

A decadal DOE plan for providing exascale applications and technologies for DOE mission needs

Rick Stevens and Andy White, co-chairs

Pete Beckman, Ray Bair-ANL; Jim Hack, Jeff Nichols, Al Geist-ORNL; Horst Simon, Kathy Yelick, John Shalf-LBNL; Steve Ashby, Moe Khaleel-PNNL; Michel McCoy, Mark Seager, Brent Gorda-LLNL; John Morrison, Cheryl Wampler-LANL; James Peery, Sudip Dosanjh, Jim Ang-SNL; Jim Davenport, Tom Schlagel, BNL; Fred Johnson, Paul Messina, ex officio



The exascale draft plan has four high-level components

- Science and mission applications
- Systems software and programming models
- Hardware technology R&D
- Systems acquisition, deployment and operations

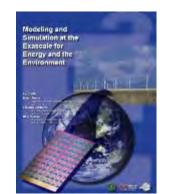


- **2018:** The plan targets exascale platform deliveries in 2018 and a robust simulation environment and science and mission applications by 2020
- **2015**: Co-design and co-development of hardware, software, programming models and applications requires intermediate platforms in 2015

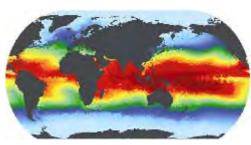


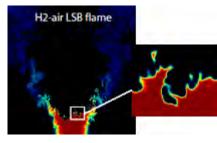
Process for identifying exascale applications and technology for DOE missions ensures broad community input

- Town Hall Meetings April-June 2007
- Scientific Grand Challenges Workshops Nov, 2008 – Oct, 2009
 - Climate Science (11/08),
 - High Energy Physics (12/08),
 - Nuclear Physics (1/09),
 - Fusion Energy (3/09),
 - Nuclear Energy (5/09),
 - Biology (8/09),
 - Material Science and Chemistry (8/09),
 - National Security (10/09)
 - Cross-cutting technologies (2/10)
- Exascale Steering Committee
 - "Denver" vendor NDA visits 8/2009
 - SC09 vendor feedback meetings
 - Extreme Architecture and Technology Workshop 12/2009
- International Exascale Software Project
 - Santa Fe, NM 4/2009; Paris, France 6/2009; Tsukuba, Japan 10/2009

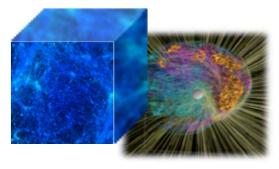


Scientific Grand Cha





MISSION IMPERATIVES

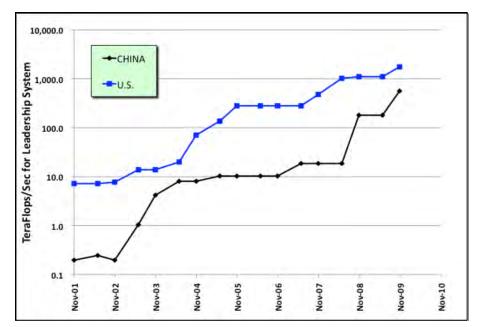


FUNDAMENTAL SCIENCE



Computational science, exascale computing & leadership in science and technology

- The future will require design and certification of complex engineered systems and analysis of climate mitigation alternatives with quantified levels of uncertainty
 - New fuels and reactors
 - Stewardship without nuclear tests
 - Carbon sequestration alternatives
 - Regional climate impacts
- Broader application of exascale computing can provide tremendous advantages for fundamental science and industrial competitiveness
 - Renewable energy and energy storage
 - Prediction and control of materials in extreme environments
 - Understanding dark energy and dark
 matter
 - Clean and efficient combustion in advanced engines

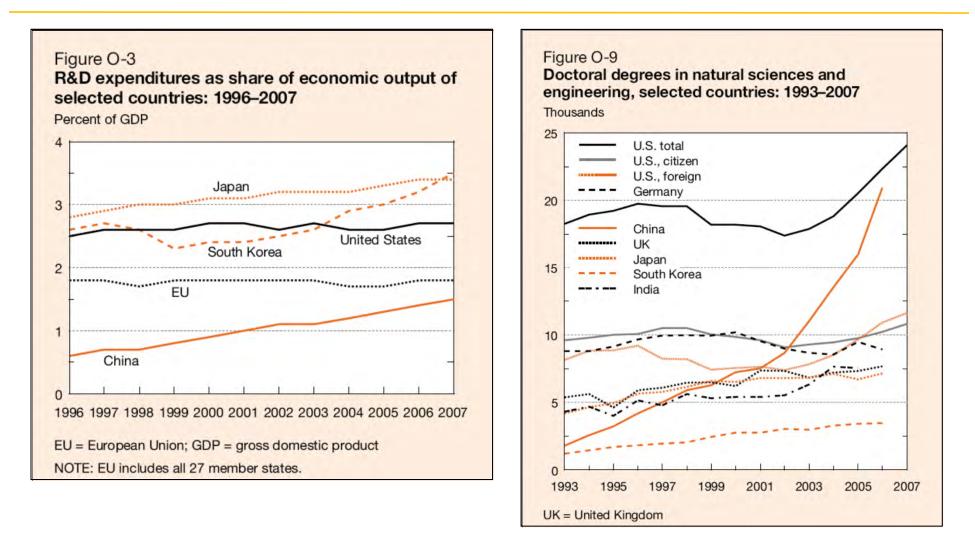


International Competition in HPC Chart shows total HPC resources available, from TOP500

"The United States led the world's economies in the 20th century because we led the world in innovation. Today, the competition is keener; the challenge is tougher; and that is why innovation is more important than ever. It is the key to good, new jobs for the 21st century." *President Barack Obama, August 5, 2009*



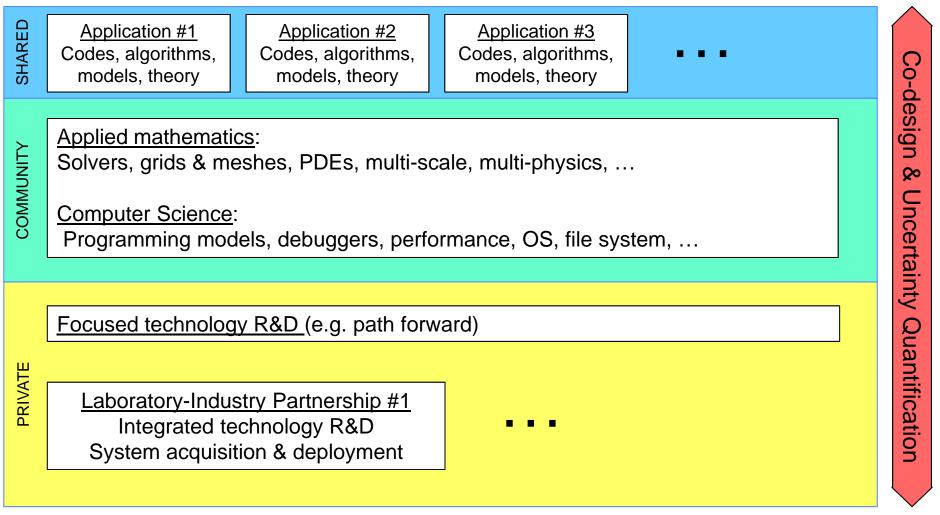
and if you think there is no competition ...



