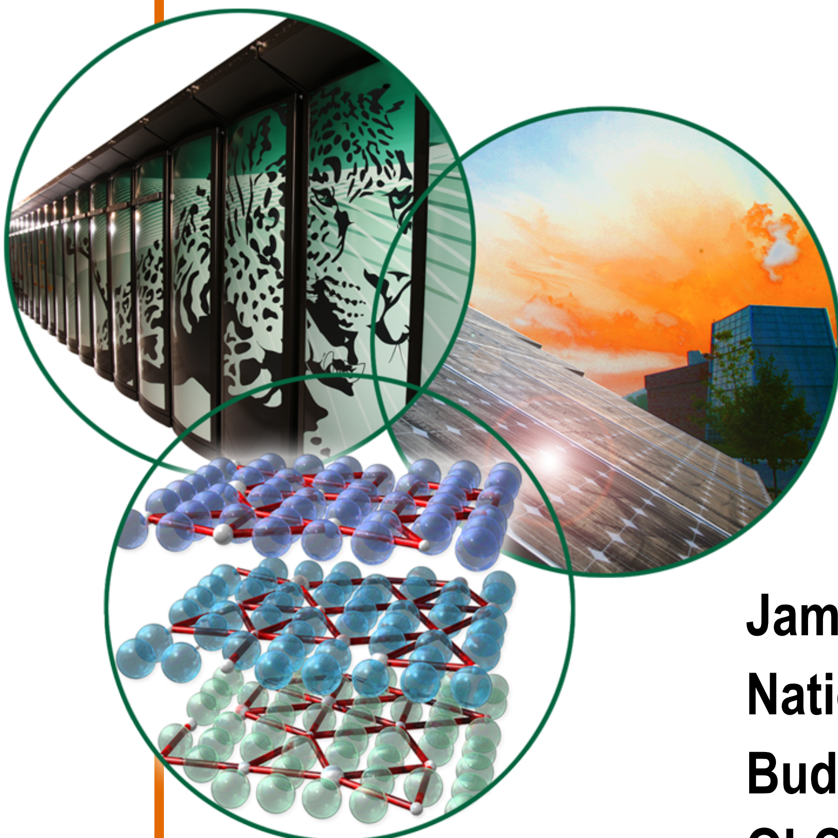
Transformational Changes Facing Climate Change Science



James J. Hack, Director
National Center for Computational Sciences
Buddy Bland
OLCF Project Director
Oak Ridge National Laboratory

Outline

Paradigm shifts in the conduct of climate change science

- **Changes in expectations**
- **New edges on science priorities**
- **Technology transformations**

Climate Science is positioned to experience a wide range of transformational changes including rapidly growing societal expectations that will accelerate the need for an improved scientific understanding of the climate system. At the same time, enabling computational technologies will be undergoing a disruptive transition to new and more complex architectures required to move high-performance scientific computing performance by another factor of 1000 before the end of this decade. This talk will touch on several of the growing range of climate science applications, some of the science that will be needed to enable those applications, and the challenges associated with navigating a complex paradigm change in computational architectures.

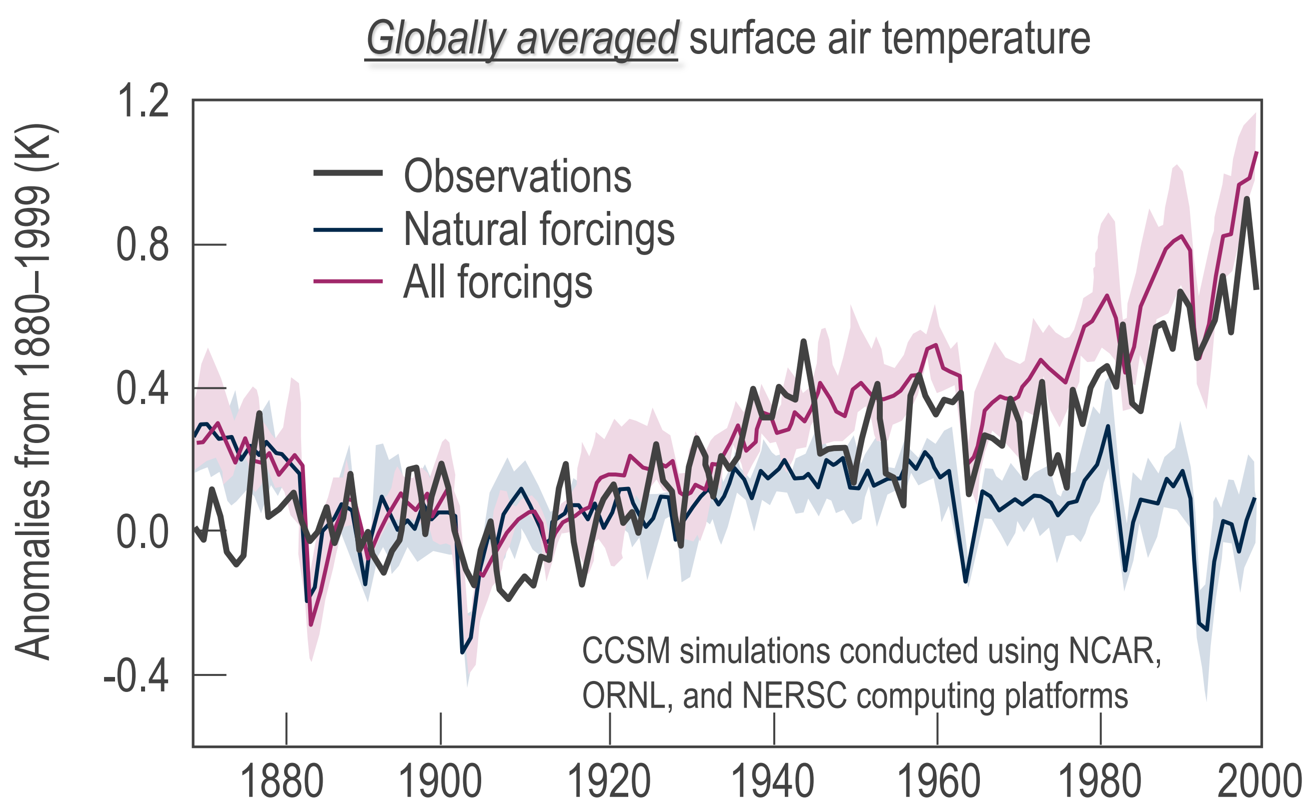
Energy is the defining challenge of our time

- The major driver for
  - *Climate change*
  - National security
  - Economic competitiveness
  - Quality of life

Global energy consumption will increase 50% by 2030

- Incremental changes to existing technologies cannot meet this challenge
  - Transformational advances in energy technologies are needed
  - Transformational adaptation strategies will need to be implemented

The IPCC AR4 Success: Attribution



Attribution **of human-induced changes to global climate**

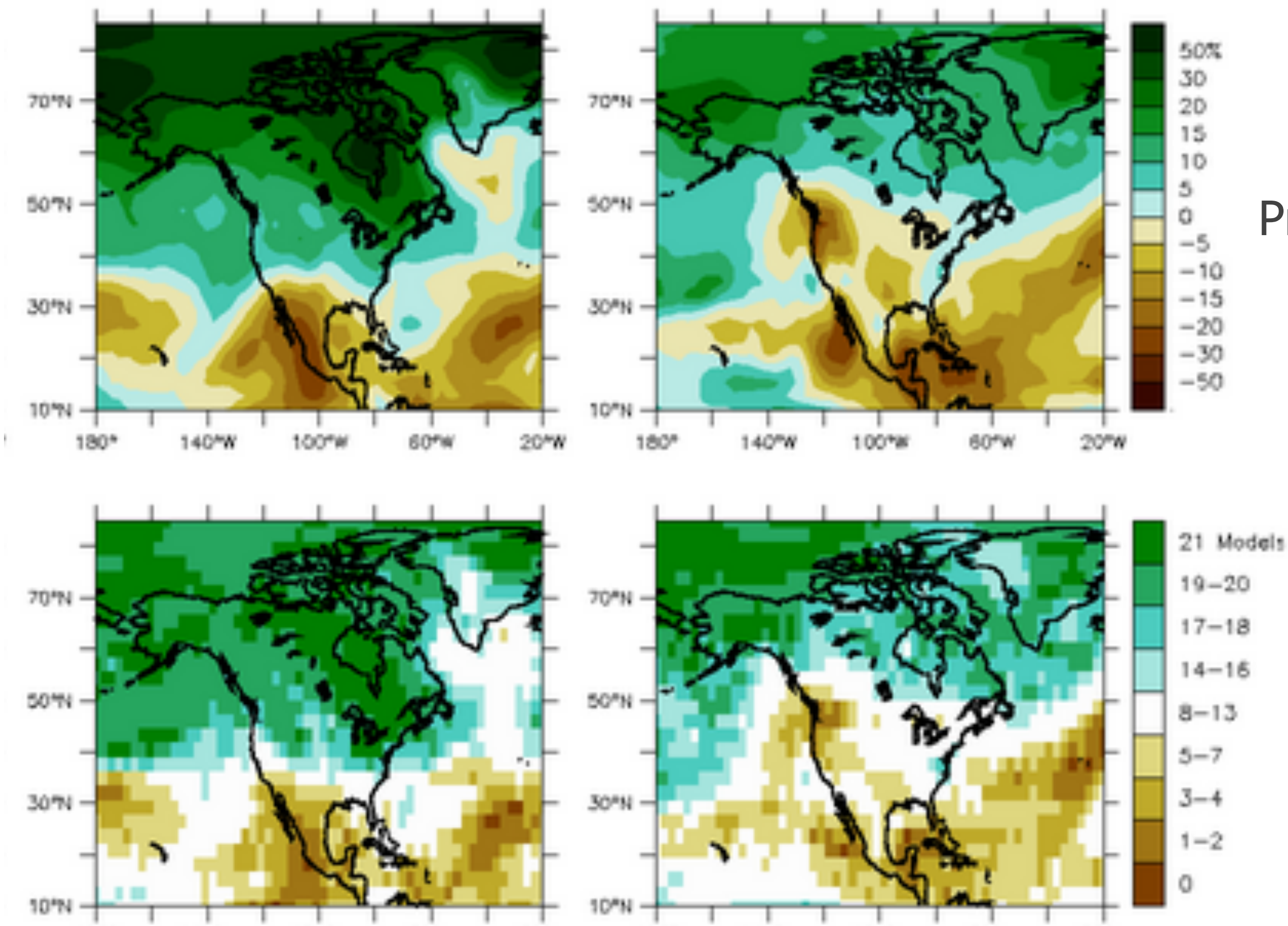
Regional Climate Change (AR4)

2080-2099 (A1B) - 1980-1999

CMIP5 models are looking very much the same!

