UPDATE ON ASCAC SUBCOMMITTEE DOCUMENTING ASCR IMPACTS

Paul Messina
Argonne National Laboratory

December 12, 2018

OUTLINE

- High-level story
- Lessons learned
- Selected challenges
- Highlights and overviews of activities funded by ASCR
- Selected achievements

HIGH-LEVEL STORY: ASCR-FUNDED RESEARCH HAS IMPACTED SIGNIFICANTLY ALMOST ALL OF THE SUPERCOMPUTING ECOSYSTEM

- A key feature that enabled the impact is the integration of the elements through collaboration among national labs, academia, and industry
  - Computational and data science, development of application codes
  - Mathematics and computer science research in collaboration with domain scientists and facilities, production of software libraries
  - Training and workforce development
  - Computing facilities that enable previously intractable research and serve as proving-grounds for advancing supercomputing and are broadly available
  - Networks that link computers, data, and people, including beyond the DOE national labs
  - Collaboration with vendors of computers and network technologies on architecture, software, experimenting with new technologies
  - Collaborations with industry, including access to DOE/SC supercomputers

EXAMPLES OF LESSONS-LEARNED WE HAVE IDENTIFIED

- Laboratory- and university-based research has been a source of innovation, both in fundamental research and applications, workforce development
- Integration of the ecosystem elements has been essential for successfully transitioning to major technology changes in
  - Computer architectures
  - Computational science and engineering applications
  - Multi-institution collaborations
  - Workflows, data-centric computing
- Open software policy and type of license have played important roles

U.S. DEPARTMENT OF ENERGY | Office of Science

EXAMPLES OF LESSONS-LEARNED WE HAVE IDENTIFIED

- Unsolicited proposals can and have led to accomplishments with big impact, sometimes over a period of decades
- Block/base funding over long periods likewise has been instrumental in achieving numerous breakthroughs
  - Base funding is essential for long-term activities, including software sustainability
- Multi-agency funding and collaborations have been fruitful, sometimes essential for involving players with the necessary expertise and facilities

SELECTED CHALLENGES

- A significant challenge going forward for ASCR (and all of international computational science) centers on more explicit emphasis on high-quality, high-performance, reusable, sustainable software itself, in order to encapsulate research advances in math, computer science and applications and enable next-generation advances toward predictive science
  - For next-generation computational science, it is not adequate for people to develop poor-quality, insufficiently tested software … and then to base scientific discoveries on this software
- Increased efforts should be supported to change the culture of computational science to fully acknowledge the important role that software plays as a foundation for CSE collaboration and scientific discovery
  - The ECP has invested in such efforts but it will end in 2021

SELECTED CHALLENGES

- Sustainability of software developed with ASCR support that is widely used for DOE missions
  - Evolution of implementations to new computer architectures when warranted
  - Inclusion of new techniques and algorithms that support new application needs or map more effectively on new computer architectures

HIGHLIGHTS AND OVERVIEWS OF AREAS IN VARIOUS STAGES OF PREPARATION

- **3.1 Computational science**
  - SciDAC story
  - Selections from 2008 Breakthroughs report, e.g.,"Modeling the molecular basis for Parkinson's Disease"
  - Taming Fire Through Extreme Scale Combustion Simulation for More Efficient Engines
  - Calculation of all possible isotopes within the limits of nuclear existence (UNEDF)
- **3.2 Applied Math**
  - Applied Analysis overview (non-numerical math)
  - Applied Math overview (numerical methods for PDEs)
  - Advanced discretization methods
  - AMR
- **3.3 Software libraries and frameworks**
  - The PACKs
  - PETSc
  - ADIFOR

HIGHLIGHTS AND OVERVIEWS OF AREAS IN VARIOUS STAGES OF PREPARATION

- **3.4 Computer science**
  - Message passing overview
  - Data and I/O Overview overview (work in progress)
  - Visualization overview (work in progress)
  - Globus technologies for Global research federation
  - Fastbit
  - ADIOS
  - Operating systems (work in progress)
- **3.5 Computer Architecture**
  - No draft yet, only various inputs
- **3.6 Facilities**
  - ASCR computing facilities overview
  - Science DMZ
  - Software defined networks
  - Network protocols

HIGHLIGHTS AND OVERVIEWS OF AREAS IN VARIOUS STAGES OF PREPARATION

▪4.0 Impact on Industry

- Improving everyday products
- Better combustion for power generation
- GE Taming Turbulence and achieving stability to generate fusion energy
- GE more efficient jet engines and wind turbines [might merge with above]
- Pratt&Whitney Improving aircraft engine combustor simulations
- Building a smart truck
- ESnet Taps Ciena for 400G Research and Education Network
- HPC-driven fuel well discovery

▪5.0 Impact on workforce

- CSGF work in progress)
- Training programs (work in progress)

RESEARCH TO APPLICATIONS: REACTING FLOW CASE STUDY

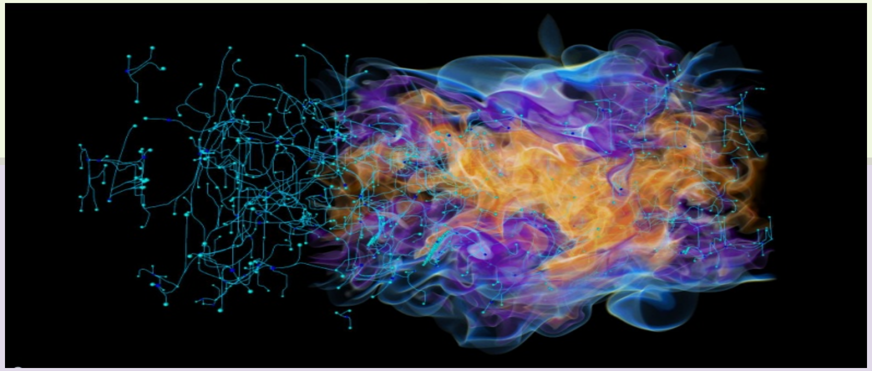
- ASCR research efforts starting in the late 1980s (base research program)
  - Applied analysis: low-Mach number asymptotics for reacting fluid flows
  - New discretization methods for fluid dynamics:
    - High-resolution methods for low-Mach number flows
    - Structured-grid adaptive mesh refinement (AMR)
    - Cut-cell methods for complex geometries
    - Semi-implicit methods for multicomponent transport and reactions, with application of robust software for ODEs to represent complex chemistry
  - Multigrid solvers for elliptic / parabolic PDE on AMR grids
- All of these research activities were undertaken with combustion as a motivating application. By the early 2000s, researchers were combining these components to simulate laboratory-scale combustion with detailed comparison to experiments (HPCC, SciDAC)
- This approach was extended to simulate type 1A supernovae (SciDAC)
- These methodologies form the core of the numerical approach for several ECP applications projects, supported by the AMREx co-design center of the ECP

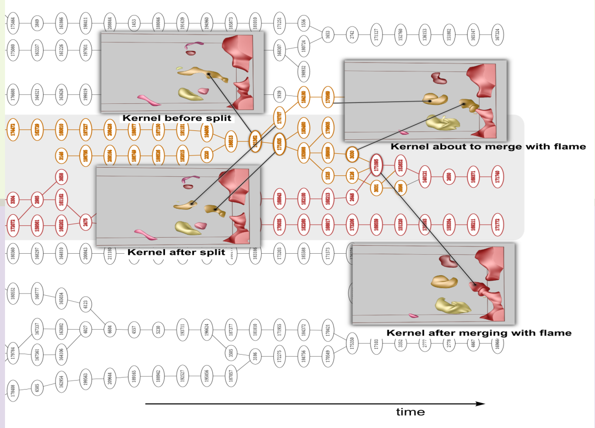
COMBUSTION SIMULATION TOWARDS EFFICIENT ENGINES ENABLED BY SCIDAC, ECP AND INCITE

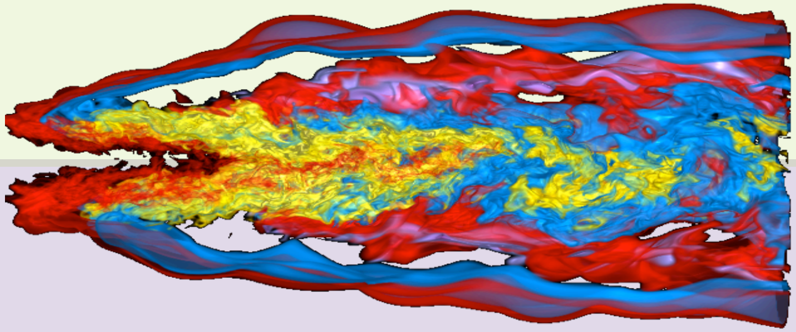
SCIDAC and INCITE

SDAV: FastBit for fast query, ADIOS I/O , Topological Feature Extraction and Tracking

Ultrascale Viz:
In situ volume and particle visualization





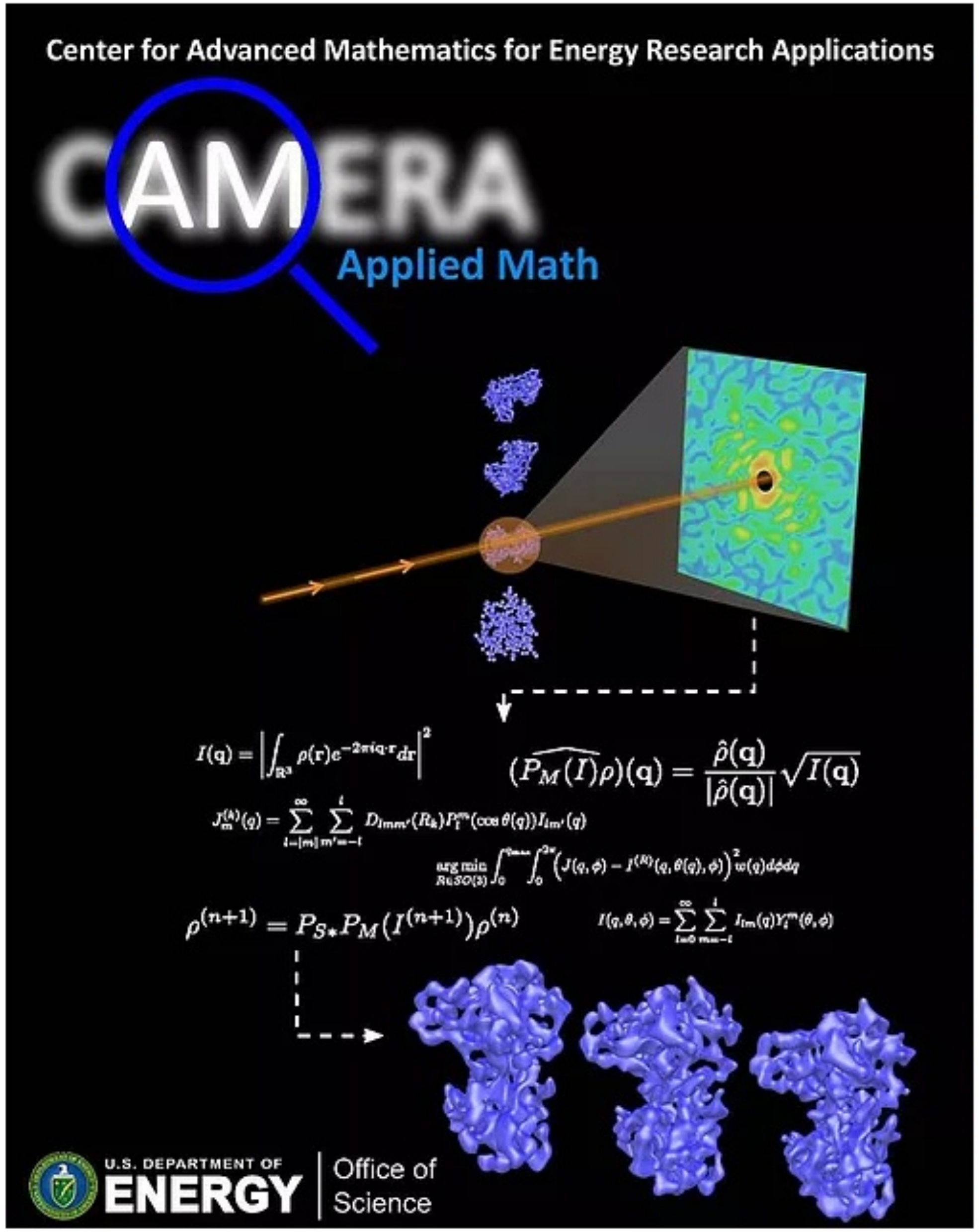


ECI: Combustion Co-Design: Legion PM, AMR, In situ Adjoint UQ & Analytics workflow

