The DOE SciDAC Institute for Scalable Data Management, Analysis, and Visualization (SDAV)

Rob Ross
Deputy Director
SciDAC Institute for Scalable Data Management, Analysis, and Visualization

Computer Scientist
Mathematics and Computer Science Division
Argonne National Laboratory
rross@mcs.anl.gov

SciDAC
Scientific Discovery through Advanced Computing

- DOE Office of Science (SC) activity, spanning multiple Office of Science programs
- SciDAC Institutes – Provide expertise and software tools in applied mathematics and computer science to advance scientific discovery through modeling and simulation
  - FASTMath – Frameworks, Algorithms, and Scalable Technologies for Mathematics
  - QUEST – Quantification of Uncertainty in Extreme Scale Computations
  - SUPER – Institute for Sustained Performance, Energy and Resilience
  - **SDAV – Scalable Data Management, Analysis and Visualization**
- SciDAC Partnerships – Partner with SC programs to combine CS and applied math with domain science expertise to target areas of strategic importance
  - Fusion Plasma Science (2 projects)
  - High Energy Physics (3 projects)
  - Nuclear Physics (3 projects)
  - Earth Systems (3 projects)
  - Chemistry and Materials (6 projects)

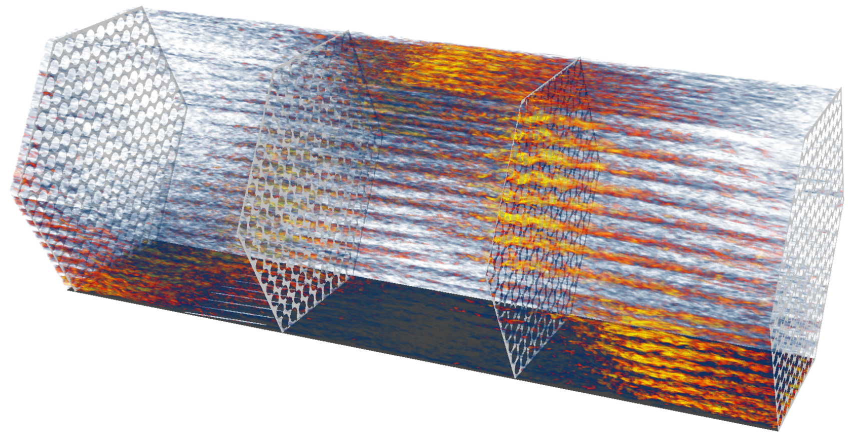
http://www.scidac.gov/

R. Laviolette and C. Susut. SciDAC Scientific Computation Application Partnerships Update. ASCAC. August 14, 2012.

Data Challenges in Computational Science

"Very few large scale applications of practical importance are NOT data intensive." – Alok Choudhary, IESP, Kobe, Japan, April 2012

- Data management and analysis plays a central role in DOE science mission
- Research challenges arise from "the 3 V's":
  - **Volume** – The application produces/ consumes terabytes or more data.
  - **Velocity** – An application has much data, moving very fast.
  - **Variety** – The application integrates data from a large variety of data sources.
- Research challenges also arise from complex system architecture demands, including heterogeneity, hierarchy, and concurrency



Visualization of coolant flows in a 217-pin nuclear reactor assembly. Visualization depicts how certain regions along the exterior (shown in yellow and red) are not as well cooled as other regions.

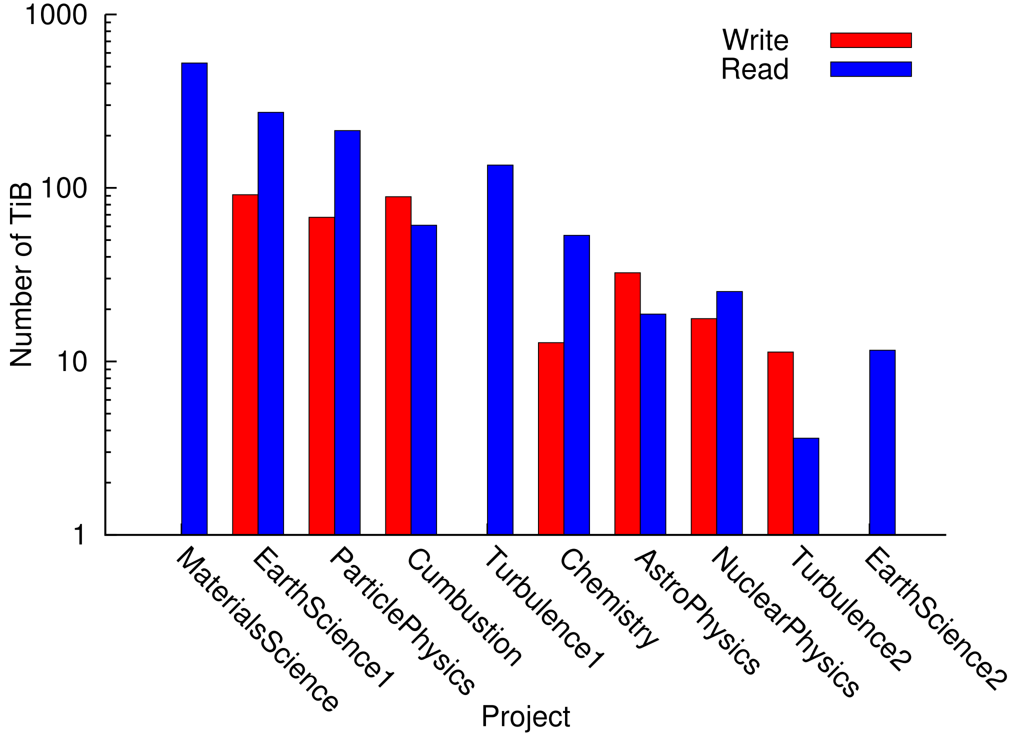
Simulation by Paul Fischer and Aleks Obabko (ANL) using 128K cores on Argonne IBM BG/P system, Nek5000 code, unstructured mesh of over 1B cells.

Visualization by H. Childs (SDAV, LBNL) using VisIt.

Data Volumes in Computational Science

Data requirements for select 2012 INCITE applications at ALCF (BG/P)

PI	Project	On-line Data (TBytes)	Off-line Data (TBytes)
Lamb	Supernovae Astrophysics	100	400
Khokhlov	Combustion in Reactive Gases	1	17
Lester	CO2 Absorption	5	15
Jordan	Seismic Hazard Analysis	600	100
Washington	Climate Science	200	750
Voth	Energy Storage Materials	10	10
Vashista	Stress Corrosion Cracking	12	72
Vary	Nuclear Structure and Reactions	6	30
Fischer	Reactor Thermal Hydraulic Modeling	100	100
Hinkel	Laser-Plasma Interactions	60	60
Elghobashi	Vaporizing Droplets in a Turbulent Flow	2	4

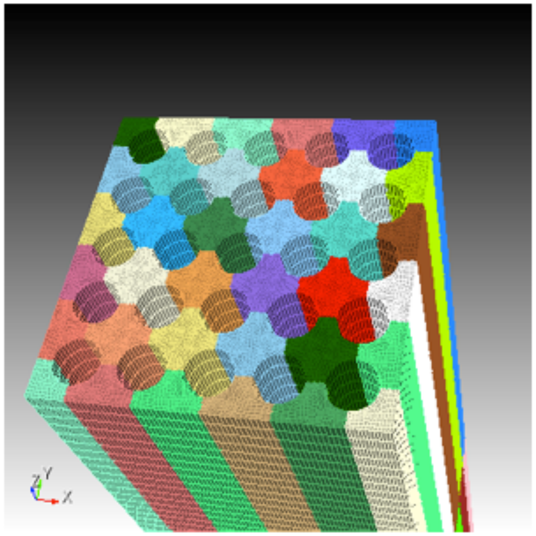


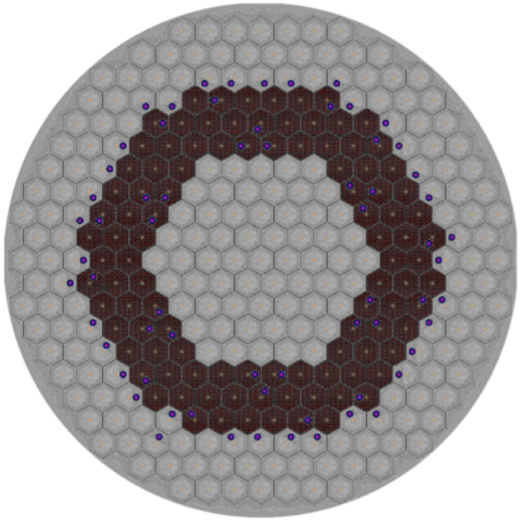
Top 10 data producer/consumers instrumented with Darshan over the month of July, 2011. Surprisingly, three of the top producer/consumers almost exclusively read existing data.

