Advanced Computing at NASA Advanced Supercomputing (NAS) Division

Dr. Piyush Mehrotra

Division Chief

NASA Ames Research Center, Moffett Field, Calif., USA!

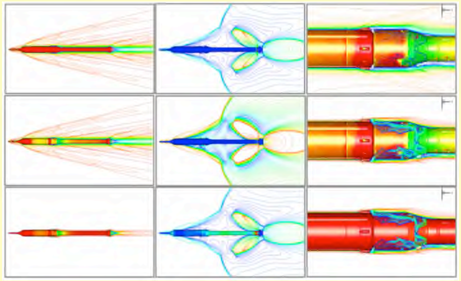
NAS Division: Advanced Computing

- •" High-End Computing Capability (HECC) Project
- •" Quantum Artificial Intelligence Lab (QuAIL)
- •" NASA Earth Exchange (NEX)
- •" Data Analytics Infrastructure for NASA
- •" Cloud Services for Science & Engineering Projects

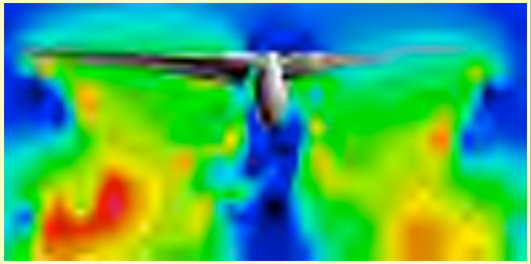
High Performance Computing

NASA's Diverse HPC Requirements

High throughput / capacity
Engineering requires HEC resources that can handle large ensembles of moderate-scale computations to efficiently explore design space



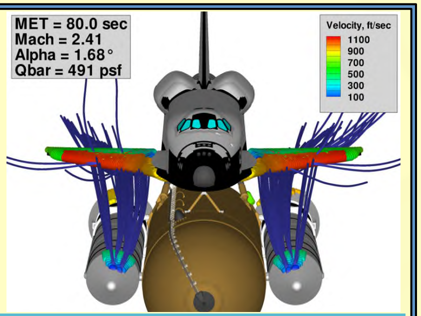
Aerodynamic database generation



Rotary wing aerodynamics

Leadership / capability Research requires HEC resources that can handle high-fidelity long-running large-scale computations to advance theoretical understanding

High availability
Time-sensitive mission-critical applications require HEC resources on demand. readily available

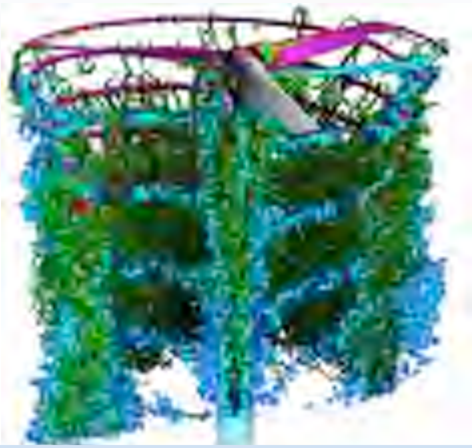


Debris transport analysis

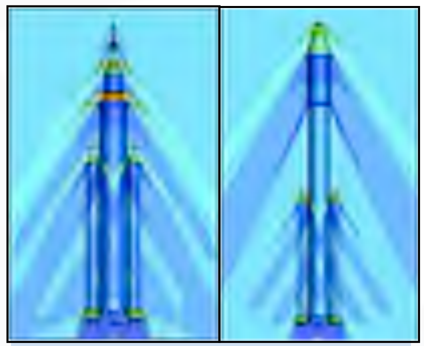
HECC Project provides the premier supercomputing resources to meet NASA's HPC requirements

High Performance Computing @ NAS

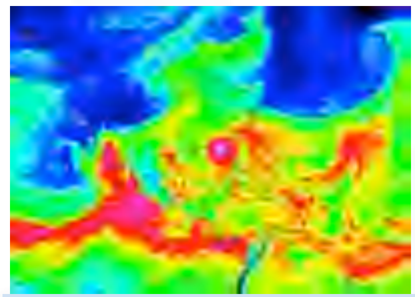
NASA's premier supercomputing facility provides HPC resources to all NASA Mission Directorates



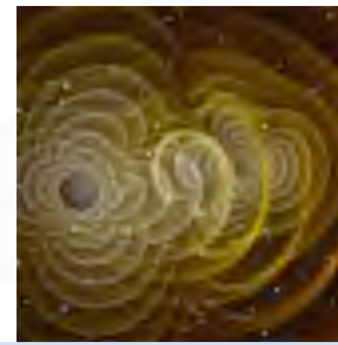
Rotary wing aerodynamics



SLS vehicle designs

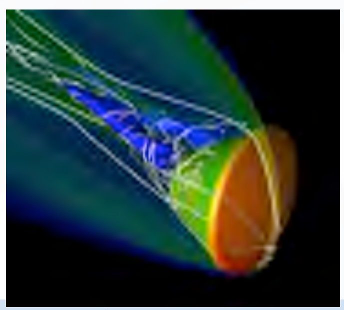


Hurricane prediction



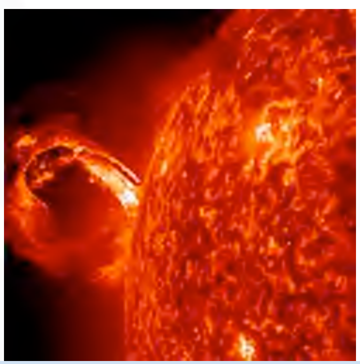
Merging black holes

Sonic boom optimization



Orion/MPCV reentry

Ocean Modeling



Solar magnetodynamics

5

High-End Computing Capability (HECC)

Supercomputing to support NASA Science & Engineering

Computing Systems

- **Pleiades** – 4.5 PF peak
  - Distributed memory cluster
  - 198,432 cores, 11232 nodes SGI Altix ICE
  - 4 generations of Intel Xeon: Westmere (6c), Sandy Bridge (8c), Ivy Bridge (10c), Haswell (12 c)
  - 162 racks, 616 TB memory
  - InfiniBand Interconnec, hypercube topology
  - #21 on TOP500 - #9 in US (06/14)
- **Endeavour** – 32 TF peak
  - Shared memory system
  - 2 SGI UV 2000 Sandy Bridge based nodes
    - 1024-core with 4 TB shared memory
    - 512-core with 2 TB shared memory

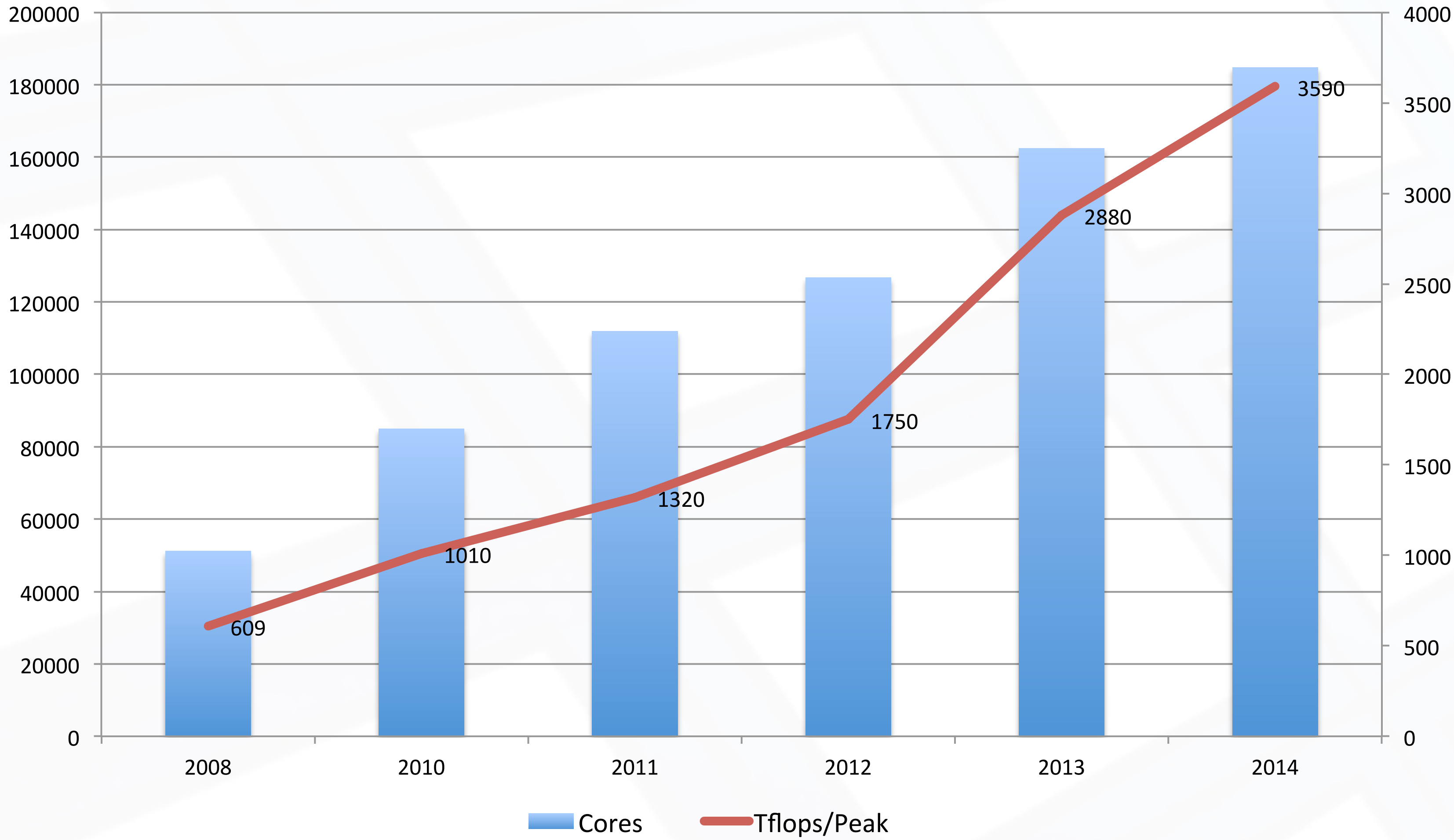
Balanced Environment

- Storage: 20.1 PB disk; 115 PB tape
  - Archiving ~1 PB/month
- WAN: 10 Gb/s external peering
  - Transferring 200 TB/month to users