ECP Pele combustion: AMR codes with multi-physics and geometry for high-fidelity engine simulations

ECI & ECP

ASCR RESEARCH AND EXPERIMENTAL SCIENCE



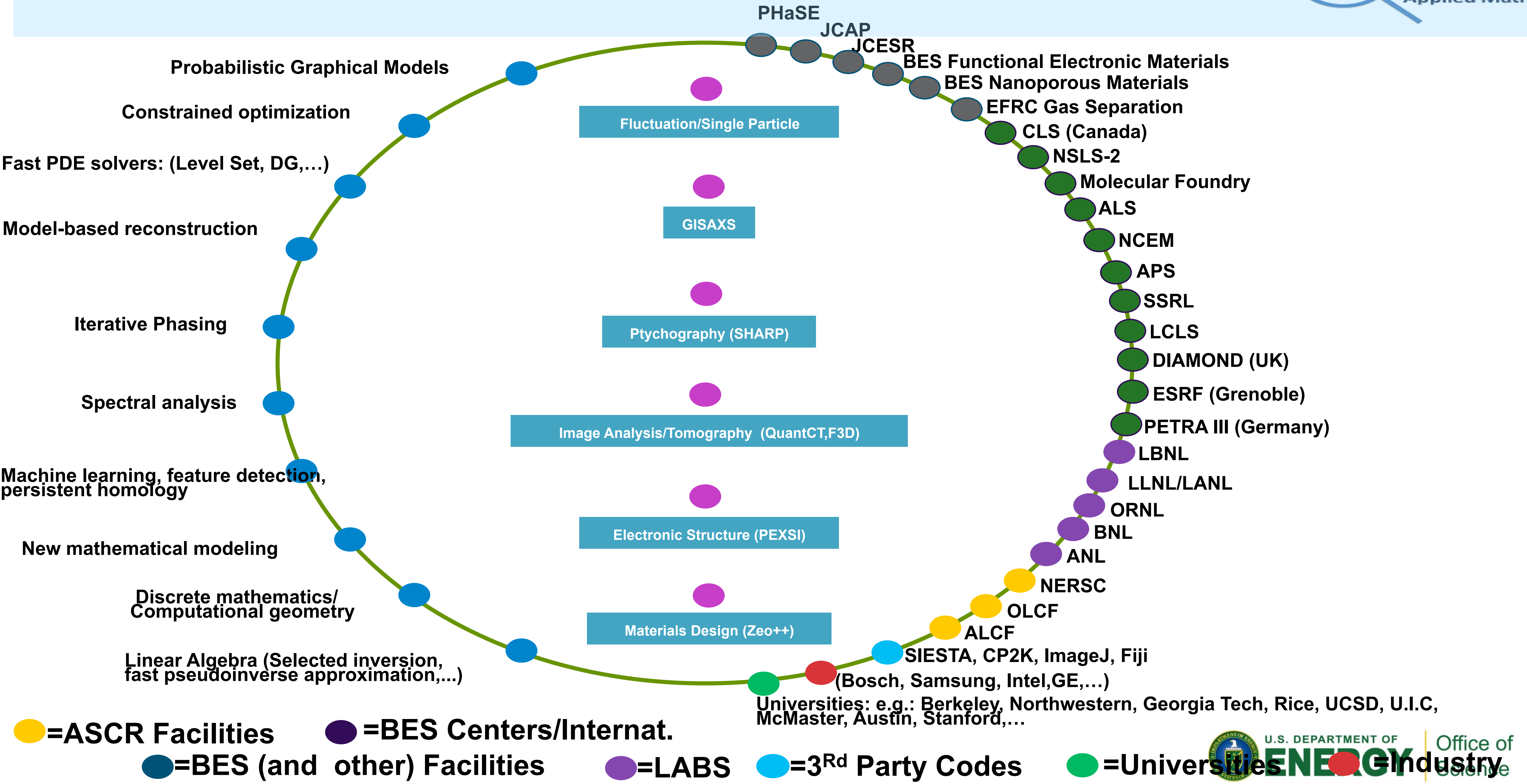
The Center for Advanced Mathematics for Energy Research Applications (CAMERA) is an integrated, cross-disciplinary center aimed at inventing, developing, and delivering the fundamental new mathematics required to capitalize on experimental investigations at scientific facilities.

Jointly funded by the Office of Advanced Scientific Computing Research (ASCR) and the Office of Basic Energy Sciences (BES) within the US Department of Energy's Office of Science, CAMERA identifi□es areas in experimental science that can be aided by new mathematical insights, develops the needed algorithmic tools, and delivers them as user-friendly so□ware to the experimental community.

Application areas include X-ray scatt□ering and ptychographic imaging, reconstruction and analysis of imaged materials, chemical informatics for analysis of crystalline porous materials, fast methods for electronic structure calculations, reconstruction methods for emerging experiments at X-ray free electron lasers, autonomous control of experiments, and real-time streaming for automatic feedback and reconstruction.

CAMERA fosters collaborations across the DOE landscape. Partners include the five DOE light sources (ALS, APS, LCLS, NSLS-II, and SSRL), with growing interactions with DOE nanoscience and supercomputer centers, other national laboratories, international partners, and universities.

CAMERA: Mathematics and Collaborations



BACKGROUND OF SCIDAC

- Context
  - Major change in computer architectures in 1990s—parallel computers
    - Software technologies needed to use parallel computers still in development
    - New mathematical algorithms required for parallel computers
    - Scientific codes not poised to take advantage of new computing technology
- Scientific Discovery through Advanced Computing
  - Support teams of computational scientists, computer scientists and applied mathematicians to create new generation of codes
  - Support development of software technologies needed by scientific applications to enable efficient and effective use use of parallel computers
  - Support development of new algorithms need to optimize performance of scientific and engineering applications on parallel computers

IMPACT OF SCIDAC

- General Observations
  - COV Report 2014:
    - "SciDAC remains the gold standard nationally and internationally for fostering interaction between disciplinary scientists and HPC."
  - Has become standard model in Office of Science for advancing the state-of-the-art in applications for computational science and engineering
- Advancements Enabled by SciDAC
  - Development of new generation of science and engineering codes along with many accompanying scientific advancements
    - BER, BES, FES, NP, and HEP
  - Development of many new software technologies for parallel computing systems
  - Development of many new mathematical algorithms for parallel computing systems

MATHEMATICS IS CENTRAL TO THE HPC ECOSYSTEM

- Mathematical models: describing science in a way that is amenable to representation of a computer
- Approximation / discretization: replacing an infinite number of degrees of freedom with a finite number of degrees of freedom
- Solvers and software: accurate and efficient representation on high-end computers

Deep mathematical questions are associated with each of these topics, and they strongly interact with one another. A major accomplishment for ASCR has been resolving these questions for a broad range of important applications.

EXAMPLES OF MATHEMATICAL ACCOMPLISHMENTS

- Mathematical models: Low Mach number asymptotics for reacting fluid flow, well-posedness of shock hydrodynamics
- Approximation / Discretization: high-order methods for PDE, adaptive mesh refinement, level-set methods for front propagation, particle methods for fluids and plasmas, simulations in complex geometries
- Solvers and software: broadly-used solvers and frameworks for HPC simulation

Scientific impacts both in DOE mission areas (combustion, fusion, defense applications, …) and beyond (aerodynamics, astrophysics / cosmology, semiconductor process design…)

HOW HAS MATH BEEN SUCCESSFUL?

- There are tradeoffs up and down the stack of models, discretization, and software, requiring actors that can look beyond their stovepipes
- Experts on crosscutting mathematical technologies working to solve a specific scientific problem often find it leads to ideas and tools that are applicable to a broad range of problems
- Long lead times (10 years or more) from first appearance of a new idea to its broad acceptance in the scientific community requires patience from funding sources

THE PACKS FOR DENSE LINEAR ALGEBRA
EISPACK, LINPACK, THE BLAS, LAPACK, AND SCALAPACK

- EISPACK 1972
  - Effort to translate Algol to Fortran
- LINPACK 1978
  - Designed to use Level 1 BLAS (vector ops)
- Level 2 & 3 BLAS 1988 & 1990
  - Higher level of granularity for matrix operations; GEMV and GEMM
- Sca/LAPACK 1992
  - Refactored to perform block algorithms for performance on “modern” architectures

MathWorks

“LINPACK and EISPACK spawned MATLAB, which changed my life. Today MATLAB has millions of users worldwide. MathWorks, the MATLAB company, has over 4,000 employees in at least a dozen countries and is poised to have sales of over one billion dollars this year.”

- Cleve Moler, MathWorks

- Matlab today has replaced LINPACK & EISPACK with LAPACK
- LAPACK is one of the most successful and influential software packages produced under the DOE‘s ASCR program
  - Used by 13+ HW vendors in their software libraries
  - Used by 17+ SW companies in their products

- Architectural changes have come every decade or so, thereby creating a need to refactor the software to the emerging architectures
- PACK's became joint efforts with DOE ASCR and other organizations
- PACK's widely embraced and helped form standards (e.g., BLAS, IEEE floating point standard, and MPI)
- DOE ASCR provided a stable organization with decades-long lifetime which can maintain and evolve the software
  - The ASCR DOE laboratories are critical elements in the development and support of HPC software
  - Funding streams of an academic institution or the NSF supercomputing centers is too unstable
- Important mechanism in this story was the existence of a block grant to do cutting-edge computer science at DOE laboratories

– Success of PACKs (and other DOE SW) in part based on the choice of the software license used

EXEMPLAR COMPUTER SCIENCE RESEARCH

- Programming models, languages, and tools
  - PGAS languages and MPI standards
  - Vectorizing and parallelizing compiler technology
  - Scalable debugging and performance tools

- System software and libraries
  - Parallel I/O libraries
  - Autotuning software
  - Scalable visualization tools
  - Operating system scaling

MESSAGE PASSING LIBRARIES

- In the late 80s and early 90s, there was no standard for doing message passing in a parallel program.
  - Each vendor of parallel systems created its own message-passing library that had its own API and was available only on that vendor’s systems
  - This led to application codes that were not portable
- Research began on two main areas: the development of “portable” message-passing libraries and on what functionality and features a message-passing library must provide to meet application needs, provide ease of programming, and the highest performance.
- ASCR-funded projects were prominent in this early stage
  - PVM, p4, Chameleon, Express
  - PVM had thousands of users all over the world. PVM worked on all platforms, won an R&D 100 Award in 1994

MESSAGE PASSING LIBRARIES

- The need for a standard, portable message passing library led to the creation of the MPI Forum in 1992
  – ASCR funded participation in the MPI forum for many researchers
  – An early commitment to producing a reference implementation (MPICH) by a group from Argonne helped demonstrate that the implementation of MPI on multiple platforms was feasible.
  – The release of the MPI standard combined with the availability of the MPICH implementation that worked on all platforms led to the quick adoption of MPI by both users and vendors
- MPICH is widely used all over the world
  – Many vendors use MPICH as the basis for their own vendor-tuned MPI implementations (e.g., Intel, Cray, IBM for the Blue Gene series, Microsoft)
  – 8 of the top 10 systems in the June 2018 edition of the Top500 list use MPICH-based MPI implementations
  – MPICH won an R&D 100 Award in 2005
  – MVAPICH, a popular MPI implementation from Ohio State Univ for InfiniBand systems is derived from MPICH
- Open MPI is another widely used MPI implementation. It has had multiple funding sources (ASCR, NNSA, NSF).