DJF

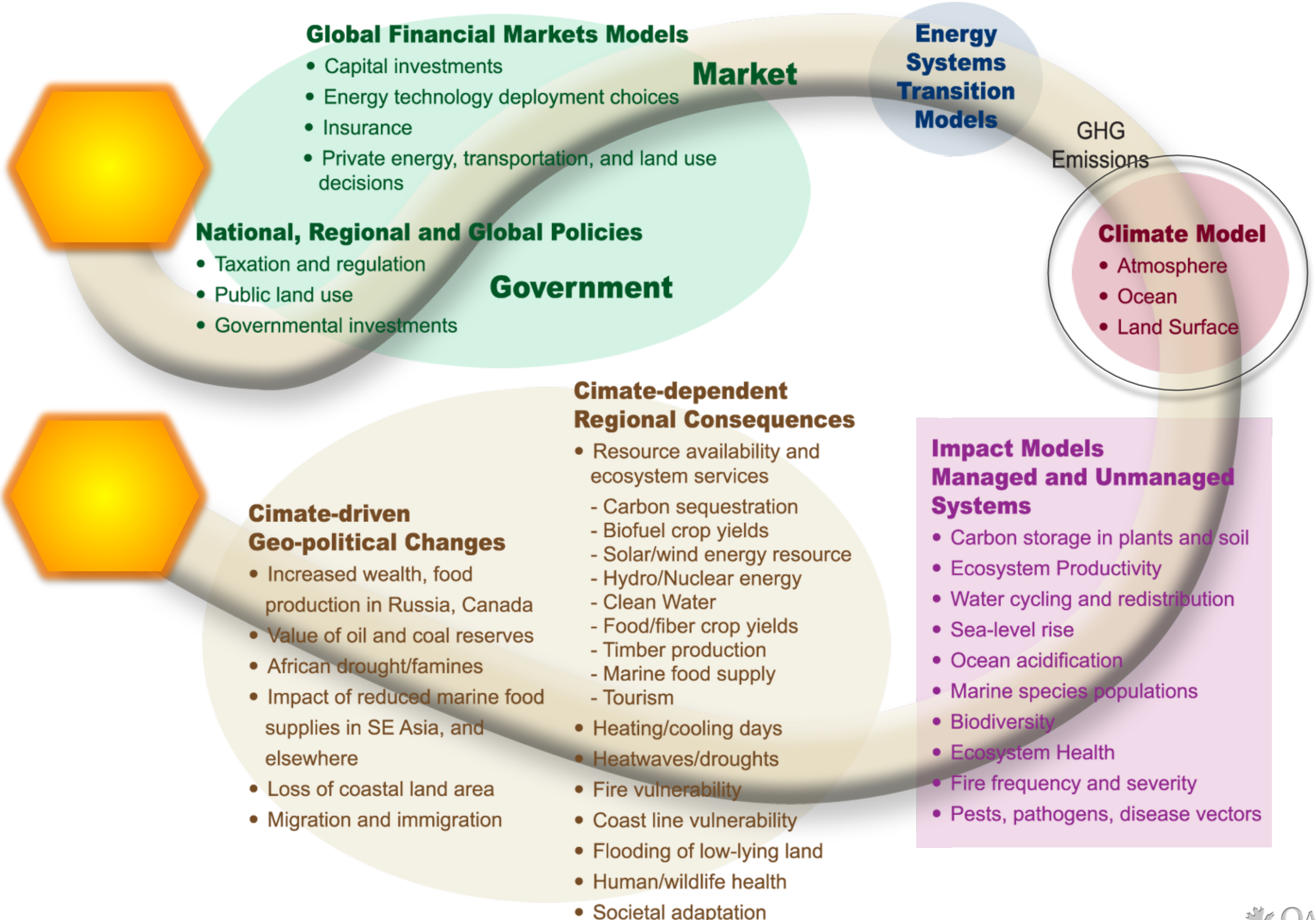
JJA



Precipitation (%)

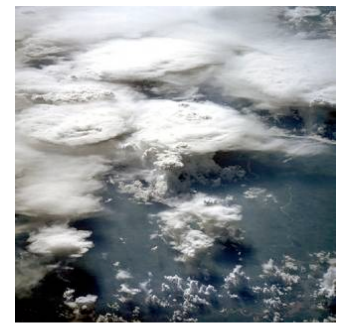
# of Models with $\Delta P > 0$

Global Change: A complex system of systems

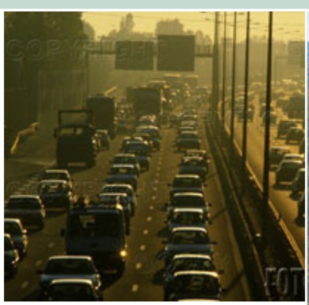


The Climate/Energy/Carbon/Water Challenge

Sustainable production and use of interrelated resources on a constrained and changing Earth



Energy	Carbon	Water
• Production, distribution, and use • Economic drivers • Environmental drivers	• Food, fiber, fuel • Ecosystem health (e.g., biodiversity) • Managing carbon for mitigation of climate change	• Quality and quantity • Energy requires water; Water requires energy • Many critical climate change impacts are water related



Climate Change Impacts Relevant to National Security

The Requirement...

From the FY08 National Defense Authorization Act:

SEC. 951. ***"DEPARTMENT OF DEFENSE CONSIDERATION OF EFFECT OF CLIMATE CHANGE ON DEPARTMENT FACILITIES, CAPABILITIES, AND MISSIONS"***

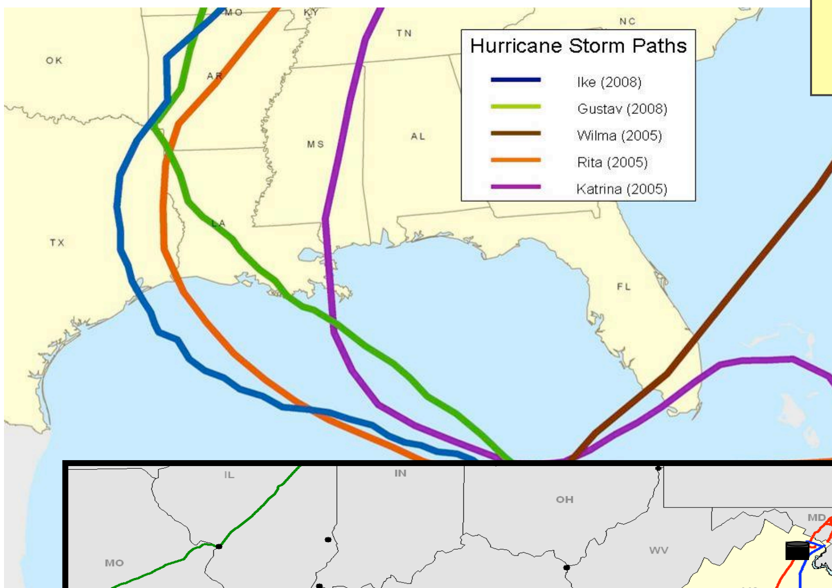
"The first ***Quadrennial Defense Review*** prepared after the date of the enactment of this subsection shall also ***examine the capabilities of the armed forces to respond to the consequences of climate change, in particular, preparedness for natural disasters from extreme weather events and other missions the armed forces may be asked to support inside the United States and overseas.*** "