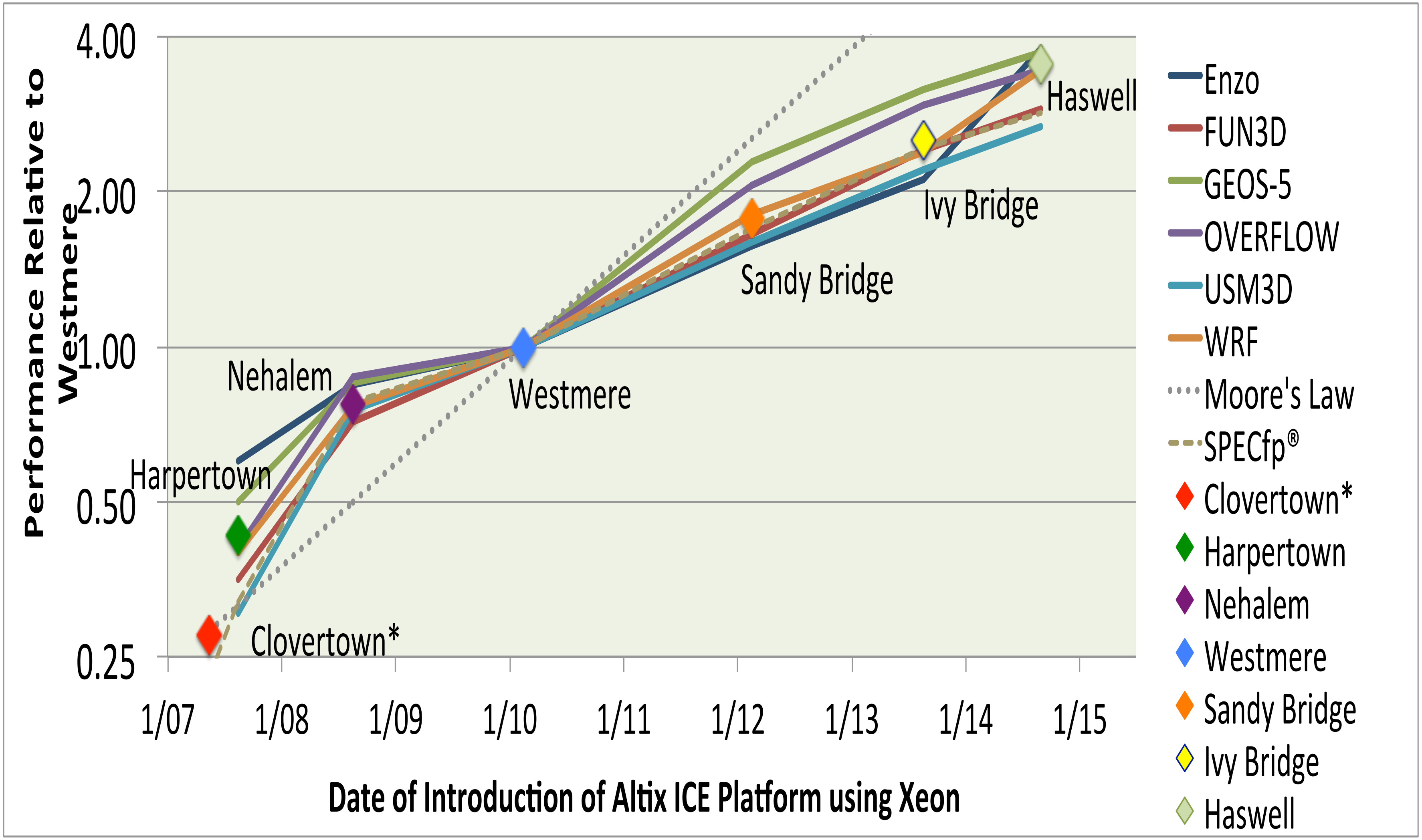
Resources enable broad mission impact

- Mission Directorates select projects, determine allocations
- Over 500 science & engineering projects
- More than 1,200 user accounts