ECOSYSTEM OF FACILITIES IS A SCIENTIFIC ENABLER

- Interconnected system of facilities provides unique capabilities
  - Facilities are more than just big computers
    - Software such as schedulers, debuggers, file system, data archives, math libraries, data analysis tools, etc.
    - Buildings, power & cooling infrastructure, operations & maintenance
    - Networks linking experimental, observational, and computational facilities
  - People who understand tools, computing, science, methods
    - User support, Computational and Data Liaisons, AI specialists
  - Bring all those resources to bear on big problems
    - Energy, Medicine, National Security, Basic & Applied Sciences, Engineering
    - Industrial competitiveness, Global challenges

DOE OFFICE OF SCIENCE COMPUTE FACILITIES

- The first supercomputers were purchased by specific programs or laboratories to address specific programs
  - Defense Programs
  - Fusion
  - High Energy Physics
  - Uranium Enrichment
- Transitioned the National Magnetic Fusion Energy Computing Center into NERSC to provide a center available to all DOE science programs.
- In 1992 created High Performance Computing Research Centers at Oak Ridge and Los Alamos to explore and scale parallel architectures and algorithms.
- In 2004 created Leadership Computing Facilities to address a broader mission of delivering the most powerful systems to government, academia, and industry
- In 2015 worked with NNSA to create the Exascale Computing Project to develop needed applications, software and technology to get to exascale systems.

COLLABORATION: VENDORS, INSTITUTIONS, AGENCIES

- Integration of research and production (e.g. Centers of Excellence)
- Synergistic relationship between the Labs, vendors, agencies
  - Co-design of systems and applications to get the requirements right
    - Government projects to create early parallel systems
    - Rapid feedback loop for innovation
    - Worked with application scientists, and industry to discover the possible
    - Then production versions were deployed
  - Examples:
    - Intel Touchstone Delta ➢ Intel Paragon, Sandia Red Storm ➢ Cray XT line
- Critically important to maintain a path to funding for important ideas
- Also important to both maintain long-term relationships with vendors and allow for innovation through competition

LARRY SMARR ON THE IMPACT OF DOE HPC LEADERSHIP

- “If you go back 30 years, you find the equivalent of what’s going on now. Back then, interest in the DOE supercomputing centers transferred to the university research community. When I wrote my proposal for NCSA in 1983, there were no multi-disciplinary nationally available supercomputers in the university sector. I proposed explicitly cloning the Lawrence Livermore National Lab system for NCSA and even used CTSS, the Cray Time Sharing System developed at LLNL. Sid Karin cloned MFECC – now known as NERSC – for the San Diego Supercomputer Center. DOE had the ideal combination of supercomputers and mass storage. “
- Smarr added. “By early adoption of the DOE approach, NSF took that proven approach in computational science and engineering and moved it in the university sector, then they trained up a good deal of American industry in supercomputing.”

FACILITIES THAT PROVIDE WORLD CLASS COMPUTING CAPABILITY

- The DOE Leadership Computing Facilities at Argonne and Oak Ridge develop and use the most advanced computing systems for the open science community, including industry
  - An important feature is that they each have teams of computational scientists and engineers who work intensively with key science teams to enable breakthrough computations
- NERSC is the high performance scientific computing resource for researchers directly supported by the Department of Energy's Office of Science (SC)
  - Its workload represents the wide variety of research performed by its users, including simulations that run at the largest scales available on the center's supercomputers and data analysis of massive datasets in support of DOE SC experimental facilities

ALLOCATION PROGRAMS THAT PROVIDE ACCESS TO WORLD CLASS COMPUTING CAPABILITY FOR SCIENTIFIC DISCOVERY

- The Innovative & Novel Computational Impact on Theory & Experiment (INCITE) program is the major means by which the scientific community gains access to the LCFs
  - Open to researchers from academia, government labs, and industry
  - Very large allocations of time are awarded
- The ASCR Leadership Computing Challenge (ALCC) is an allocation program for projects of interest to the DOE with an emphasis on high-risk, high-payoff simulations and for broadening the community of researchers capable of using leadership computing resources
  - Open to researchers from academia, government labs, and industry
  - Allocates time on all three ASCR supercomputing facilities: ALCF, OLCF and NERSC
- Allocations from both programs have yielded major achievements in computational science and engineering, including ones that earned Nobel Prizes

HIGH PERFORMANCE NETWORKS INTERCONNECT SCIENCE

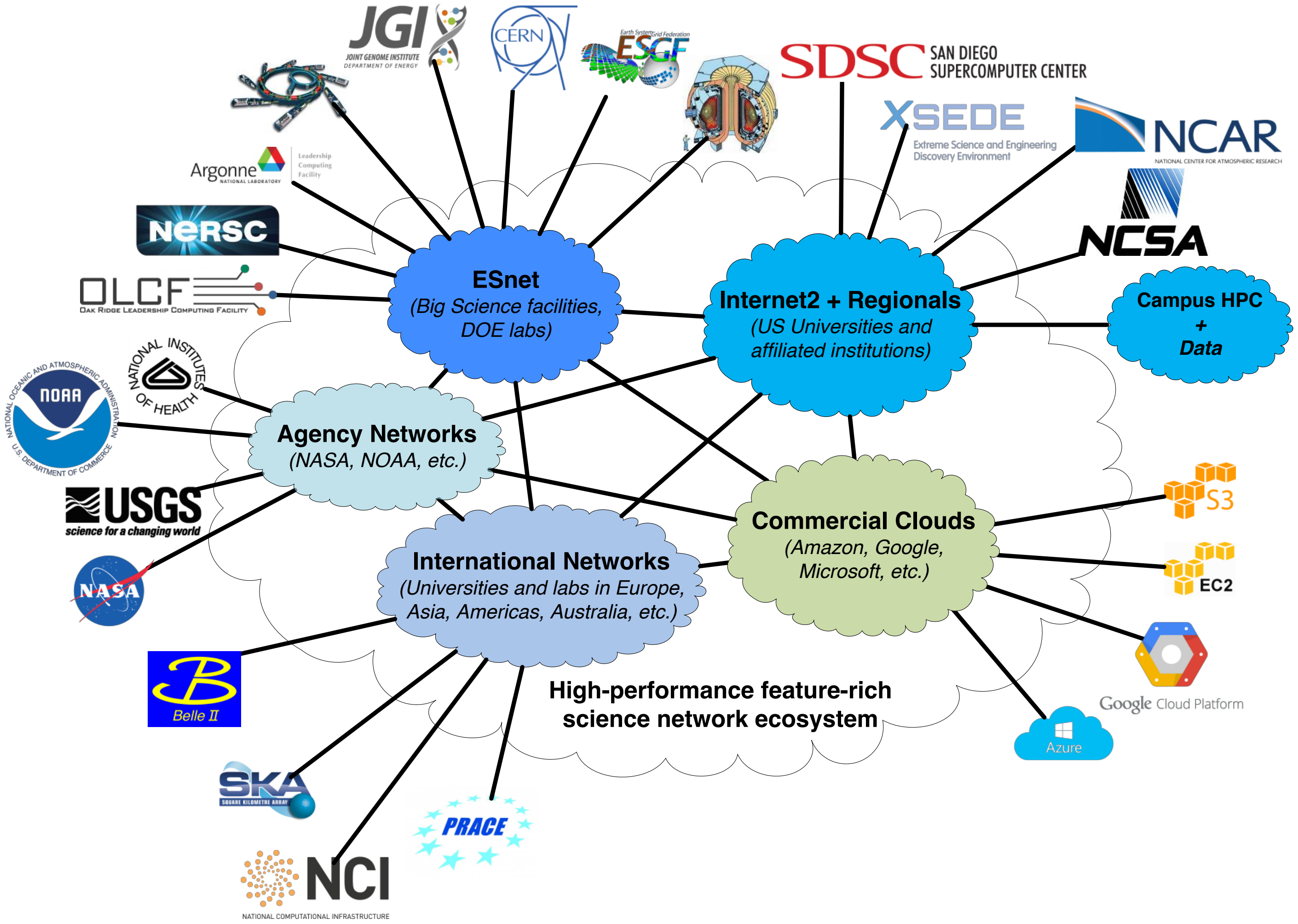
- Interconnected facilities key to scientific advancements
  - Data circulatory system for DOE
  - Data-intensive computing at HPC facilities
  - Distributed data analysis, e.g. LHC experiments
  - Support for integrating HPC with experiment and observation
- ESnet is the ASCR facility that provides high performance networking for DOE/SC
  - Global science connectivity in support of DOE/SC research
  - ESnet collaboration and leadership in networking benefits DOE users both at Labs and at other institutions globally
- National Labs have a long history of network research
  - Fix Internet congestion collapse in 1986
  - Science DMZ
  - Software Defined Networking, including OSCARS system and new ESnet6 network in development today

LARRY SMARR ON THE IMPACT OF ESNET LEADERSHIP

- Fast forward 30 years and ESnet, with federal stimulus funding, built a 100Gbps national footprint
- “They defined the Science DMZ and took it to the DOE science community. NSF has now cloned this approach through the CC-NIE program over the past three years,” Smarr said. “It’s been built out on over 100 campuses and these Science DMZs are all based on ESnet’s approach.”

ESNET CONNECTS USERS GLOBALLY

- ESnet connects and peers with regional, national and international networks
- Supports access to DoE resources for thousands of scientists, engineers, faculty, staff and students at leading academic institutions and other research organizations



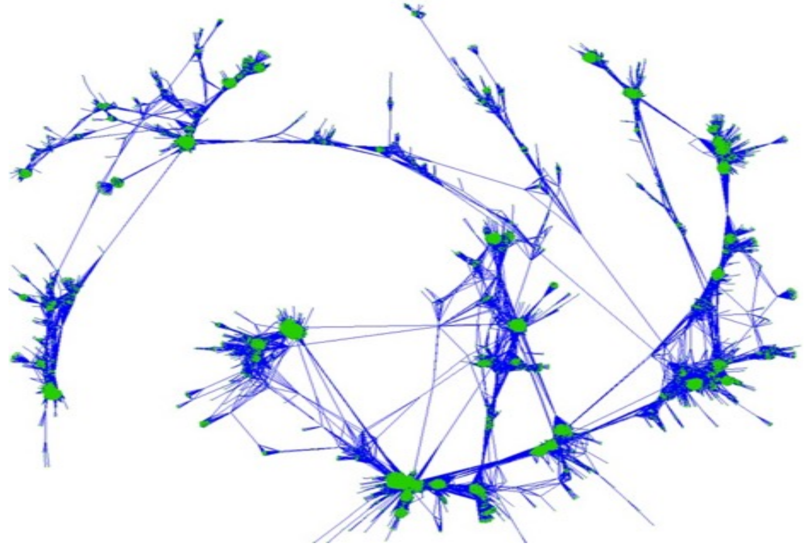
SELECTED ACHIEVEMENTS – MORE IN PROGRESS

COMPUTING GENES TO SUPPORT LIVING CLEANLY

ORNL RESEARCHERS RECOGNIZED AS GORDON BELL WINNER FOR BREAKTHROUGH GENOMIC DATA SCIENCE

The Science

An ORNL team led by computational systems biologist Dan Jacobson and computational scientist Wayne Joubert developed a genomics algorithm capable of using mixed-precision arithmetic to attain a speedup of more than 20,000-fold over the previous state of the art. On Summit, the team's Combinatorial Metrics application achieved a peak throughput of 2.36 exaops—or 2.36 billion billion calculations per second, the fastest science application ever reported. Jacobson's work compares genetic variations within a population to uncover hidden networks of genes that contribute to complex traits, including diseases. One condition Jacobson's team is studying is opioid addiction, which was linked to the overdose deaths of more than 49,000 people in the United States in 2017.



One component of a correlation network mapping variations in single nucleotides that occur at the same location in the genome across a population. These correlations can be used to identify genetic markers linked to complex observable traits.