Need to Characterize Extreme Events

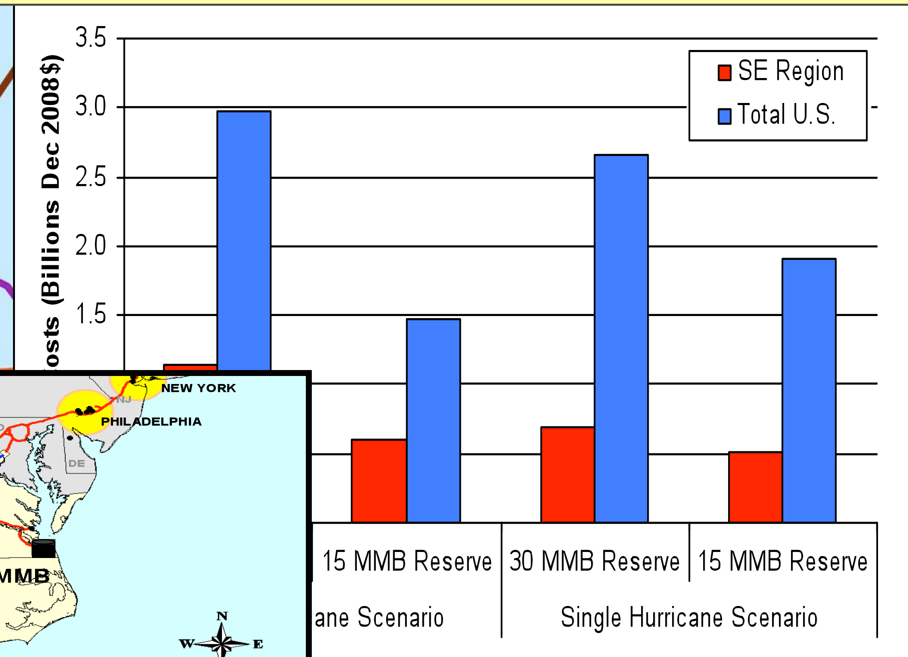
Storms, Floods, Droughts, Cyclones

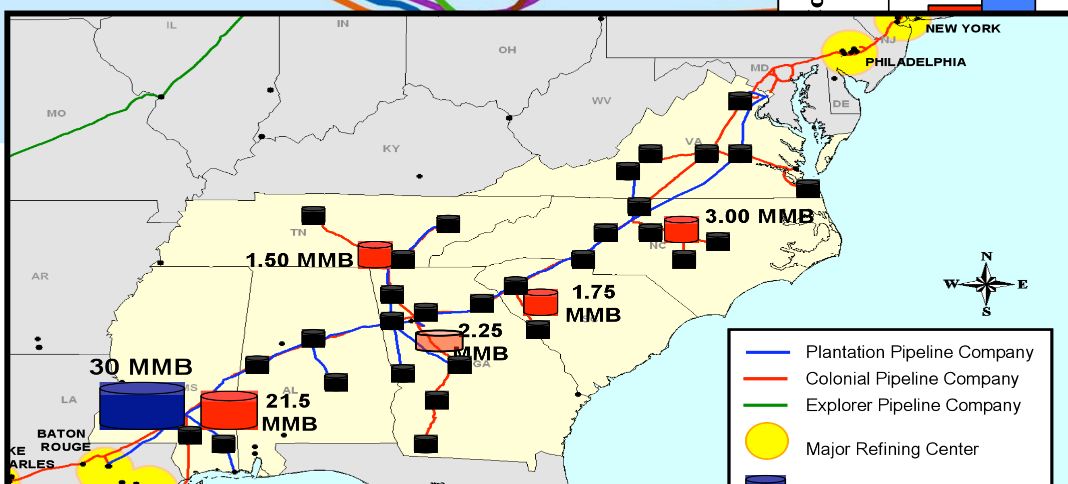
- **More frequent droughts and periods of intense precipitation**
- **Direct loss of life and injury**
- **Indirect effects**
  - **Loss of shelter**
  - **Population displacement**
  - **Contamination of water supplies**
  - **Loss of food production**
  - **Increased risk of infectious disease epidemics (diarrhoeal and respiratory)**
  - **Damage to infrastructure for provision of health services**

Preparation and Response to Severe Weather Events



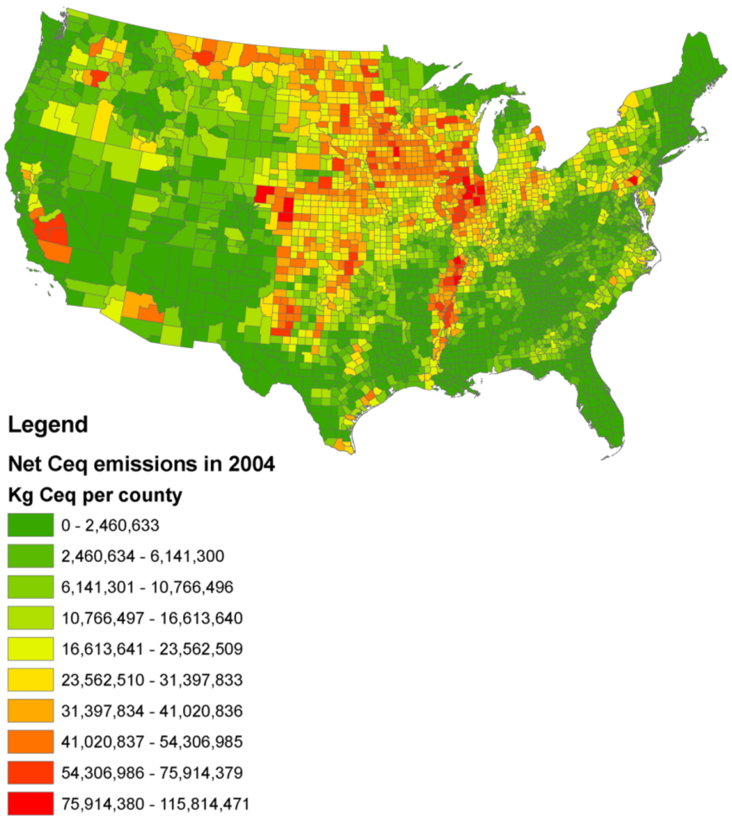
Evaluate the Benefits of Individual Options against Costs



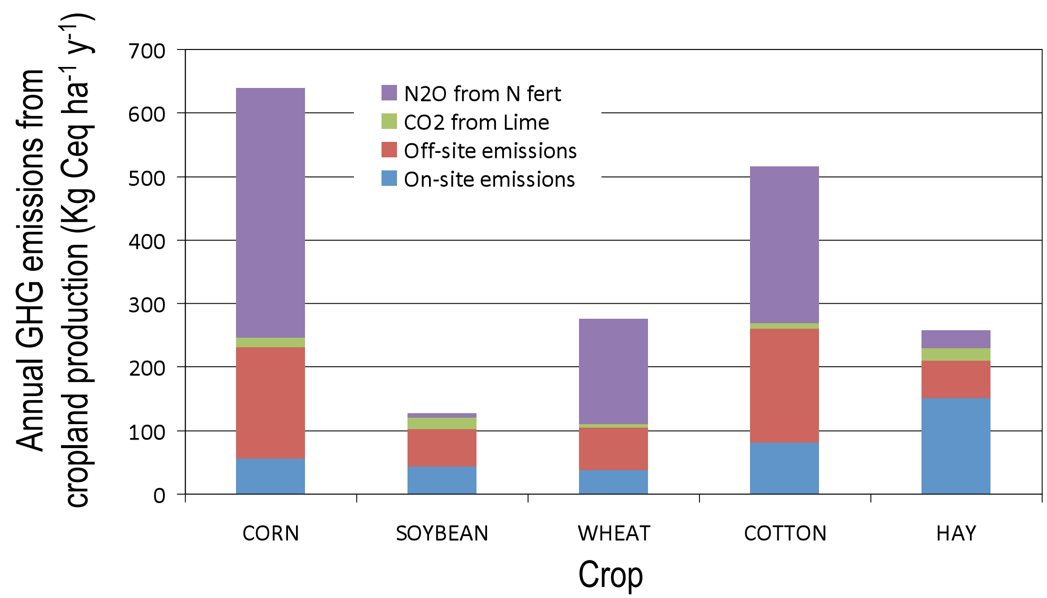


Evaluate Options for Replacing Supply Losses

What are current cropland emissions based on land management?



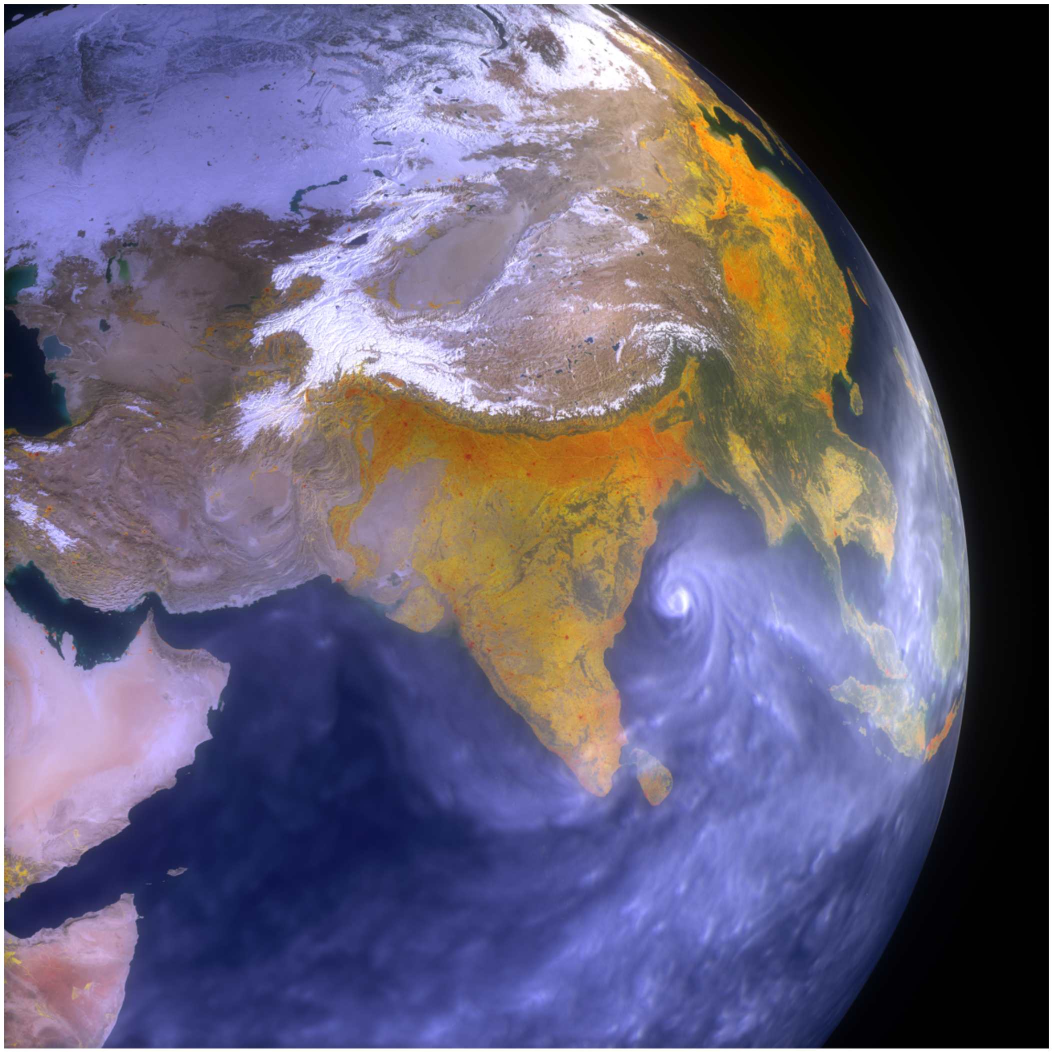
Energy and CO_2-equivalent emissions associated >3500 land management options, representing >96% cropland management in the United States.