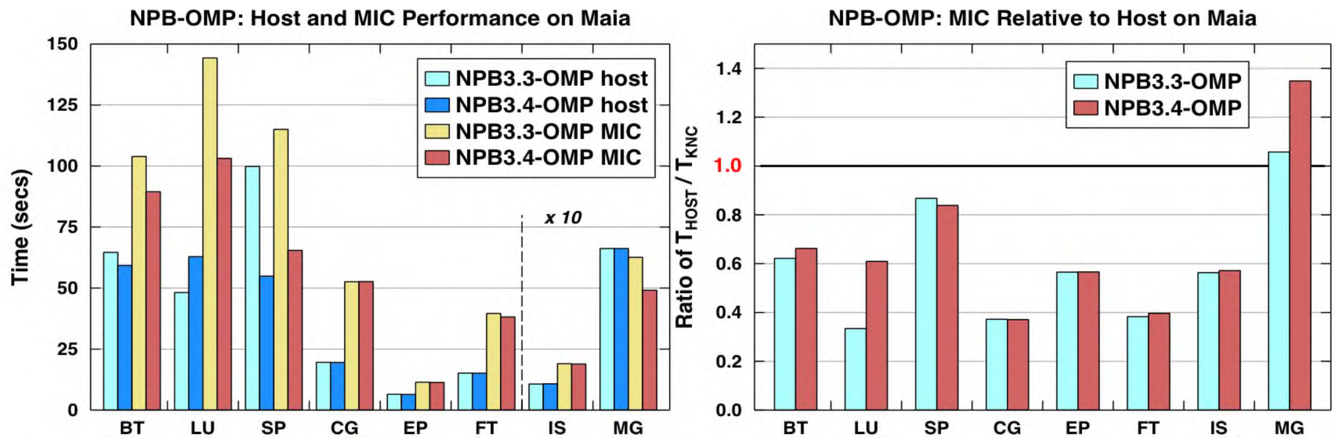
Evolution of Pleiades



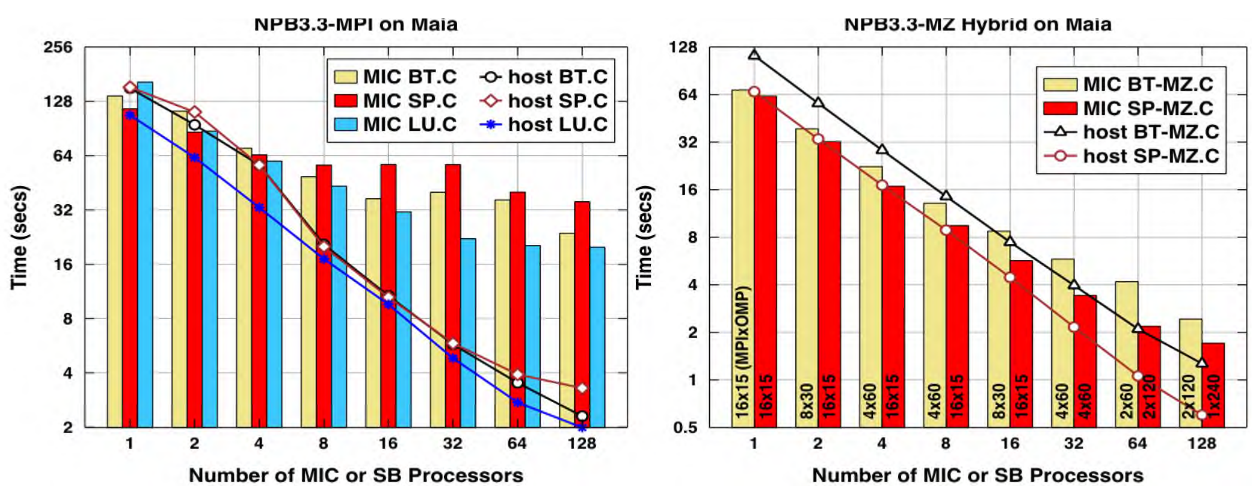
Tracking Xeon Performance



NPBs

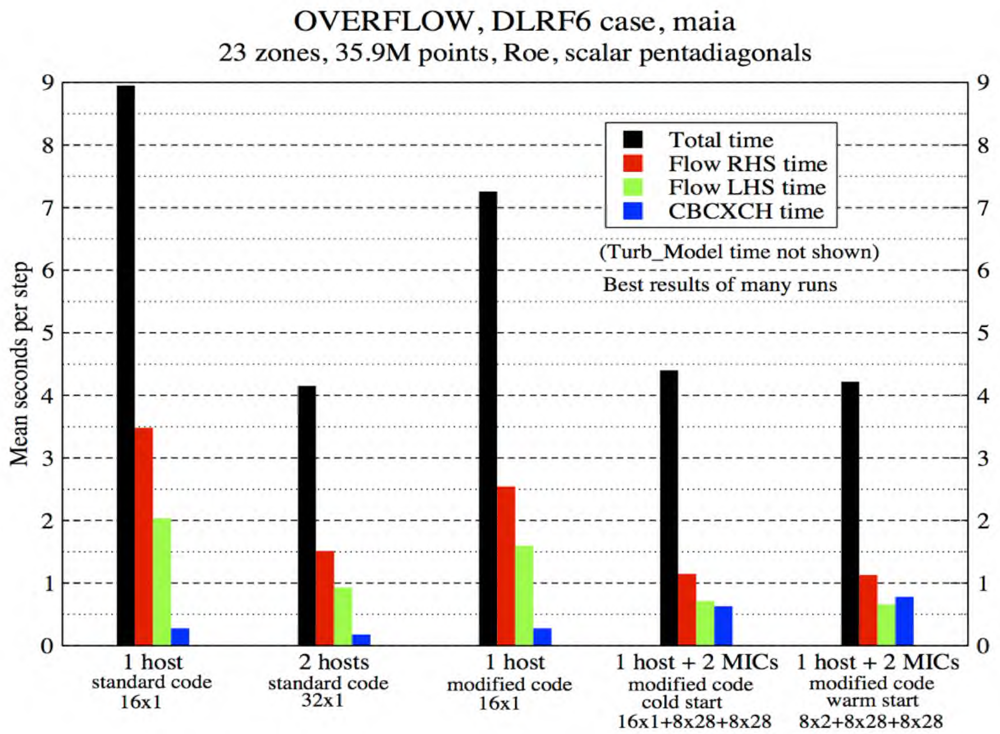


Performance comparison of the OpenMP versions of NPBs on a single node.



Performance comparison of the MPI and hybrid versions of NPBs on multi nodes.

Overflow



Break-down time of OVERFLOW on host+MIC runs compared to host only runs.

Overflow

mxn	*standard code*	*line-oriented code*	*%imp*
2x112	6.78	2.34	65.9
3x72	5.49	2.32	58.4
4x56	5.11	2.21	57.3
6x36	4.87	2.08	57.2
7x32	4.74	2.04	57
8x28	4.49	2.06	54.1
14x16	3.77	2.01	46.5

mxn	*standard code*	*line-oriented code*	*%imp*
1x16	1.82	1.36	24.7
2x8	1.67	1.28	22.1
4x4	1.55	1.26	17.5
8x2	1.47	1.24	14.1
16x1	1.33	1.26	3.9

Performance comparison of OVERFLOW on a single node.

Accelerator Performance

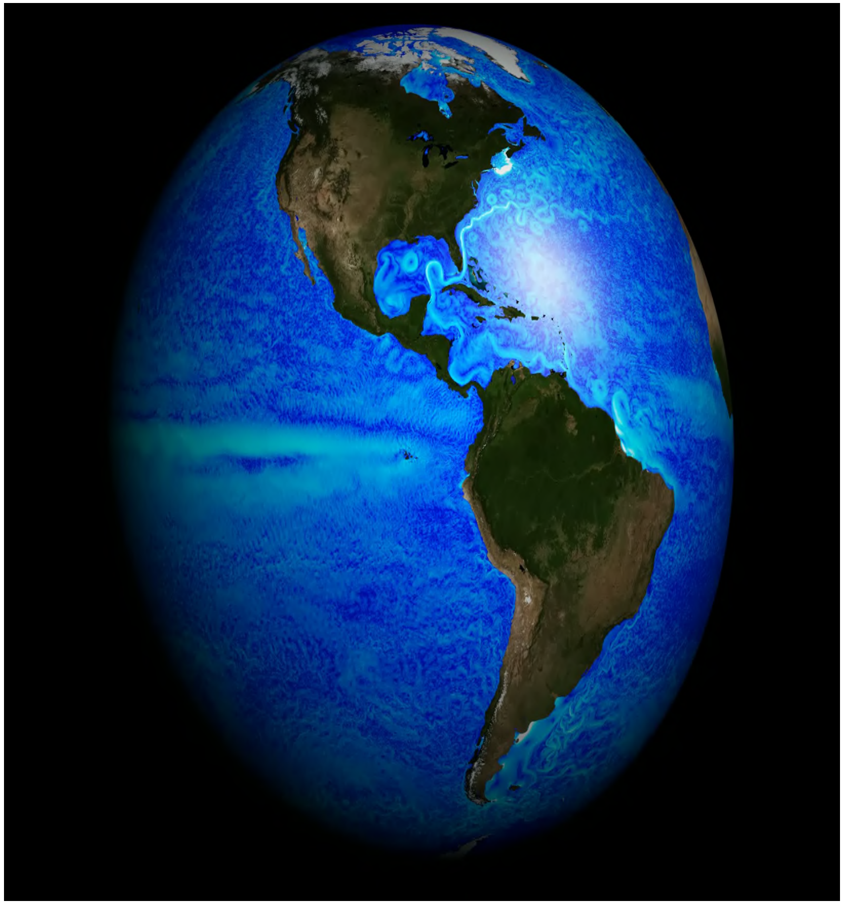
- Lack of system software commonality across host and coprocessor was a hindrance
- Many performance issues in all modes of use
  - Offload, Native & Symmetric
- Hybrid code in symmetric mode yields best performance

Bottom line: *Not ready for prime-time, but next version looks more promising:*

- *Several architectural improvements*
- *No host requirement*

Global Ocean Modeling

- ECCO project: a quantitative depiction of the time-evolving global ocean state
- MITgcm – a global circulation model used in conjunction with diverse observational data to force and tune the model
- 1/48° resolution; 22,5B grid points
- Runs up to 70K+ cores ; ~3 PB of data simulating ~4.5 months
- Innovative visualization techniques to allow scientists to interact with the data at different values and scalar values



Challenges of Running at Scale

- The Goal: Scale the MITgcm code up to ~70,000 cores for runs at 1/48° resolution.

- I/O Bottleneck –
  - Single MPI rank handled all I/O
  - Complete re-design of I/O - added auxiliary I/O processes (*m x n* communication t computational and I/O ranks)
  - I/O ranks communicate among themselves to assemble the desired output datasets maximizing overlap of I/O with computation
  - ~ 10% of resources dedicated to I/O ranks

Challenges of Running at Scale (contd.)

- Communication Issues:
  - Corrected Ivy Bridge node BIOS problem with QPI coherency operations and PCI writes
  - Updated FDR switch firmware with improved link level recovery algorithm
  - Corrected FDR HCA firmware problem with RC transport drops
  - Resolved several issues with SGI's MPT (Message Passing Toolkit)
    - Rewrote packet retry logic
    - Updated intercommunicator implementation
    - Adjusted several interdependent network timeout values
    - Controlled Delay (CoDel) to meter packet injection rates onto network

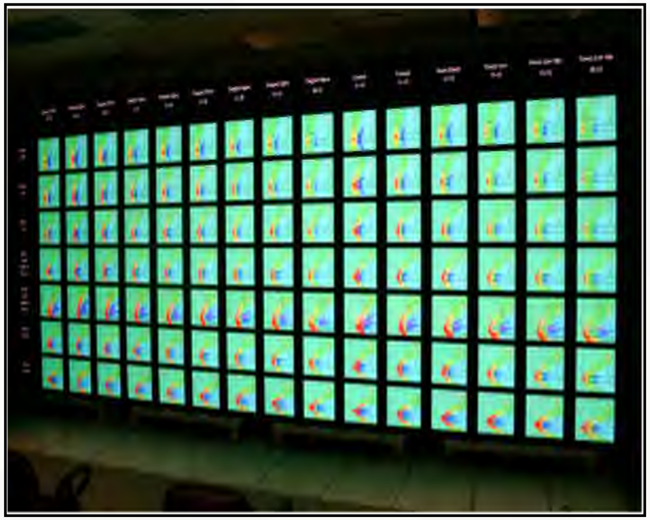
- *Running at scale allowed us to diagnose and fix problems that small-scale runs ride through;*
- *updates resolved issues affecting many supercomputer sites*

Advanced Visualization

- ***Supercomputing-scale visualization system*** to handle massive size of simulation results and increasing complexity of data analysis
  - 8x16 LCD display (23 feet x 10 feet)
  - 245 million pixels

- ***Two primary modes***
  - Single large high definition image
  - Sets of related images (e.g. parameter study)

- ***High-bandwidth t to HEC resources***
  - *Traditional Post-Processing*: Direct read/write access to Pleiades file systems eliminates need for copying large datasets
  - *Concurrent Visualization*: Runtime data streaming allows visualization of every simulation time step - ultimate insight into simulation code without increase in traditional disk I/O