Science and Engineering Indicators [2010], National Science Board



A Taxonomy for DOE Exascale Initiative





Outline of talk

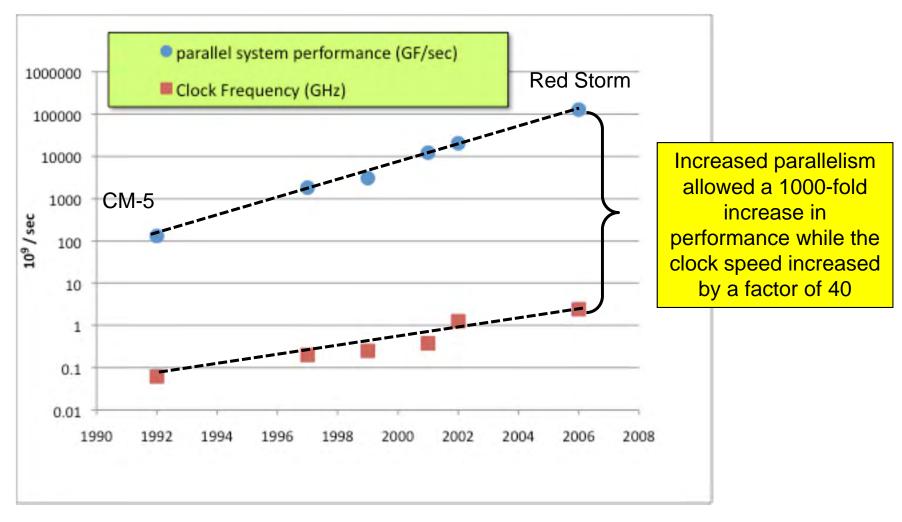
- Technology needs
- Co-design
- Mission and science needs



TECHNOLOGY NEEDS



Concurrency is one key ingredient in getting to exaflop/sec



and power, resiliency, programming models, memory bandwidth, I/O, ...



What are critical exascale technology investments?

- **System power** is a first class constraint on exascale system performance and effectiveness.
- Memory is an important component of meeting exascale power and applications goals.
- **Programming model**. Early investment in several efforts to decide in 2013 on exascale programming model, allowing exemplar applications effective access to 2015 system for both mission and science.
- **Investment in exascale processor design** to achieve an exascale-like system in 2015.
- **Operating System strategy for exascale** is critical for node performance at scale and for efficient support of new programming models and run time systems.
- **Reliability and resiliency are critical at this** scale and require applications neutral movement of the file system (for check pointing, in particular) closer to the running apps.
- **HPC co-design strategy and implementation** requires a set of a hierarchical performance models and simulators as well as commitment from apps, software and architecture communities.



Notional System Architecture Targets and "swim lanes"

System attributes	2010	"2015"		"2018"	
System peak	2 Peta	200 Petaflop/sec		1 Exaflop/sec	
Power	6 MW	15 MW		20 MW	
System memory	0.3 PB	5 PB		32-64 PB	
Node performance	125 GF	0.5 TF	7 TF	1 TF	10 TF
Node memory BW	25 GB/s	0.1 TB/sec	1 TB/sec	0.4 TB/sec	4 TB/sec
Node concurrency	12	O(100)	O(1,000)	O(1,000)	O(10,000)
System size (nodes)	18,700	50,000	5,000	1,000,000	100,000
Total Node Interconnect BW	1.5 GB/s	150 GB/sec	1 TB/sec	250 GB/sec	2 TB/sec
MTTI	day	O(1 day)		O(1 day)	



Swim lanes affect the number of threads that the system needs to support.

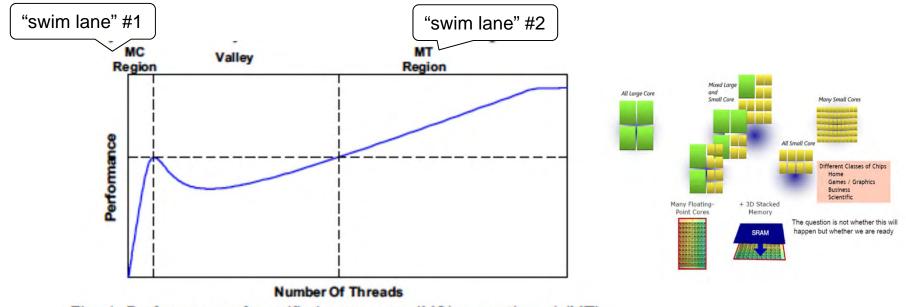


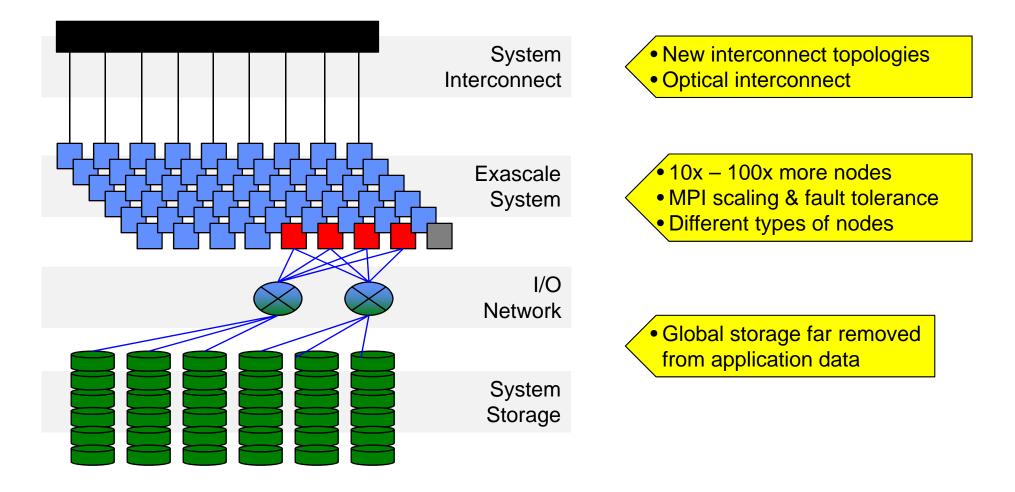
Fig. 1. Performance of a unified many-core (MC) many-thread (MT) machine exhibits three performance regions, depending on the number of threads in the workload.

There are currently two basic design points for achieving high performance in technical applications. In the future it is expected that these design points may (or may not) become more Integrated.

Many-core vs. many-thread machines: stay away from the valley, IEEE 2009 Exascale Initiative Steering Committee

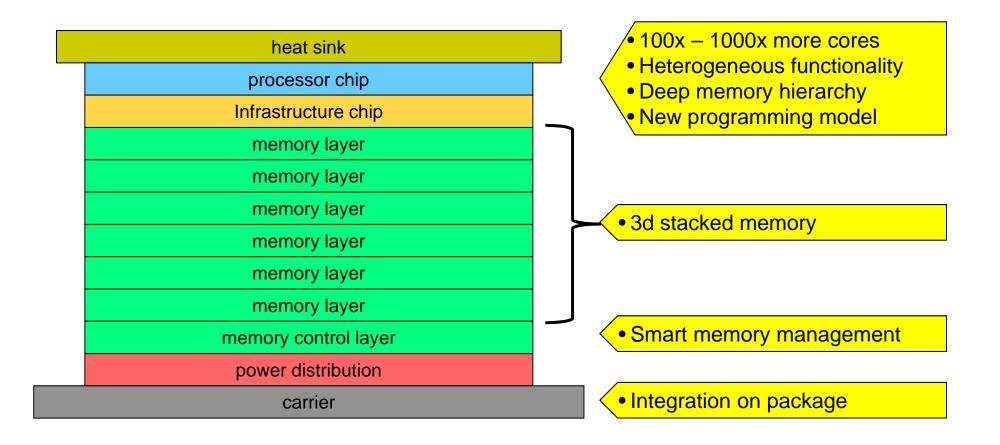


The high level system design may be similar to petascale systems



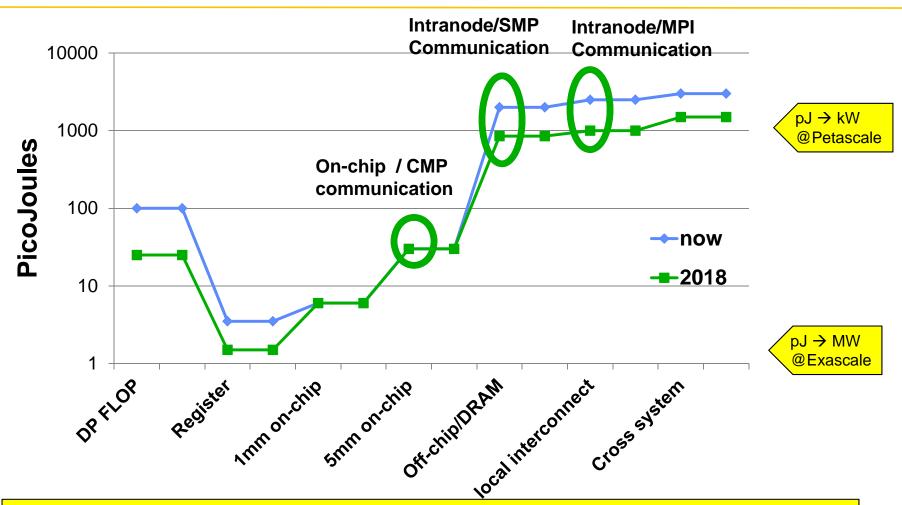


The node is the key for exascale, as well as for ~ exascale.