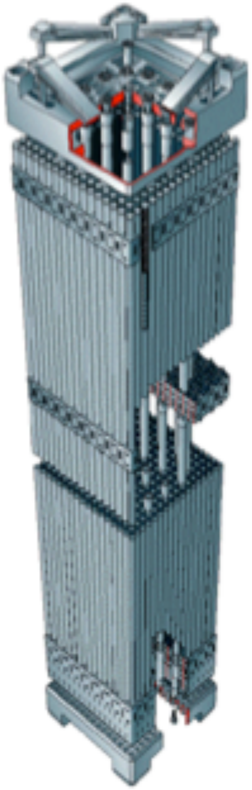
Dataset Complexity in Computational Science

Complexity is an artifact of science problems and codes:

- Coupled multi-scale simulations generate multi-component datasets consisting of materials, fluid flows, and particle distributions.
- Example: thermal hydraulics coupled with neutron transport in nuclear reactor design
- Coupled datasets involve mathematical challenges in coupling of physics over different meshes and computer science challenges in minimizing data movement.







Model complexity: Spectral element mesh (top) for thermal hydraulics computation coupled with finite element mesh (bottom) for neutronics calculation.

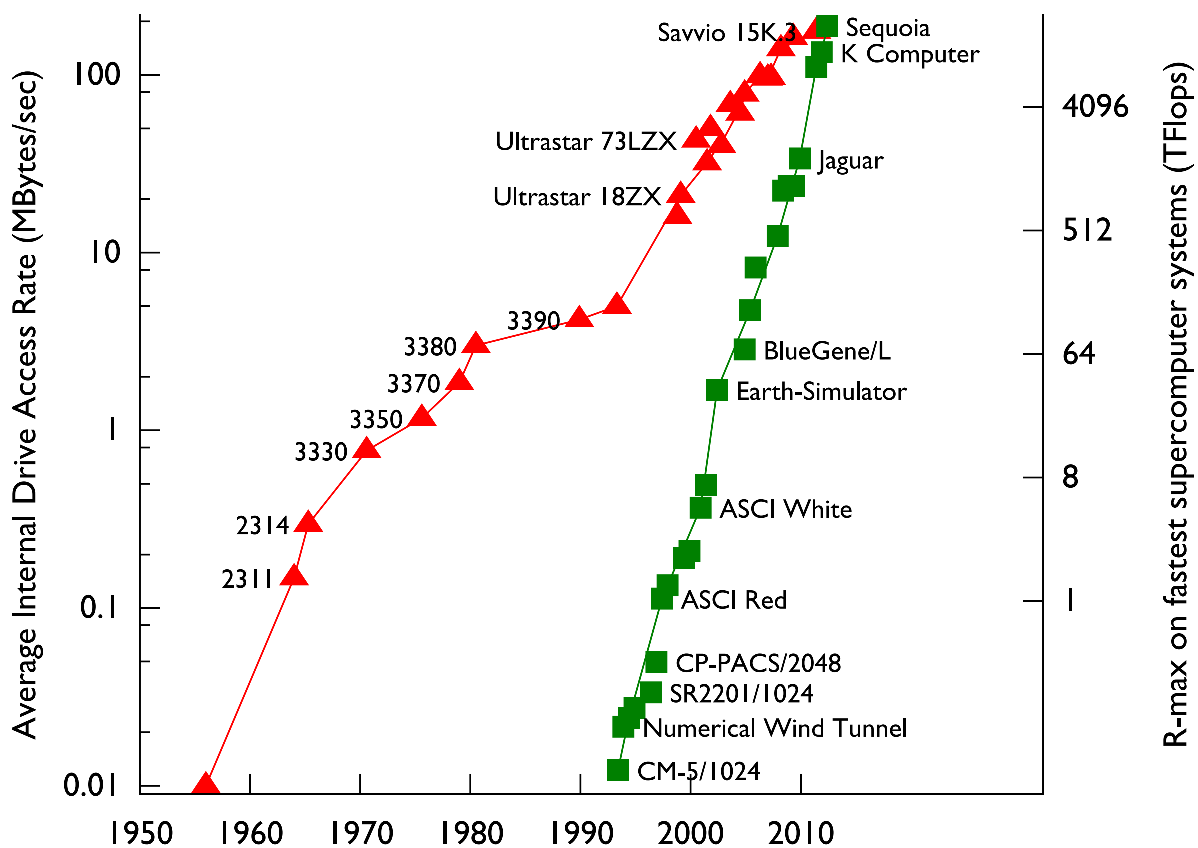
Scale complexity: Spatial range from the reactor core in meters to fuel pellets in millimeters.

Images from T. Tautges (ANL) (upper left), M. Smith (ANL) (lower left), and K. Smith (MIT) (right).

Data, Velocity, and System Architectures

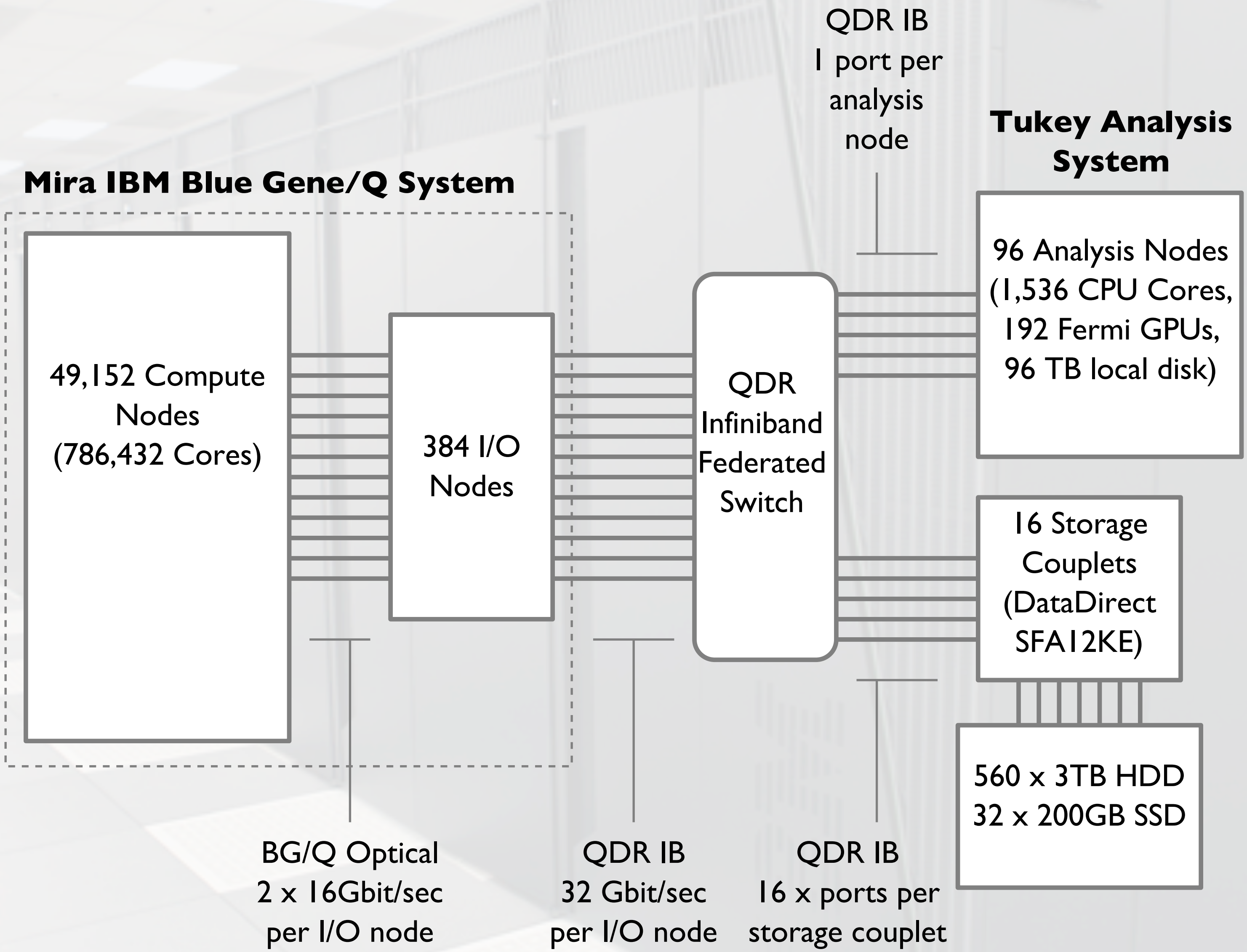
Data Velocity in Computational Science

Data from computational science applications comes in bursts that must be absorbed quickly to maintain high system utilization. Storage systems must serve unprecedented numbers of clients and incorporate massive numbers of devices to meet requirements.