The Impact

Exascale-level performance allows researchers to analyze datasets composed of millions of genomes—a problem size that was previously intractable. Combining clinical and genomic data with machine learning and Summit's advanced architecture will allow researchers to gain new insight into things like the genetic factors that contribute to conditions such as cardiovascular disease, prostate cancer, Alzheimer's disease, and opioid addiction. This kind of knowledge can inform medical treatment and improve patient outcomes.

PI(s)/Facility Lead(s): Dan Jacobson
ASCR Program/Facility: OLCF
ASCR PM: Christine Chalk
Publication(s) for this work: Wayne Joubert, et al. "Attacking the Opioid Epidemic: Determining the Epistatic and Pleiotropic Genetic Architectures for Chronic Pain and Opioid Addiction," *Proceedings of SC18* (2018).

Oak Ridge National Laboratory

GORDON BELL WINNERS: SUSTAINED PERFORMANCE PRIZE

Employing Supercomputers to Combat the Opioid Epidemic
Paper Title: “Attacking the Opioid Epidemic: Determining the Epistatic and Pleiotropic Genetic Architectures for Chronic Pain and Opioid Addiction"
Prize Category: Sustained Performance Prize
Team: Oak Ridge National Laboratory

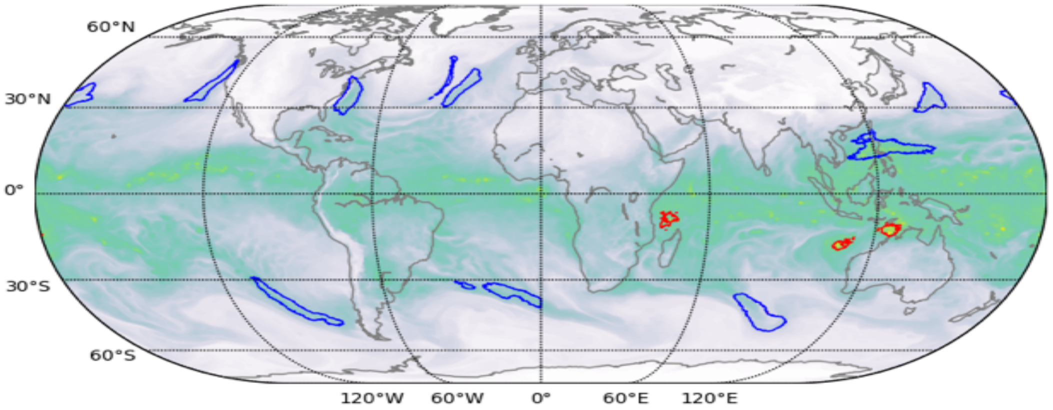
U.S. DEPARTMENT OF ENERGY | Office of Science

IDENTIFYING EXTREME WEATHER PATTERNS AT EXASCALE SPEEDS

LBNL-LED TEAM BREAKS DEEP-LEARNING EXASCALE BARRIER, SOLVES SEGMENTATION PROBLEMS IN CLIMATE SCIENCE

The Science

A team from Lawrence Berkeley National Laboratory led by data scientist Prahbat used the OLCF's Summit supercomputer to train a deep neural network to identify extreme weather patterns from high-resolution climate simulations. By tapping into Summit's multiprecision capabilities, the researchers achieved a peak performance of 1.13 exaops and a sustained performance of 0.999 exaops—the fastest deep-learning algorithm reported to date. Traditional image segmentation tasks work on three-channel red/blue/green images. But scientific datasets often comprise many channels; in climate, for example, these can include temperature, wind speeds, pressure values, and humidity. Running the optimized neural network on Summit allows the team to use all 16 channels and dramatically improve accuracy.



High-quality segmentation results produced by deep learning on climate datasets.

The Impact

Extracting pixel-level classifications of extreme weather patterns could aid in the prediction of how extreme weather events are changing as the climate warms. Though the team applied its work to climate science, many of its innovations, such as pattern-recognition algorithms, high-speed parallel data staging, an optimized data ingestion pipeline, and multichannel segmentation, lay the groundwork for future exascale deep-learning applications.

PI(s)/Facility Lead(s): Prahbat
ASCR Program/Facility: OLCF, NERSC
ASCR PM: Carolyn Lauzon, Christine Chalk
Publication(s) for this work: Thorsten Kurth, et al. "Exascale Deep Learning for Climate Analysis," *Proceedings of SC18* (2018).

OAK RIDGE National Laboratory | NERSC | NVIDIA

GORDON BELL WINNERS: SCALABILITY AND TIME TO SOLUTION

Employing Deep Learning Methods to Understand Weather Patterns
Paper Title: “Exascale Deep Learning for Climate Analytics"
Prize Category: Scalability and Time to Solution
Team: Lawrence Berkeley National Laboratory

U.S. DEPARTMENT OF ENERGY | Office of Science

Modeling the Molecular Basis of Parkinson's Disease

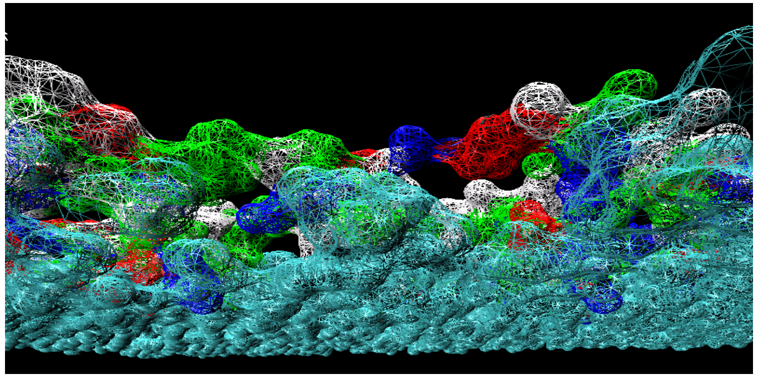
⇨ Igor Tsigelny (SDSC)
⇨ Eliezer Masliah (UCSD)
⇨ Stanley Opella (UCSD)

Accomplishment

- Elucidated the molecular mechanism of the progression of Parkinson's disease.
- Insights will help focus the search for treatment.

Broader Implications

- Provided a test bed for identifying possible therapeutic interventions through computational modeling.
- Overall approach has broad applicability to other diseases.



Support: INCITE

I.F. Tsigelny et al. 2007. Dynamics of α-synuclein aggregation and inhibition of pore-like oligomer development by β-synuclein, *FEBS J.*, **274**, 1862-1877.

Prediction and Design of Macromolecular Structures and Functions

Accomplishment

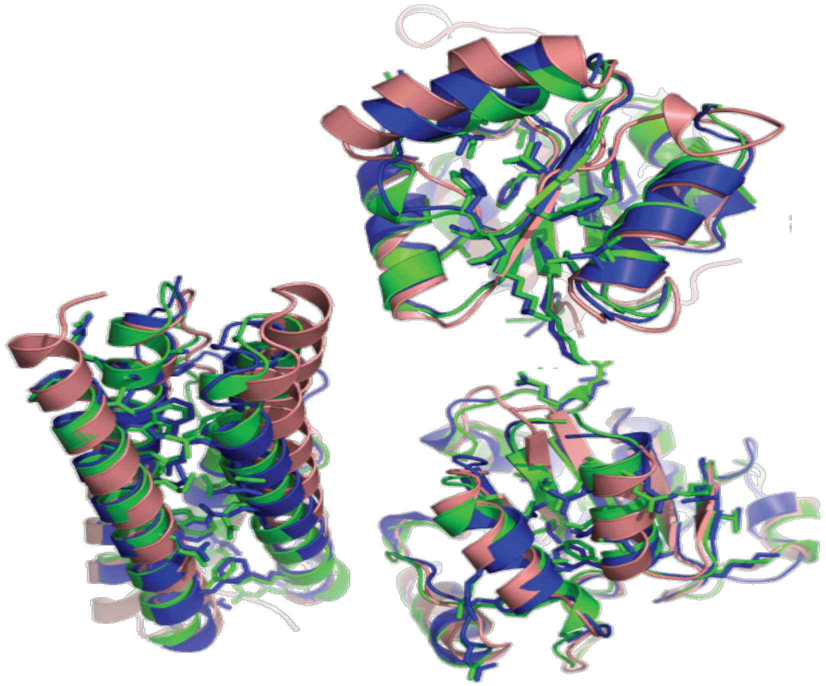
First predictions of protein structures with atomic-level accuracy.

Atomic level protein structure prediction has been a Holy Grail of molecular biology for over thirty years.

Broader Implications

Accurate protein structure prediction will greatly speed the interpretation of the genome sequence information generated in ongoing large-scale sequencing projects such as the Human Genome Project.

Support: INCITE



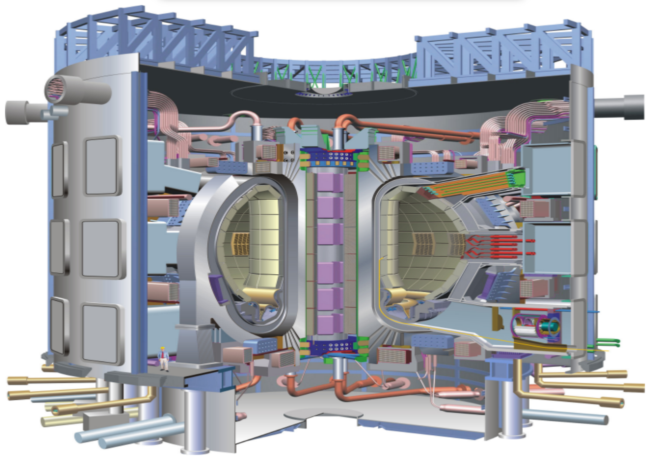
1. Röthlisberger D, Khersonsky O, Wollacott AM, Jiang L, Dechancie J, Betker J, Gallaher JL, Althoff EA, Zanghellini A, Dym O, Albeck S, Houk KN, Tawfik DS, Baker D., Kemp elimination catalysts by computational enzyme design, ***Nature***. 2008 Mar 19; [Epub ahead of print]
2. Jiang L, Althoff EA, Clemente FR, Doyle L, Röthlisberger D, Zanghellini A, Gallaher JL, Betker JL, Tanaka F, Barbas CF 3rd, Hilvert D, Houk KN, Stoddard BL, Baker D., De novo computational design of retro-aldol enzymes, ***Science***. 2008 Mar 7;319(5868):1387-91.
3. Qian B, Raman S, Das R, Bradley P, McCoy AJ, Read RJ, Baker D. High-resolution structure prediction and the crystallographic phase problem, **Nature**. 2007 Nov 8;450(7167):259-64. Epub 2007 Oct 14.

U.S. DEPARTMENT OF ENERGY | Office of Science

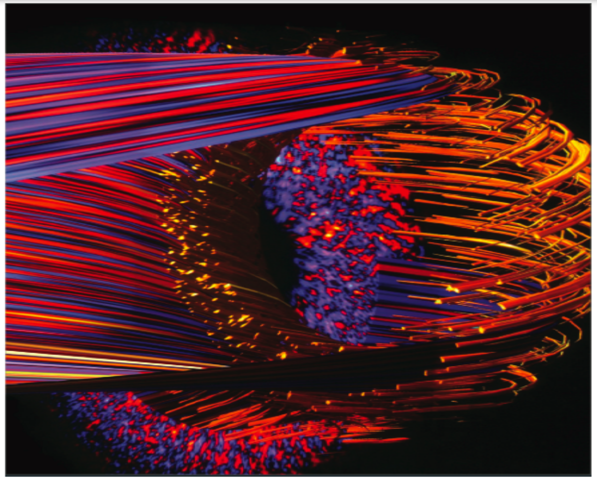
New Insights from LCF-Enabled Simulations of Global Turbulence in Fusion Systems

PPPL, UCI, ORNL, Columbia, UCSD

ITER



FUSION TURBULENCE SIMULATION



Accomplishment

3D simulations of unprecedented resolution led to a fundamentally new understanding of thermal energy loss in magnetically confined toroidal plasmas, such as in fusion tokomaks.

Broader Implications

Critical to the design of ITER.

