2023 NETL Workshop on Multiphase Flow Science

End-to-end Interactive Feature Analysis in Large-scale Multiphase Flow Simulations using In Situ Feature Extraction and Post-hoc Visual Analytics

Presented by: Soumya Dutta

Assistant Professor, Indian Institute of Technology Kanpur (IIT Kanpur)

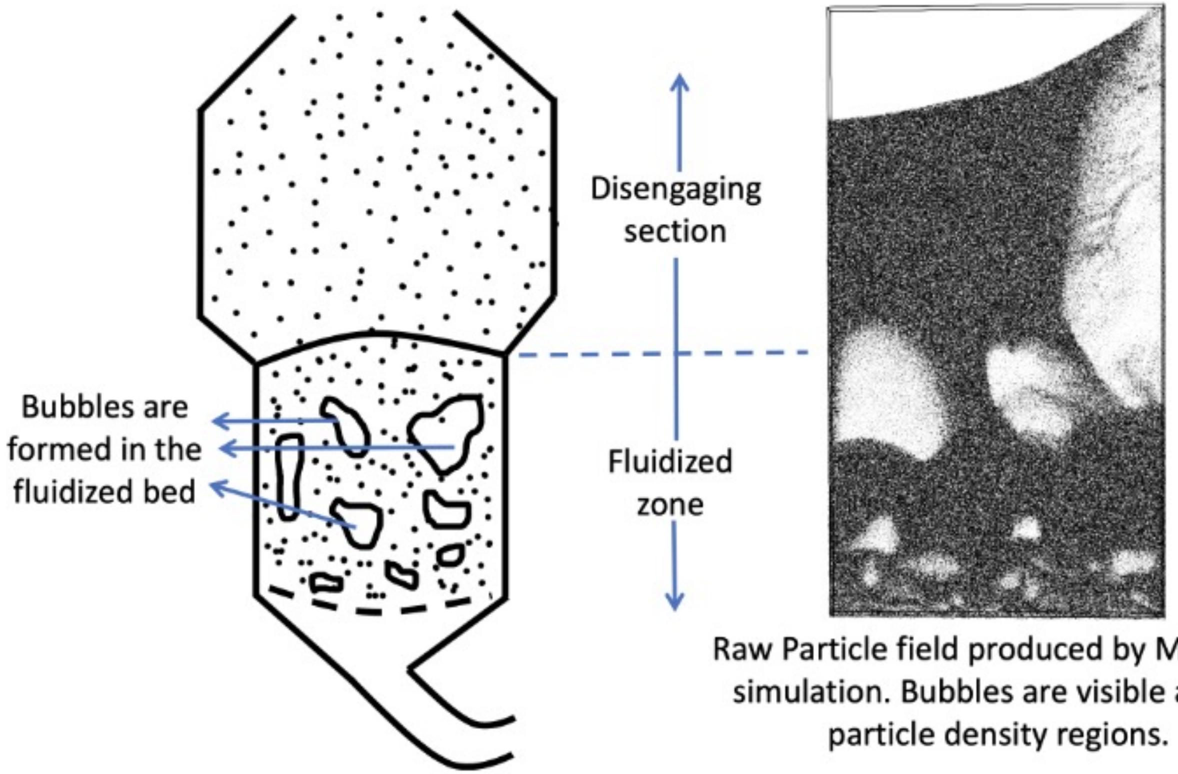
Department of Computer Science and Engineering

Joint work with:

- Terece Turton, David Rogers, and James Ahrens from LANL
- Jordan M. Musser from NETL
- Ann S. Almgren from LBNL

In Situ Feature-driven Analysis at Exascale

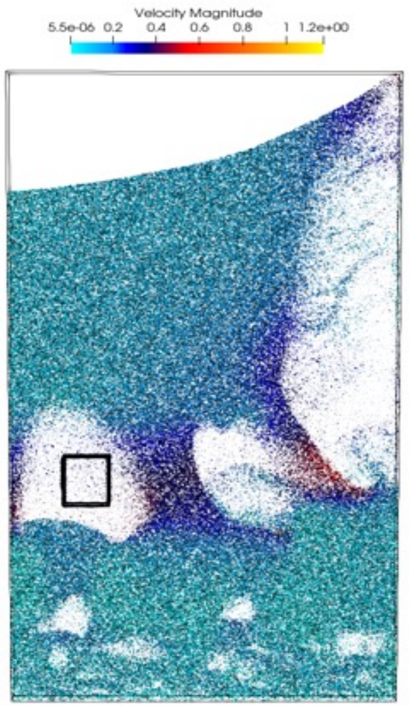
- Exploration of scientific features is essential for understanding the simulated physical phenomenon
  - Bubbles (voids) in multiphase flow modeling
  - Halos in Cosmology simulations
- Accurate extraction of features at Exascale can be challenging
  - Complexity of the modelled physical phenomenon
  - Features may lack precise descriptors
    - e.g. Hard to define a bubble in a particle data set



