Representation



Particle-based (Lagrangian) Representation





Simulations and analyses that utilize both data types...



Particle Data

Field Data

Need a way to better support such tasks in a large-scale setting







- 1. How can we combine the data types into a single representation?
- 2. What efficient data operations does this representation enable?



Analysis

Methods: Data Organization





particles with grid locations

Advantages

- The particle and field data is pre-correlated
- The unit cells can be treated individually
 - Can operate on unit cells in parallel
 - Can operate on groups of unit cells out-of-core
- All particle and field information is locally available









Preprocessing

- Organization scheme involves a preprocessing cost
 - One time cost
 - Enables fast interactive exploration
- Can move preprocessing in situ
 - Data already distributed
 - Little communication costs





Methods: Data Operations





Data Operations

- Conditional Unit Cell Queries
- Transforming Variables Between Reference Frames
- Flow-aware Trajectory Construction
- Multi-resolution Sampling





Conditional Unit Cell Queries





Multi-resolution Sampling



Unbiased Sampling



Biased Sampling



Test Datasets and Performance Results





Test Datasets and Sizes

	Fusion	Combustion	Cosr
Grid Type	Unstructured	Structured	Unstr
Num. of Grid Points	3.5 M	1.3 B	7
Num. of Particles	40 K	40 M	

Compare various operations using the *joint representation*, an octree, and a *k*-d tree...



mology ructured '.9 K 2 M

Fusion (unstructured, 3.5M grid points, 40K particles)



Perform conditional unit cell queries based on plasma density (Eulerian variable)



Fusion (unstructured, 3.5M grid points, 40K particles)

Joint representation performs the best

- Particle and grid points are pre-correlated
- Much less information needs to be loaded from disk









Combustion (structured, 1.3B grid points, 40M particles)

Conditional Unit Cell Query

- *Joint*: 202.2 s
- Octree: 211.1 s
- *k-d*: 220.3 s

No significant speedup in structured grids





Cosmology (unstructured, 7.9K halos, 2M particles)





Cosmology (unstructured, 7.9K halos, 2M particles)

Unbiased Sampling

 All representations perform the same

Biased Sampling

Joint representation performs the best

Cosmology Dataset: Biased Sampling







Discussion and Summary





Discussion

- Specifically designed to support certain data operations
- Cases where this scheme is less effective
 - E.g., cell location problem
 - Octree and k-d tree very effective
- Future:
 - Hybrid organization
 - More tests on moving preprocessing in situ



Summary

- Utilize a unit cell format to combine data representations
- Apply indexing techniques to organize the data
- Describe joint data operations that can efficiently use this format
- Demonstrate its effectiveness using performance tests in multiple application areas



Thank You – Questions?

This work is funded in part by the U.S. Department of Energy through grants DE-SC0007443 and DE-SC0012610