Investments in architecture R&D and application locality are critical.



"The Energy and Power Challenge is the most pervasive ... and has its roots in the inability of the [study] group to project any combination of currently mature technologies that will deliver sufficiently powerful systems in any class at the desired levels." *DARPA IPTO exascale technology challenge report*



Embedded processor example: it's about architecture and moving data.

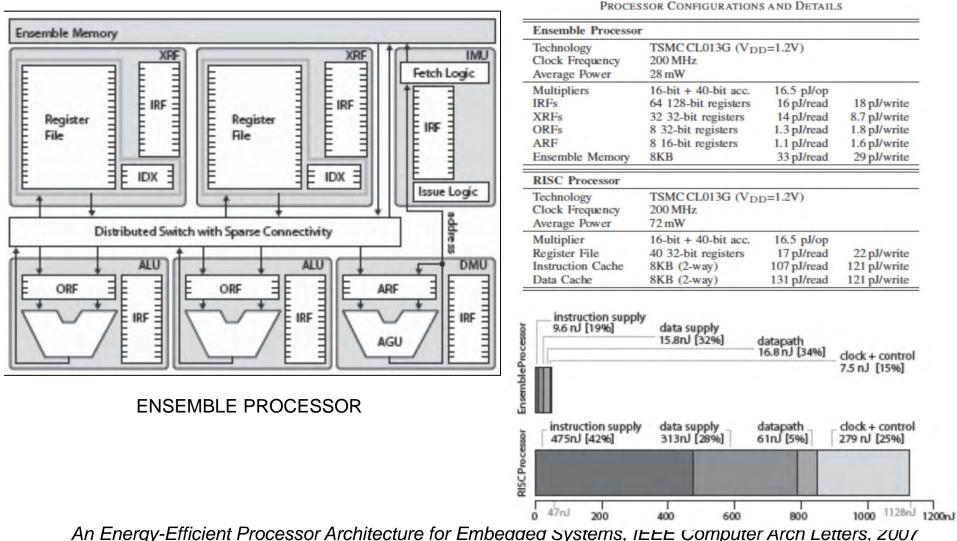
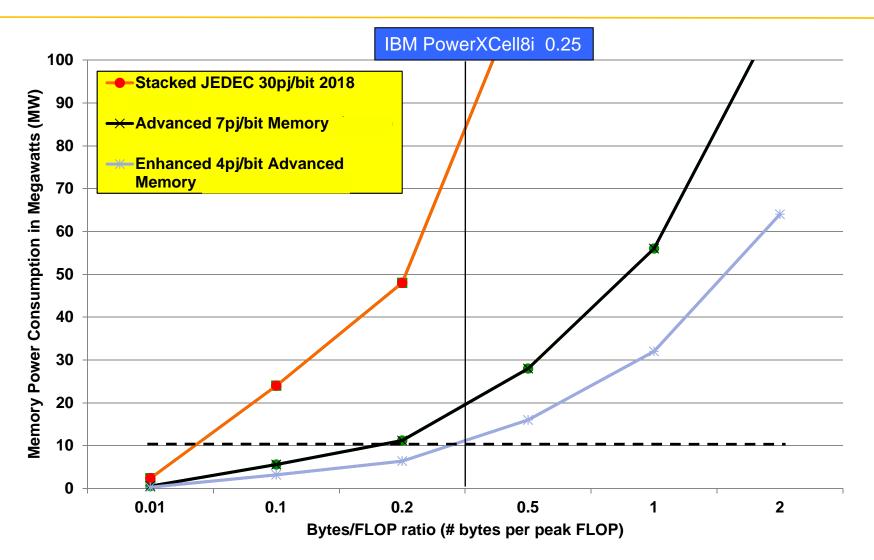


TABLE I



Investments in memory technology mitigate risk of narrowed application scope.

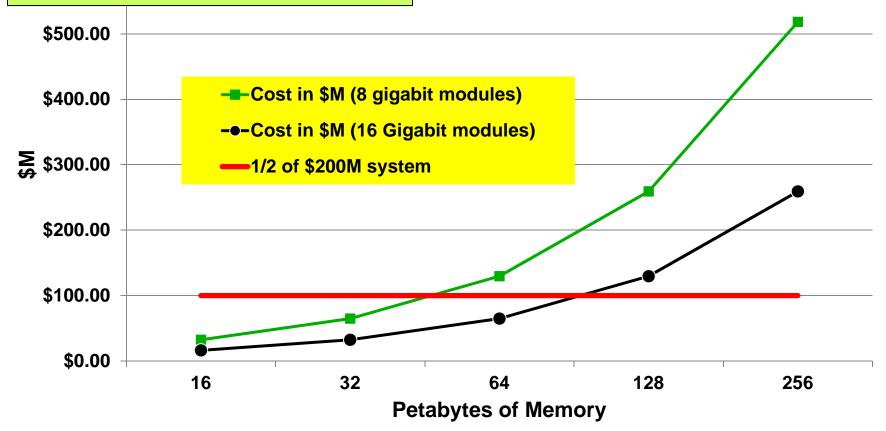




Cost of Memory Capacity for two different potential memory Densities

- Memory density is doubling every three years; processor logic, every two
 - Project 8Gigabit DIMMs in 2018
 - 16Gigabit if technology acceleration

- Storage costs are dropping gradually compared to logic costs
 - Industry assumption is \$1.80/memory chip is median commodity cost





We need solutions to handle decreased reliability and a new model for resiliency

• Barriers

- System components, complexity increasing
- Silent error rates increasing
- Reduced job progress due to fault recovery if we use existing checkpoint/restart

• Technical Focus Areas

- Local recovery and migration
- Development of a standard fault model and better understanding of types/rates of faults
- Improved hardware and software reliability
 - Greater integration across entire stack
- Fault resilient algorithms and applications

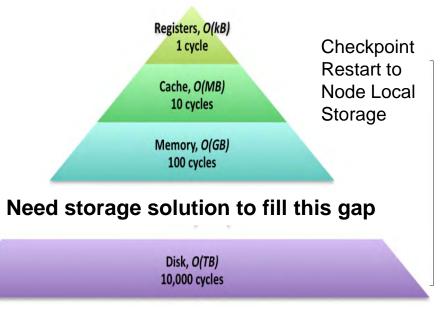
Technical Gap

•

- Maintaining today's MTTI given 10x 100X increase in sockets will require:
 - 10X improvement in hardware reliability 10X in system software reliability, and 10X improvement due to local recovery and migration as well as research in fault resilient applications

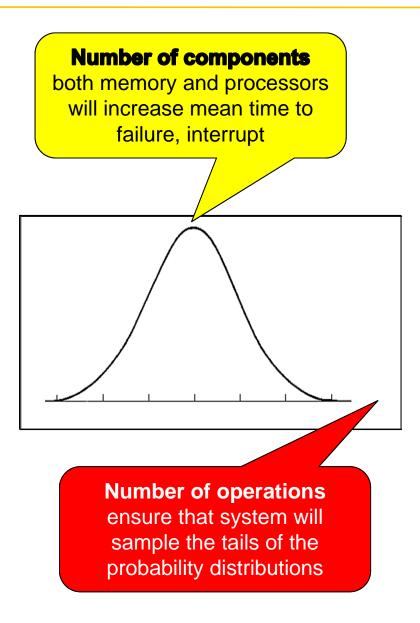
Taxonomy of errors (h/w or s/w)

- Hard errors: permanent errors which cause system to hang or crash
- **Soft errors**: transient errors, either correctable or short term failure
- Silent errors: undetected errors either permanent or transient. Concern is that simulation data or calculation have been corrupted and no error reported.