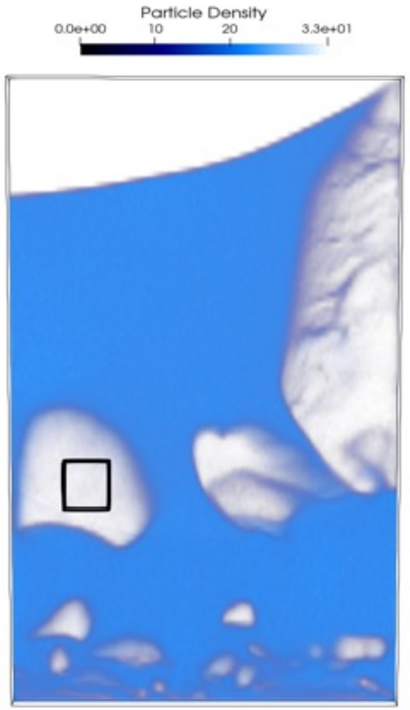
Raw Particle field produced by MFiX-Exa simulation. Bubbles are visible as low particle density regions.

Use Distributions as a Statistical Feature Descriptor

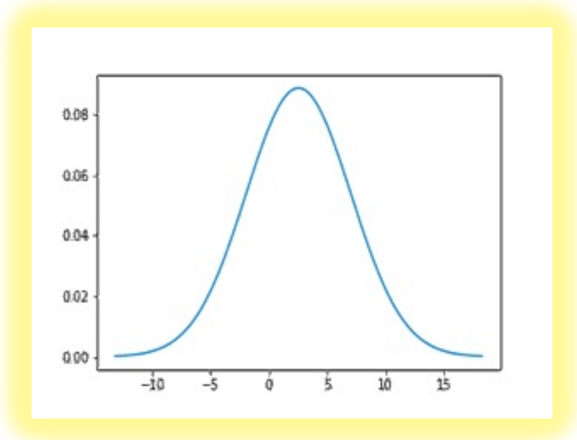
- Probability distributions as feature descriptors
  - Highlight the feature of interest in data directly
  - No precise descriptor/definition is needed