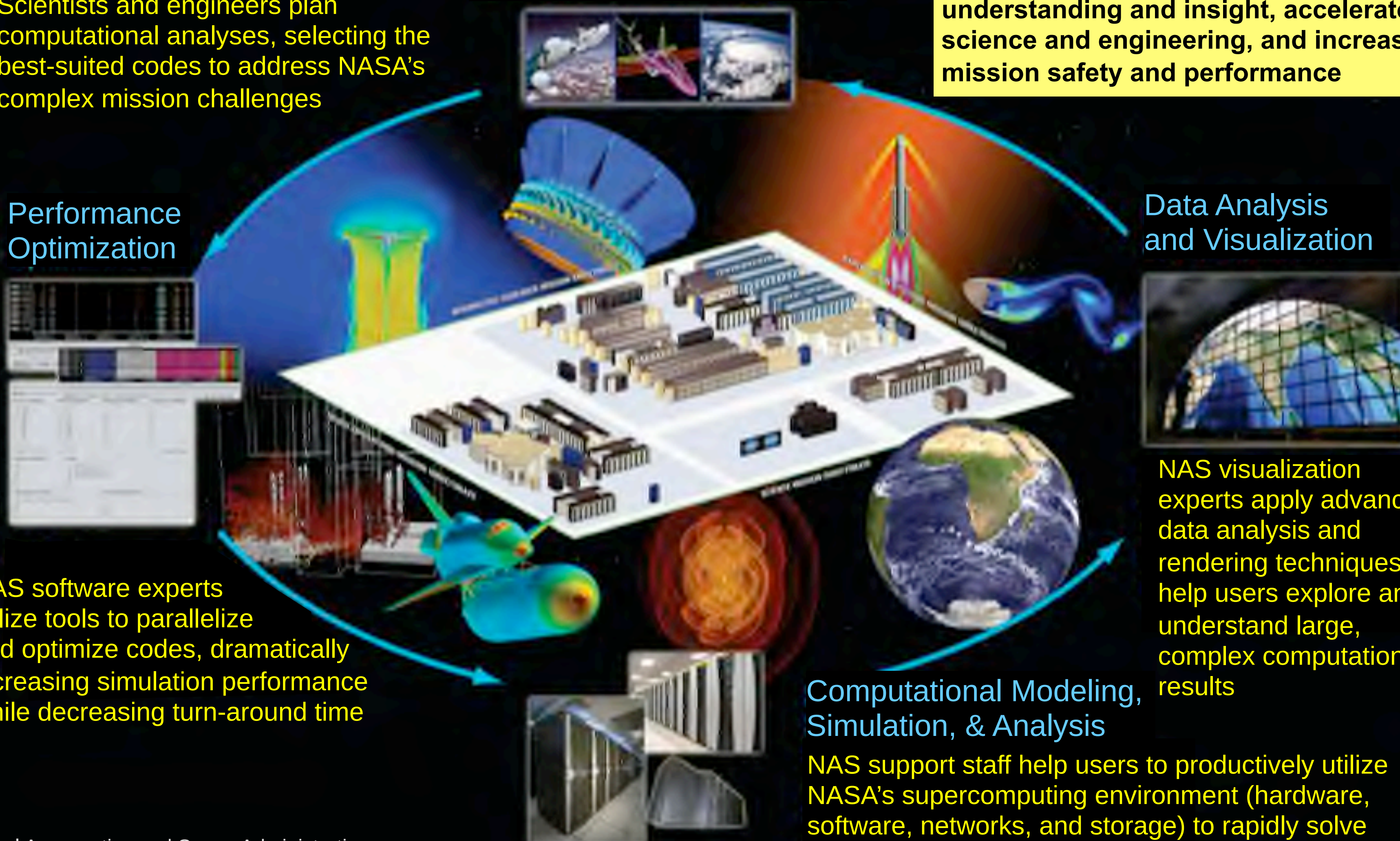
Integrated Spiral Support for MS&A

Develop and deliver the most productive integrated supercomputing environment in the world, enabling NASA to extend technology, expand knowledge, and explore the universe

Outcome: Dramatically enhanced understanding and insight, accelerated science and engineering, and increased mission safety and performance

NASA Mission Challenges

Scientists and engineers plan computational analyses, selecting the best-suited codes to address NASA's complex mission challenges



Performance Optimization

NAS software experts utilize tools to parallelize and optimize codes, dramatically increasing simulation performance while decreasing turn-around time