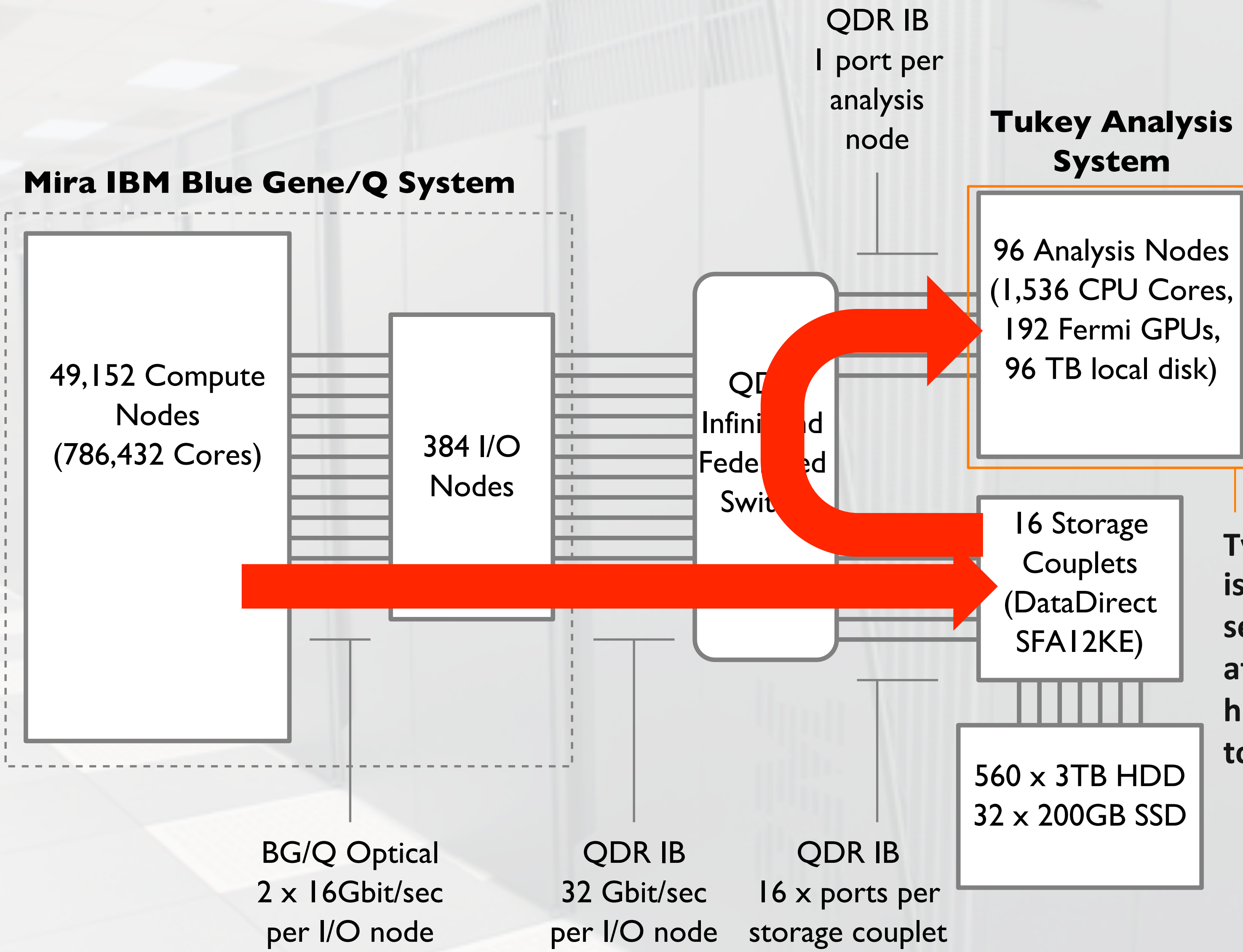
- Trajectory of disk access rate improvements has led to more disks at each HPC system generation
- Projections indicate disk-only storage for exascale would require ~175K disks
- NVRAM helps, but analysis approaches must adapt as well

An Example Leadership System Architecture



High-level diagram of 10 Pflop IBM Blue Gene/Q system at Argonne Leadership Computing Facility

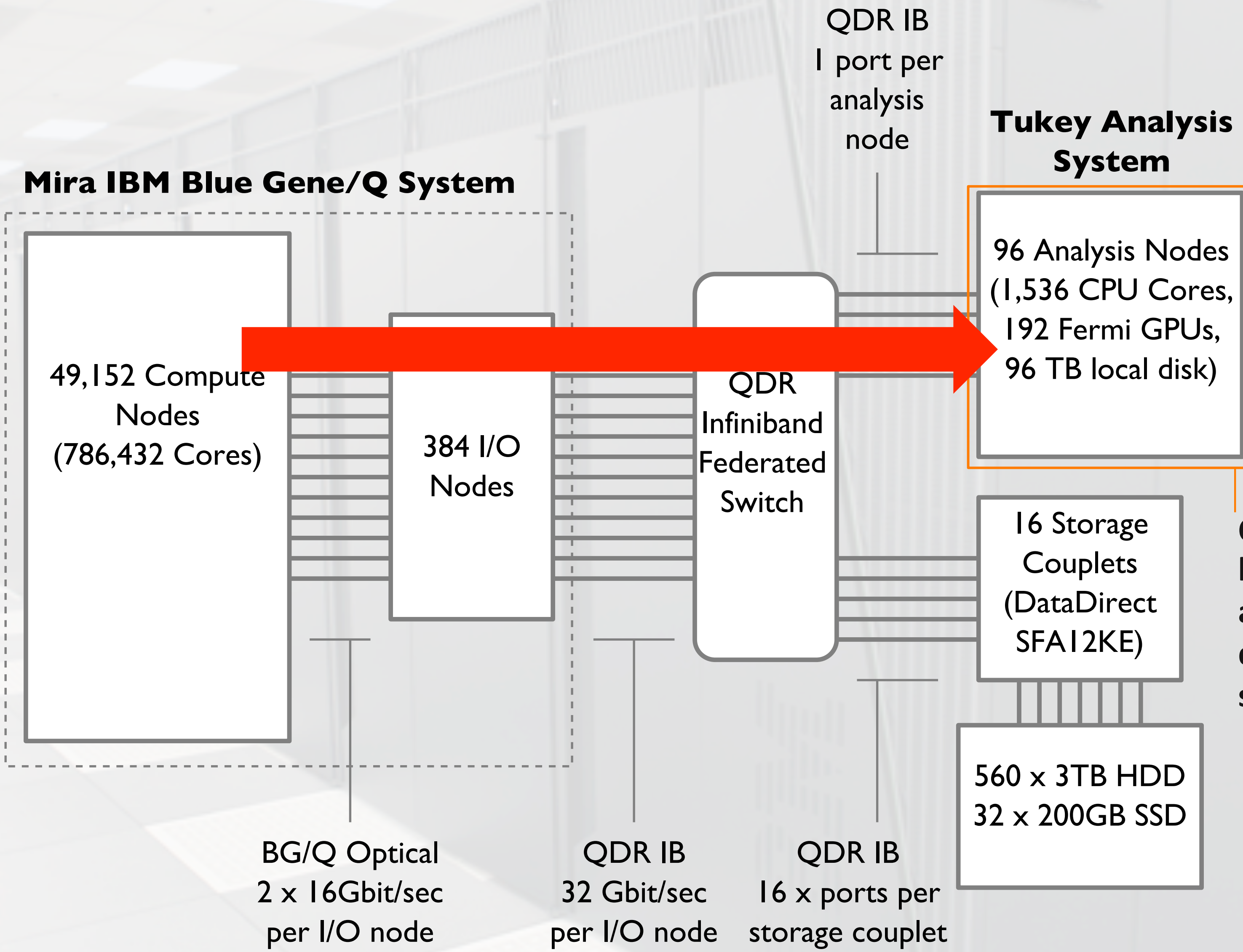
Analyzing Data: Traditional Post-Processing



Typically analysis is performed on a separate cluster, after simulation has written to data to disk.

