Scalable Data Management, Analysis and Visualization of Particle Accelerator Simulation Data

Oliver Rübel*, Cameron G. R. Geddes†, Min Chen†, Estelle Cormier-Michel‡, Ji Qiang†, Robert D. Ryne†, Jean-Luc Vay†, David Grote†§, Homa Karimabadi$, Vadim Roytershteyn$, William S. Daughton++, Jerry Chou*, Kesheng Wu*, Suren Byna*, Mark Howison*, Prabhat*, Brian Austin*, Arie Shoshani*, and E Wes Bethel*

*Computational Research Division, LBNL | †Accelerator & Fusion Research Division, LBNL | ‡Tech-X Corp. | §Fusion Energy Sciences Program, LLNL | $University of California – San Diego | ++Los Alamos National Laboratory

High-performance Query-based Analysis and Visualization

Science Problem:

- Simulations of particle acceleration processes require the use of large numbers of particles, making analysis and visualization challenging task:
  - Simulations of electron LINACs for a proposed next-generation x-ray free electron laser (FEL) at LBNL require $>10^8$ macroparticle to control macroparticle shot noise and avoid overestimation of the microbunching instability.
  - Magnetic reconnection processes are initiated in the small scale around individual electrons but eventually lead to large-scale reconfigurations of the magnetic field, e.g., in the Earth's magnetosphere. Detailed modeling of electron motions in large-scale fields requires the use of $>10^{12}$ macroparticles.

Simulation Codes:

- Impact-T, Impact-Z, VPIC

Solution:

- **I/O:** Integrated H5Part with the Impact and VPIC simulation suite, enabling high-performance parallel data write and storage of large particle data sets.
- **Data Management:** Developed FastQuery, making distributed, hybrid parallel (MPI+pthreads) index/query technology based on FastBit accessible to scientific file formats enabling:
  - Fast generation of indexes (256s for 1billion particles using 10,000 cores) and evaluation of data queries (Fig. 1),
  - Repeated, complex, large-scale query-based analysis of massive datasets, which would otherwise be impractical with respect to both time as well as computational cost (Fig 2.).
- **Visualization:** Integrated FastQuery with the parallel visualization system VisIt for fast, query-based data exploration.

Application Example I: Query-based Beam Diagnostics for LINACs

- Use advanced queries to identify and quantify the extend of the halo of particles beam in a LINAC simulation (Fig. 2)

Application Example II: Magnetic Reconnection

- Combine query-based data analyses and cross-mesh field evaluation to study the relationship between particle subsets of interest (scattered data) and the magnetic/electric fields (structured data)
  - Study the global and local distributions of accelerated particles to investigate, e.g., the spatial localization patterns of particles (Fig.3)
  - Compare accelerated particles with different principle motion with respect to the magnetic field to study the alignment of accelerated particles with the magnetic field (Fig. 4a,b)