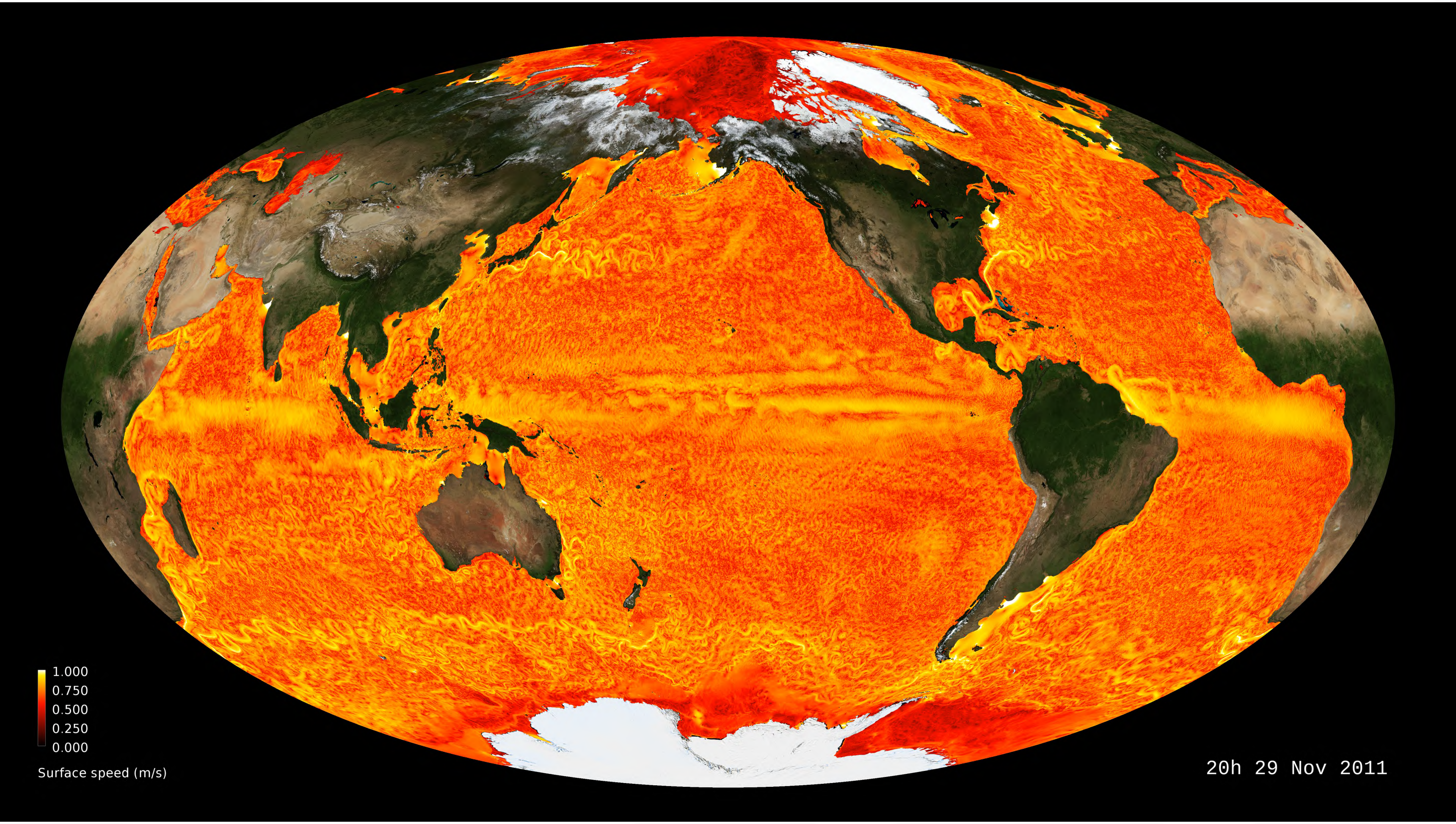
Data Analysis and Visualization

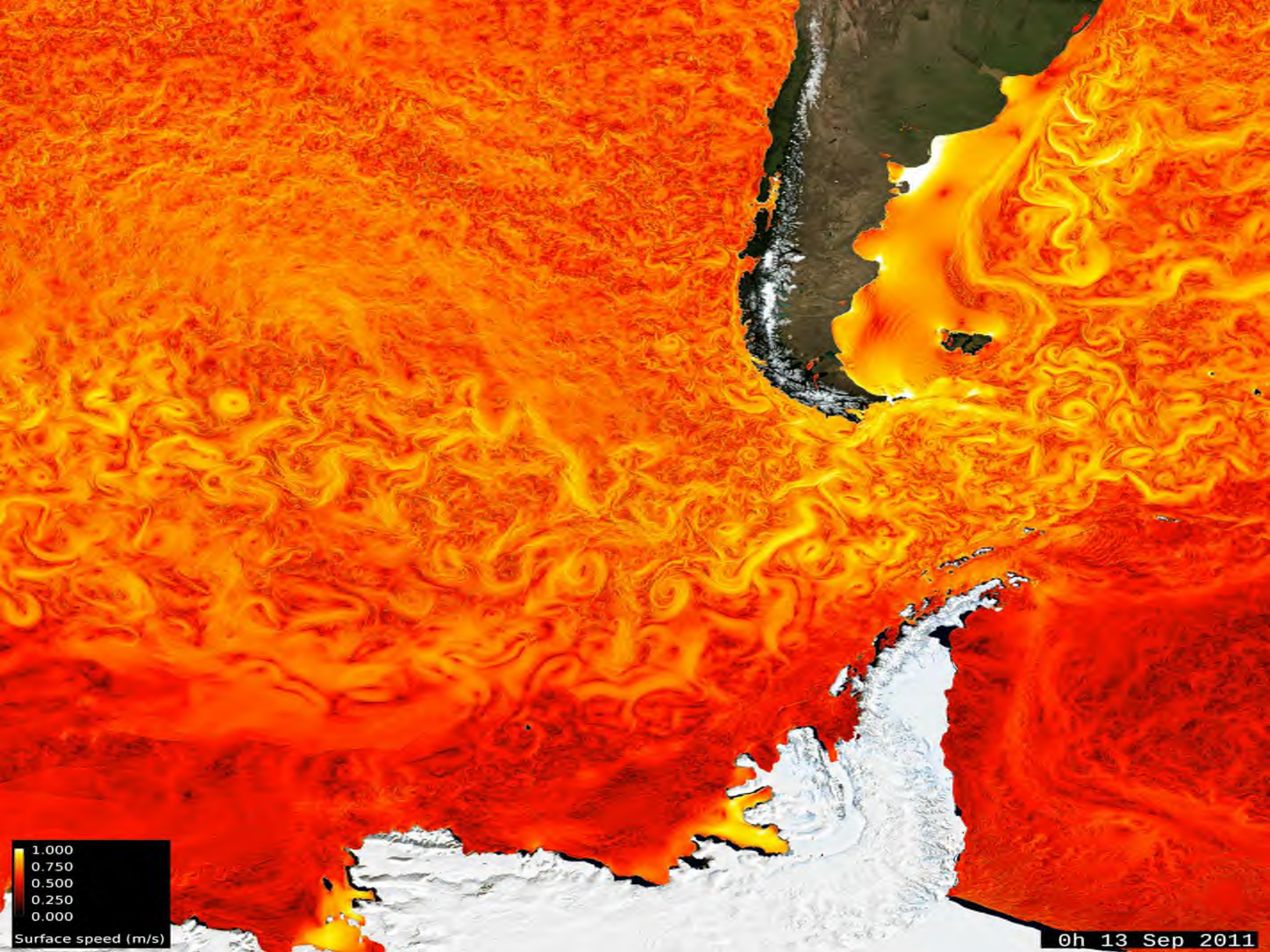
NAS visualization experts apply advanced data analysis and rendering techniques to help users explore and understand large, complex computational results

Computational Modeling, Simulation, & Analysis

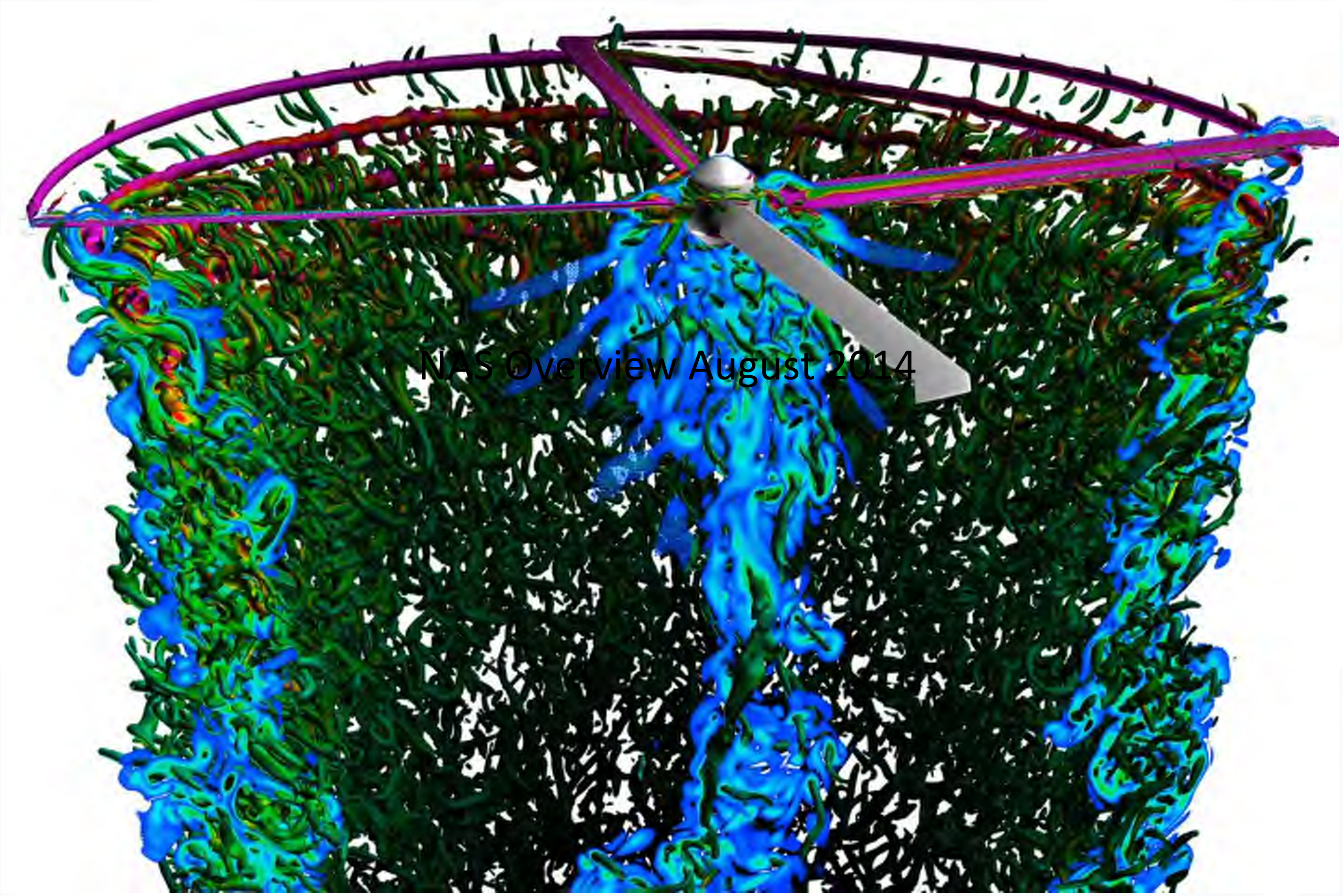
NAS support staff help users to productively utilize NASA's supercomputing environment (hardware, software, networks, and storage) to rapidly solve large computational problems