High-level diagram of 10 Pflop IBM Blue Gene/Q system at Argonne Leadership Computing Facility

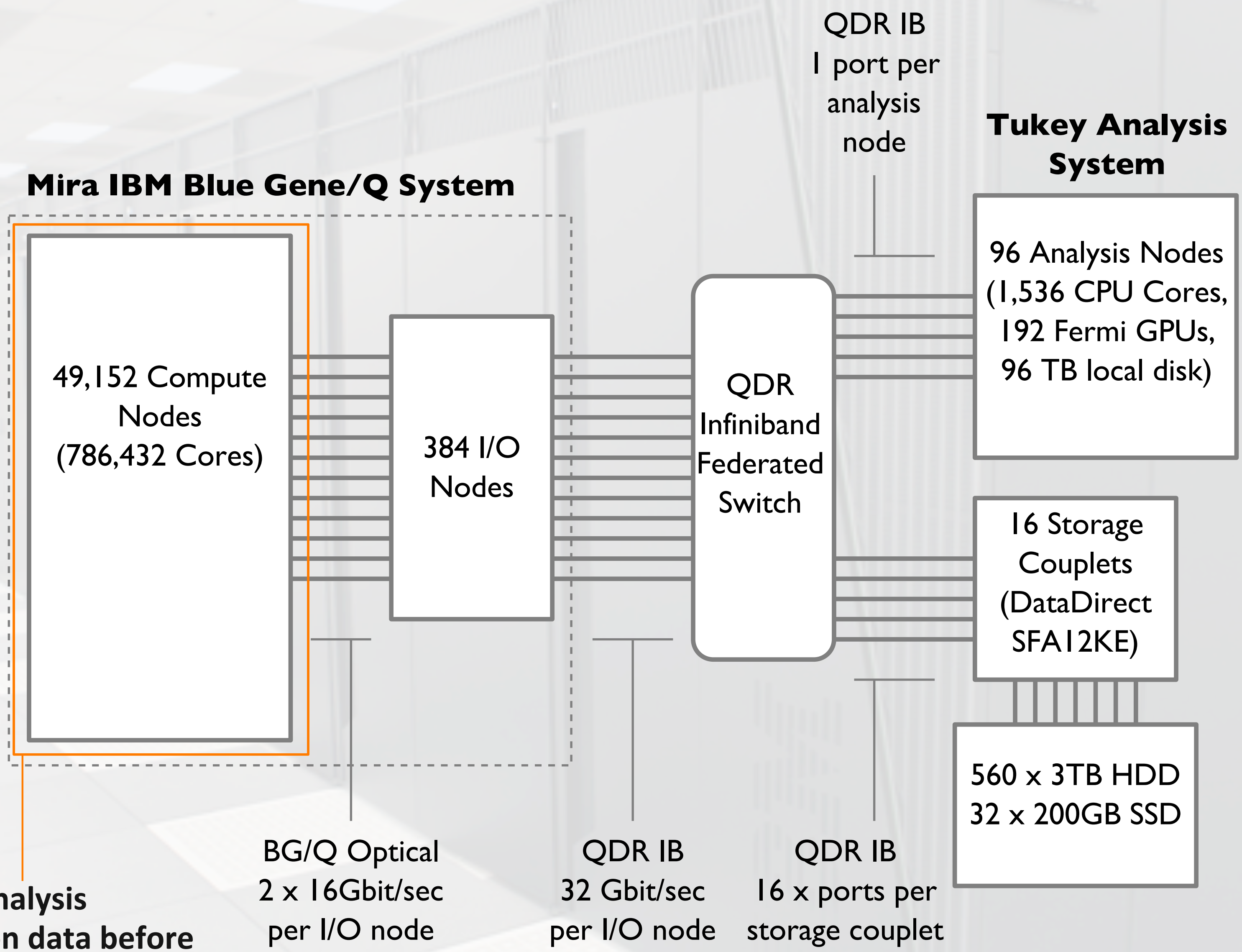
Analyzing Data: Co-Analysis



Co-analysis bypasses storage and processes data while simulation runs.

High-level diagram of 10 Pflop IBM Blue Gene/Q system at Argonne Leadership Computing Facility

Analyzing Data: In Situ Analysis



"In situ" analysis operates on data before it leaves the compute nodes.

High-level diagram of 10 Pflop IBM Blue Gene/Q system at Argonne Leadership Computing Facility

Exascale Systems: Potential Architecture

Systems	2009	2018*	Difference
System Peak	2 Pflop/sec	1 Eflop/sec	O(1000)
Power	6 Mwatt	20 Mwatt	
System Memory	0.3 Pbytes	32-64 Pbytes	O(100)
Node Compute	125 Gflop/sec	1-15 Tflop/sec	O(10-100)
Node Memory BW	25 Gbytes/sec	2-4 Tbytes/sec	O(100)
Node Concurrency	12	O(1-10K)	O(100-1000)
Total Node Interconnect BW	3.5 Gbytes/sec	200-400 Gbytes/sec	O(100)
System Size (Nodes)	18,700	O(100,000-1M)	O(10-100)
Total Concurrency	225,000	O(1 billion)	O(10,000)
Storage	**15 Pbytes**	**500-1000 Pbytes**	**O(10-100)**
I/O	**0.2 Tbytes/sec**	**60 Tbytes/sec**	**O(100)**
MTTI	Days	O(1 day)	

From J. Dongarra, "Impact of Architecture and Technology for Extreme Scale on Software and Algorithm Design," Cross-cutting Technologies for Computing at the Exascale, February 2-5, 2010.

The SDAV Institute

SDAV
Scalable Data Management, Analysis, and Visualization

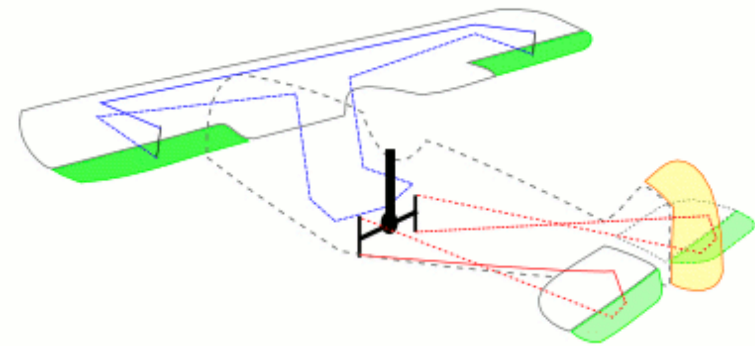
Goal is to assist application scientists in using state-of-the-art data management, analysis, and visualization techniques to make new science discoveries:

- **Data Management** – infrastructure that captures the data models used in science codes, efficiently moves, indexes, and compresses this data, enables query of scientific datasets, and provides the underpinnings of in situ data analysis
- **Data Analysis** – application-driven, architecture-aware techniques for performing in situ data analysis, filtering, and reduction to optimize downstream I/O and prepare for in-depth post-processing analysis and visualization
- **Data Visualization** – exploratory visualization techniques that support understanding ensembles of results, methods of quantifying uncertainty, and identifying and understanding features in multi- scale, multi-physics datasets