Specify bubble feature in particle field

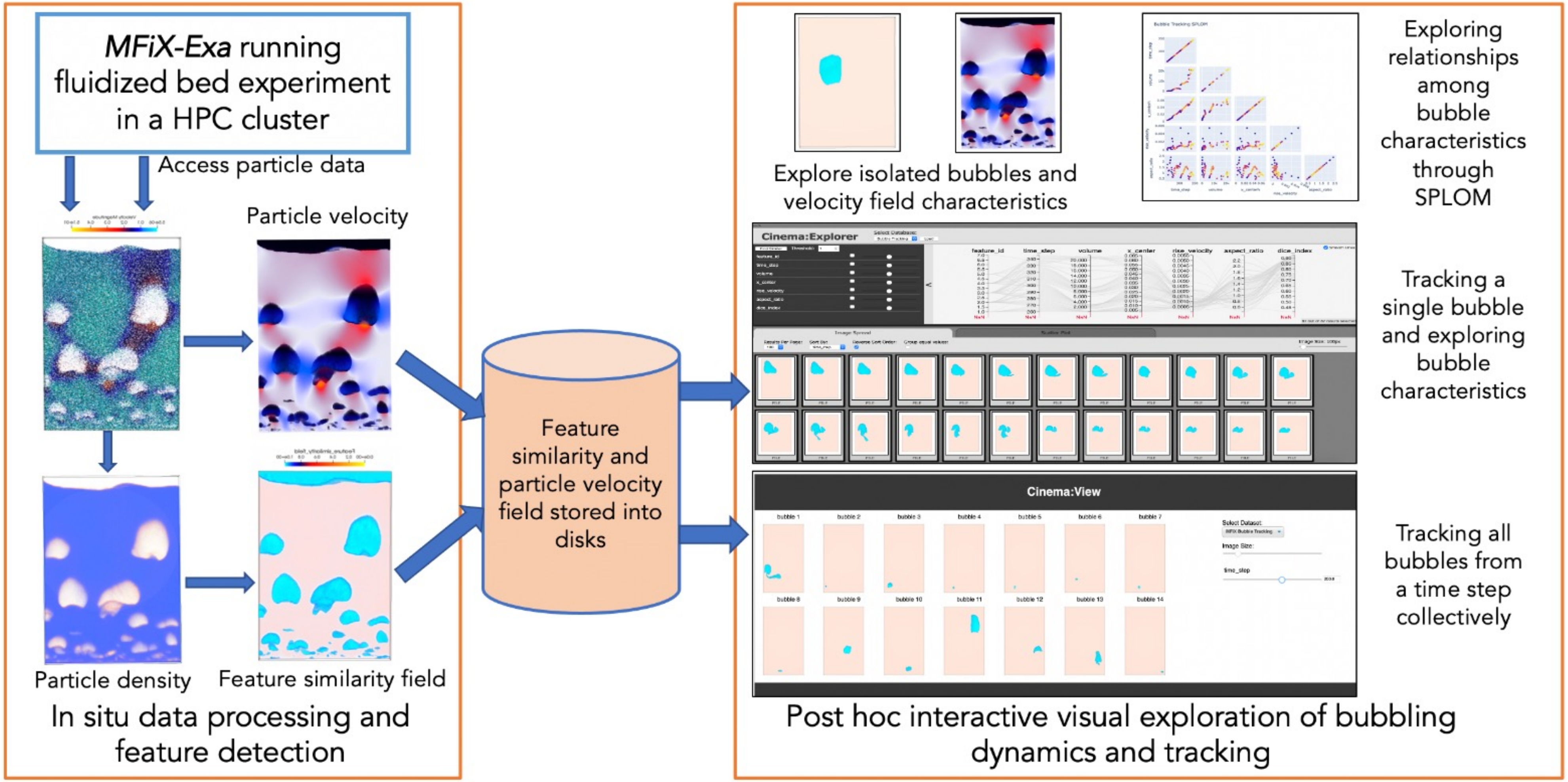


Highlighted bubble feature in intermediate particle density field



Distribution represents the selected region as the target feature

An Overview of the End-to-end Analysis Pipeline



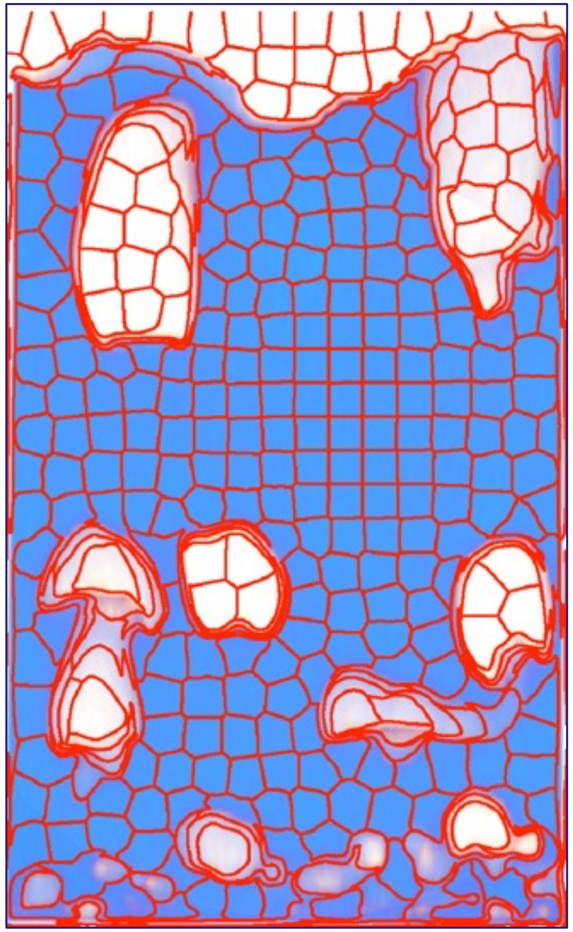
In Situ Statistical Feature Detection Pipeline

- Compute the particle density field from unstructured particle data
  - A 3D histogram of particle locations that captures particles densities in 3D simulation domain
  - Convert the 3D histogram into a regular grid scalar data
    - Bin frequencies represent particle density