Factors affecting resilience @ exascale

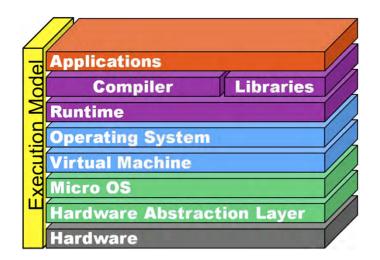


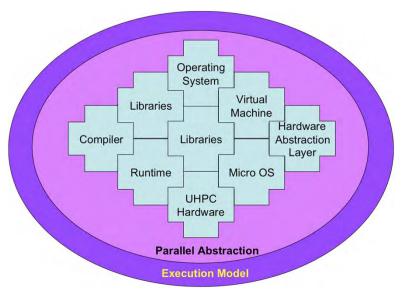
- Smaller circuit sizes, running at lower voltages to reduce power consumption, increases the probability of switches flipping spontaneously due to thermal and voltage variations as well as radiation, increasing soft errors
- Heterogeneous systems make error detection and recovery even harder, for example, detecting and recovering from an error in a GPU can involve hundreds of threads simultaneously on the GPU and hundreds of cycles in drain pipelines to begin recovery.
- Increasing system and algorithm complexity makes improper interaction of separately designed and implemented components more likely.
- "Cost" to add additional HW detection and recovery logic right on the chips to detect silent errors. Because it will increase power consumption by 15% and increase the chip costs.



System software as currently implemented is not suitable for exascale system.

- Barriers
 - System management SW not parallel
 - Current OS stack designed to manage only O(10) cores on node
 - Unprepared for industry shift to NVRAM
 - OS management of I/O has hit a wall
 - Not prepared for massive concurrency
- Technical Focus Areas
 - Design HPC OS to partition and manage node resources to support massively concurrency
 - I/O system to support on-chip NVRAM
 - Co-design messaging system with new hardware to achieve required message rates
- Technical gaps
 - 10X: in affordable I/O rates
 - 10X: in on-node message injection rates
 - 100X: in concurrency of on-chip messaging hardware/software
 - 10X: in OS resource management





Software challenges in extreme scale systems, *Sarkar, 2010*



Factors Leading to Gap

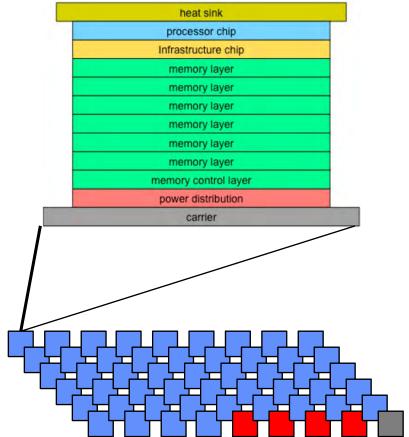
It is not only the massive increase in concurrency, but also the change of architecture

- OS
 - Current OS designs focus on homogeneous cores, memory structures, and tasks
 - Designs to manage 256 "full" cores or efficiently coordinate thousands of stream processors for HPC applications do not exist
 - HW makers are moving to on-chip page-mapped memory, but no OS features have been developed to leverage this for HPC applications
- I/O
 - Current designs are primarily file based, and cannot efficiently optimize HPC workloads for aggregation, ordering, and patterns
 - Chip makers are putting NV RAM on or close to die, but file-based I/O paradigms are not suited to leverage this development
 - Currently, I/O is "far", through many hardware layers (torus, I/O forwarding, Infiniband, RAID controller, SCSI, etc). I/O balance guickly falling behind – new integrated design approach required
- Messaging & Run-time Systems
 - HW put/get message queues for interconnect must rapidly evolve to support massive concurrency – however SW community has not explored how to manage millions of msg endpoints, dynamic mapping of memory buffers, or fault resilience at that scale 22



Programming Model Approaches

- Hierarchical approach: intra-node + inter-node
 - Part I: Inter-node model for communicating between nodes
 - MPI scaling to millions of nodes: Importance high; risk
 low
 - One-sided communication scaling: Importance medium; risk low
 - Part II: Intra-node model for on-chip concurrency
 - Overriding Risk: No single path for node architecture
 - OpenMP, Pthreads: High risk (may not be feasible with node architectures); high payoff (already in some applications)
 - New API, extended PGAS, or CUDA/OpenCL to handle hierarchies of memories and cores: Medium risk (reflects architecture directions); Medium payoff (reprogramming of node code)
- Unified approach: single high level model for entire system
 - High risk; high payoff for new codes, new application domains

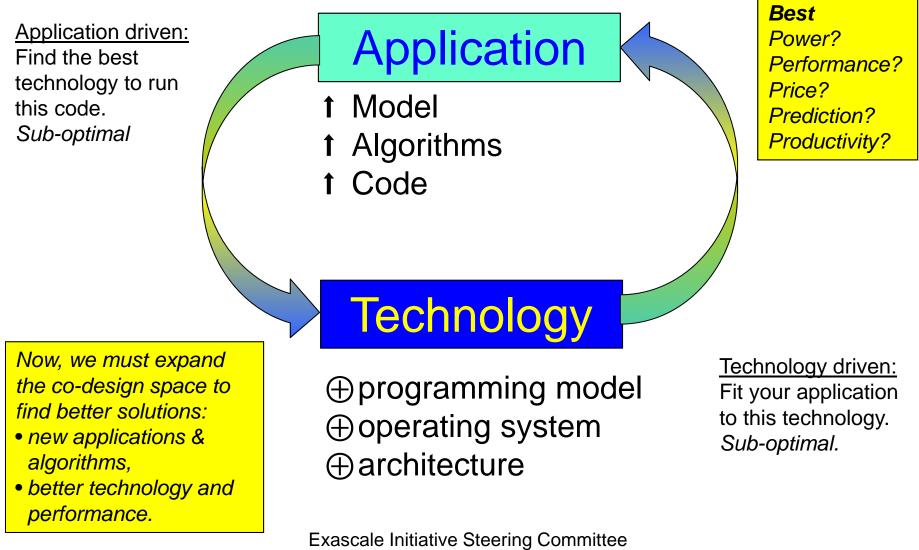




CO-DESIGN



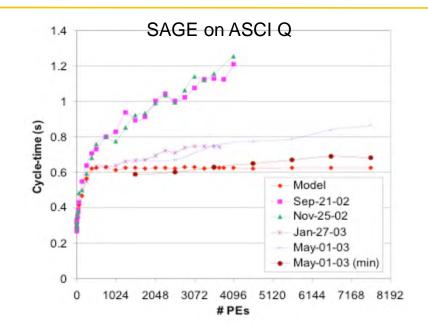
Co-design expands the feasible solution space to allow better solutions.