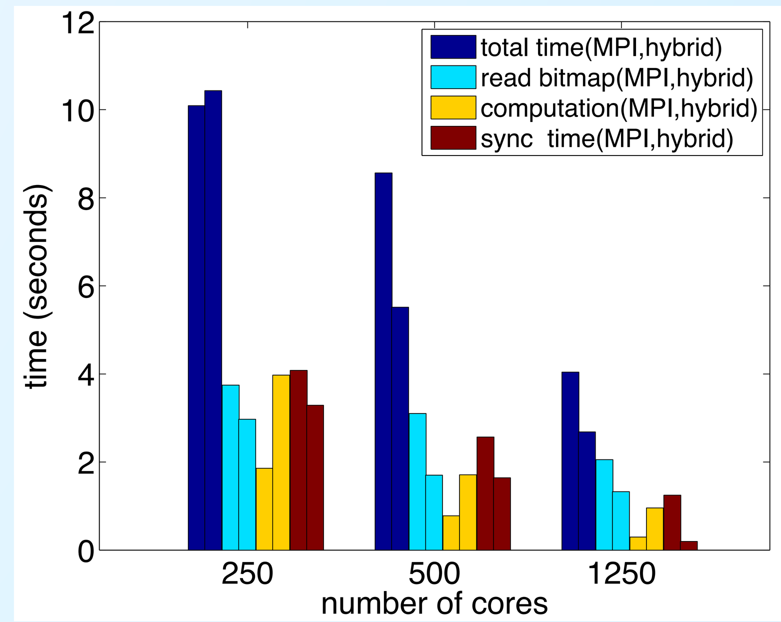

Fig. 1: Time for querying 1 trillion particle dataset with different number of cores.

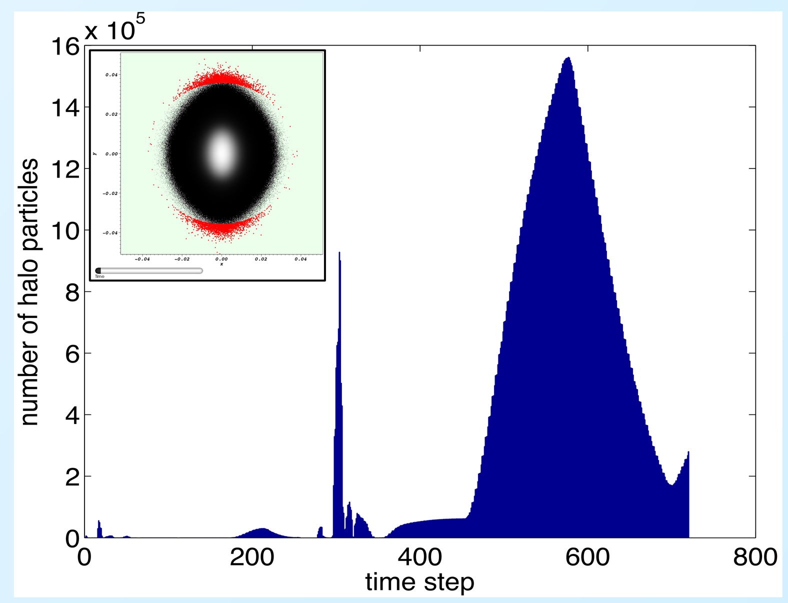

Fig. 2: Number of halo particles per time step and halo query at timestep 20 (inset) of an IMPACT LINAC simulation.

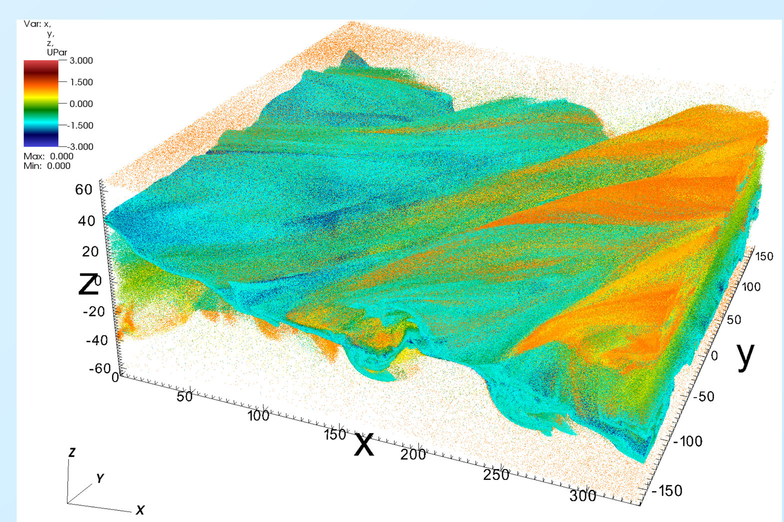

Fig. 3. Plot showing all energetic particles in a trillion electron VPIC simulation colored by UPar.