Must understand current carbon sources and sinks to (a) quantify the impact of climate change on these sources and sinks, and to (b) change management for mitigation and adaptation purposes

New Scientific Opportunities

Continued scientific and computational science research coupled with advances in computational technology opens new opportunities

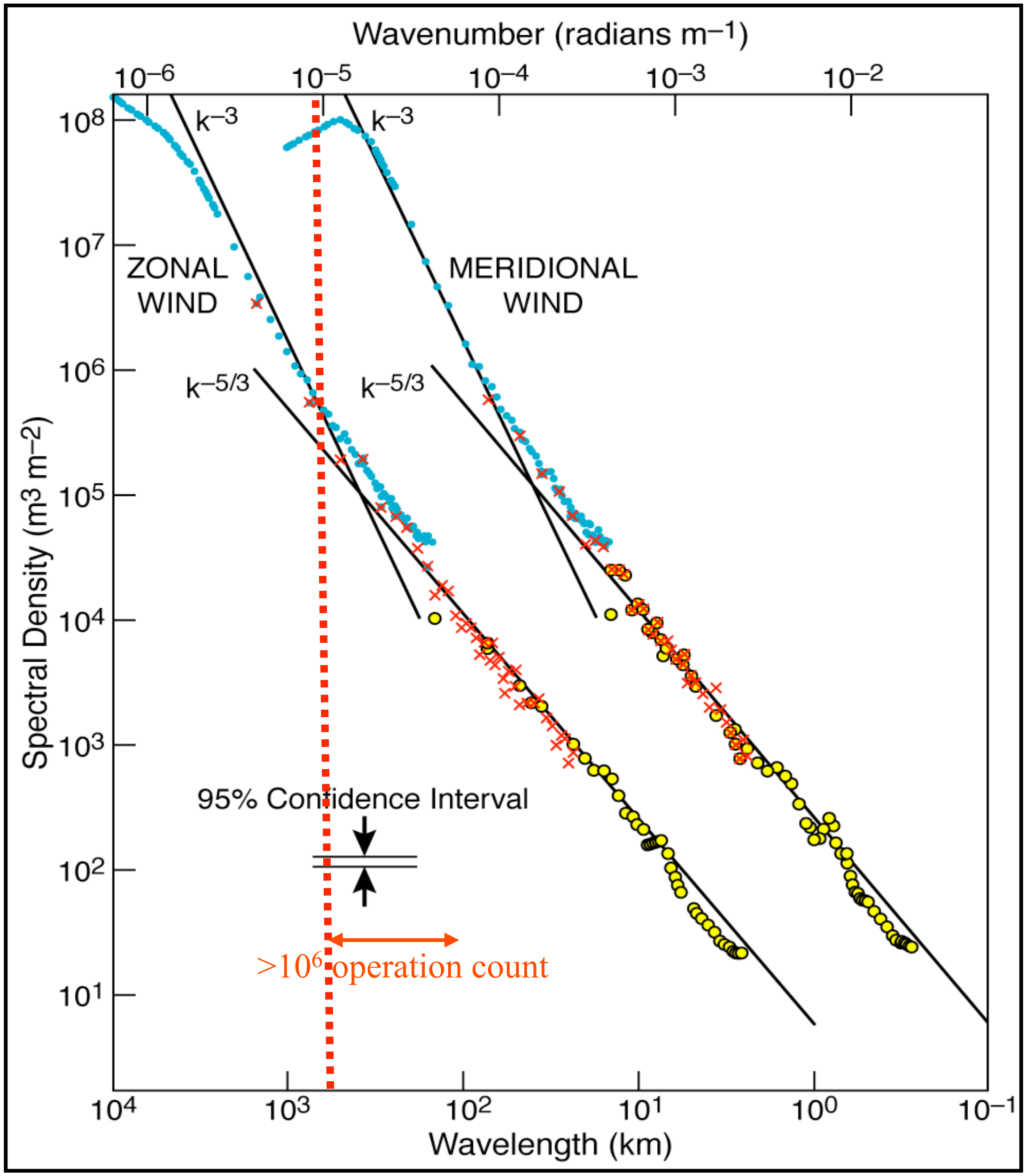


Exploring population dynamics in the context of adaptation to water availability, coastal vulnerability using Landscan population dataset

Examples of climate consequences questions

- **Water Resources**
  - management and maintenance of existing water supply systems, development of flood control systems and drought plans
- **Agriculture and food security**
  - Erosion control, dam construction (irrigation), optimizing planting/harvesting times, introduction of tolerant/resistant crops (to drought, insect/pests, etc.)
- **Human health**
  - Public health management reform, improved urban and housing design, improved disease/vector surveillance and monitoring
- **Terrestrial ecosystems**
  - Improvement of management systems (deforestation, reforestation,…), development/improvement of forest fire management plans
- **Coastal zones and marine ecosystems**
  - Better integrated coastal zone planning and management
- **Human-engineered systems**
  - Better planning for long-lived infrastructure investments

Kinetic energy spectra observations



Nastrom and Gage, 1985

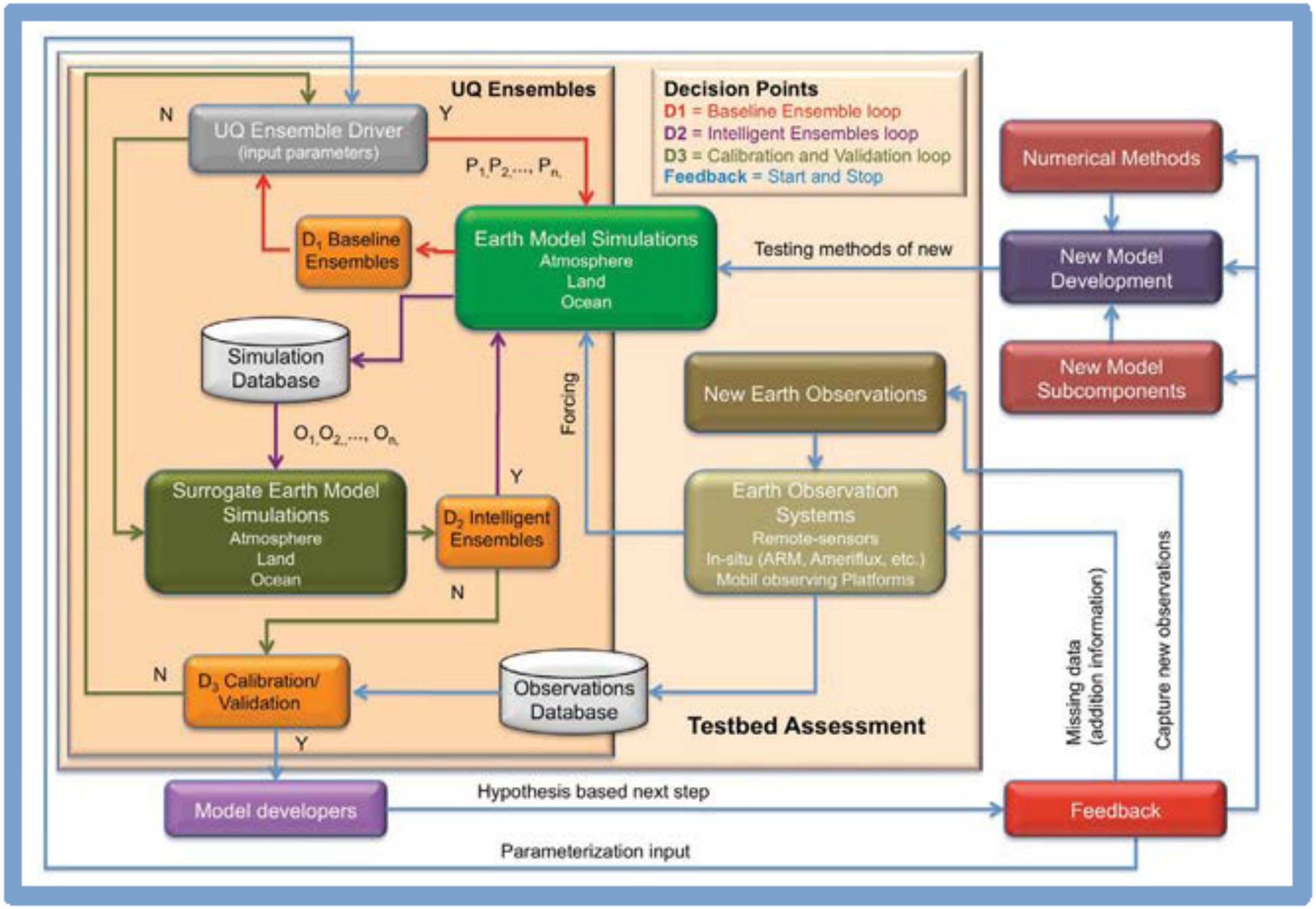
Unresolved processes

- Traditionally assume that effects of unresolved scales can be represented through parameterizations based on bulk formulae.
- Inherently assumes that there is no coupling between dynamics and physics on these unresolved scales.
- Essentially ignores upscale energy cascades