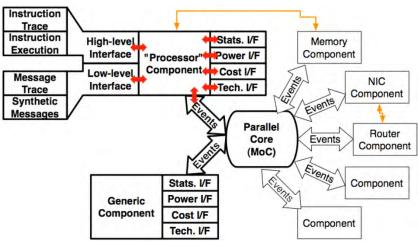




Hierarchical {application, s/w, h/w} cosimulation is a the key for co-design

- Hierarchical co-simulation capability
 - Discussions between architecture, software and application groups
 - System level simulation based on analytic models
 - Detailed (e.g. cycle accurate) cosimulation of hardware and applications
- Opportunity to influence future architectures
 - Cores/node, threads/core, ALUs/thread
 - Logic layer in stacked memory
 - Interconnect performance
 - Memory/core
 - Processor functionality
- Current community efforts must work together to provide a complete co-design capability







A first step toward co-design was February exascale workshop.

 The approach will be to engage experts in computational science and computer science with the goal of

Cross-cutting Technologies for Computing at the Exascale February 2-5, 2010 · Washington, D.C.

- Producing a first cut at the characteristics of systems that (a) could be fielded by 2018 and (b) would meet applications' needs
- Outlining the R&D needed for "co-design" of system architecture, system software and tools, programming frameworks, mathematical models and algorithms, and scientific application codes at the exascale, and
- Exploring whether this anticipated phase change in technology (like parallel computing in 1990s) provides any opportunities for applications. That is, whether a requirement for revolutionary application design allows new methods, algorithms, and mathematical models to be brought to bear on mission and science questions.

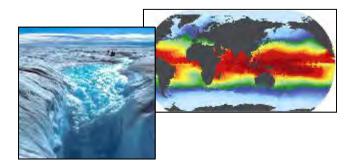


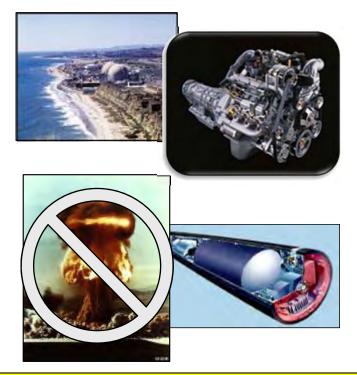
MISSION & SCIENCE NEEDS



DOE mission imperatives require simulation and analysis for policy and decision making

- Climate Change: Understanding, mitigating and adapting to the effects of global warming
 - Sea level rise
 - Severe weather
 - Regional climate change
 - Geologic carbon sequestration
- *Energy*: Reducing U.S. reliance on foreign energy sources and reducing the carbon footprint of energy production
 - Reducing time and cost of reactor design and deployment
 - Improving the efficiency of combustion energy systems
- National Nuclear Security: Maintaining a safe, secure and reliable nuclear stockpile
 - Stockpile certification
 - Predictive scientific challenges
 - Real-time evaluation of urban nuclear
 detonation





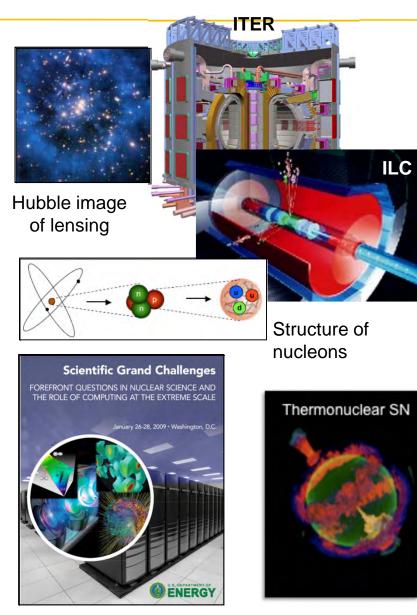
Accomplishing these missions requires exascale resources.



Exascale simulation will enable fundamental advances in basic science.

- High Energy & Nuclear Physics
 - Dark-energy and dark matter
 - Fundamentals of fission & fusion reactions
- Facility and experimental design
 - Effective design of accelerators
 - Probes of dark energy and dark matter
 - ITER shot planning and device control
- Materials / Chemistry
 - Predictive multi-scale materials
 modeling: observation to control
 - Effective, commercial technologies in renewable energy, catalysts, batteries and combustion
- Life Sciences
 - Better biofuels
 - Sequence to structure to function

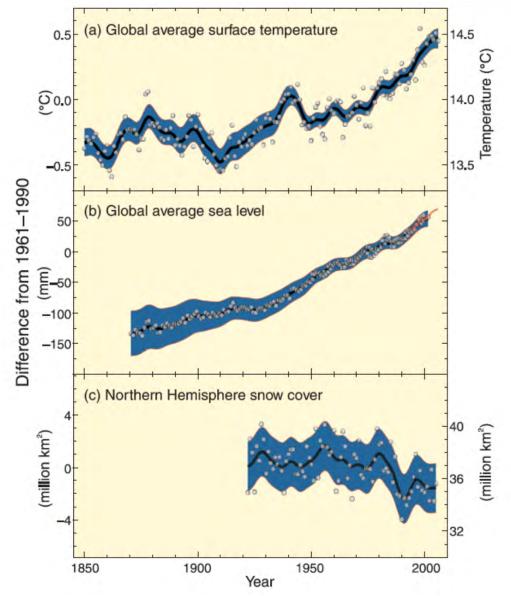
These breakthrough scientific discoveries and facilities require exascale applications and resources.





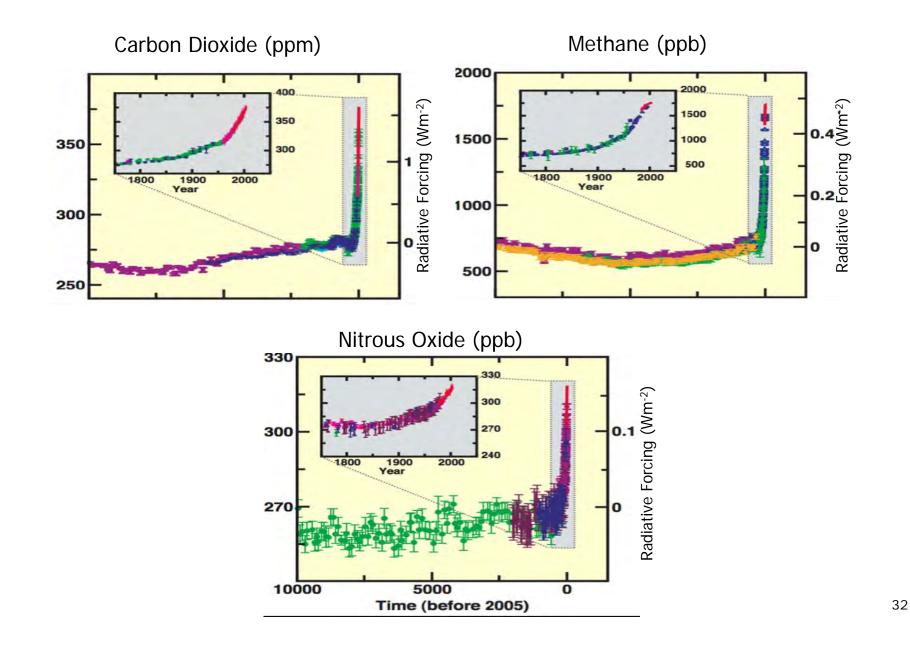
Climate Change 2007

"Warming of the climate system is unequivocal, as is now evident from observations of increases in global average air and ocean temperatures, widespread melting of snow and ice and rising global average sea level."