⇨ Size and cost depend on balance between losses and self heating.

Support: FES Base Program, SciDAC, and INCITE

1. Z. Lin, I. Holod, L. Chen, P. H. Diamond, T. S. Hahm, and S. Ethier, *Phys. Rev. Lett.* **99**, 265003 (2007)
2. Wang W.X., Hahm T.S., Lee W.W., et al., *Phys. Plasmas* **14**, 072306 (2007)

High-Transition Temperature Superconductivity

D. Scalapino (UCSB)
T. Maier, P. Kent, and T. Schulthess (ORNL)
M. Jarrell and A. Macridin (University of Cincinnati)
D. Poilblanc (Laboratoire de Physique Théorique, CNRS and Université de Toulouse)

Accomplishment

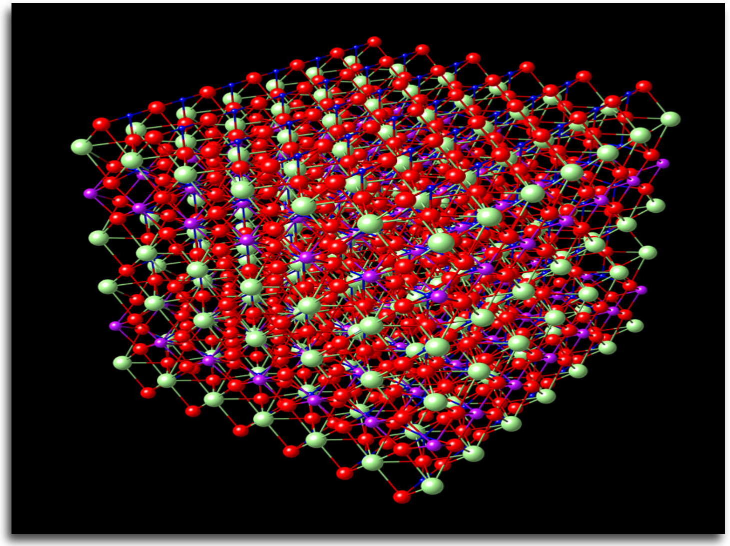
- Given new numerical techniques and leadership-class resources, 2D Hubbard Model solved computationally.
- Proved model does in fact describe high-temperature superconductivity.
  - ⇨ *Settled a debate that raged for two decades.*

Broader Implications

- Provides a deeper understanding of high-T superconductivity.
- Key step toward the identification of high-T superconducting materials, *the Holy Grail of superconductivity research*.

Support: BES Base Program, SciDAC

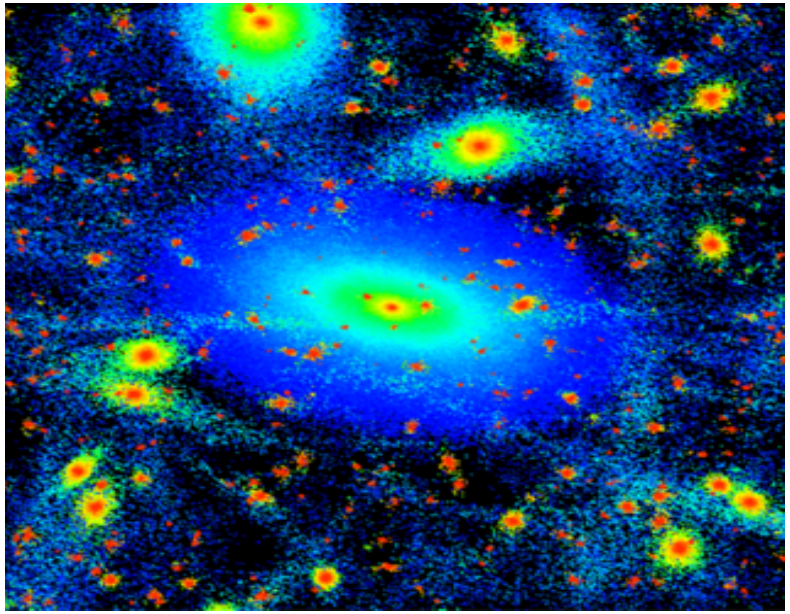
Maier, T.A., A. Macridin, M. Jarrell, and D. J. Scalapino, "Systematic analysis of a spin-susceptibility representation of the pairing interaction in the two-dimensional Hubbard model" *Phys. Rev. B*. **76**, 144516 (2007).
Maier, T.A., M. Jarrell, and D. J. Scalapino, "Spin susceptibility representation of the pairing interaction for the two-dimensional Hubbard model" *Phys. Rev. B*. **75**, 134519 (2007).



Model of a $YBa_2Cu_3O_7$ high-temperature superconductor crystal

Via Lactea II: A Billion Particle Simulation of the Dark Matter Halo of the Milky Way Galaxy

⇨ Piero Madau (UCSC)
⇨ Juerg Diemand (UCSC)
⇨ Michael Kuhlen (IAS)
⇨ Marcel Zemp (UCSC)



<u>Accomplishment</u>

Largest simulations ever performed of the formation of the dark matter halo of the Milky Way galaxy.

⇨ *Resolved and predicted structure at small scales.*

<u>Broader Implications</u>

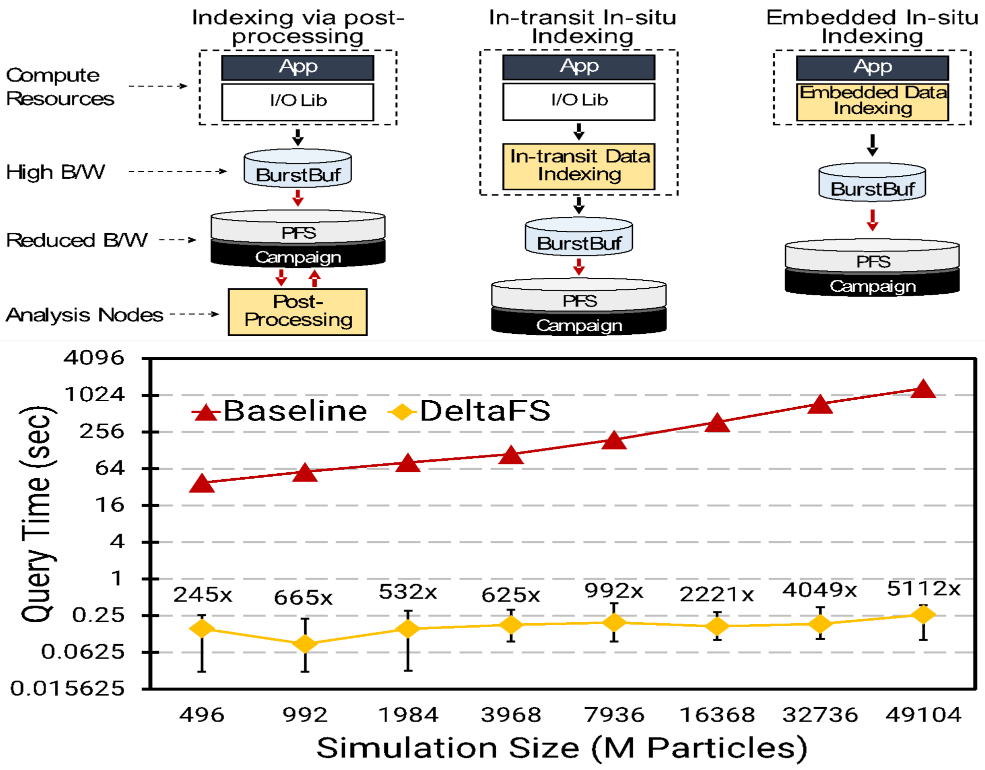
Structures predicted for small scales have observable gamma-ray signatures for certain classes of dark matter candidates.

✶ *Simulations may play an important role in identifying the dark matter in the Universe.*

Support: INCITE

Diemand, Kuhlen, Madau, Zemp, Moore, Potter, and Stadel 2008, Clumps and Streams in the Local Dark Matter Distribution, *Nature*, in press.

VPIC IN SITU ANALYSIS CAPABILITY FOR KINETIC-PLASMA SIMULATIONS



Indexing during post-processing is becoming increasingly time-consuming for exascale applications. In-transit indexing avoids post-processing but requires extra resources to process data in-situ. The DeltaFS embedded in-situ indexing shifts from using dedicated to using spare resources on the compute nodes. The result is an improvement of over 3 orders of magnitude in query speed.

Scientific Achievement

The Mochi project has developed a new in situ analysis capability for VPIC, a plasma-kinetic simulation code, enabling a 5000x speedup in data discovery.

Significance and Impact

Provides more insight to scientists supporting the Stockpile Stewardship Program (SSP).

Research Details

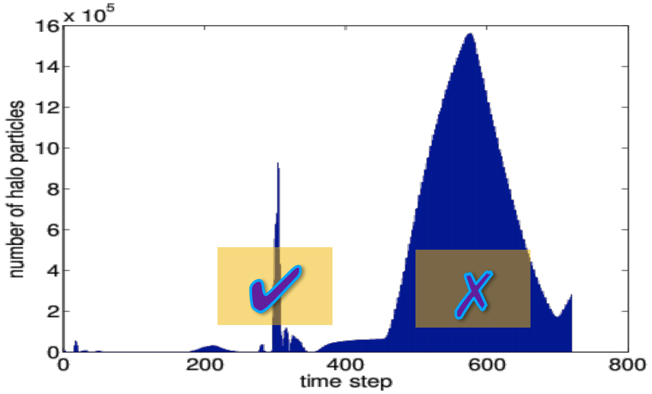
- VPIC is a first-principles, kinetic-plasma simulations code frequently run at large scale on Advanced Technology Systems, e.g., Trinity (ATS-1).
- VPIC simulations often model trillions of particles and require months of continuous run time.
- The experimental DeltaFS file system allows in-situ data indexing to efficiently search for particle data.

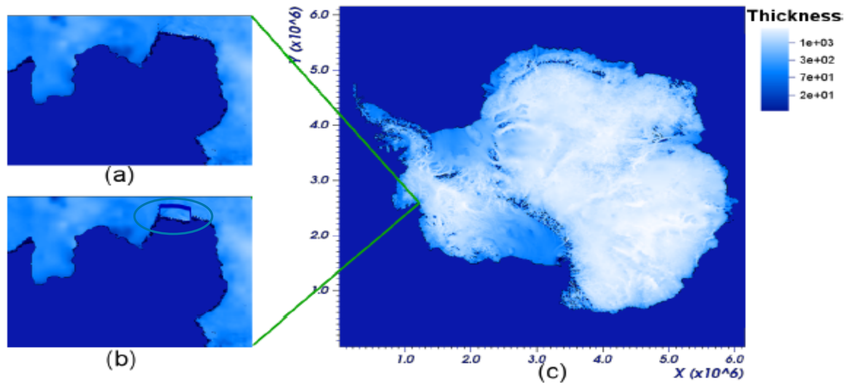
Wired Magazine highlighted this work, including it's world record file creation rate, in https://www.wired.com/story/this-bomb-simulating-us-supercomputer-broke-a-world-record/

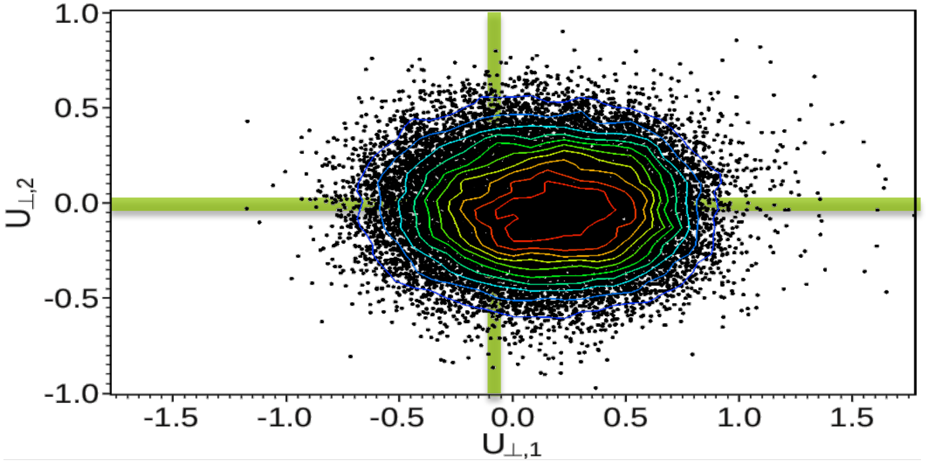
Los Alamos NATIONAL LABORATORY EST. 1943 · The HDF Group · Argonne NATIONAL LABORATORY · U.S. DEPARTMENT OF ENERGY · Office of Science