- Funded by the DOE Office of Science Advanced Scientific Computing Research Program
- Lead by Arie Shoshani (LBNL)
- Focus is on users of largest DOE/ASCR computational resources
- http://www.sdav-scidac.org

Improving Aircraft Designs

- Goal is to reduce fuel consumption, noise, and drag in commercial aircraft:
  - Redesigning the vertical tail of a commercial jet could reduce jet fuel use by 0.5%, resulting in annual savings of $300 million.
  - One new aircraft control method employs *synthetic jets*; understanding the behavior (e.g., frequency, amplitude, location) is critical for future aircraft wing design
- Synthetic jet simulations are conducted using the PHASTA CFD solver and are being correlated with experimental data
  - Collaboration between Univ. of Colorado, Boulder, Rensselaer Polytechnic Institute, and Boeing
  - Adaptive unstructured mesh code, has scaled to 4.3 billion mesh elements and 160K cores on ALCF Blue Gene/P



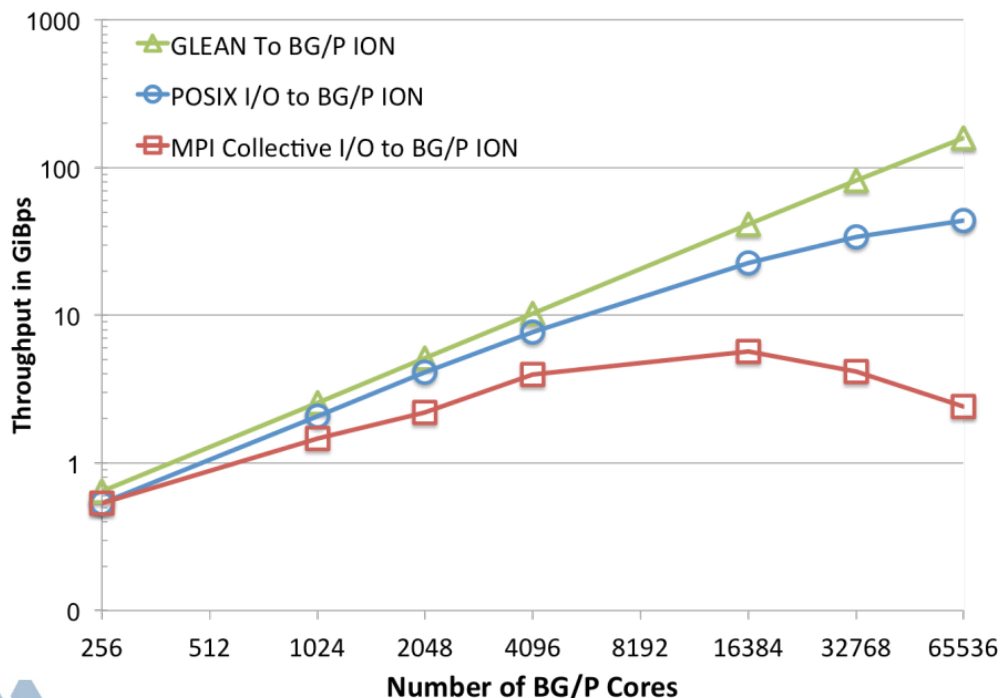
The conventional fix-wing aircraft control surface design is over a century old. New control methods are being developed that improve efficiency and enable new aircraft designs.

Image by Piotr Jaworski, released under GNU Free Documentation License.

Thanks to V. Vishwanath (ANL) for providing this material.

Streamlining Data Movement in Airflow Simulation

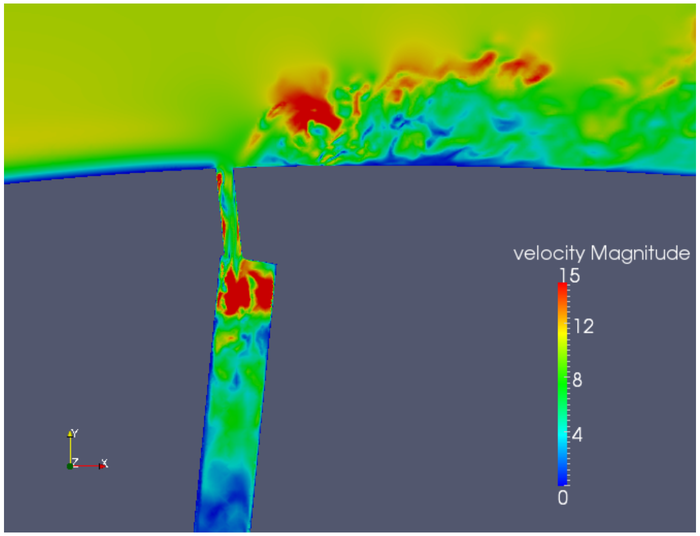
- PHASTA CFD simulations produce as much as ~200 GB per time step
  - Rate of data movement off compute nodes determines how much data the scientists are able to analyze
- GLEAN is a flexible and extensible framework for simulation-time data movement and analysis
  - Accelerating I/O via topology awareness, asynchronous I/O
  - Enabling in situ analysis and co-analysis

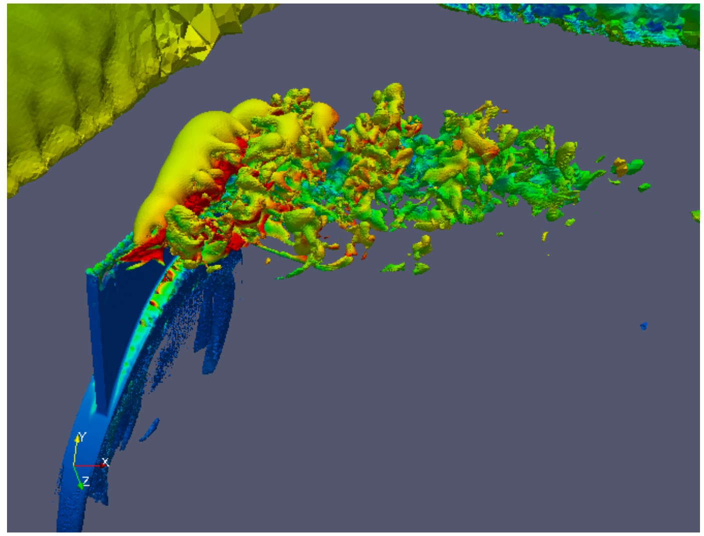


Strong scaling performance for 1GB data movement off ALCF Intrepid Blue Gene/P compute nodes. GLEAN provides 30-fold improvement over POSIX I/O at large scale. Strong scaling is critical as we move towards systems with increased core counts.

Thanks to V. Vishwanath (ANL) for providing this material.

Observing Simulated Synthetic Jet Behavior





Cut plane through synthetic jet (left) and isosurface of vertical velocity (right) colored by velocity (both for 3.3 billion element mesh). Analysis performed with ParaView.

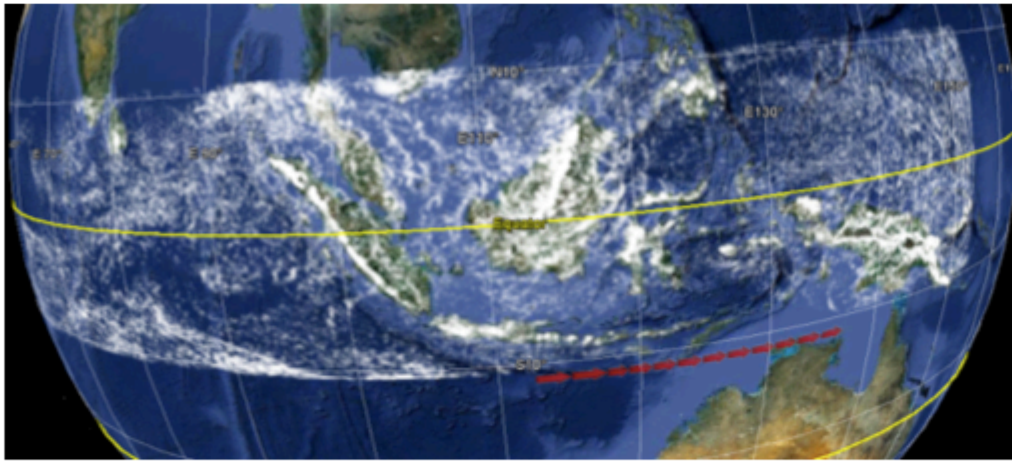
Thanks to K. Jansen (UC Boulder) for these images.