In Situ Statistical Feature Detection Pipeline

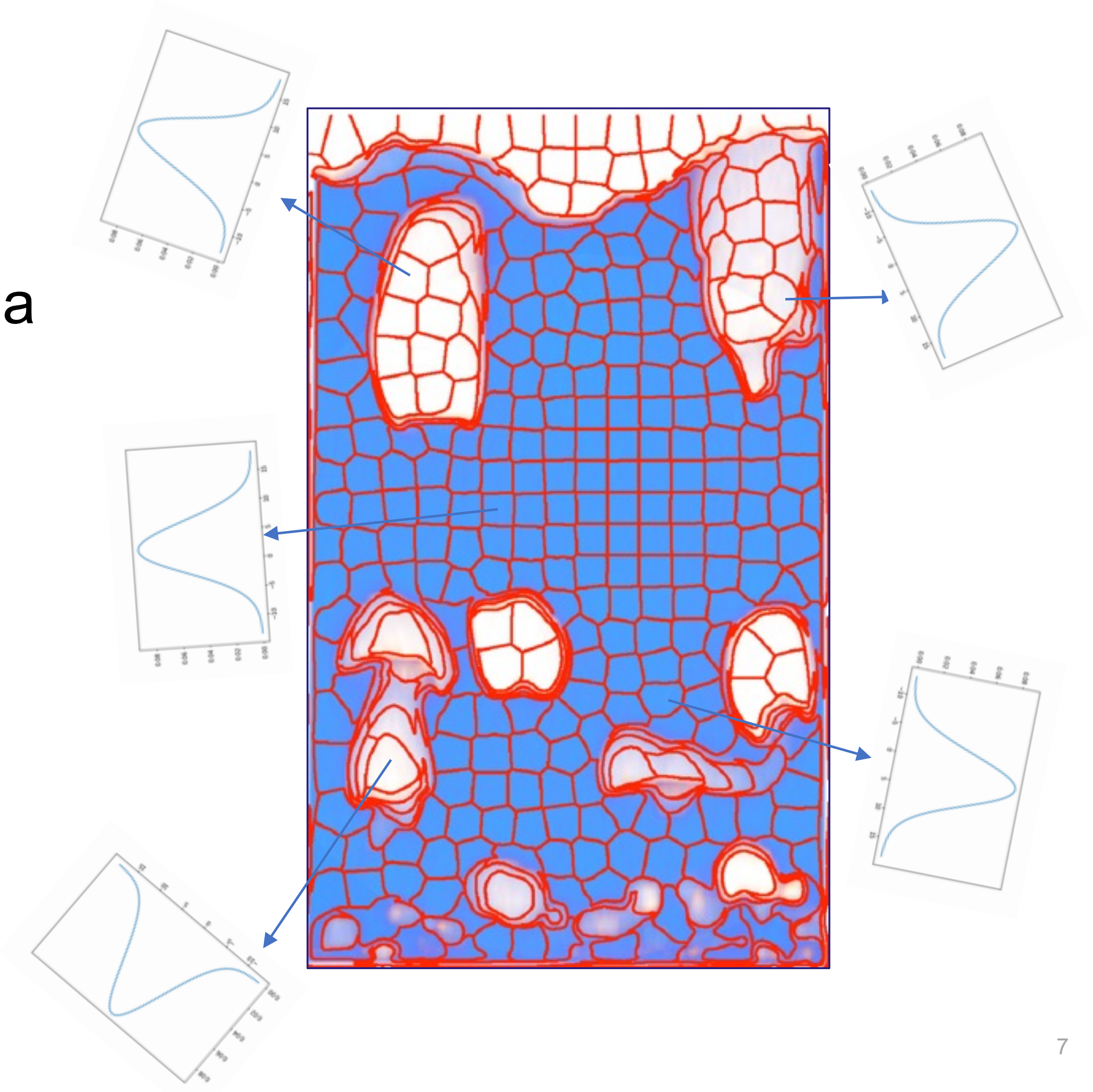
- Compute the particle density field from unstructured particle data
- Partition the density field based on homogeneity
  - Simple Linear Iterative Clustering (SLIC) for generating partitioning
    - A fast local K-means algorithm that generates super-pixels/voxels
    - Value variation inside partitions is minimized
    - The distance function considers both data and spatial coherence

$$D(x, y) = \gamma \cdot ||c_x - p_y||_2 + (1 - \gamma) \cdot |v_x - v_y|$$



In Situ Statistical Feature Detection Pipeline

- Compute the particle density field from unstructured particle data
- Partition the density field based on homogeneity
- Model data values at each partition using a Gaussian distribution