FAST QUERY FOR DATA DRIVEN SCIENCE

- **Challenge**: given a large dataset, quickly find records satisfying user-specified conditions
  - Example: count the particles far away from center of an accelerator beam
- **Solutions**
  - **Algorithmic research**: developed new indexing techniques and a new compression method, achieved 10-100 fold speedup compared with existing methods
  - **Efficient software implementation:** available open source from http://sdm.lbl.gov/fastbit/ (>25000 downloads), received a R&D 100 Award
- **Enabled science**
  - **Particle Accelerator modeling** (left): count the number of halo particles to reveal a potential problem in an accelerator design
  - **Ice Sheet modeling** (middle): capture ice calving much faster to improve ice sheet model
  - **Space Weather modeling** (right): verify a particle acceleration mechanism known as anisotropy by analyzing trillions of particles, made possible by our querying capability





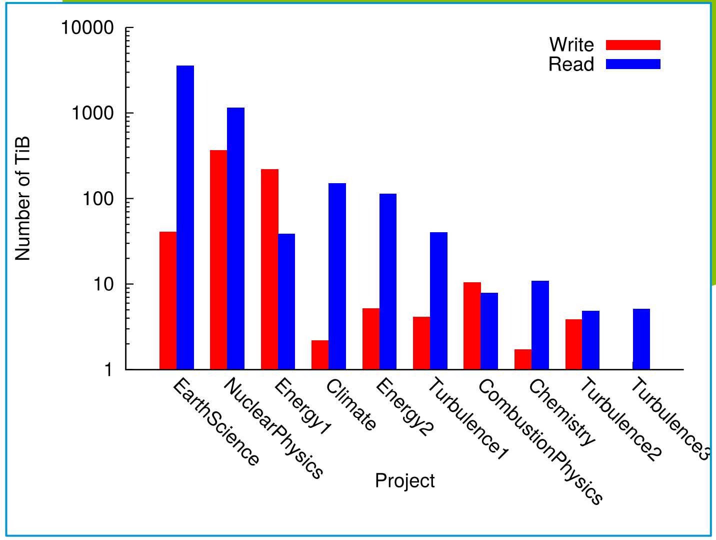


DARSHAN: I/O CHARACTERIZATION FOR DATA-INTENSIVE SCIENCE

SDAV
Scalable Data Management, Analysis, and Visualization

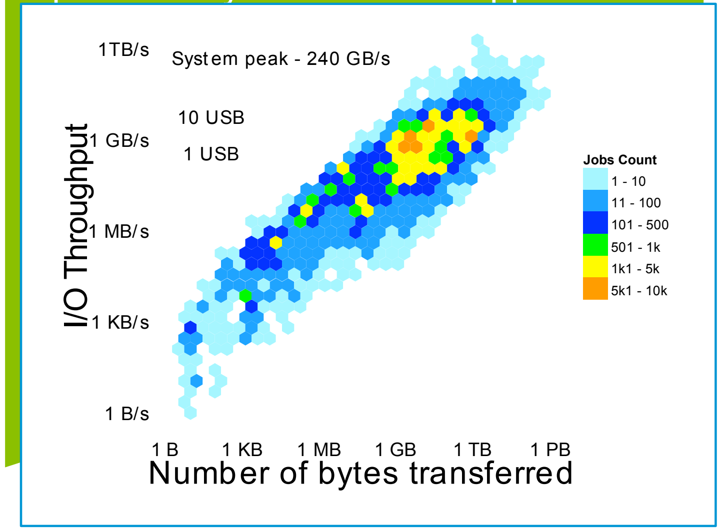
ASCR Base (2008-2011)

- Darshan was conceived to address the need for greater understanding of I/O behavior in diverse scientific applications
- Enabled unprecedented insight into the behavior of the most data-intensive scientific applications at Argonne National Laboratory



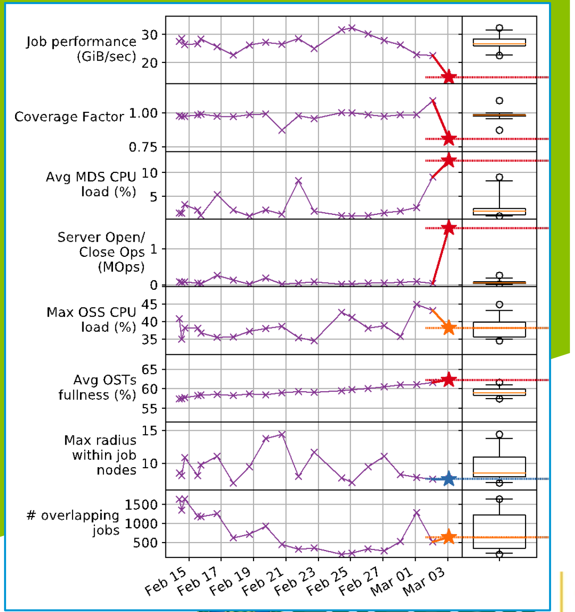
SciDAC (2012-2017)

- Darshan was generalized and ported to multiple computational platforms (IBM BG/Q, Cray XE and XC, Linux clusters) and deployed at every major ASCR facility
- Widespread deployment enabled both cross-platform studies and targeted optimizations to improve the scientific productivity of numerous applications



Impact Going Forward

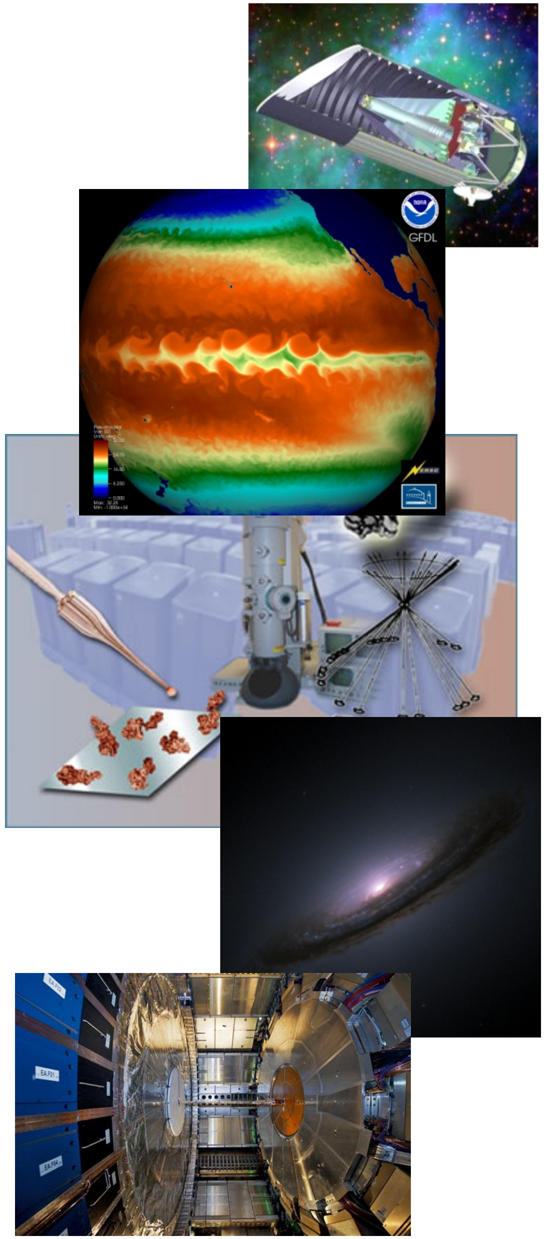
- Darshan is supported by the ALCF, NERSC, and OLCF computing facilities on their largest systems
- Vendors such as Intel are contributing major features

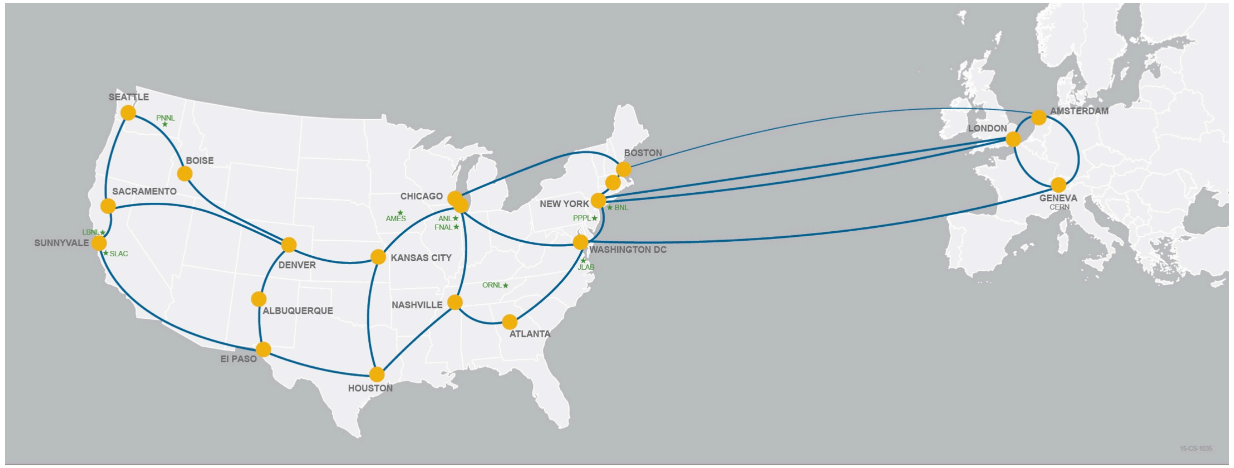


ENERGY | Office of Science

DOE OFFICE OF SCIENCE AND ESNET – THE ESNET MISSION

- **ESnet - the Energy Sciences Network - is an Office of Science user facility whose <u>primary mission is to enable large-scale science</u> that depends on**:
  - Multi-institution, world-wide collaboration
  - Data mobility: sharing of massive amounts of data
  - Distributed data management and processing
  - Distributed simulation, visualization, and computational steering
  - Collaboration with the US and International Research and Education community
- ESnet traffic reflects this: Growth of about 10x every 4 years, and currently moving about 80 Petabytes/month
- ESnet is a multi-hundred gigabit/sec network whose backbone covers the US and Europe in order to connect the DOE National Laboratories, science instruments and user facilities to each other and to collaborators worldwide





LHC EXPERIMENTS – INTERNATIONALLY CONNECTED

- Distributed, interconnected set of LHC sites bring computing, data and people together
- Ecosystem of connected sites accomplishes things unachievable at a single university or Laboratory
- Higgs Boson discovery

