RAPIDS

Resource and Application Productivity through computation, Information, and Data Science

Application Engagement & Community Outreach
Tiger Teams
Liaisons | Outreach

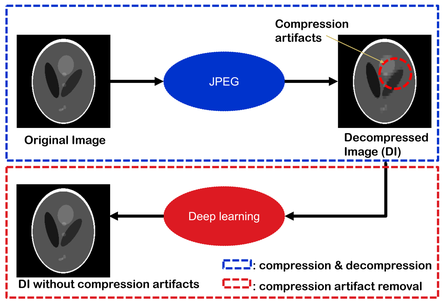
Data Understanding | Platform Readiness | Scientific Data Management

RAPIDS Data Understanding

D. Morozov, P. Balaprakash, J. Ahrens, U. Ayachit, W. Bethel, E. Brugger, A. Buluc, B. Geveci, P. Grosset, H. Guo, C. Harrison, W. Liao, S. Madireddy, J.H. Park, J. Patchett, T. Peterka, S. Philip, D. Pugmire, O. Ruebel, H.-W. Shen, C. Steed, G. Weber, S. Yoo

Machine Learning

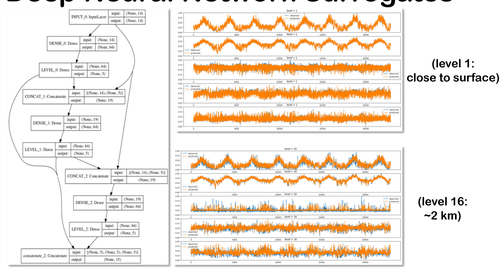
Compression Artifact Removal



- Compression techniques such as JPEG can significantly reduce the size of images, but there are two main challenges for scientific image compression: blocking artifacts and blurring.
- We employed deep learning techniques (EDSR and SRGAN) to remove these artifacts.

Collaborations: FASTMATH (R. Archibald)

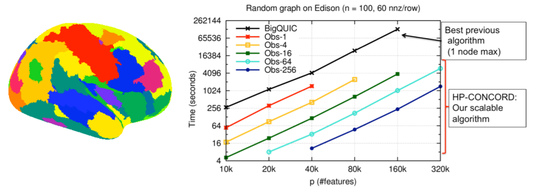
Deep Neural Network Surrogates



- Simulation-enabled domain-informed learning.
- Developed deep neural network model to accelerate the physical processes simulation in weather research forecasting simulation on leadership-class machines.
- A factor of 10 reduction in the model simulation time would increase the ensemble size into hundreds of individual realizations as opposed in the tens that are feasible now.

P. Balaprakash, J. Wang, R. Kotamarthi. Deep Neural Network Emulation of a Planetary Boundary Layer Parameterization in Weather Research Forecasting model.

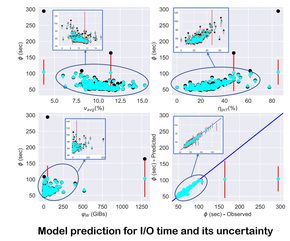
HP-CONCORD:
Graphical Model Structure Learning at Unprecedented Scale



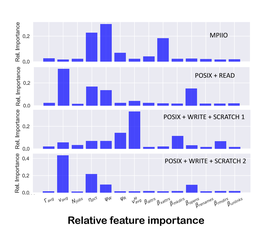
- Developed HP-CONCORD, a statistically grounded and extremely scalable unsupervised learning method, able to sift through trillions of pairwise relationships to find the most prominent ones.
- HP-CONCORD bridges a computational scalability gab between statistically sounds methods and practical usability for some of the largest modern datasets.

Software: bitbucket.org/penpornk/spdm3-hpconcord

Parallel I/O Predictive Modeling: A Case Study on Lustre File Systems



Model prediction for I/O time and its uncertainty



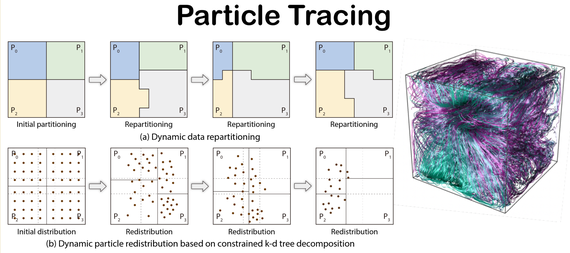
Relative feature importance

- Developed a sensitivity-based robust Gaussian process (GP) regression approach that explicitly treats the variability in the data and groups applications with similar characteristics.
- The robust GP approach provides significant predictive accuracy improvements in the presence of outliers and quantifies the uncertainty in the model prediction.