MITgcm Result



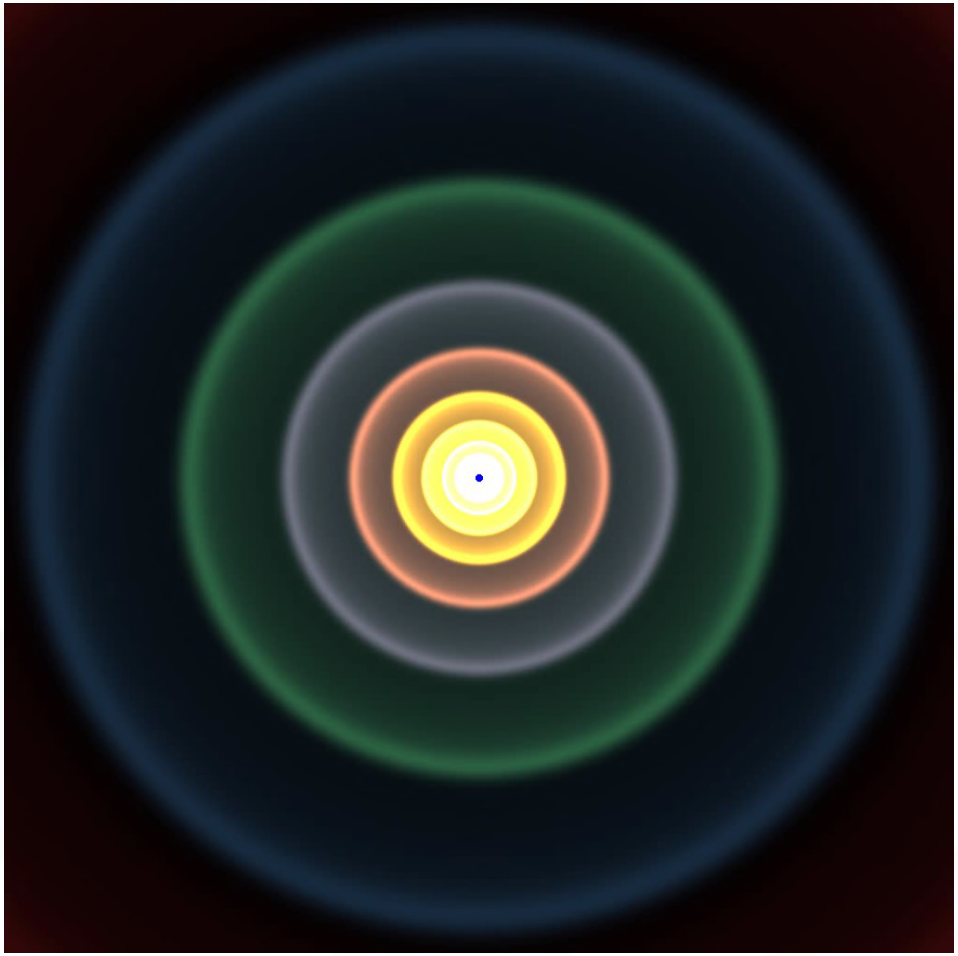

1.000
0.750
0.500
0.250
0.000
Surface speed (m/s)
0h 13 Sep 2011

Rotorcraft Video



Flame Trench Video

Star Formation Video



Data Intensive Supercomputing Environment

Big Data @ NAS

NASA has enormous collections of observational and model data

- *Model/Simulation Data*
  - NAS has 20+ PB unique data stored; 115 PB archive storage; archiving 1+ PB per month
  - MITgcm 35K core@ $1/48^{th}$ degree resolution run produced 1.44 PB in its 5 day run; full run will produce 9-18 PB
- *Observational Data*
  - Tens of satellites and telescopes producing multi-petabytes of data per year
  - Solar Dynamics Observation (SDO) satellite produces 1 GB per minute => > 1/2 PB/year; ~3 PB in its 5-year life cycle
  - NASA Earth Science operates 12 DAACs (archive centers); National Space Science Data Center

Fun Fact: The term “Big Data” was first used by Michael Cox & David Ellsworth of NAS in a paper: “*Visualizing flow around an airframe*” Visualization 97, Phoenix AZ.

- Biggest data set considered 7.5GB; high-end analysis machines had less than 1GB memory

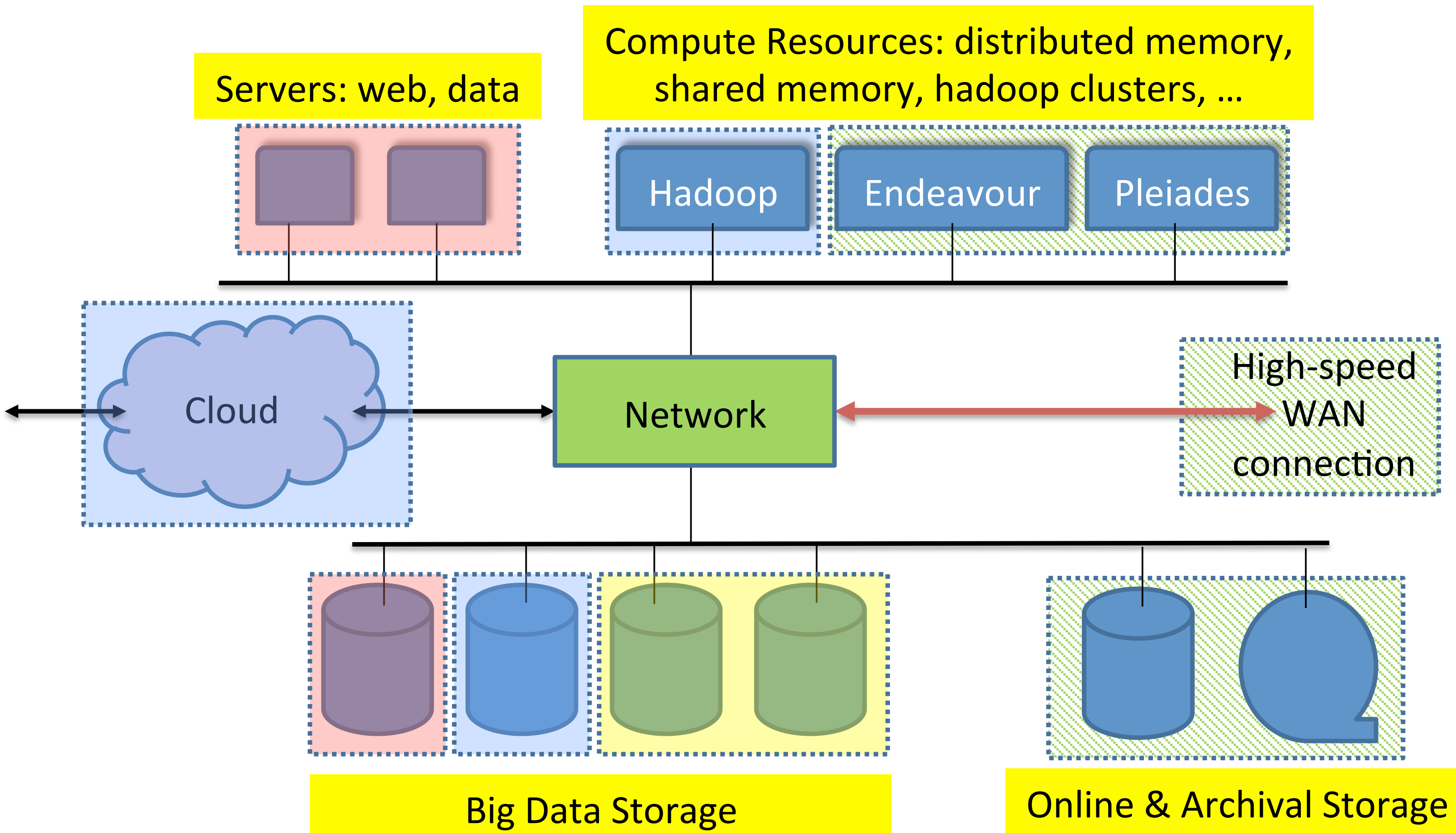
Big Data Challenges for Users

Current Focus: *Develop and implement a roadmap for an infrastructure to support large-scale science data analysis & analytics*

- Conducted survey of NASA projects dealing with big data to understand what users are doing and what infrastructure is needed to serve them - Tech report available at the NAS website
- Continuing deep dives in use cases including prototype implementations
- Note: we already perform visualization based analysis of large datasets

- ***Data Discovery*** *– finding what data is available and where?*
  - Indexing/archiving, metadata – requires federated metadata service and semantic reasoning
- ***Data management*** *–managing/ transferring very large data sets from archives to computational resources*
  - Distributed file system support
  - Increased WAN bandwidth
  - Fault tolerant and resilient hardware/software infrastructure
- ***Tools/models/algorithms*** *- developing analytics/analysis software at scale*
  - *NASA datasets generally consist of structured gridded/geo-spatial data*
- ***Analysis workflow*** *– increasing complexity of processing pipelines have multiple components requiring heterogeneous resources*
  - Software for workflow description and management to tie all components together and facilitate re-use
- ***Analysis Infrastructure*** *– inadequacy of available resources*
  - Large I/O pipelines
  - Large memory spaces for in-core analysis
  - Support for the heterogeneous requirements across the workflow, e.g., Distributed memory & shared memory systems, hadoop cluster with local storage, etc.
- ***Collaboration environments*** *– difficult to share knowledge across a wider community*
  - Support for dissemination and sharing of code, data products, results, etc.

Proposed DICE Environment



NEX: NASA Earth Exchange

NASA Earth Exchange (NEX)

NEX

A collaborative environment that brings scientists and researchers together in a knowledge-based social network along with tools, computing power and data to accelerate research, innovation and provide transparency.

VISION

To provide "*science as a service*" to the Earth Science community addressing global environmental challenges.

GOAL

To improve efficiency and expand the scope of NASA Earth science technology, research and applications programs.

Project Manager: Piyush Mehrotra
Principal Scientist: Ramakrishna Nemani

NEX Environment

Collaboration Portal
(Over 400 members)