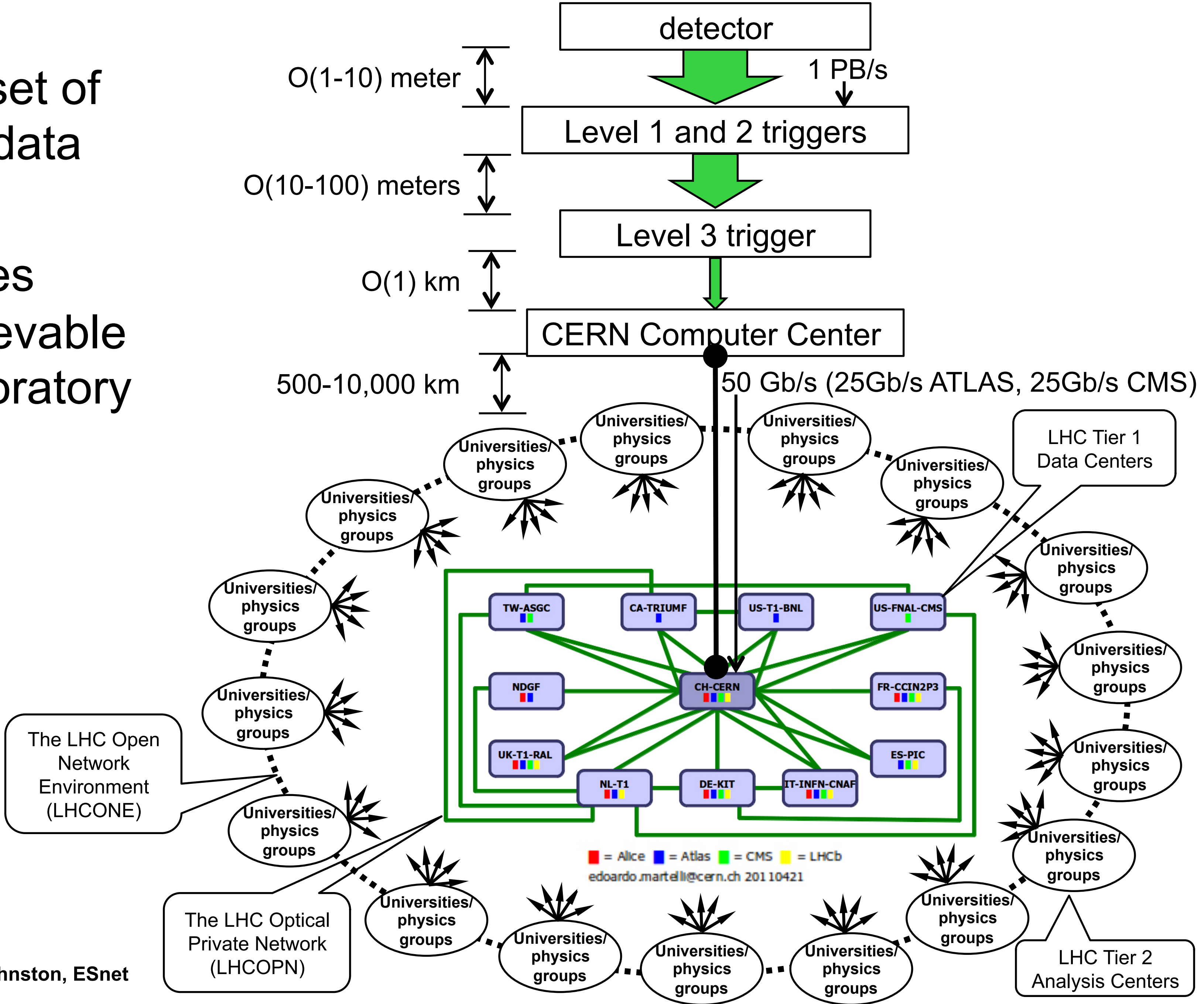


Image source: Bill Johnston, ESnet

FIXING INTERNET CONGESTION COLLAPSE

- Van Jacobson at LBNL, 1986
  - 1000x decrease in throughput between LBNL and UCB
  - Several key improvements to TCP
    - Slow start
    - Congestion window
    - Others
  - Fixed the throughput problem and laid the foundation for dramatic Internet growth
- Critical architectural improvements served as a platform for multiple technologies

Figure 3: Startup behavior of TCP without Slow-start

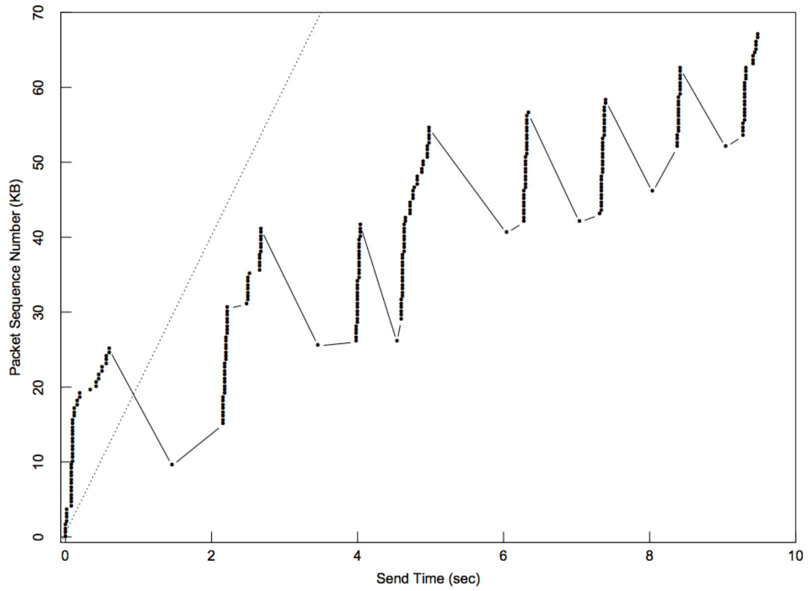
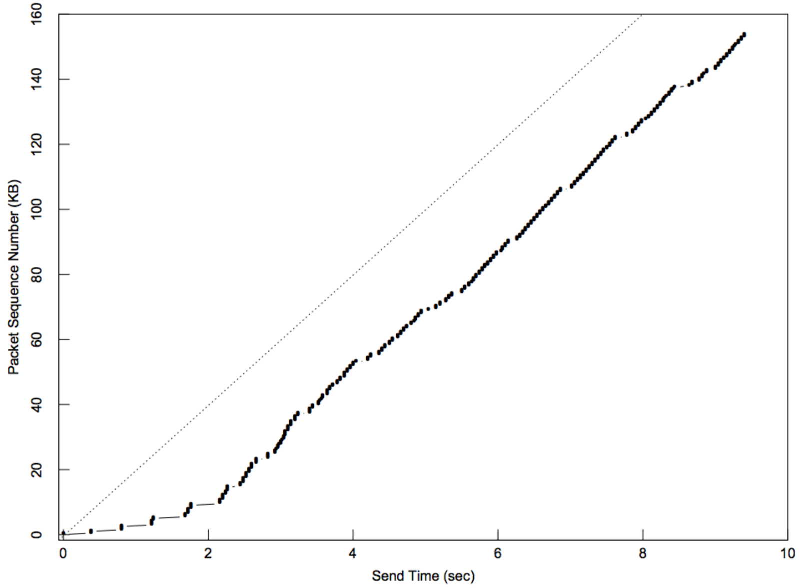


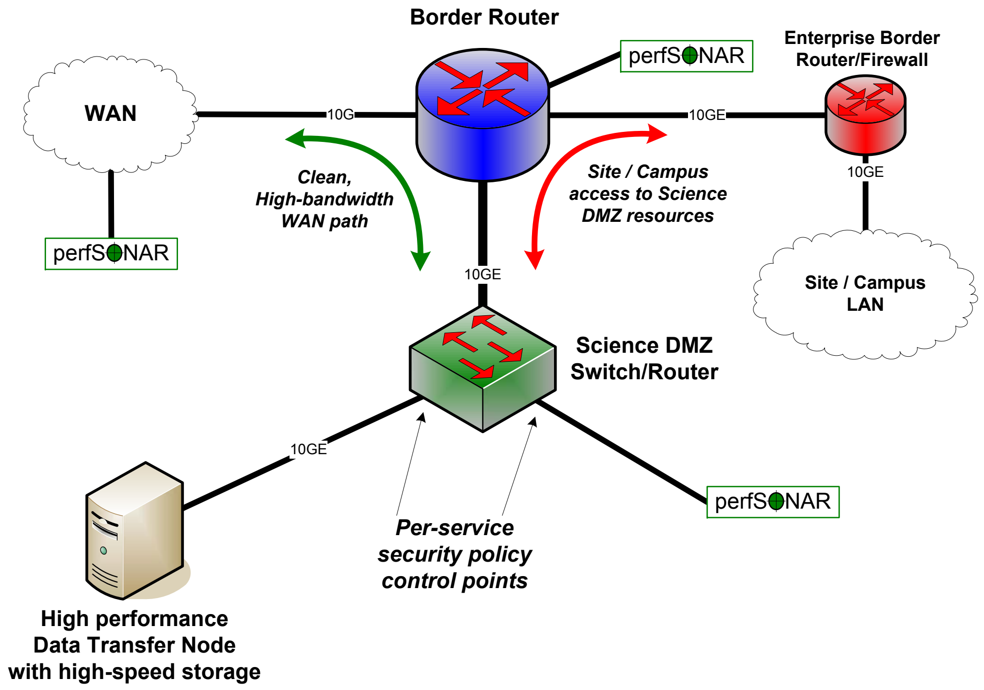
Figure 4: Startup behavior of TCP with Slow-start



Images: JACOBSON, V. Congestion avoidance and control. In Proceedings of SIGCOMM ’88 (Stanford, CA, Aug. 1988), ACM.

SCIENCE DMZ DESIGN PATTERN

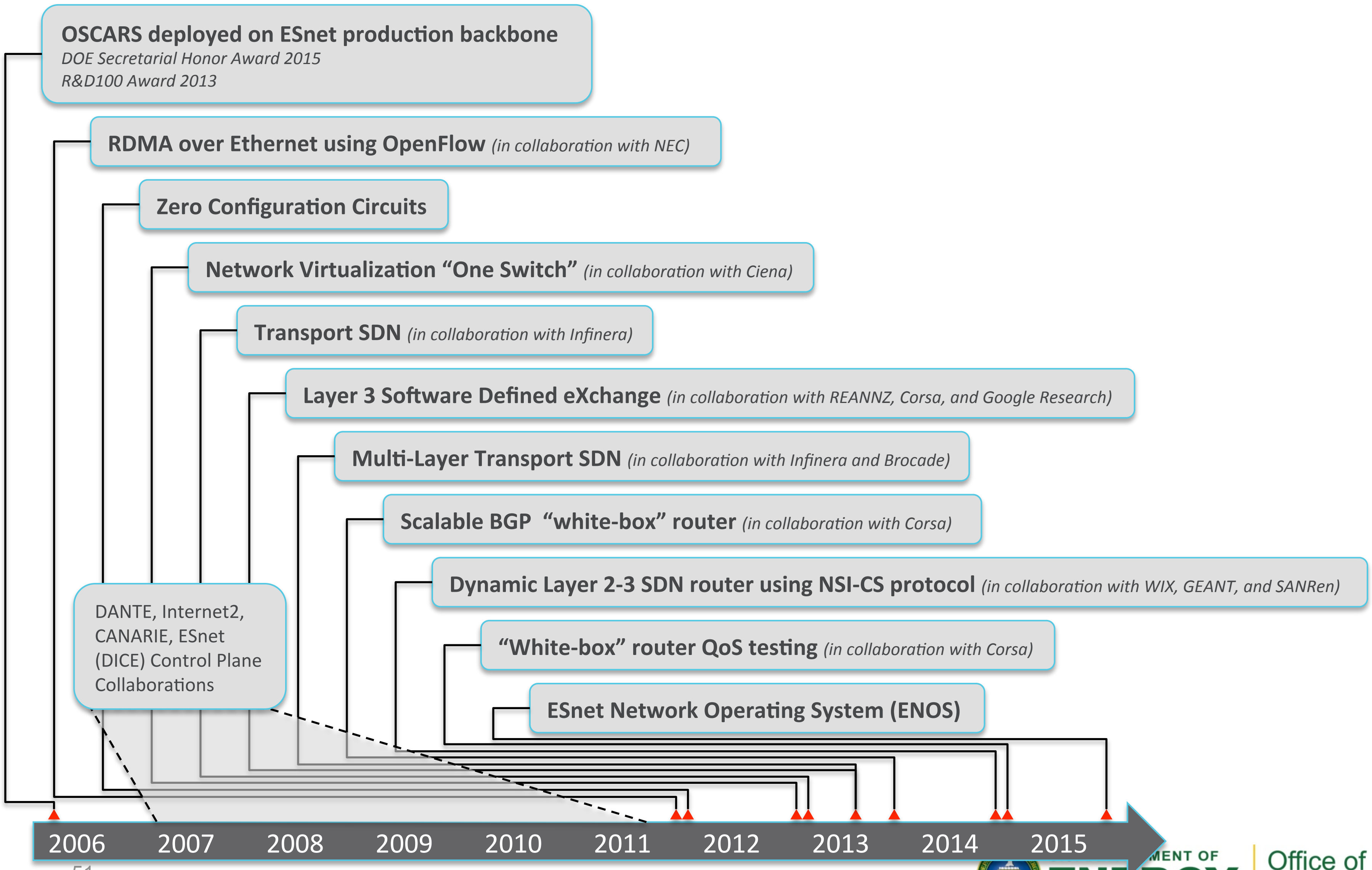
- High-speed interface between site/ campus network and wide area science network
- Allows for easy deployment of scalable data transfer services
- Data placement at scale allows HPC to be brought to bear on previously intractable problems
- Working toward a future where all major scientific facilities have high speed connections to capable science networks