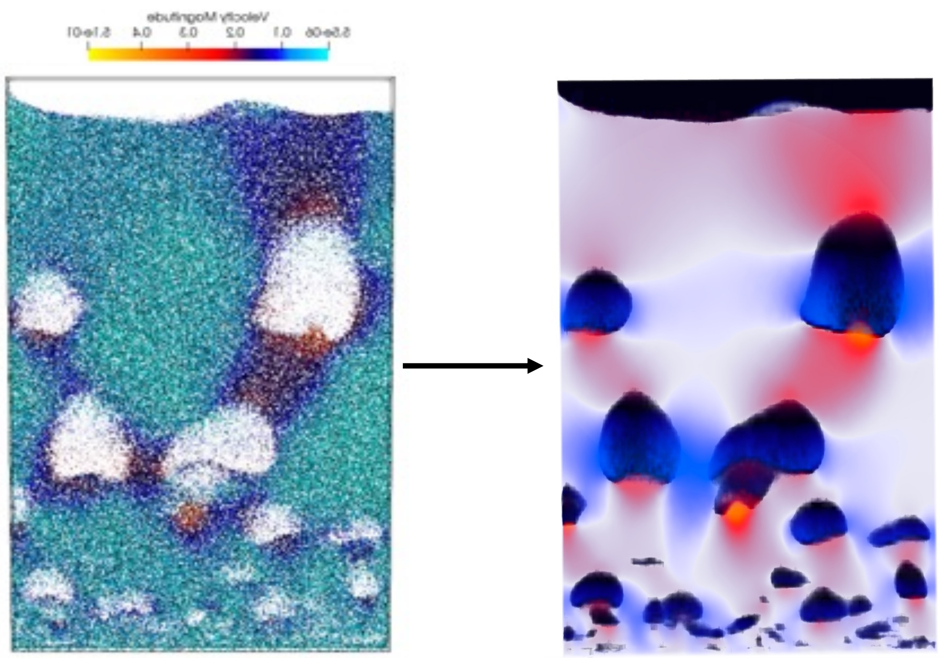


In Situ Statistical Feature Detection Pipeline

- Compute the particle density field from unstructured particle data
- Partition the density field based on homogeneity
- Model data values at each partition using a Gaussian distribution
- Compute statistical similarity of each partition with the user provided distribution of interest
  - Use distribution similarity measures
    - e.g.: Bhattacharya distance, KL-divergence
- Compute a new scalar field where each location indicates how similar the point as to the target distribution

In Situ Statistical Feature Detection Pipeline

- Compute the particle rise velocity field from unstructured particle data
  - A 3D histogram of particle x-velocity component
  - Convert the 3D histogram into a regular grid scalar data