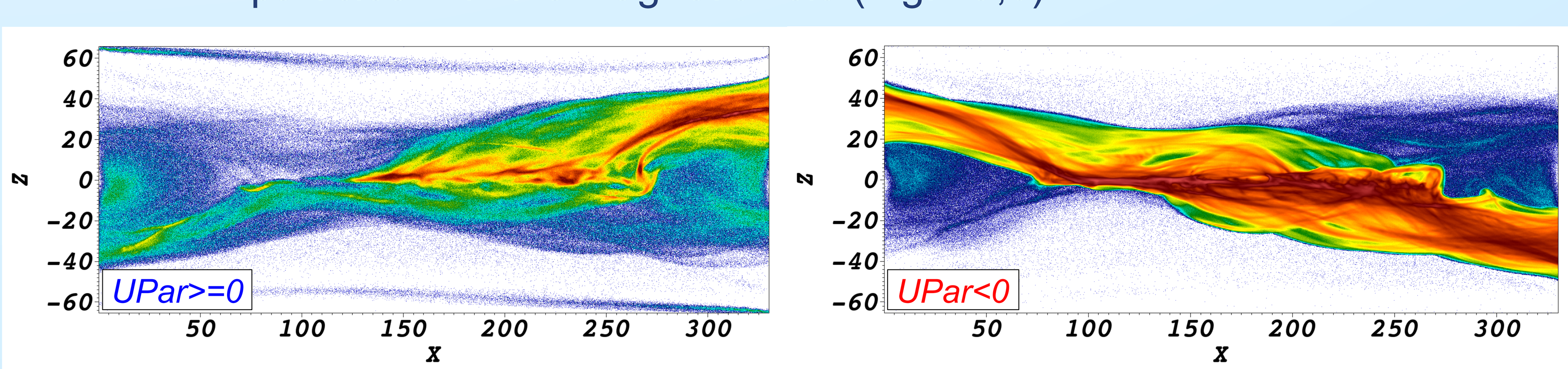

Fig. 4a: Comparison of the spatial distribution of particles with positive (left) and negative (right) Upar values in a trillion electron VPIC simulation.

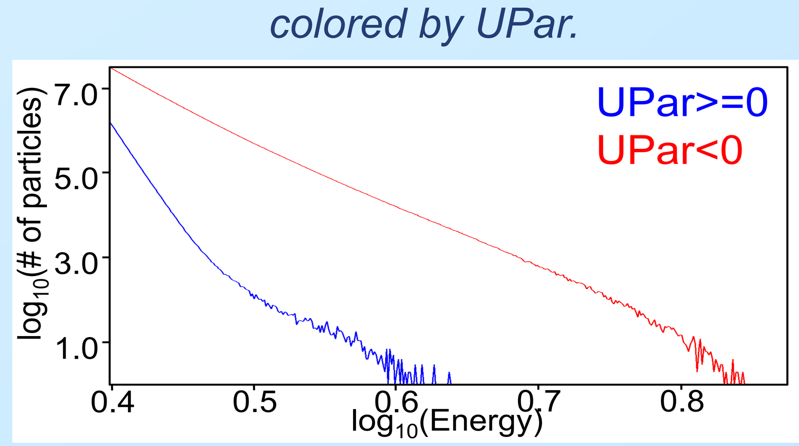

Fig.4b: Energy distributions for the two particle subsets shown in Fig.4a.