Original image from Julia Slingo

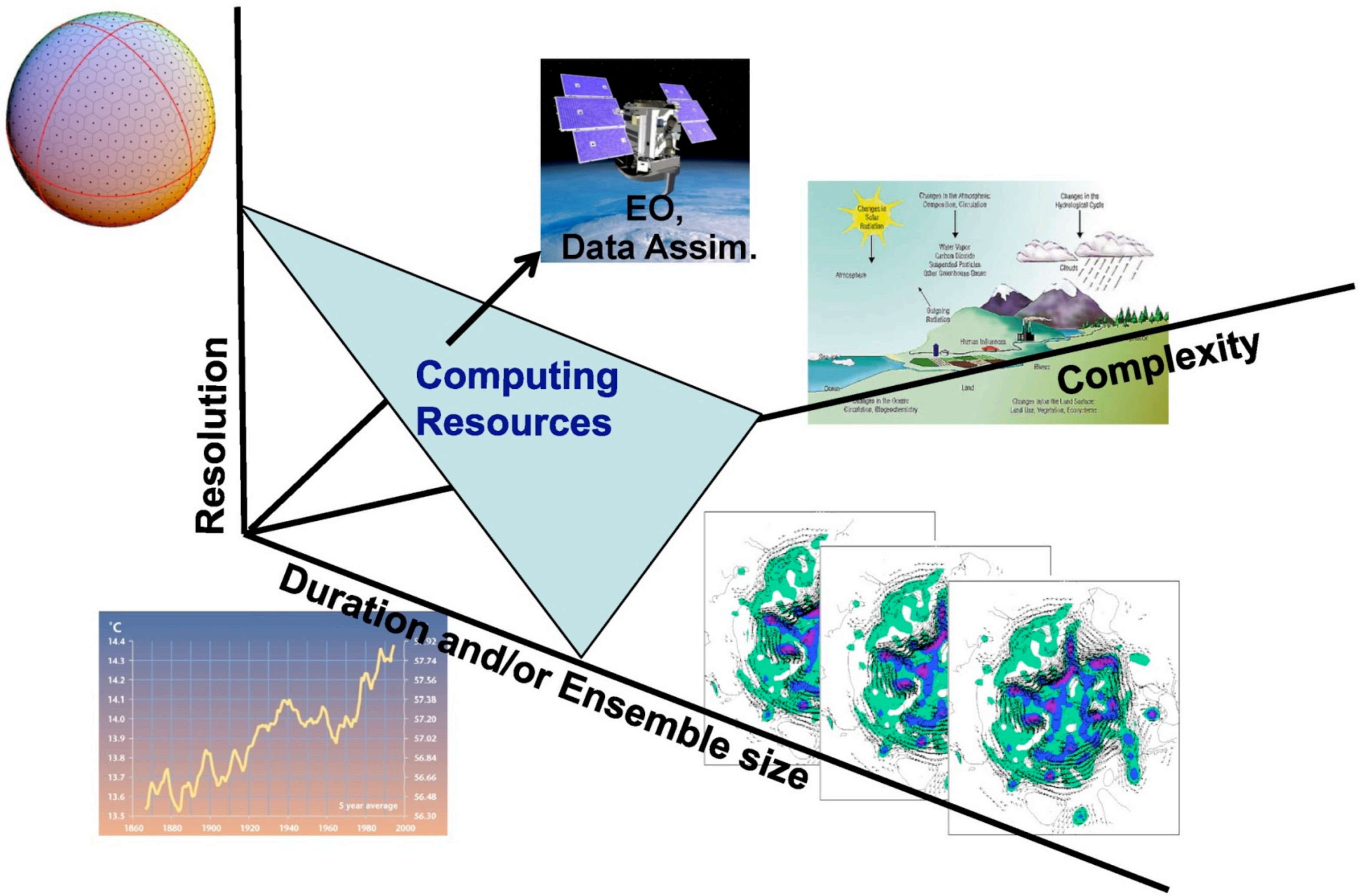
CSSEF Testbed

Computationally demanding, including workflow



The CSSEF testbed and development system was designed by the crosscutting Data Infrastructure and Testbed team based on requirements set by CSSEF modelers and observational scientists.

Advances in Predictive Science will be enabled by advanced computational facilities



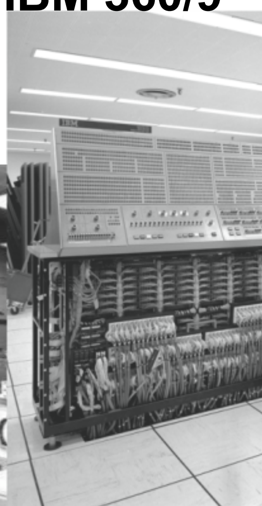
OAK RIDGE
National Laboratory

ORNL has a long history in High Performance Computing

ORNL has had 20 systems on the TOP500® SUPERCOMPUTER SITES lists

1954
ORACLE

1969
IBM 360/9



1985
Cray X-MP

1992-1995
Intel Paragons

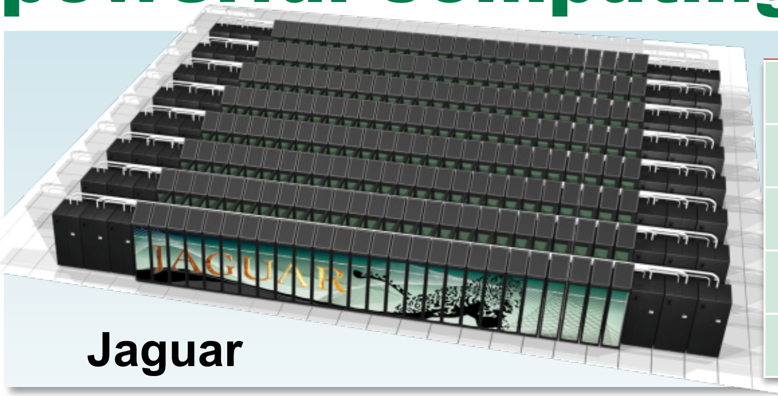
1996-2002
IBM Power 2/3/4

2007
IBM Blue Gene/P

2003-2005
Cray X1/X1E

OAK RIDGE National Laboratory

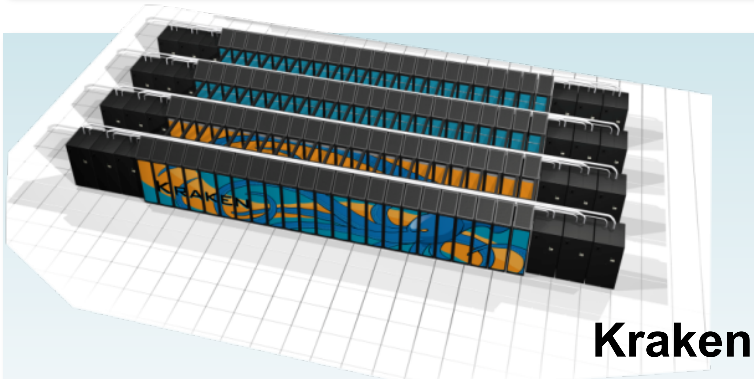
Today, ORNL has one of the world's most powerful computing facilities



Peak performance	3.27 PF/s
Memory	600 TB
Disk bandwidth	> 240 GB/s
Square feet	5,000
Power	7 MW

TOP500 SUPERCOMPUTER SITES #6

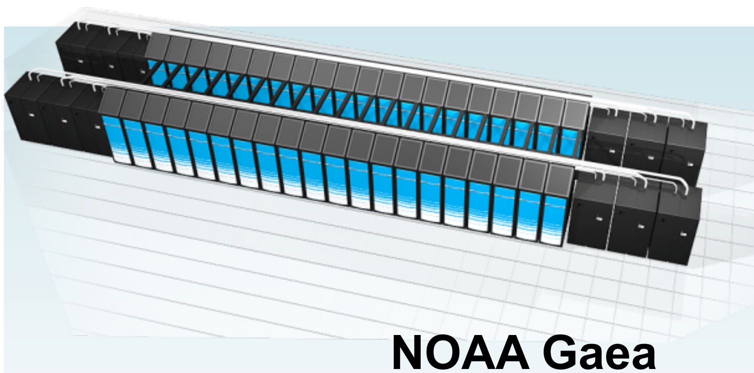
Dept. of Energy's most powerful computer



Peak performance	1.17 PF/s
Memory	151 TB
Disk bandwidth	> 50 GB/s
Square feet	2,300
Power	3.5 MW

TOP500 SUPERCOMPUTER SITES #21

National Science Foundation's most powerful computer



Peak Performance	1.1 PF/s
Memory	248 TB
Disk Bandwidth	104 GB/s
Square feet	1,600
Power	2.2 MW