S. Madireddy, P. Balaprakash, P. Carns, R. Latham, R. Ross, S. Snyder, and S. M. Wild. Analysis and Correlation of Application I/O Performance and System-Wide I/O Activity, The 12th International Conference on Networking, Architecture, and Storage, Shenzhen, 2017, pp. 1-10.
S. Madireddy, P. Balaprakash, P. Carns, R. Latham, R. Ross, S. Snyder, and S. M. Wild, Machine Learning Based Parallel I/O Predictive Modeling: A Case Study on Lustre File Systems, In: Yokota R., Weiland M., Keyes D., Trinitis C. (eds) High Performance Computing. ISC High Performance 2018.

Feature Analysis

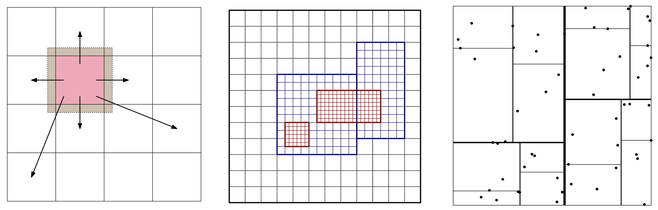
Dynamic Load Balancing for Parallel Particle Tracing



- Particle tracing is a fundamental technique for visualizing and analyzing flow fields. Characterized by greatly imbalanced workload.
- Two solutions:
  - Dynamic load balancing via data repartitioning
  - Dynamic load balancing via k-d trees to periodically evenly redistribute particles
- Evaluated on Vesta, BlueGene/Q at ANL. Significant improvement compared to baseline method.

Jiang Zhang, Hanqi Guo, Xiaoru Yuan, and Tom Peterka. Dynamic Data Repartitioning for Load-Balanced Parallel Particle Tracing. In Proceedings of IEEE Pacific Visualization Symposium (PacificVis '18), pages 86-95, Kobe, Japan, April 10-13, 2018.
Jiang Zhang, Hanqi Guo, Fan Hong, Xiaoru Yuan, and Tom Peterka. Dynamic Load Balancing Based on Constrained K-D Tree Decomposition for Parallel Particle Tracing. IEEE Transactions on Visualization and Computer Graphics (VIS '17), 24(1):954-963, 2018.

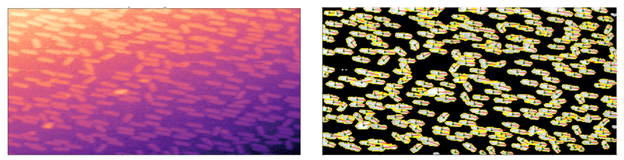
DIY: Software Infrastructure for Data Analytics



- Data-parallel and out-of-core library: supports different communication patterns, domain decompositions, algorithms (k-d tree decomposition, sorting), I/O, etc.
- New communication patterns:
  - rexchange (distributed consensus for remote communication)
  - iexchange (dynamic interleaving of computation and communication)
- Added support for AMR grid data
- Better integration with VTK-m and ParaView (Kitware)

Software: github.com/diatomic/diy

Improved Understanding of Particle Interactions



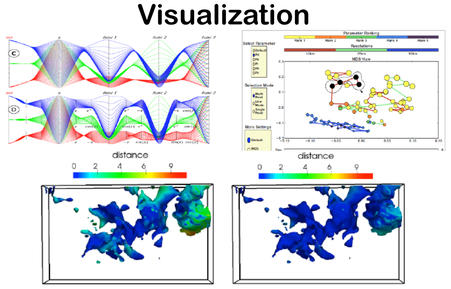
Application: Characterize the roles of interfacial chemistry and structure in particle-particle and particle-solvent interactions in complex chemical environments using high-resolution, time-varying liquid-phase Atomic Force Microscopy (AFM).

Goal: Automate AFM image analyses to characterize microscopic system dynamics, enable analysis of high-speed AFM images consisting of 1000s of timesteps per dataset, and inform experimenters with respect to event and parameter selection.

Preliminary Results: Developed prototype pipelines for automatic analysis of protein nanorods to evaluate the application of various image segmentation, morphology, and filtering methods and to automate estimation of feature characteristics (e.g., size, orientation etc.). Challenges in automation of the analysis include among others stripe noise and varying characteristics (e.g. size) of individual features and feature cluster.