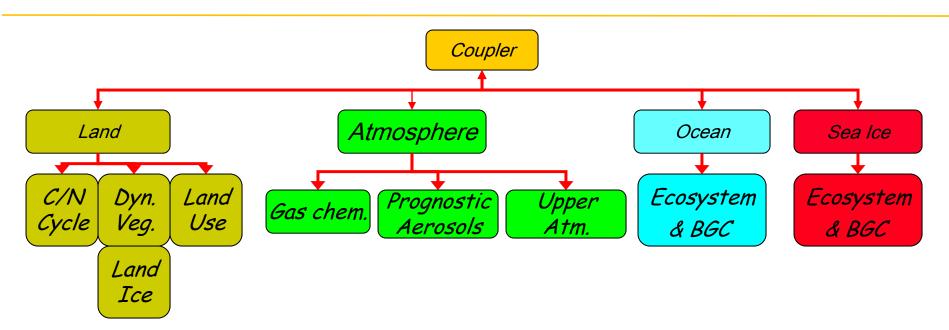


Changing atmospheric composition are indicators of human influence





Evolution toward an Earth System Model



- Coupled climate-chemistry model in the immediate future
 - Terrestrial and oceanic biogeochemical models
 - Ocean ecosystem and terrestrial C/N models
 - Ability to simulate interactions of aerosols with water and biogeochemical cycles
- Explore and understand importance of upper atmospheric process and sun-weather relationships
- Land use and land cover change

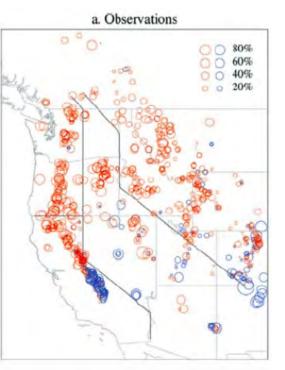


The workshop identified potential impacts of exascale on climate assessments

- Drivers
 - Carbon, methane, and nitrogen cycles
 - Local and regional water, ice, and clouds
 - Distribution of extreme weather events
 - Sea level and ocean circulation
- Petascale climate simulations are uncertain due to parameterization of clouds, convection, and ocean eddies.
- Need for exascale now:
 - Simulating clouds at their native scale for global climate is inherently an exascale problem.
 - Modeling fully turbulent exchange of heat and gases between the atmosphere and ocean is an exascale problem.

"Given these drivers [anticipation, adaptation and mitigation] ... it is clear that extreme scale computers and ultra fast networks, data systems and computational infrastructure will be required by 2020."

Challenges in Climate Change Science and the Role of Computing at Extreme Scale, November, 2008



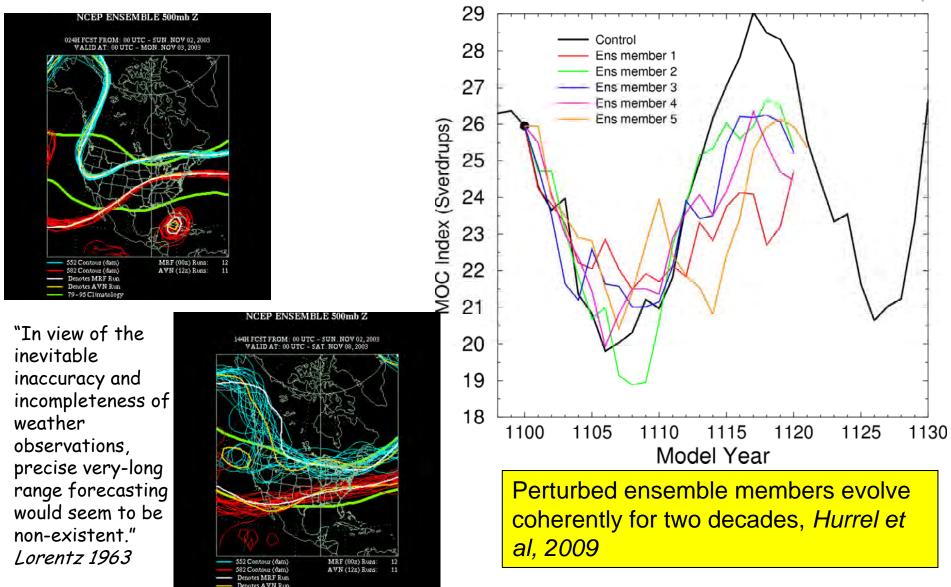
Observed change in snow pack, 1950-1997 *Mote et al,* 2005

Pine Island Glacier, past its tipping point? Katz et al, 2010





Is the climate predictable on decadal time scales?

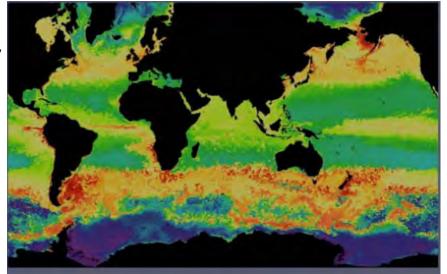




Exascale resources are required for predictive climate simulations.

- Finer resolution
 - Provide regional details
- Higher realism, more complexity
 - Add "new" science
 - Biogeochemistry
 - Ice-sheets
 - Up-grade to "better" science
 - Better cloud processes
 - Dynamics land surface
- Scenario replication, ensembles
 - Range of model variability
- Time scale of simulation
 - Long-term implications

Adapted from *Climate Model Development Breakout Background* Bill Collins and Dave Bader, Co-Chairs



Ocean chlorophyll from an eddy-resolving simulation with ocean ecosystems included

It is essential that computing power be increased substantially (by a factor of 1000), and scientific and technical capacity be increased (by at least a factor of 10) to produce weather and climate information of sufficient skill to facilitate regional adaptations to climate variability and change.

World Modeling Summit for Climate Prediction, May, 2008



America's Energy challenges have far reaching effects.

- National Security
 - dependence on unreliable sources
- Economic Security
 - need for reliable supplies at affordable prices

Environmental Security

 obtaining energy in ways that does not harm the environment





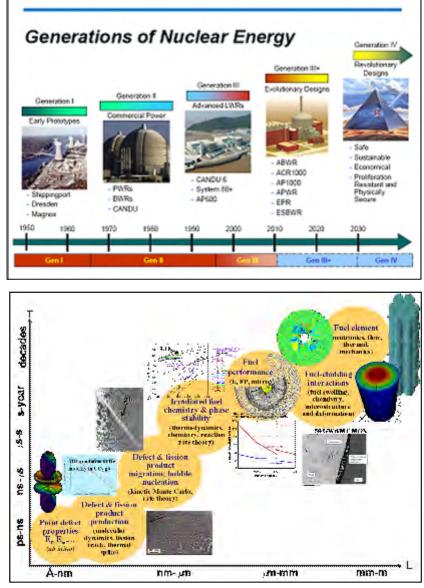
- Nuclear energy can be a safe, reliable and carbon-free source of energy
- Transportation sector (automobiles and trucks) accounts for 2/3 of petroleum use and 1/4 of CO₂ emissions
- Fusion
 - Inertial confinement
 - Magnetic confinement
- Renewable
 - Solar
 - Wind
 - Batteries



Modeling and simulation for nuclear energy supports four integrated codes

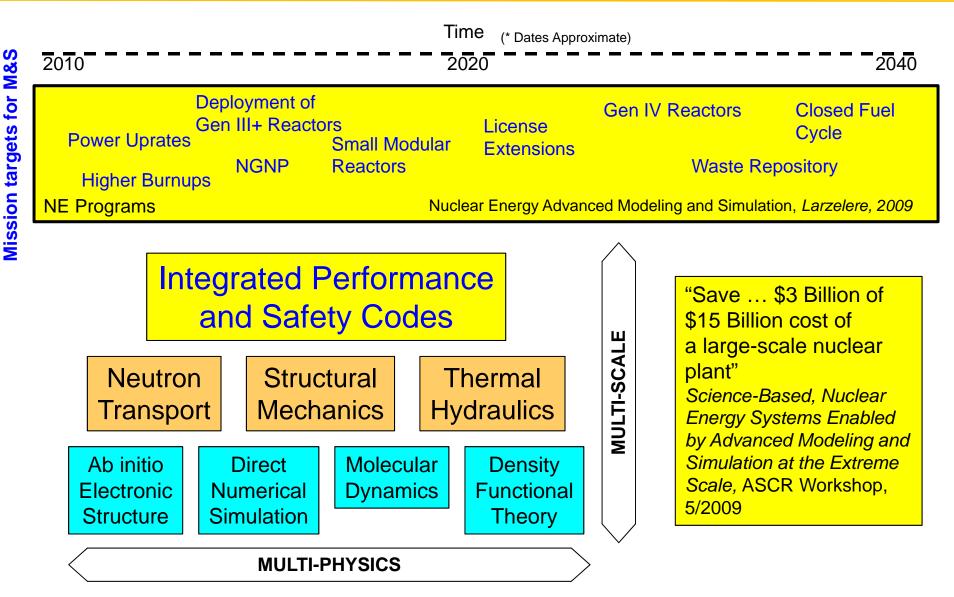
- Reactors: Predict performance and safety of fast reactors over 40 – 60 year lifetime
- Nuclear Fuels: Develop a coupled, predictive three-dimensional, predictive computational tool to predict the performance of nuclear fuel pins and assemblies, applicable to both existing and future advanced nuclear reactor fuel design, fabrication
- Safeguards and Separations: Provide coupled performance of safeguards and separation systems
- Waste: Predict the performance of waste forms under repository conditions for their expected lifetimes (potentially up to a million years)