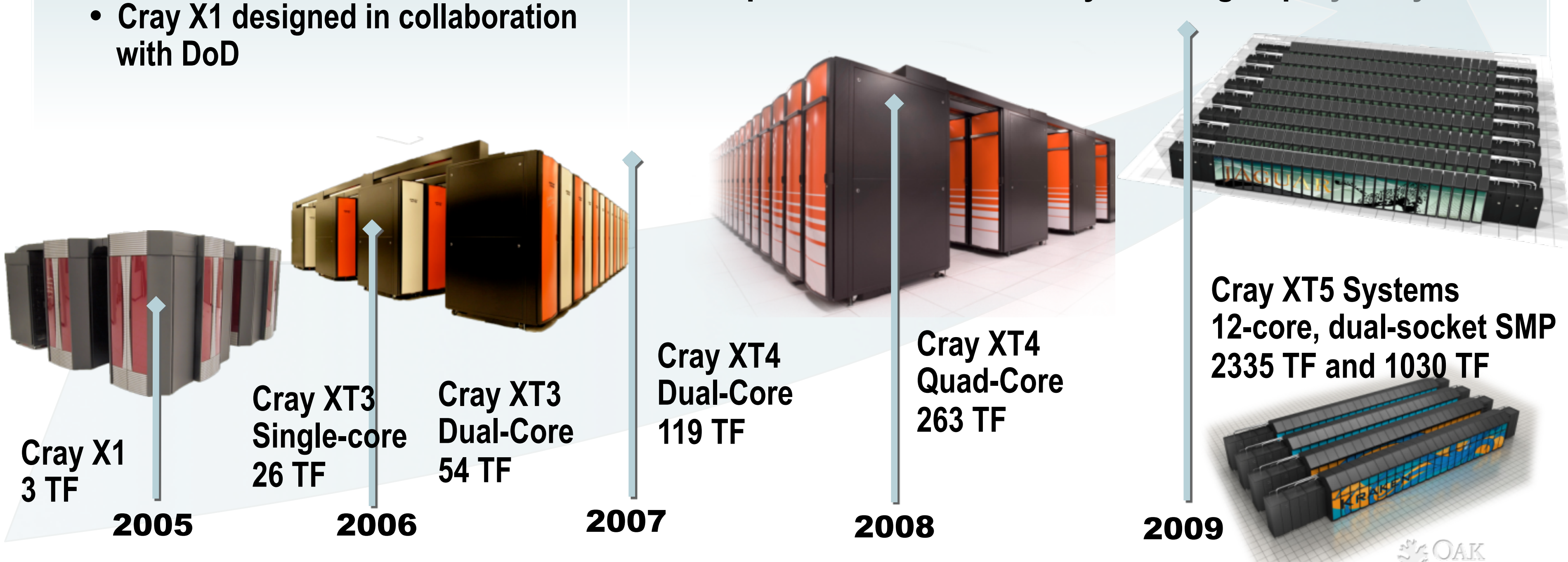
TOP500 SUPERCOMPUTER SITES #33

National Oceanic and Atmospheric Administration's most powerful computer

Systems have increased system performance by 1,000 times since 2004

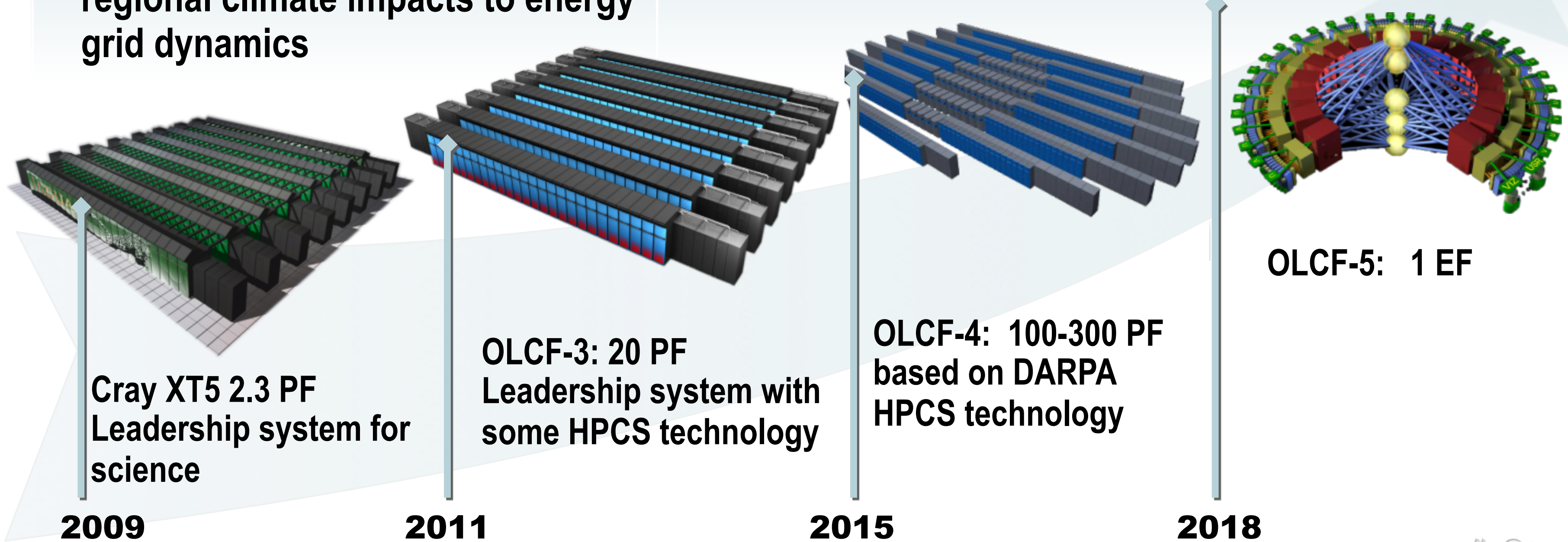
Hardware scaled from single-core through dual-core to quad-core and dual-socket , 12-core SMP nodes	Scaling applications and system software is the biggest challenge

• NNSA and DoD have funded much of the basic system architecture research
  • Cray XT based on Sandia Red Storm
  • IBM BG designed with Livermore
  • Cray X1 designed in collaboration with DoD
	• DOE SciDAC and NSF PetaApps programs are funding scalable application work, advancing many apps
	• DOE-SC and NSF have funded much of the library and applied math as well as tools
	• Computational Liaisons key to using deployed systems



Climate Science requires 1000x advances in computational capability over next decade

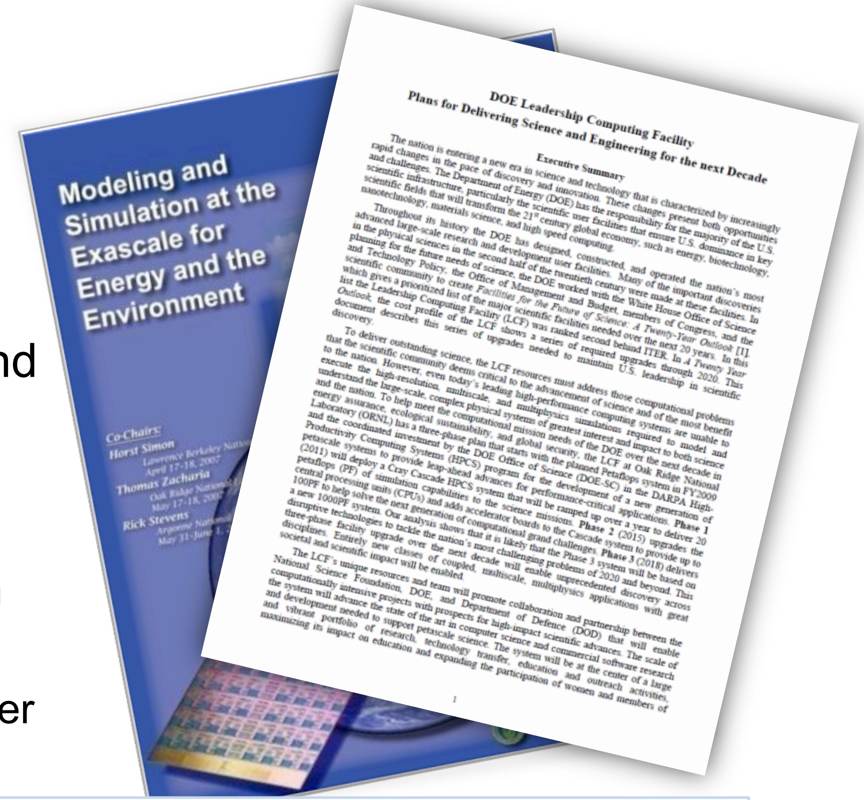
Mission: Deploy and operate the computational resources required to tackle global challenges	Vision: Maximize scientific productivity and progress on the largest scale computational problems
• Deliver transforming discoveries in climate, materials, biology, energy technologies, etc. • Ability to investigate otherwise inaccessible systems, from regional climate impacts to energy grid dynamics	• Providing world-class computational resources and specialized services for the most computationally intensive problems • Providing stable hardware/software path of increasing scale to maximize productive applications development

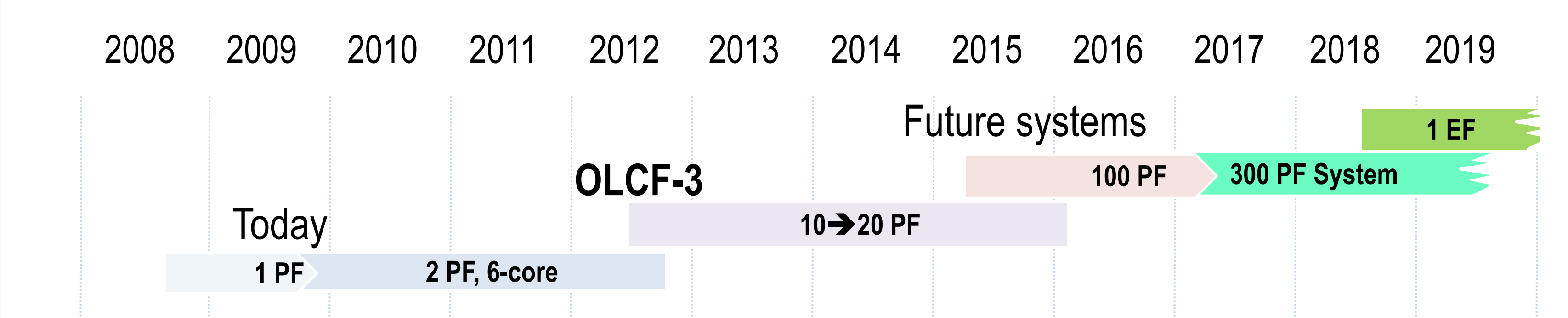


OAK RIDGE National Laboratory

10 Year Strategy: Moving to the Exascale

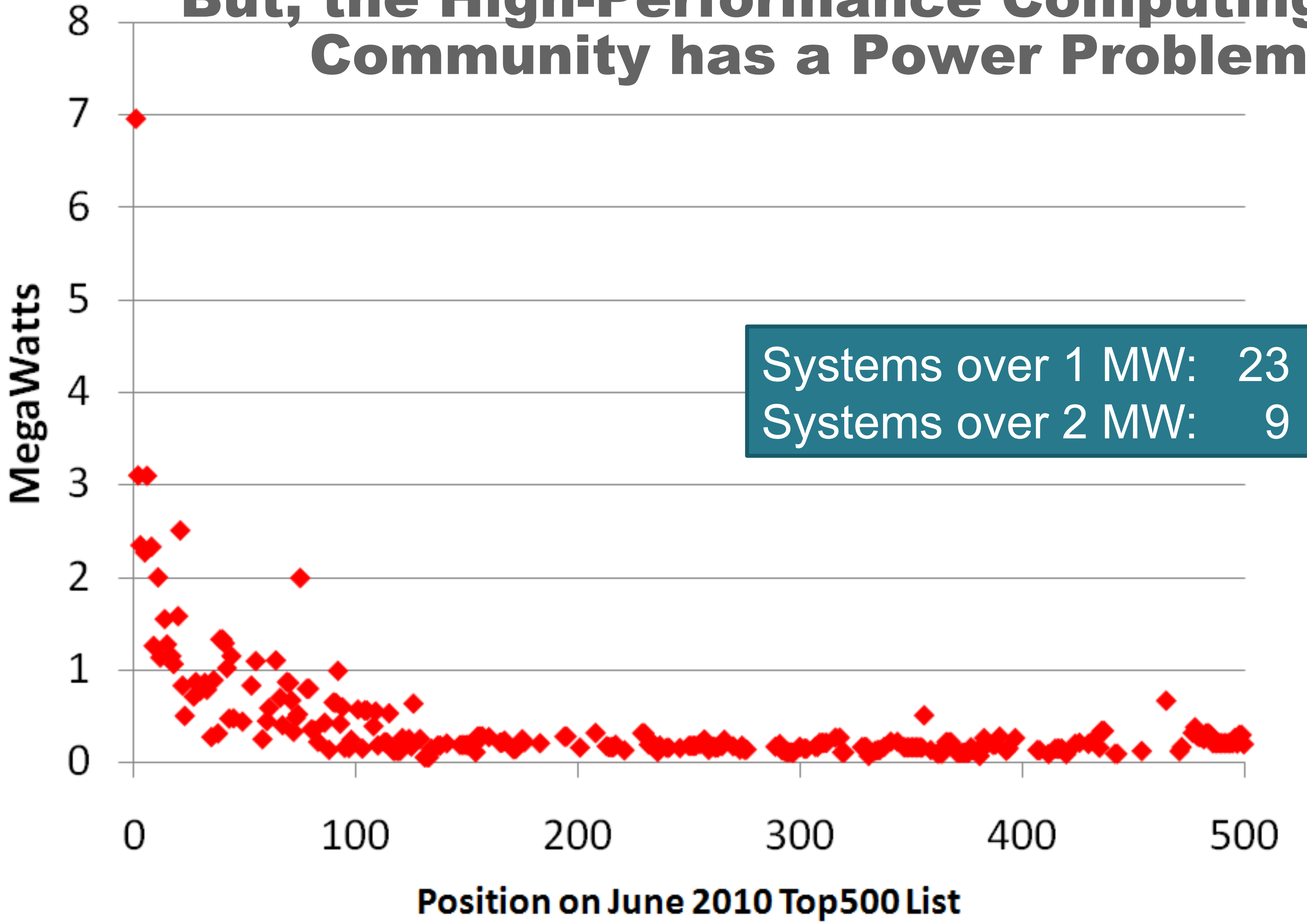
- The U.S. Department of Energy requires exaflops computing by 2018 to meet the needs of the science communities that depend on leadership computing
- Our vision: Provide a series of increasingly powerful computer systems and work with user community to scale applications to each of the new computer systems
  - **OLCF-3 Project**: New 10-20 petaflops computer based on early hybrid multi-core technology