Tools
Models
Workflows
Papers
Data

Compute Resources

Sandbox for experimentation

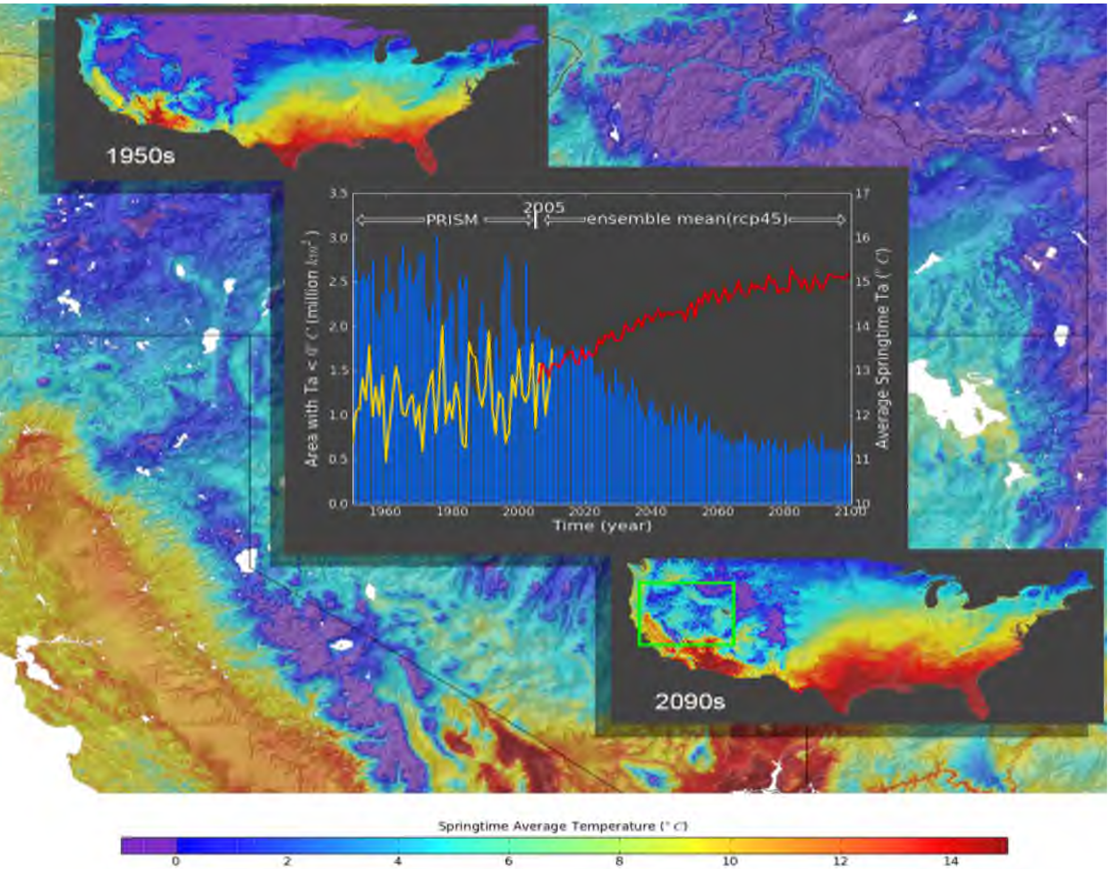
HPC

Centralized Data Repository
(over 1.3 PBs of Data)

Working copies of observational data
Project data

- Collaboration portal open to all Earth Scientists
- Sandbox currently available to a subset with NASA credentials
- HPC resources available only to approved projects with allocation
- OpenNEX a collaboration with Amazon to provide access to some NEX data
- In the process of setting up an OpenSandbox to allow access to all NEX datasets along with some compute resources to a wider Earth Science community

Creating Downscaled Climate Projections for National Climate Assessment



Changes in springtime (March-April-May) mean temperature over the conterminous US from 1950 to 2099.

- Historical and 4 possible future scenarios based on the latest climate models for the conterminous U.S.
- ~800m grid cell size, monthly time-step, 1950-2099
- Includes maximum/minimum temperature and precipitation
- Statistical downscaling (bias-corrected spatial disaggregation; Maurer et al., 2007)

Thrasher et al., 2013, ***Eos***

Fallowed Area Mapping for Drought Impact Reporting and Decision Making

PI: *James Verdin, USGS / NIDIS*

Highlight: Working together, USGS, USDA, and NASA have demonstrated the feasibility of applying satellite data to provide early-season and within-season fallowed area maps and acreage estimates using NEX to accelerate prototyping and workflow development.

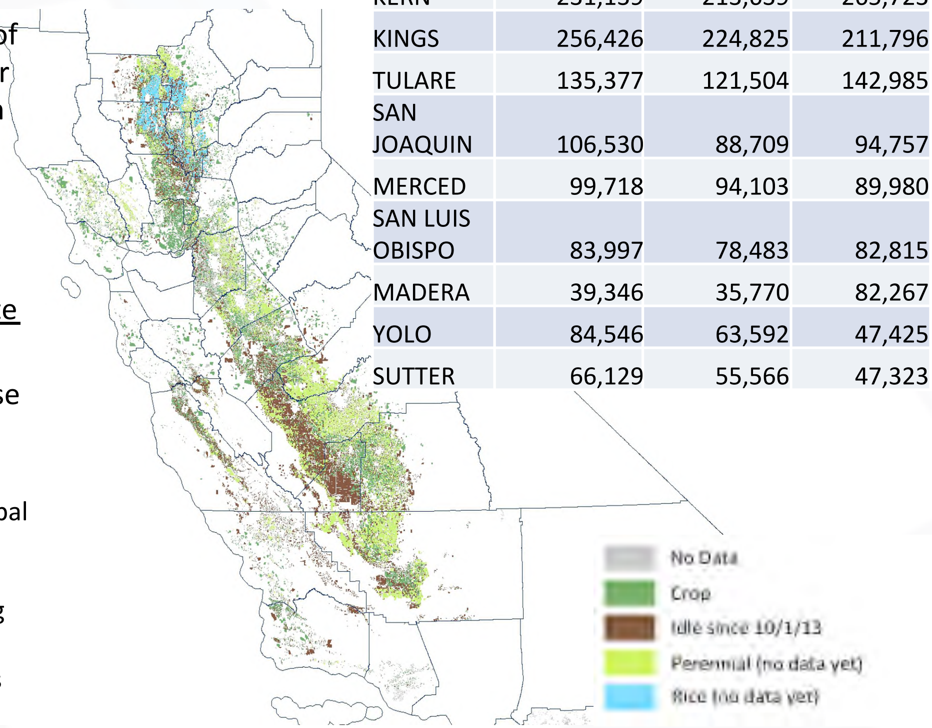
- User requirements specified by the California Department of Water Resources (CDWR) included an accuracy of +/- 25% or better, and data availability on a monthly basis, beginning in April of each year with a latency of less than 2 weeks.
- Using NEX, team developed early season classification algorithms with accuracy of >90% and met all other CDWR requirements. USDA National Agricultural Statistics Service accelerated production of their idle lands maps from satellite data by 7 months.
- Data under review by CDWR and scheduled for public release in Sept 2014.

Relevance: Shortage of water for irrigation and crop production is a principal impact of drought in the Central Valley of California, yet there has not previously been a source of timely, objective information on the extent of fallowed acreage during the main growing season to guide decision making with respect to requests for local water transfers, county drought disaster designations, or state emergency proclamations. Widely varying estimates are put forward by interested parties, though typically without objective supporting evidence.

Role of NEX / ESD Data: Use of NEX and access to Landsat/MODIS archives accelerated application prototyping by >9 months. The project is mapping fallowed agricultural lands using satellite data from MODIS, Landsat TM, Landsat ETM+, and Landsat OLI, as well as data from the DMC constellation.

Total Idle Acres by Month, Top 10 Counties in CA

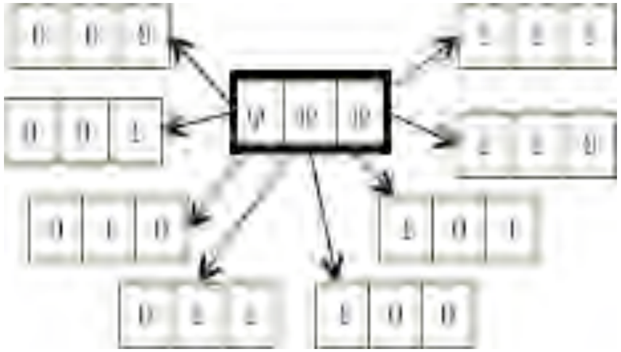
	Month (2014)		
County	March	April	May
FRESNO	387,607	344,631	410,840
KERN	231,139	213,639	265,723
KINGS	256,426	224,825	211,796
TULARE	135,377	121,504	142,985
SAN JOAQUIN	106,530	88,709	94,757
MERCED	99,718	94,103	89,980
SAN LUIS OBISPO	83,997	78,483	82,815
MADERA	39,346	35,770	82,267
YOLO	84,546	63,592	47,425
SUTTER	66,129	55,566	47,323



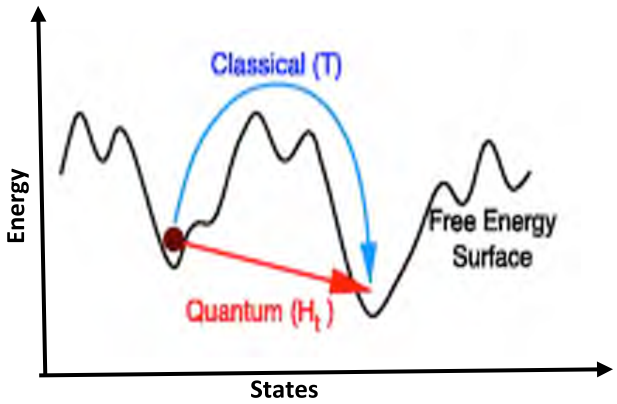
Fallowed area map for May 2014 showing acreage that has been idle since Oct 1, 2013, and county-level acreage totals for the top 10 counties in California by idle acreage through May 31.

Quantum Computing

Quantum Computing

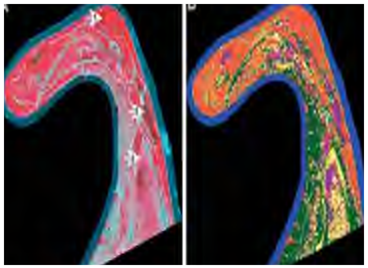


- A quantum computer uses quantum-mechanical phenomena (superposition, tunneling, entanglement, …) to solve problems!
- *Qubits* (quantum bits) can be in superpositions of 0 and 1 states at the same time (a digital bit can be in only one state: either 0 or 1) !
    - *n* qubits can represent 2^n states simultaneously.!