Nuclear Energy Advanced Modeling and Simulation, Larzelere, 2009





Multi-scale, multi-physics modeling and simulation is critical for nuclear energy.

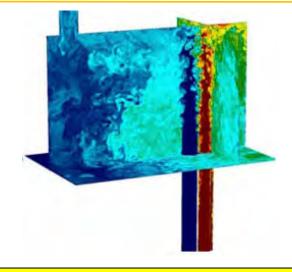




Verification, validation and uncertainty quantification in nuclear energy

- Predict with confidence, accounting for errors and uncertainties
 - Modeling;
 - Algorithmic;
 - Software;
 - Epistemic (unknown unknowns);
 - Aleatoric (random phenomena)
 - Initial and boundary conditions
- Probabilistic risk assessment
 - Chance of high-consequence, rareoutcome event sequences
- Experiments
 - Unit & integrated experiments
- Extrapolation
 - Accident analysis
 - Aging of materials
 - Long-term repository performance
- Aggregation

Exascale Initiative Steering Committee

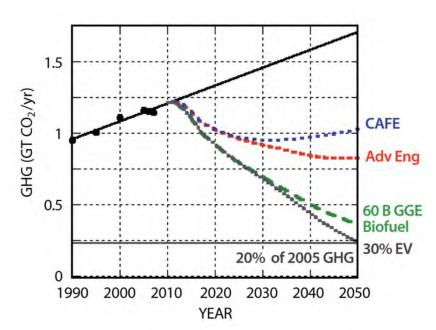


"For nuclear energy systems, there are two motivations to perfect Verification, Validation and Uncertainty Quantification (VU). The most obvious is to improve the confidence users have in simulations' predictive responses and our understanding of prediction uncertainties in simulations. Additionally, scientists must also perform VU for nuclear energy systems because the USNRC, the licensing body, requires it." *Science-based nuclear energy systems enabled by advanced modeling and simulation at extreme scale*, May 11-12, 2009



Automobiles and trucks account for 2/3 of petroleum use and 1/4 of CO₂ production

- Realizing thermodynamic limits with clean operation
 - Target: 30% more efficient engines
- Maintain clean fleet operation with evolving fuel streams
 - Low net carbon fuels, 25% reduction in CO₂ emission
- Optimize IC engines for PHEV use
 - Electrification, 25% reduction in CO₂ emission
 - PHEV 30% of fleet
 - EV 10% of fleet



Goals:

 20% reduction in GHG emissions by 2020, 80% by 2050.

• 3.5 M bbl/day reduction in petroleum usage

Predictive Simulation of Combustion Engine Performance in an Evolving Fuel Environment, 3/2010



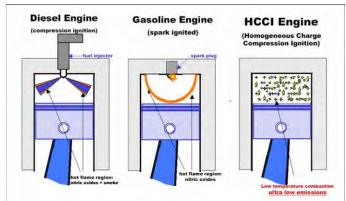
Fuel streams and engine technologies are rapidly evolving.

Workshops 1/19/2010 & 2/11/2010

Predictive Simulation of Combustion Engine Performance in an Evolving Fuel Environment

How can the current and future advances in scientific understanding and simulation of the combustion process be translated into the design and production of commercial, costeffective advanced engines so as to make a <u>significant and timely</u> contribution to President's goals?

- Fuel streams are rapidly evolving
 - Heavy hydrocarbons
 - Oil sands
 - Oil shale
 - Coal
 - Carbon-neutral renewable fuel sources
 - Iso-butanol
 - Biodiesel
- New engine technologies
 - Homogeneous Charge Compression Ignition (HCCI)
 - Low-temperature combustion

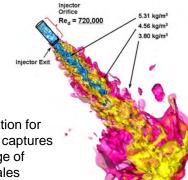


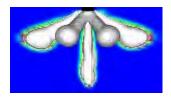


The use of computational science tools for engine design is rapidly evolving

- Current tools for engineering
 - Reynolds-Averaged Navier Stokes
 - Calculate bulk effects of turbulence
 - Turbulent combustion submodels calibrated over narrow range
- Current tools for science
 - Large Eddy Simulation (LES) with simple chemistry, normal pressures
 - Capture transient behavior
 - Physics-based turbulent combustion submodels for standard fuels

- Future tools for engineering are similar to those used for science
 - LES with complex chemistry, high pressures, multiphase transport
 - Cycle-to-cycle transient behavior for control strategy design
 - Physics-based turbulent combustion submodels for biofuels





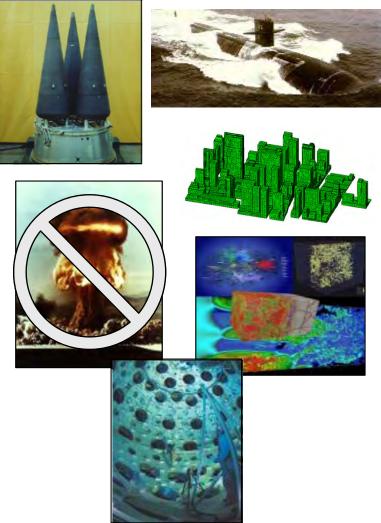
RANS calculation for fuel injector captures bulk behavior

LES calculation for fuel injector captures greater range of physical scales



National Nuclear Security

- U.S. Stockpile must remain safe, secure and reliable without nuclear testing
 - Annual certification
 - Directed Stockpile Work
 - Life Extension Programs
- A predictive simulation capability is essential to achieving this mission
 - Integrated design capability
 - Resolution of remaining unknowns
 - Energy balance
 - Boost
 - Si radiation damage
 - Secondary performance
 - Uncertainty Quantification
 - Experimental campaigns provide critical data for V&V (NIF, DARHT, MaRIE)
- Effective exascale resources are necessary for prediction and quantification of uncertainty

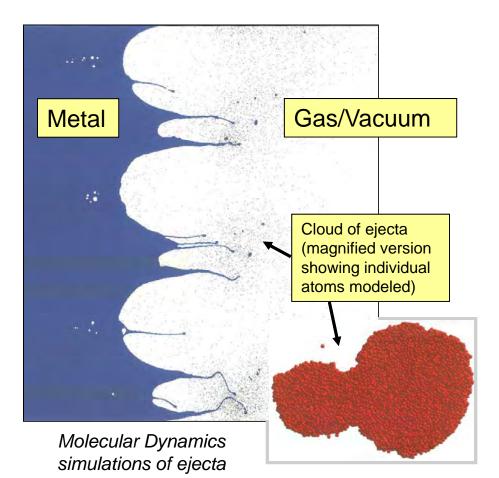




Increased computational power is driven by requirements to reduce uncertainty.