Automatic Beam Detection for Laser Plasma Accelerator Simulations

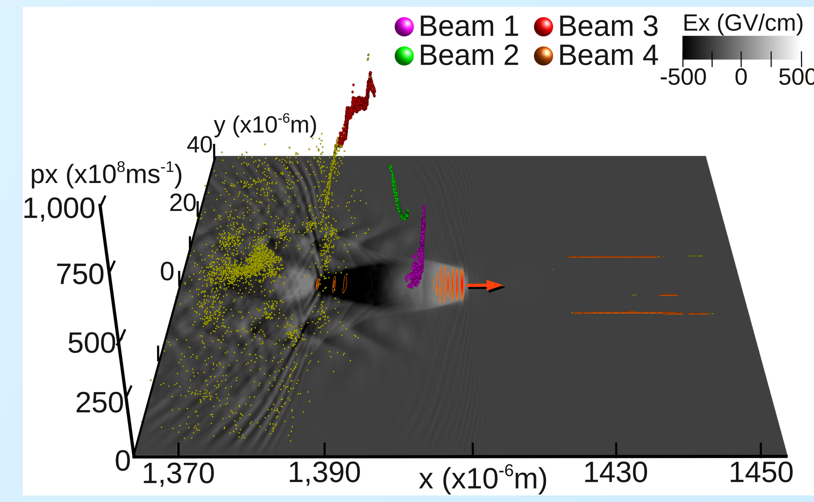

Fig 5a: Single laser pulse

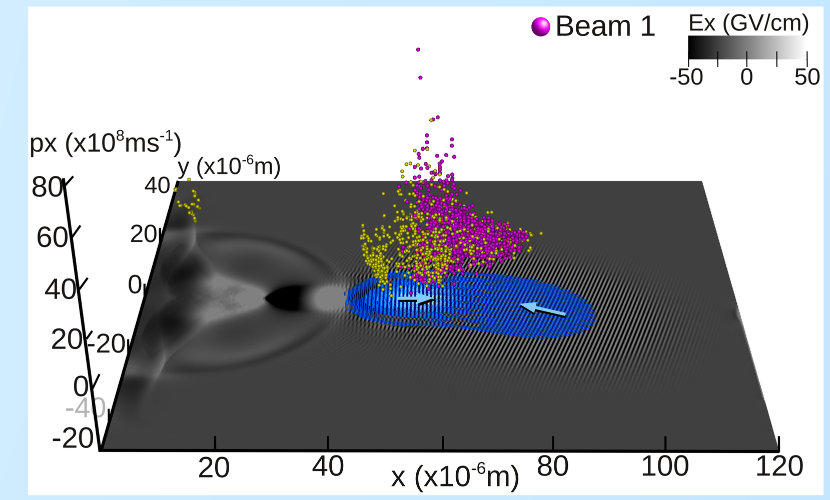

Fig 5b: Dual colliding laser pulse

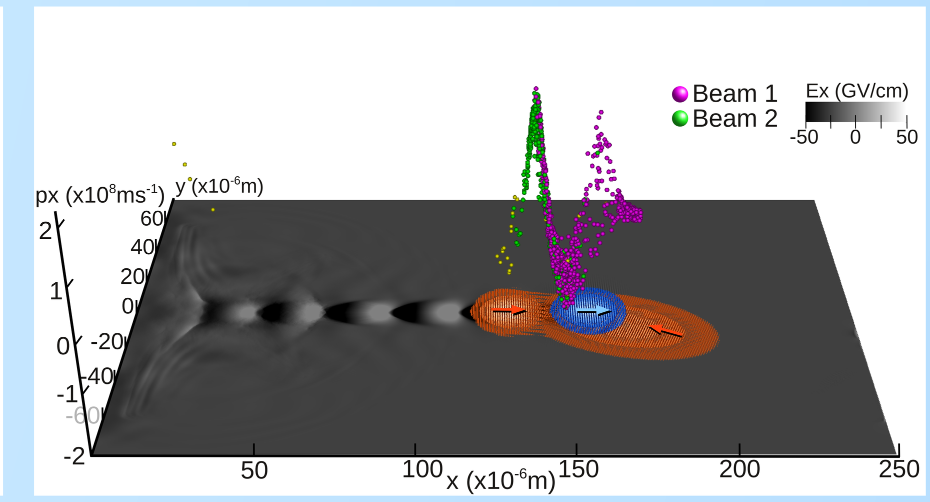

Fig 5c: Triple colliding laser pulse

Science Problem:

- Develop new compact, high-quality accelerator technology based on laser-plasma acceleration.
- Datasets are extremely large, heterogeneous and of varying spatial and temporal resolution.
- Particle beams are small (<1% of the data) and beam formation and acceleration are highly complex processes.
- Diverse physical settings, including single laser pulse and multiple colliding laser pulse accelerator designs (Fig. 5).