Collaborations: IDREAM EFRC (PI: James De Yoreo)

Uncertain Data Analysis and Visualization

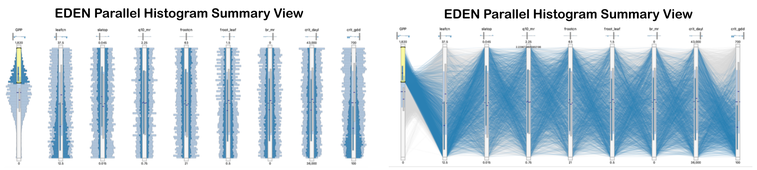


- Representation of large scale uncertain data
  - Compact data representations
  - Surface density estimation
- Analysis of ensemble and uncertain features
  - Uncertain isocontours
  - Uncertain flow features
  - Domain specific uncertain features
- Exploration of parameter space for ensemble simulations
  - Visual Exploration Interface
  - Sensitivity studies

Software: sites.google.com/site/gravityvisdb/edda

Visualization

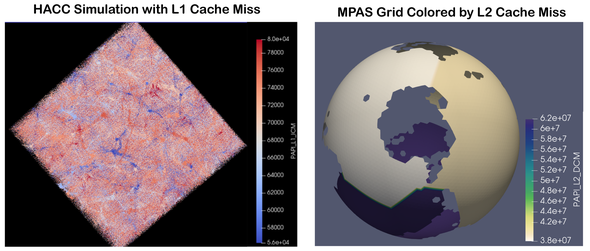
Multivariate, Temporal Visual Analytics for Climate Analysis



- EDEN enables exploratory data analysis for new DOE E3SM climate simulation and observational data using techniques that combine interactive data visualization and statistical analytics.
- EDEN gives climate scientists the ability to consider more variables from large scale, land model parameter sensitivity analyses and ultimately improve DOE model accuracy.

Collaborations: An Integrated System for Optimization of Sensor Networks to Improve Climate Model Predictions (PI: Ricciuto); Software: github.com/ornl/edenfx

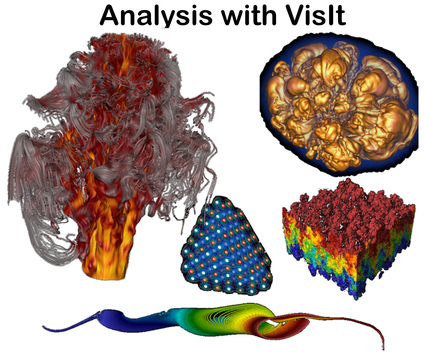
Performance Visualization



- Allow scientists to overlay performance data on top of simulation
- Allow scientists to see how "performance hot spots" in simulation are linked to simulation features
  - No need to use another instrumentation tool
  - Performance tools only show "hot spots" related to function calls

Collaborations: HACC (PI: Habib), MPAS Ocean (PI: Jones)

Scalable HPC Visualization and Data Analysis with VisIt

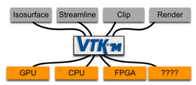


Recent accomplishments: Improved scalability of ghost zone generation for AMR meshes. Memory usage for a large mesh was reduced by 3 Gbytes/core allowing visualization to be performed much more frequently due to the reduced core counts needed to visualize the data

Software: visit.llnl.gov

Visualization Services using VTK-m

Portable Performance across Heterogeneous Architectures using VTK-m

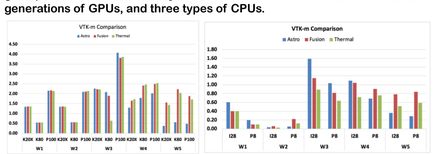


Decoupling of algorithm from hardware allows flexibility for in situ visualization workflows

Portable performance for particle advection

Particle advection is a foundational algorithm for the analysis and visualization of flow. Our performance-portable implementation using VTK-m allows flexibility in performing flow analysis in an in situ workflow.

Comparisons between VTK-m and hardware-specific implementations shows good performance on a variety of vector fields and workloads across three generations of GPUs, and three types of CPUs.



D. Pugmire, A. Yenpure, M. Kim, J. Kress, R. Maynard, H. Childs, B. Hentschel, "Portable-Performance Particle Advection with VTK-m" Eurographics Syposium on Parallel Graphics and Visualization (EGPGV), June 2018, Brno, Czech Republic. (Finalist for Best Paper)

Argonne NATIONAL LABORATORY | BROOKHAVEN NATIONAL LABORATORY | BERKELEY LAB Lawrence Berkeley National Laboratory | The HDF Group | University of Delaware | Kitware | Los Alamos NATIONAL LABORATORY EST.1943 | Lawrence Livermore National Laboratory | Northwestern University | OAK RIDGE National Laboratory | The Ohio State University | University of Oregon | Rutgers | The University of Utah | U.S. Department of Energy | Office of Science