- Improved GEOMETRIC fidelity
 - Safety, surety & security features
 - Design features
 - UQ methodologies
 - Naturally 3D phenomena e.g. turbulence, material failure
- Improved NUMERICAL fidelity
 - TBI uncovers potentially important phenomena at 10x standard resolution
 - Bridging strongly-coupled multi-scale
 phenomena
 - Need to perform UQ studies over greater variable counts
 - Weapons science simulations displaying convergence at very large coverage
 - atoms or dislocations or
- Improved PHYSICS fidelity
 - Energy balance
 - Boost
 - Si radiation damage
 - Secondary performance

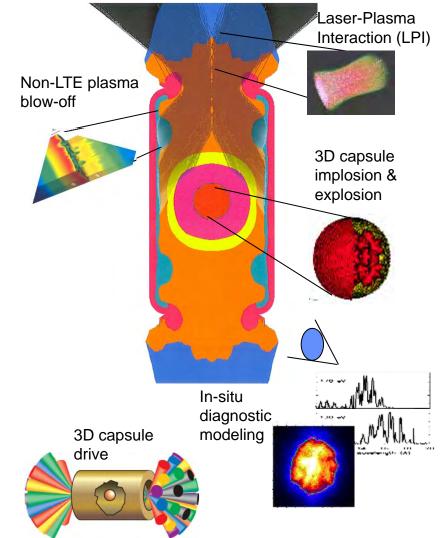
3D simulations of ductile spall failure with predictive potentials will require exascale resources





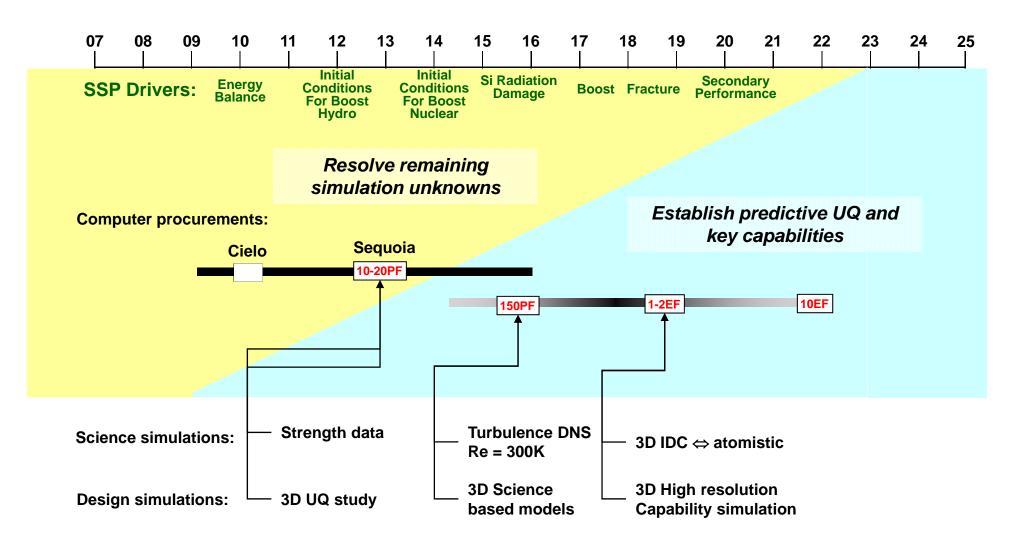
Validation demands HED experiments requiring exascale computing.

- Predictive design, diagnosis & interpretation of complex HED experiments demand detailed 3D multi-physics simulations
- Three major requirements drive computational costs for these simulations to the exascale level
 - Improved Physics
 - Laser beam propagation, LPI , plasma blowoff and effect on drive and symmetry
 - Predictive models of capsule implosion and explosion symmetry
 - Detailed atomic physics and line radiation transport to model spectroscopic output
 - Improved resolution
 - Enhanced spatial, temporal, and spectral resolution required to converge the models
 - Improved understanding
 - Large sensitivity studies to quantify the uncertainties of aggregating key physics





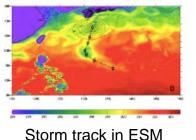
Stockpile stewardship requires exascale resources to establish predictive 3D UQ

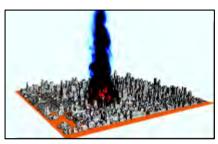




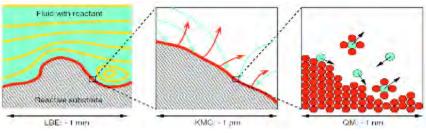
Cross-cutting capabilities are critical to success for DOE mission and science

- Uncertainty quantification
 - Predict climate response to energy technology strategies
 - Assessment of safety, surety and performance of the aging/evolving stockpile without nuclear testing
 - Energy security
 - Responding to natural and manmade hazards
- Multi-scale, multi-physics modeling
 - Multiple physics packages in earth system model: ocean, land surface, atmosphere, ice
 - Multiple physics packages in modeling reactor core: neutronics, heat transfer, structures, fluids
- Statistics of rare events
 - Severe weather and surprises in climate system
 - Accident scenarios in nuclear energy
 - Nucleation of cracks and damage in materials

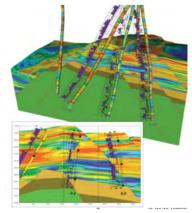




urban fire



Multi-scale physics modeling

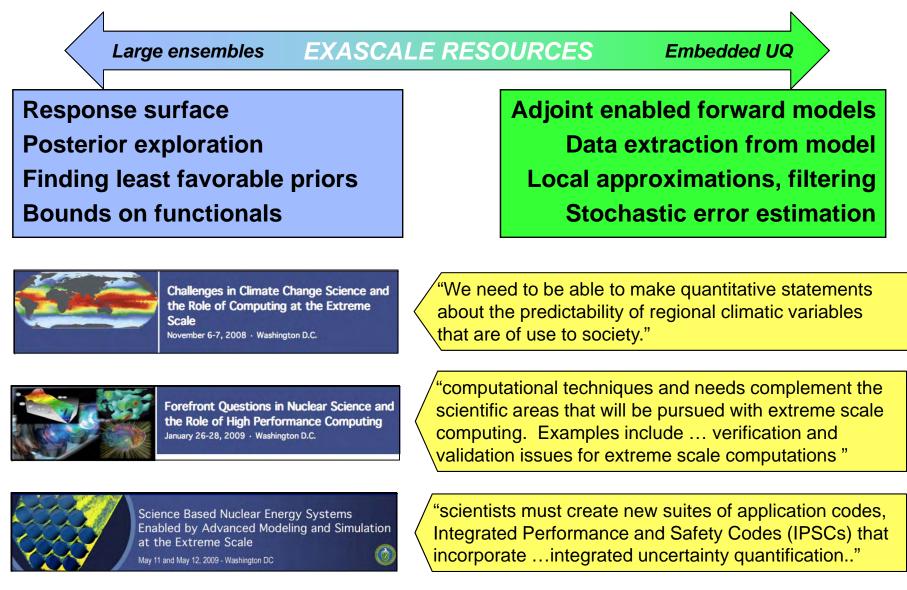




Geologic sequestration



Uncertainty quantification is critical and requires exascale resources.





Uncertainty comes in a variety of shapes and sizes

	Parametric	Structural	Relational
Theory & models	•Calibrated parameters in models	•Unknown effects omitted from models •Extrapolation	 Multi-scale, multi- physics effects
Algorithms	•Discretization error	•Extrapolation	•Multiple time scales in operator split algorithms
Applications code	•Convergence criteria	•Errors in apps code	•Data mapping among different components
Computation and communication	•Rounding errors	•Silent data corruption	•Race conditions among separate components of system
Operating system & environment	•ECC error rates (chip bit errors)	 System parameters set incorrectly Chip temperature excursions 	•System policy mis- match (e.g. memory management)
Observations & data assimilation	 Statistical variation in experimental data 	•Unknown systematic errors in data	•Contextual mismatch of observational and computational data



Scope of a DOE Exascale Initiative