Simulation Code:

- VORPAL

Solution:

- Integrate high-performance data query with the visualization and analysis to accelerate beam detection and selection enabling:
  - Fast tracing of particles over time within seconds rather than hours,
  - Analysis of even large 3D laser plasma particle accelerator simulations.
- Develop advanced algorithms for automatic detection of particle beams and their sub-structures based on the full temporal particle traces enabling:
  - Automatic detection of all particle beams and subtle temporal beam substructures not easily detected otherwise,
  - Efficient processing of large numbers of particle datasets.
- Link automated analysis with the visualization enabling:
  - Effective exploration of the space of particle features,
  - More accurate and efficient analysis of particle beams and beam substructures then previously possible.

In Situ Visualization and Analysis using Warp and VisIt

Science Problem:

- The available I/O bandwidth and data storage capabilities are decreasing relative to computation, making *in situ* visualization and analysis indispensable.
- The current approach for *in situ* visualization and analysis using Warp has several limitations:
  - The legacy system OpenDX is no longer supported,
  - Visualization is performed by a single compute core,
  - The data often needs to be reduced for the visualization.

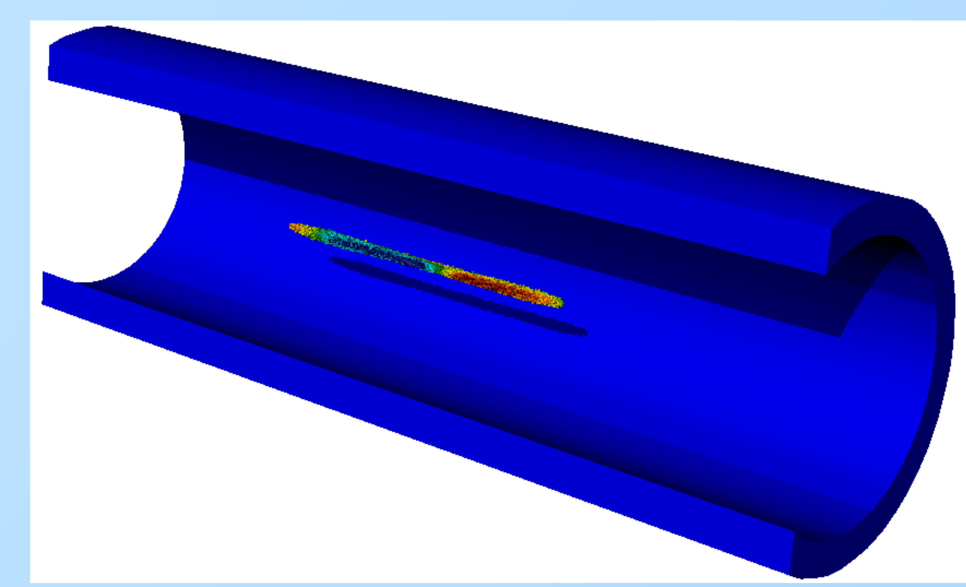

Fig. 6: Example in situ visualization of Warp simulation using VisIt's libsim simulation interface

Solution:

- Couple general purpose, state-of-the-art *in situ* technology using VisIt with Warp
- Using VisIt for *in situ* data analysis and visualization:
  - Enables *in situ* processing of the complete data in parallel,
  - Makes new advanced analysis capabilities accessible to Warp,
  - Avoids reduction and communication of data for the purpose of visualization as the data can be processed in place.

SDAV

Scalable Data Management, Analysis, and Visualization