- Using GLEAN, scientists are able to use co-analysis to observe simulation behavior at run time and avoid storage bottlenecks
  - In co-analysis, data is moved from compute to analysis resources without first being stored on disk
  - Reduces storage requirements, overlaps analysis with simulation, and achieves very data throughput (48 GiBps)
- This enables the scientists to better understand the temporal characteristics of the synthetic jet
  - Cost of analyzing a timestep is much lower, so scientists can view results at a higher temporal fidelity than was feasible before (approx. every 10 timesteps)

Understanding the Madden-Julian Oscillation (MJO)

- MJO is a 30-60 day oscillation of enhanced and suppressed rainfall near the Indian and western Pacific Oceans
  - MJO can be thought of as a wavefront indicating how the cloud system is moving
  - Understanding the phenomenon helps explain tropical weather variations
  - Also related to summer precipitation patterns in North America
- Simulation performed by R. Leung and S. Hagos (PNNL) using ARW-WRF3.1
  - 2700 x 600 x 27 (vertical) curvilinear mesh
  - 480 timesteps representing 120 days (multiple iterations of phenomenon)
  - 3GB per timestep

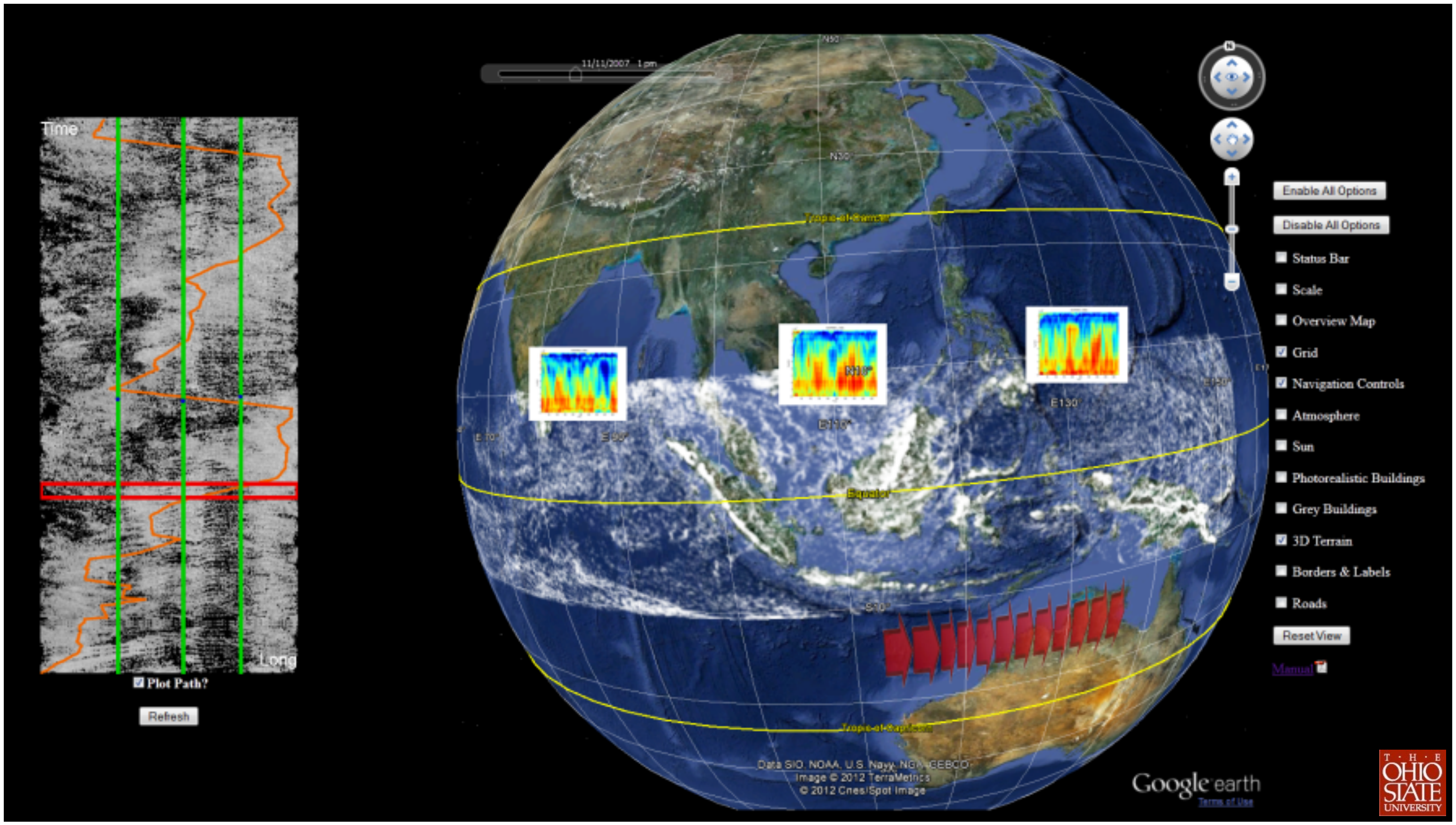
The image shows a rendering of clouds on a virtual globe interface developed by SDAV researchers at Ohio State University, in collaboration with P. C. Wong, S. Hagos, and R. Leung (PNNL).



S. Hagos, L. R. Leung, and J. Dudhia. Thermodynamics of the Madden-Julian oscillation in a regional model with constrained moisture. Journal of Atmospheric Sciences, 68:1974–1989, 2011.

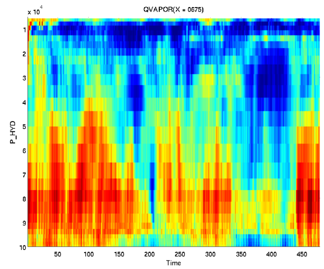
Interactively Exploring the MJO Phenomenon

SDAV members H. Shen and T. Lee (OSU), with collaborator P.C. Wong (PNNL) developed an interface to assist scientists in identifying and exploring the MJO phenomenon in simulations.