E. Dart, L. Rotman, B. Tierney, M. Hester and J. Zurawski, "The Science DMZ: A network design pattern for data-intensive science," *SC '13: Proceedings of the International Conference on High Performance Computing, Networking, Storage and Analysis*, Denver, CO, 2013, pp. 1-10. doi: 10.1145/2503210.2503245

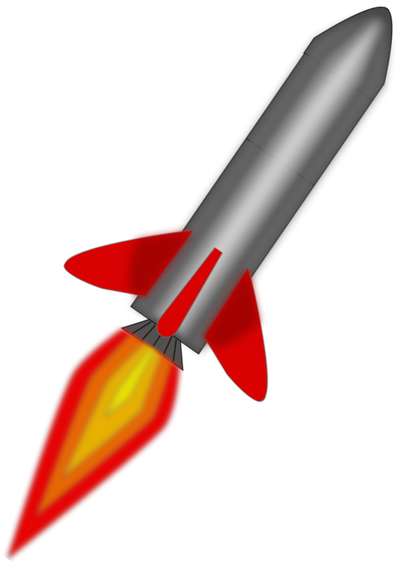
ESNET DRIVING SOFTWARE-DEFINED NETWORKING CONCEPTS ACROSS DOE AND INDUSTRY SINCE 2006

- SDN will do for networking in the future what modern orchestration and service constructs have done for computing
  - More responsive, self-healing, more efficient, higher performance
  - The APIs, interfaces, and services are an area of active research
- Getting the architecture right for multiple technology generations with new ESnet6 network



PROCTER&GAMBLE …

Our Technical Challenges lead to collaboration with the best Scientists



- **P&G & LANL… Since 1993**
  - dozen+ CRADA's … T, D, ESA, Center for Nano.
  - > $20MM Funds in, + equal matching in-kind.
  - LANL Ind. Fellow @ P&G
  - P&G employee holds Q- as visiting scientist.

- **P&G & SNL… since 1999**
  - Umbrella CRADA …2003
    - Foam mechanics, Porous media flow, Sierra Suite
  - R&D Dir on ESRF Ind. Adv. since 2005

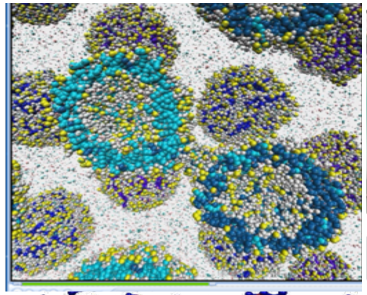
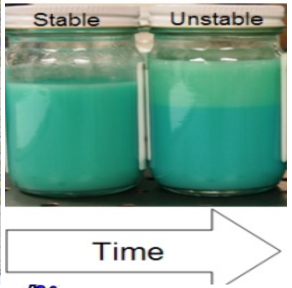
- **P&G & ORNL…Since 2009**
  - Comp Chem of soft materials,
  - Mfg. Systems
- **P&G & ANL…**
  - Agent Based Market Modeling
  - INCITE on Comp Chem
- **P&G & PNNL**
  - Materials from renewable sources

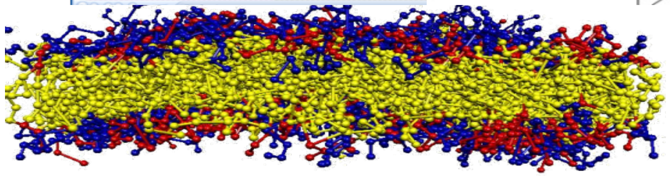
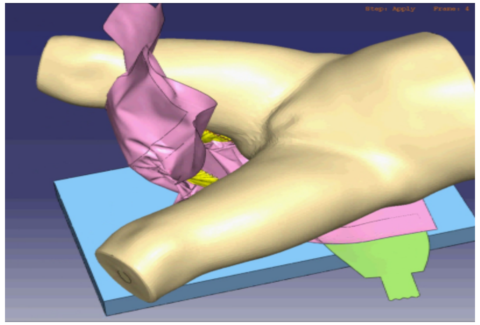
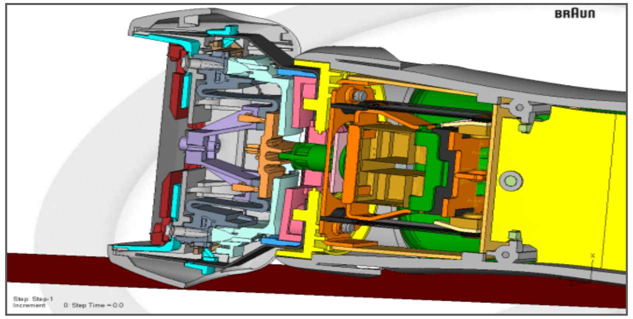
- **P&G & NCSA**
  - Industrial User Facility & Advisory Board

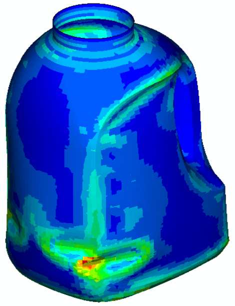
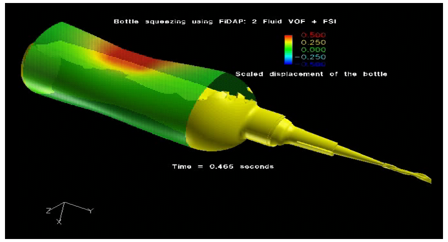
- **P&G & NSF ERC's**
  - Biomechanics, Structured Organic Powders

… P&G AND ATOMS TO THE ENTERPRISE

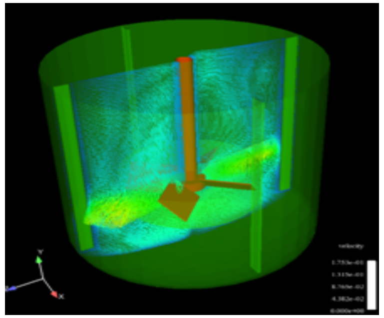
Product/ Materials/ Device/ Package

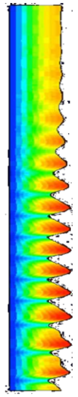
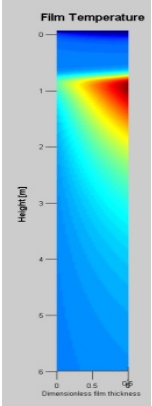



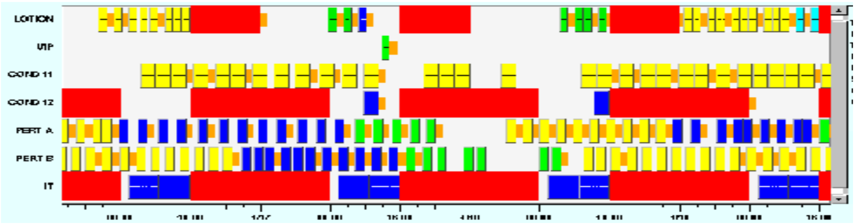



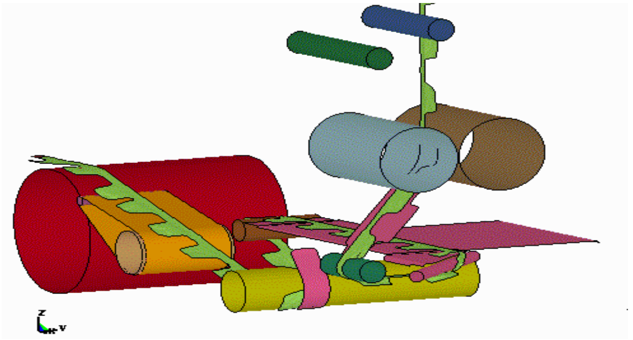
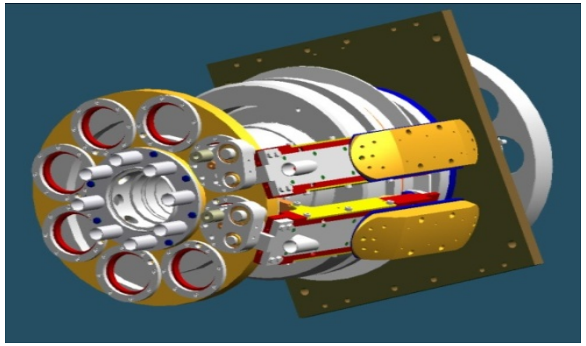
Process



Production Plan & Schedule, Reliability



Converting & Machines

ADDITIONAL MATERIAL

EXEMPLAR DOE SUPPORTED RESEARCH COMPUTERS

- Caltech Cosmic Cube (Chuck Seitz/Geoffrey Fox)
  - Lattice QCD calculations using commodity microprocessors
- Illinois Cedar multiprocessor (David Kuck)
  - Shared memory computing and compiler technology
- NYU Ultracomputer (Jack Schwartz)
  - Shared memory fetch & add intrinsics
- Argonne Advanced Computing Research Facility
  - Parallel computing research experiments plus education
  - Alliant FX/8, Denelcor HEP, Encore Multimax
  - Sequent Balance, Intel iPSC, AMT DAP

- Ecosystems that shaped computing
  - Vendor relations and new computer architecture co-design
  - Software systems and broad uptake and deployment
  - Staff and researcher training and education

APPLIED MATHEMATICS

- Over the years, the ASCR office and its predecessors have made significant contributions to the discipline of Applied Mathematics by supporting research activities at the national laboratories and universities.
- Mathematics and applied analysis research supported by ASCR that is not computational includes
  - Formulating mathematical models of physical systems
  - Analysis of the models to ascertain its properties: is the model complete in the sense that it has a solution? Are there multiple solutions? What are the properties of the solution?
  - Simplification of the model by reducing its dimensionality, using fewer state variables or introducing simplifying assumptions
  - Determining whether it is possible to view the model as a perturbation of a simpler model that can be analyzed

APPLIED MATHEMATICS

- ASCR has been in the forefront of funding research in several areas in applied mathematics, especially
  - asymptotic analysis and
  - the modeling of multiscale multiphysics problems
- Major advances have been achieved in the equations of fluid flow, neutron diffusion, combustions and materials science
- Applied mathematics is more than modeling, analysis, and computation. Data have become as important for the discipline as models, methods and algorithms. Data assimilation (i.e., the integration of data and models) is already an integral component of applied mathematics and an active area of current research.

HIGHLIGHT SLIDES ON ADDITIONAL TOPICS ARE UNDER PREPARATION

- Visualization
- CSGF
- Training programs
- Computational science and engineering achievements
- …

SUBCOMMITTEE MEMBERS

- Buddy Bland, ORNL
- Jackie Chen, SNL
- Phil Colella, LBNL
- Jack Dongarra, University of Tennessee and ORNL
- Thom Dunning, PNNL
- Wendy Huntoon, KINBER
- Bill Johnston, LBNL (ret.)
- Paul Messina, ANL, Chair
- Jim Pool, Caltech (ret.)
- Dan Reed, University of Iowa
- John Sarrao, LANL

- Considerable contributions from Eli Dart and Inder Monga

THANK YOU!