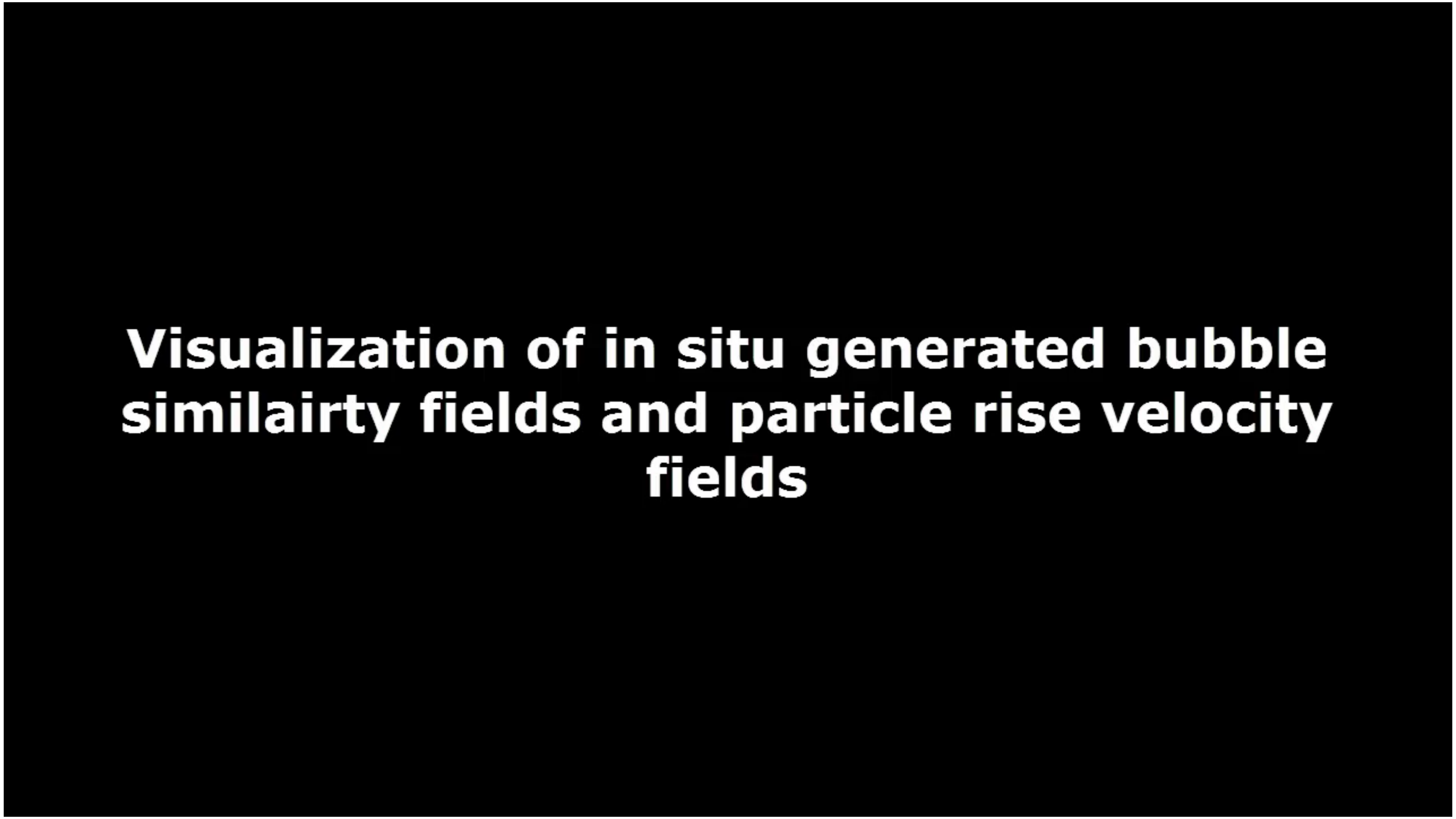


Raw particle data | Particle rise velocity field

- **<u>Storage reduction</u>**: Raw particle data ~310GB vs 224MB storage needed by our method for 100 timesteps
- We have a VTK filter that runs in situ with MFIX-Exa simulation using ParaView Catalysts
- A GPU-accelerated VTK-m filter version of this algorithm has also been developed to deploy it in Exascale machines

LA-UR-21-23925

Post Hoc Interactive Bubble Analysis



More details about this visualization tool: https://github.com/cinemascience/cinema_viewers

In Situ Performance Evaluation

	Density and velocity field	SLIC	Similarity field	Total in situ computation	Total simulation time	In situ I/O	Simulation I/O
2048 MPI processes	2.58	124.37	1.44	128.39	4408.6	14.60	504.85

In situ processing and I/O times (in seconds) taken by the proposed method compared to the simulation time

Density and velocity field	SLIC	Similarity field	Total I/O
39420.87	59.27	144.57	3364.43

Post hoc timings (in seconds) for different steps of our pro-posed algorithm. By processing data in situ, timings shown in this table can be saved

Acknowledgments

- This research was supported by the Exascale Computing Project (17-SC-20-SC)
- This research used resources of the Oak Ridge Leadership Computing Facility at the Oak Ridge National Laboratory which is supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC05-00OR22725

Reference

- For more details about this work, please refer to:
  - **Soumya Dutta**, Terece Turton, David Rogers, Jordan M. Musser, James Ahrens, Ann S. Almgren, *In situ feature analysis for large-scale multiphase flow simulations*, Journal of Computational Science, Volume 63, 2022, 101773, ISSN 1877-7503, https://doi.org/10.1016/j.jocs.2022.101773.
  - **Soumya Dutta**, Dan Lipsa, Terece L. Turton, Berk Geveci, and James Ahrens, In Situ Analysis and Visualization of Extreme-Scale Particle Simulations, *WOIV: 6th International Workshop on In Situ Visualization*, 2022 (co-located with ISC High Performance Conference), https://link.springer.com/chapter/10.1007/978-3-031-23220-6_19