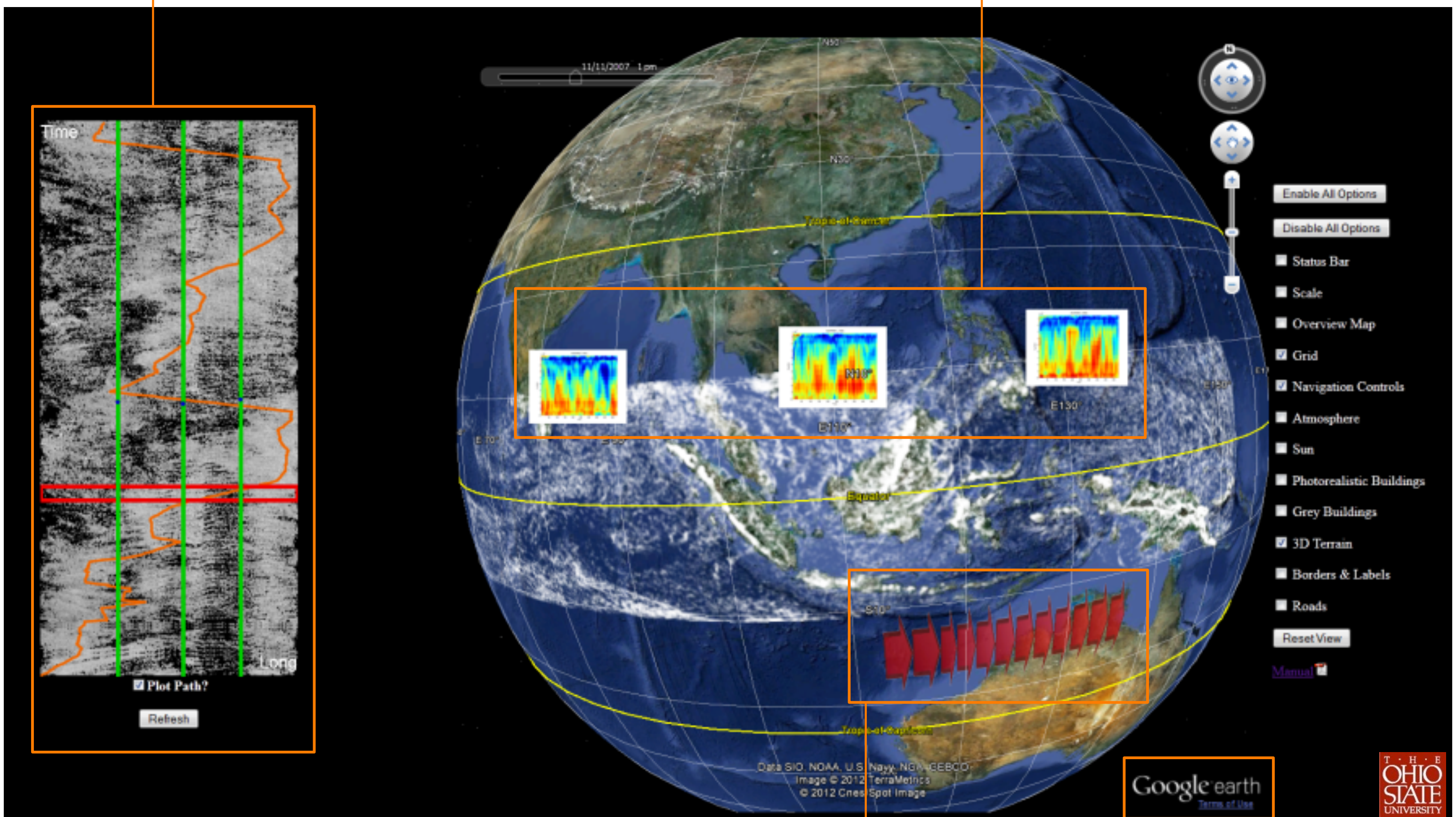


A typical method of viewing water vapor mixing ratio is via a timeline view, with time on the Y axis and longitude on the X axis (known as a Hovmoller diagram).

Orange line tracks highest water vapor mixing ratio over time (MJO path), while red box selects an interval of time for visualization on the right. Green lines show heatmap locations.



Heatmap views are generated on demand when user selects a longitude. Heatmap shows water vapor mixing ratio by altitude (Y axis) over time (X axis).



Red arrows indicate direction of cloud system movement in the selected time period.

Builds off the widely-available Google Earth platform, can be embedded in web pages.

Thanks to H. Shen and T. Lee (OSU) for providing this material.

Understanding How a Laser Pulse Propagates Through a Hydrogen Plasma

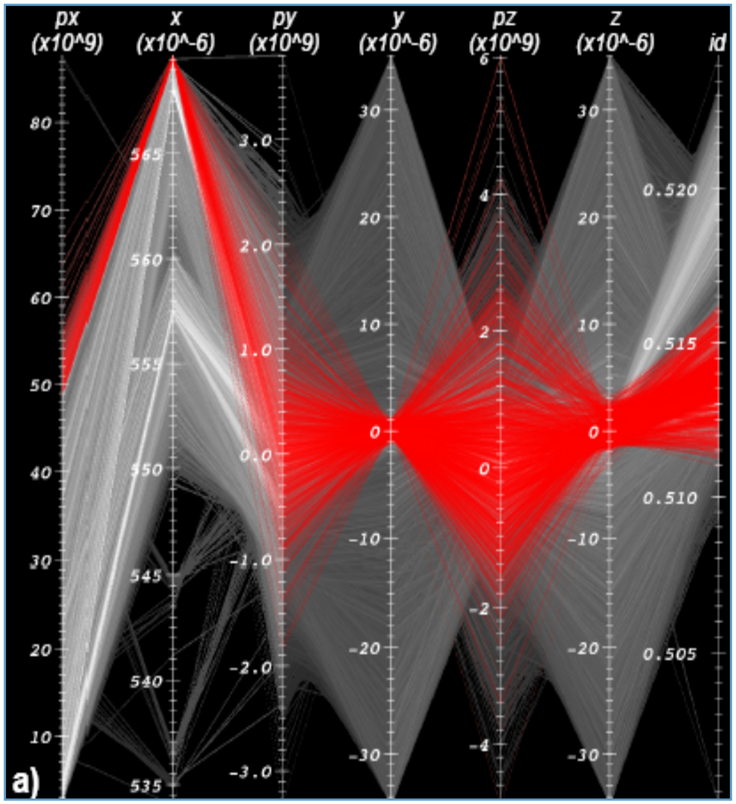
- VORPAL code used to simulate laser wakefield particle accelerator
  - 3D simulation
  - 30 timesteps
  - 90 million particles per timestep, ~5 Gbytes of data per timestep
- Questions:
  - Which particles become accelerated? How are they accelerated?
  - How did the beam form? How did it evolve?
- Data management, analysis, and visualization:
  - **Data model support** – HDF5, H5Part to store data with appropriate metadata
  - **Indexing** – FastBit to enable quick identification of particles of interest, associate particles between timesteps
  - **Visualization** – Parallel coordinates view to help user select particles, VisIt as deployment vehicle

Rubel et al. High performance multivariate visual data exploration for extremely large data. SC08. November, 2008.

Beam Selection

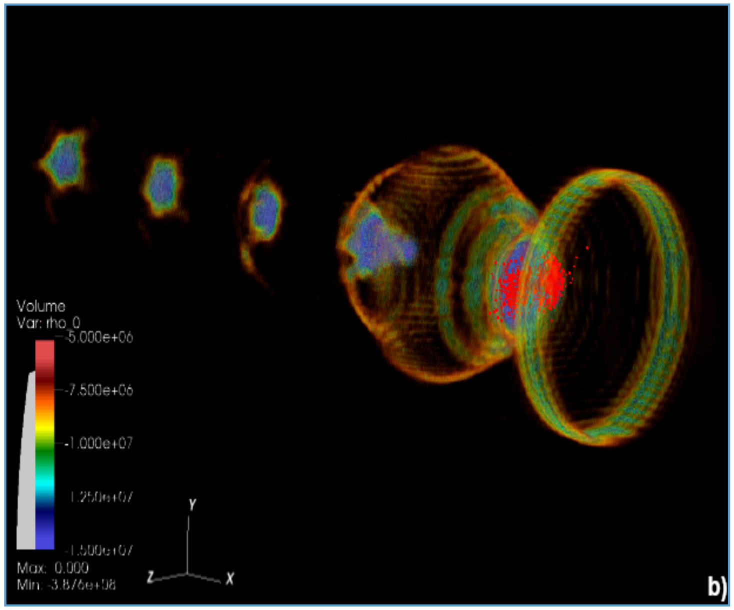
Parallel coordinates view of *t* = 12

- Grey particles represent initial selection ($px > 2*10^9$)
- Red particles represent "focus particles" in first wake period following pulse ($px > 4.856*10^{10}$) && ($x > 5.649*10^{-4}$)



Volume rendering of plasma density with focus particles included in red (*t* = 12)