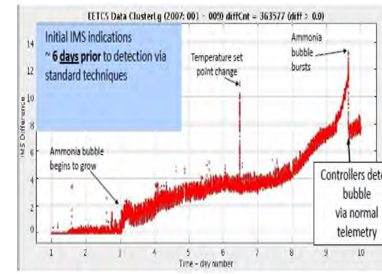
- Quantum Annealing is a technique to solve combinatorial optimization problems!
    - Find the deepest valley in a terrain!
- Provides a probabilistic answer after examining all the states !

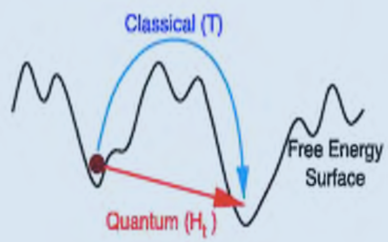
NASA and Quantum Computing



Data Analysis and Data Fusion

Anomaly Detection and Decision Making

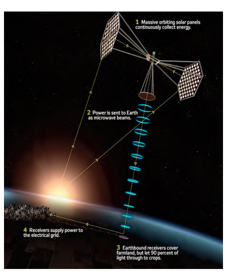


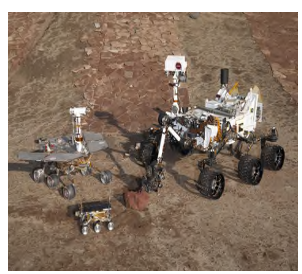


Air Traffic Management

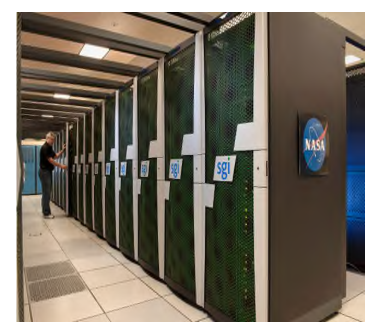
V&V and optimal sensor placement

Mission Planning and Scheduling, and Coordination



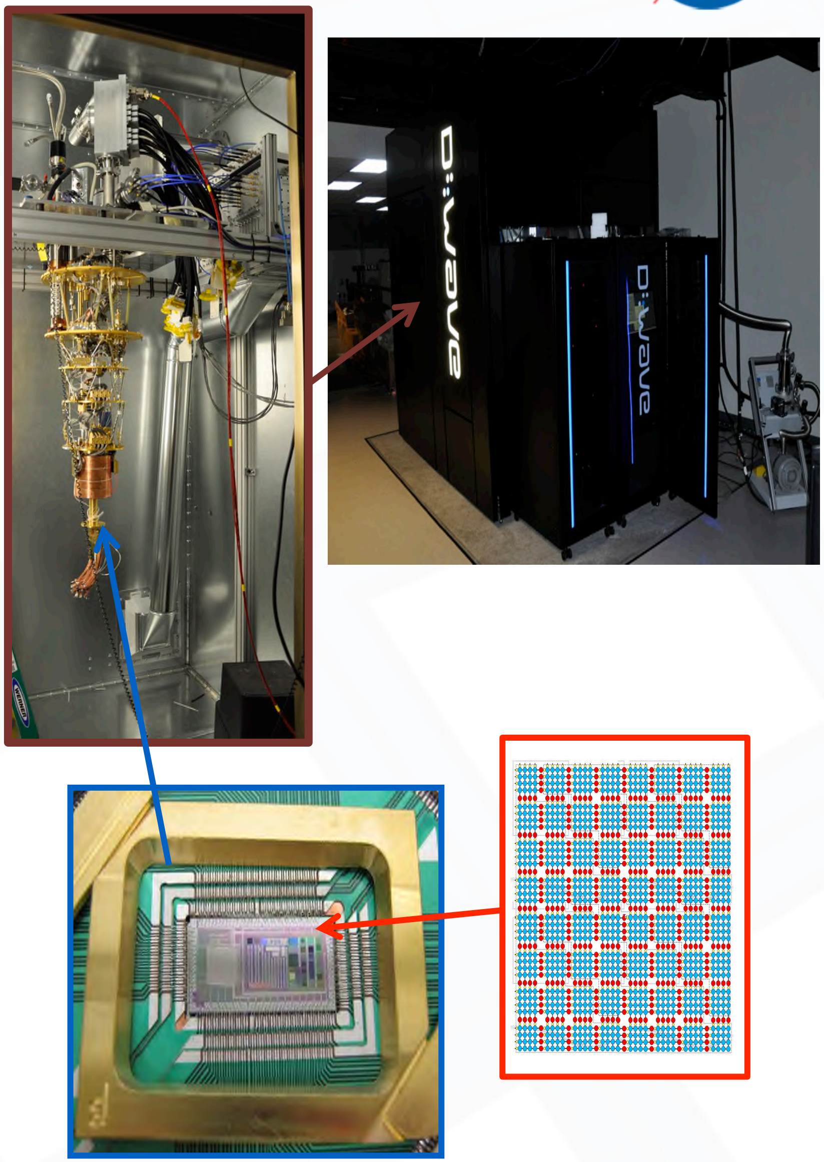


Topologically aware Parallel Computing



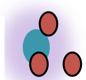
D-Wave Two™ System

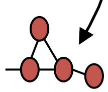
- Collaboration between NASA / Google / USRA!
- D-Wave Two system installed at NAS!
  - Vesuvius processor – 512 qubits (quantum bits – niobium superconducting loops encoding 2 magnetic states)!
  - Physical characteristics:!
    - 10 kg of metal in vacuum at 15° mK (outer space is at 3°K)!
    - Magnetic shielding to 1 nanoTesla (50,000x less than Earth's magnetic field) !
    - Isolated from the building to reducing vibration impact!
    - Uses 12 kW electrical power!



How to Program D-Wave Two

1 Map the discrete combinatorial optimization problem into QUBO, Q_{ij} (Quadratic Unconstrained Binary Optimization) form

$\alpha_{ijk} z_i z_j z_k$ 

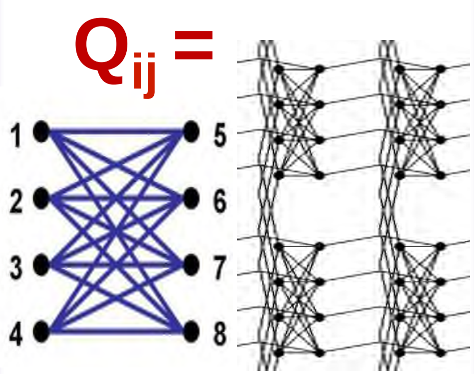
$\alpha_{ijk} y_{ij} z_k +$ 

$\beta_{ijk}\left(3y_{ij} - 2z_i y_{ij} - 2z_j y_{ij} + z_i z_j\right)$

$$E(\{z\}) = \sum Q_{ij} z_i z_j$$

2 Embed the QUBO coupling matrix in the hardware graph of interacting qubits

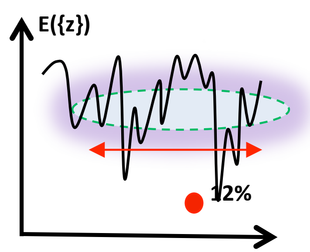
The D-Wave hardware qubit connectivity is a "Chimera Graph," so embedding methods mostly based on heuristics



Note: D-Wave provides a heuristic blackbox compiler that bypasses embedding

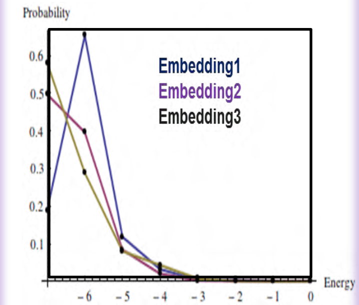
3 Initialize the system in a minimum energy state with know Q' and {z'}. Adiabatically change Q' ➔ Q while maintaining minimum energy of the system.

Solution is the configuration {z} that produces the minimum E given Q



4 \Run the problem many times and collect statistics

Use symmetries, permutations, and error correction to eliminate the systemic hardware errors and check the solutions



Set Partitioning Problem

Given a set of integers: $S = \{n_1, n_1, \dots, n_m\}$,
find two partition subsets R and T such that:

- $S = R \cup T$ and $R \cap T = \emptyset$
- $\sum_{n_i \in R} n_i = \sum_{n_i \in T} n_i$

QUBO Form:

Consider $z_i \in \{0, 1\}, i = 1, \dots, m$ such that if

- $z_i = 0$ then $n_i \in R$
- $z_i = 1$ then $n_i \in T$

Find z_i for which

$$min\, E(S) = \left(\sum_{i=1}^{m} n_i \quad (2z_i - 1) \right)^2$$

$$\dots$$

$$= constant\ terms + \sum_{\forall\ (i,j) i \neq j} Q_{ij}(n_i, n_j) z_i z_j$$

NASA FOCUS

Assess the feasibility of utilizing quantum systems, such as the D-Wave Two, to solve NASA's optimization problems.

Current work:

- Planning and scheduling: small problems inspired by Rover mission planning
- Fault diagnosis in Electrical systems
- Quantum program compilation: Optimization of mapping and embedding processes for quantum annealing hardware
- Protein folding
- Combinatorial geometry optimization: Polyhedron terahedralization for CFD mesh generation

Questions?