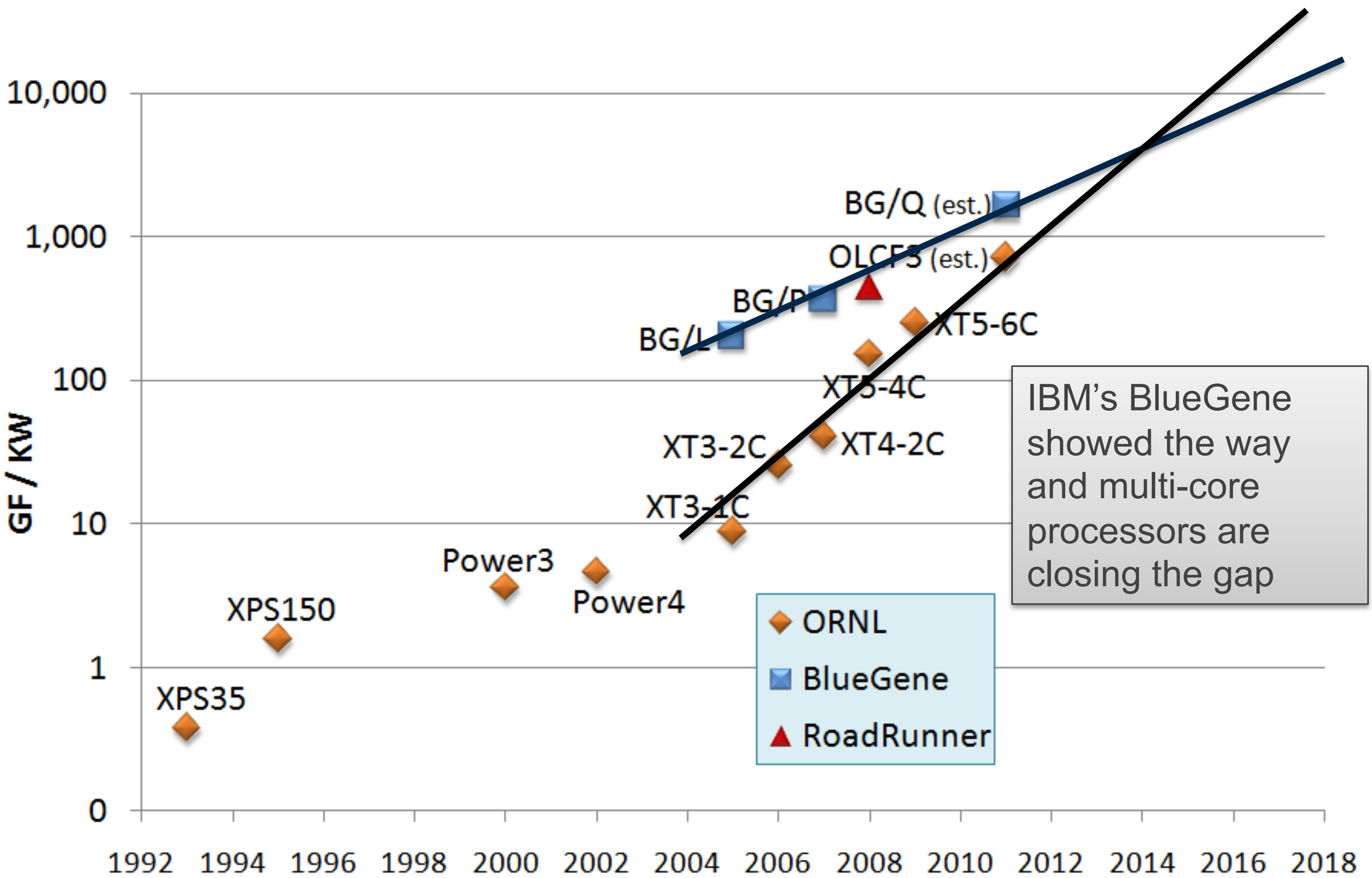
OLCF Roadmap from 10-year plan

But, the High-Performance Computing Community has a Power Problem!

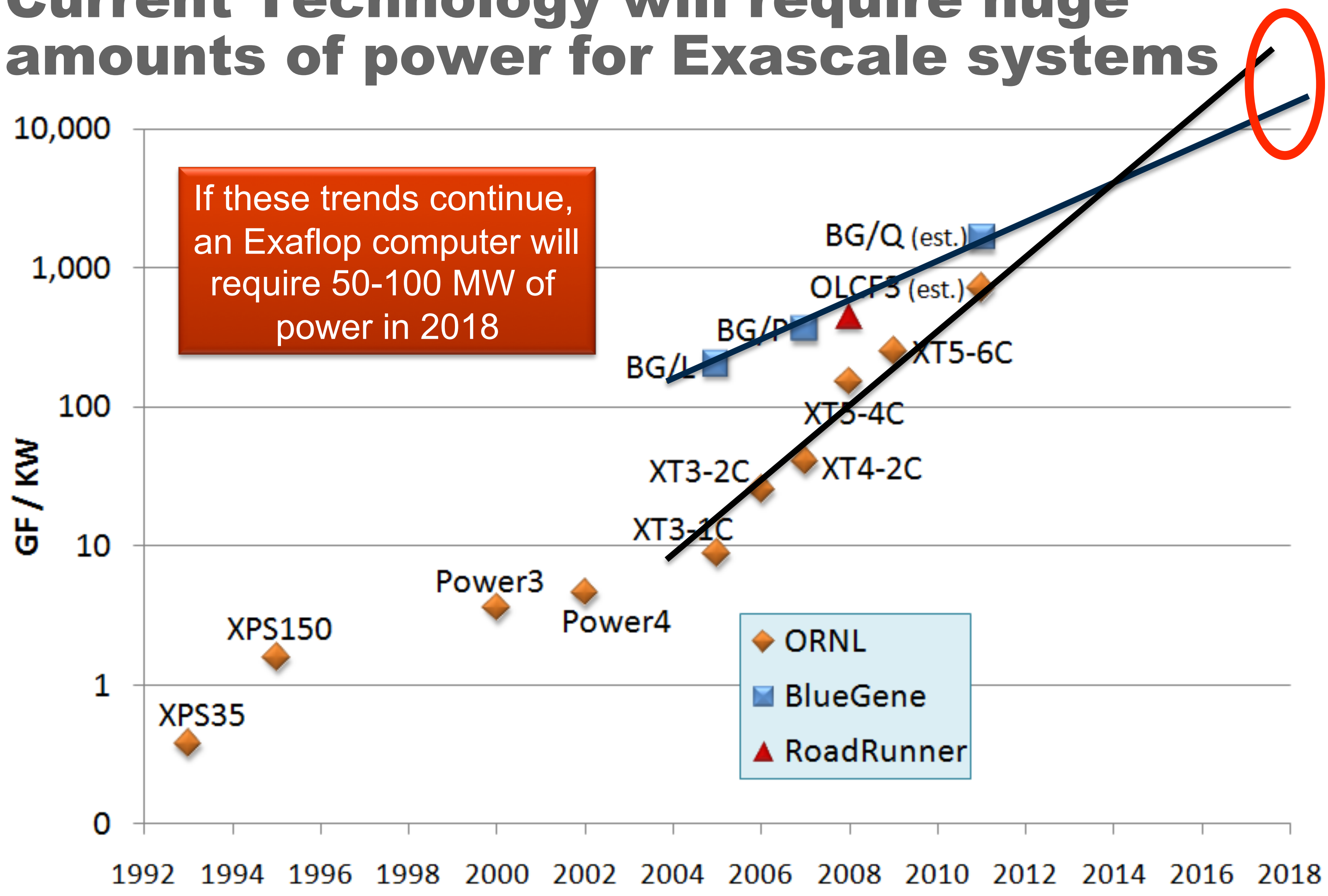


Data from http://www.Top500.org

Trends in power efficiency



Current Technology will require huge amounts of power for Exascale systems

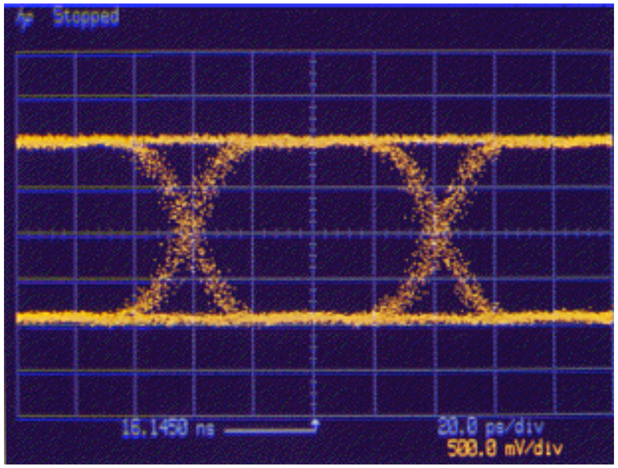


OAK RIDGE National Laboratory

Why has clock rate scaling ended?