- Helps locate beam within wake

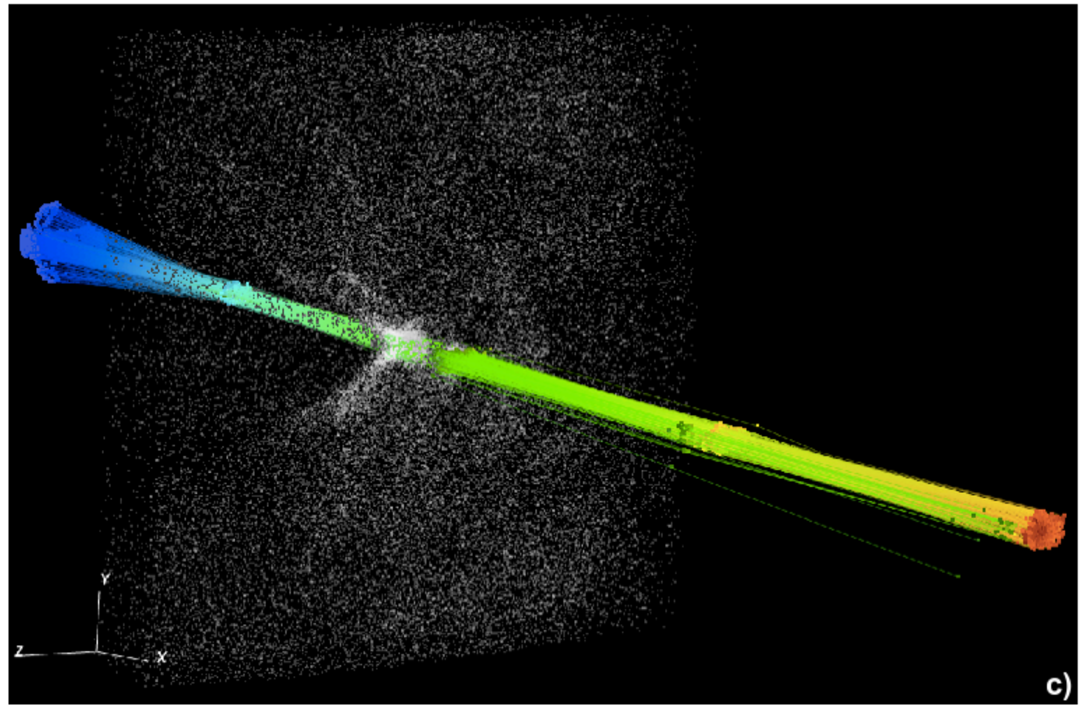


Thanks to E. Wes Bethel (LBNL) for providing this material.

Tracing Particles over Time

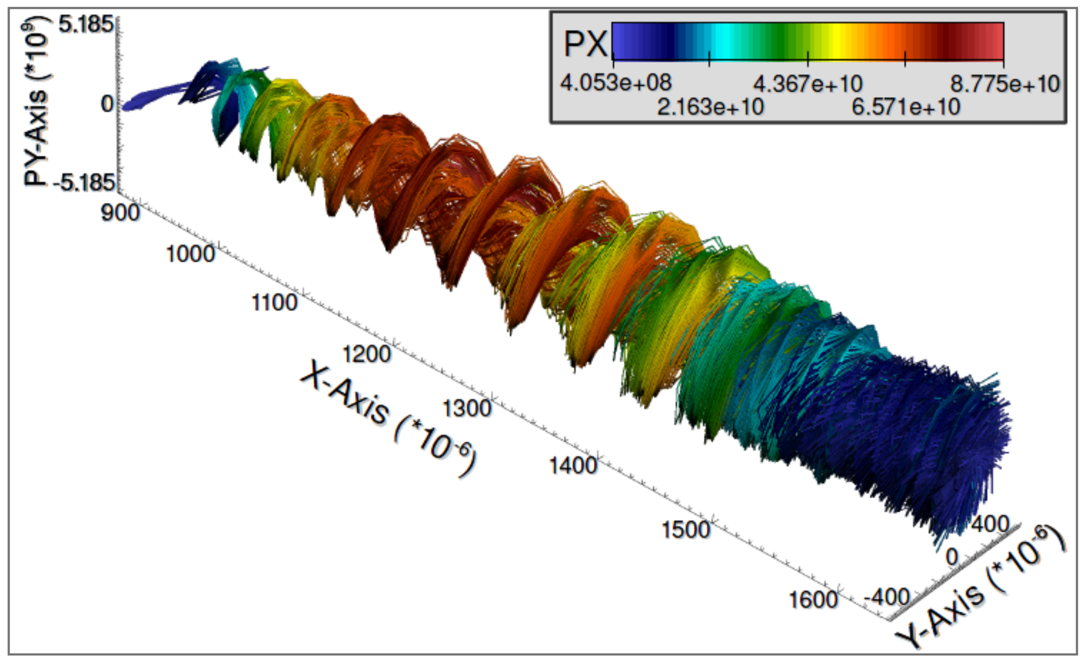
Tracing particles back to *t* = 9 and forward to *t* = 14 allows scientist to see acceleration over time:

- Heatmap shows particles constantly accelerated over time (increase in *px*, left to right).
- Grey particles show initial selection (for reference).



More recent work shows:

- Particles start out slow (blue, left), undergo acceleration (reds), then slow again as the plasma wave outruns them (blue, right).
- Spiral structure shows particles oscillating transversely in the focusing field (new science).



Thanks to E. Wes Bethel (LBNL) for providing this material.

SDAV Technology Use in Leadership Applications

Application	Code	Contact	Allocation (M node hours)	SDAV Technologies
Astrophysics	Chimera	T. Mezzacappa	60	ADIOS, VisIt, Ultravis-V
Astrophysics	FLASH	D. Lamb	80	PnetCDF, GLEAN, ROMIO, VisIt, VTK
Astrophysics	Maestro	J. Bell	50	VisIt
Astrophysics	Enzo	M. Norman	35	ParaView, VisIt
Biology	Nektar	G. Karniadakis	50	ParaView
Climate	POP	P. Jones	110	PnetCDF, ParaView, ROMIO
Combustion	S3D	J. Chen	60	ADIOS, Dataspaces, Ultravis-V, Ultravis-P, ViSUS IDX, Topologika
Combustion	Boxlib	J. Bell	60	VisIt, ADIOS, Topologika
Combustion	Nek5000	C. Frouzakis	150	VisIt
Cosmology	HACC	S. Habib	150	ParaView, ROMIO, Ultravis-P
Fusion	GTC	Z. Lin	35	ADIOS, DataTap, FastBit, Ultravis-V
Fusion	XGC	C.S. Chang	50	ADIOS, Dataspaces, FastBit, Ultravis-V, VTK
Fusion	GTC-P	W. Tang	58	ADIOS, Ultravis-V, Ultravis-P
Plasma	VPIC	B. Daughton	30	PnetCDF, ParaView, ROMIO
Nuclear	Nek5000	P. Fischer	25	ROMIO, VisIt

Final Comments: Accomplishing Our Goal

A mix of activities contribute to the success:

- **Community Engagement –** Actively engaging application teams running on leading DOE computing systems, our sibling Institutes, and DOE computing facility personnel over the lifetime of the Institute.
- **Technology Deployment –** Working with application scientists so that they can use state of the art tools and techniques to support their needs in data management, analysis, and visualization tasks.
- **Research Integration –** Incorporating ASCR basic research results into our portfolio and developing new technologies as needed to meet the needs of application scientists over the next five years.
- **Software Support –** Performing quality software deployment, maintenance, and support to ensure the success of our tools.

Computational science applications are data intensive. SDAV is assisting scientists in using state-of-the-art tools and techniques to manage this data and glean new science discoveries.

Acknowledgments and SDAV Participants

This work was supported by the Director, Office of Advanced Scientific Computing Research, Office of Science, of the U.S. Department of Energy under Contract No. DE-AC02-06CH11357, through the Scientific Discovery through Advanced Computing (SciDAC) Institute of Scalable Data Management, Analysis and Visualization.

Institute Director:
Arie Shoshani, LBNL

Deputy Director:
Robert Ross, ANL

Executive Council:
Arie Shoshani, LBNL (chair)
James Ahrens, LANL
Wes Bethel, LBNL
Hank Childs, LBNL
Scott Klasky, ORNL
Kwan-Liu Ma, UC Davis
Valerio Pascucci, U Utah
Robert Ross, ANL

Key Personnel:
Sean Ahern, ORNL
James Ahrens, LANL
Wes Bethel, LBNL
Peer-Timo Bremer, LLNL
Eric Brugger, LLNL
Phil Carns, ANL
Hank Childs, LBNL
Alok Choudhary, Northwestern
Berk Geveci, Kitware
Charles Hansen, U Utah
Chris Johnson, U Utah
Kenneth Joy, UC Davis
Scott Klasky, ORNL
Robert Latham, ANL
Kwan-Liu Ma, UC Davis
Anatoli Melechko, UT Knoxville
Kenneth Moreland, SNL
Michael Papka, ANL
Manish Parashar, Rutgers
Valerio Pascucci, U Utah
Tom Peterka, ANL
Norbert Podhorszki, ORNL
David Pugmire, ORNL
Nagiza Samatova, NC State
William Schroeder, Kitware
Karsten Schwan, GA Tech
Han-Wei Shen, OSU
Venkat Vishwanath, ANL
Matthew Wolf, GA Tech
Jonathan Woodring, LANL
John Wu, LBNL