Power ∝ Capacitance * Frequency * Voltage^2 + Leakage

- Traditionally, as Frequency increased, Voltage decreased, keeping the total power in a reasonable range
- But we have run into a wall on voltage
  - As the voltage gets smaller, the difference between a “one” and “zero” gets smaller. Lower voltages mean more errors.
  - While we like to think of electronics as digital devices, inside we use analog voltages to represent digital states. But this is imperfect!



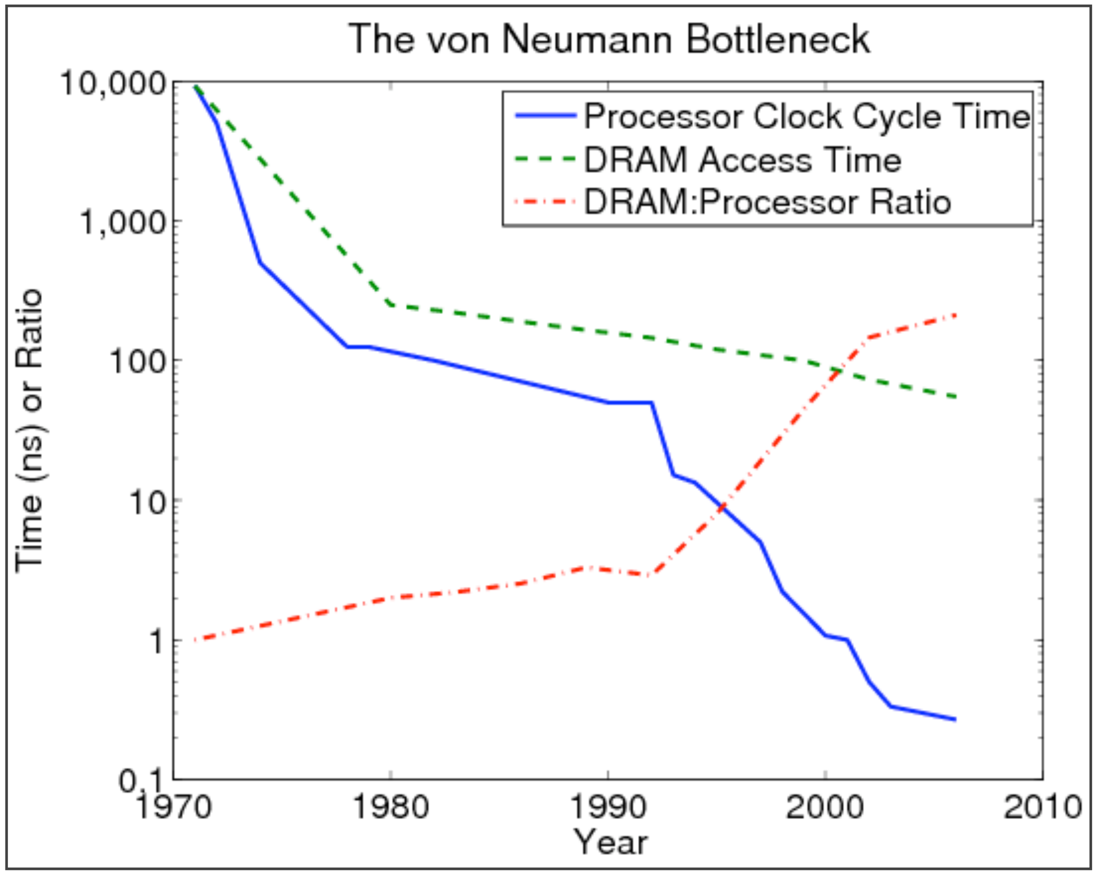
- Capacitance increases with the complexity of the chip
- Total power dissipation is limited by cooling

And Then There's the Memory Wall

"FLOPS are 'free'. In most cases we can now compute on the data as fast as we can move it." - Doug Miles, The Portland Group

What we observe today:

- Logic transistors are free
- The von Neumann architecture is a bottleneck
- Exponential increases in performance will come from increased concurrency not increased clock rates if the cores are not starved for data or instructions



Power to move data

$$Energy\_to\_move\_data \propto bitrate * length^2 / cross\_section\_area\_of\_wire$$

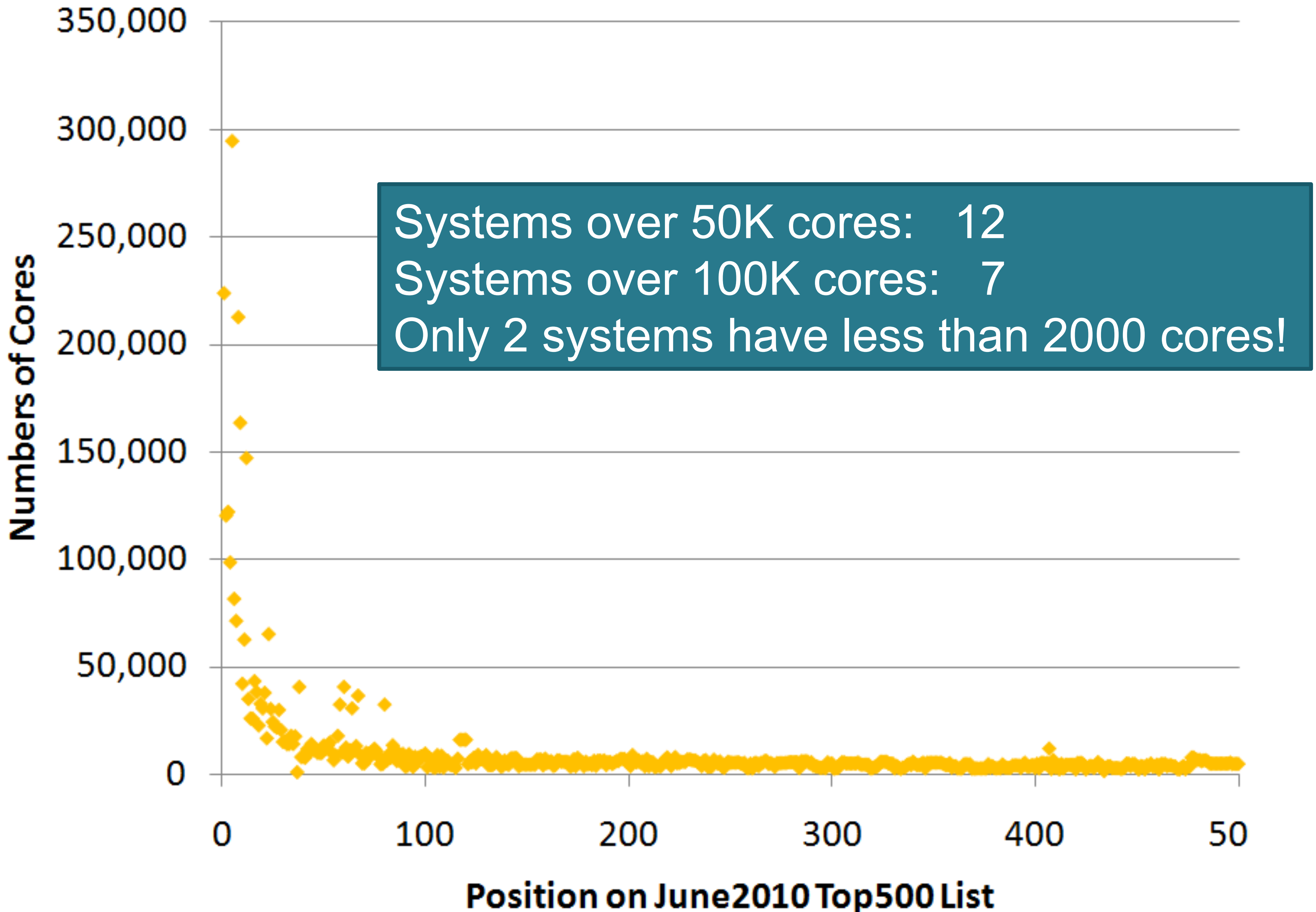
- **The energy consumed increases proportionally to the bit-rate, so as we move to ultra-highbandwidth links, the power requirements will become an increasing concern.**
- **The energy consumption is highly distance-dependent (the square of the length term), so bandwidth is likely to become increasingly localized as power becomes a more difficult problem.**
- **Improvements in chip lithography (making smaller wires) will not improve the energy efficiency or data carrying capacity of electrical wires.**

D. A. B. Miller and H. M. Ozaktas, "Limit to the Bit-Rate Capacity of Electrical Interconnects from the Aspect Ratio of the System Architecture," Journal of Parallel and Distributed Computing, vol. 41, pp. 42-52 (1997) article number PC961285.

Implications for future systems

- **Clock rates will stay largely the same as today, increasing the parallelism of systems to improve performance**
- **Energy cost of moving data is very large. We will have to explicitly manage data locality to limit power consumption**

We also have a scaling problem!



What do we need to go beyond Jaguar

- Weak scaling of apps has run its course
- Need more powerful nodes for strong scaling
  - Faster processors but using much less power per GFLOPS
  - More memory
  - Better interconnect
- Hierarchical programming model to expose more parallelism
  - Distributed memory among nodes   100K – 1M way parallelism
  - Threads within nodes   10s – 100s of threads per node
  - Vectors within the threads   10s – 100s of vector elements/ thread

1 Billion way parallelism needed for exascale systems

Exascale Systems

- 1-10 billion way parallelism
  - Requires hierarchical parallelism to manage
    - MPI between nodes
    - OpenMP or other threads model on nodes
    - SIMD / Vector parallelism within thread
- Power will dominate architectures
    - Takes more power to go to memory for data than to recompute it
- Traditional “balance ratios” can’t continue to be met
    - Memory size is not growing as fast as processor performance
    - Memory bandwidth is growing even more slowly
    - So compute becomes relatively cheaper while memory and data movement becomes the expensive resource in a system

Industry Hardware Trends

Not clear how industry trends will meet HPC application needs

- Semi-conductor industry trends
  - Clock speed now constrained by power and cooling limits
  - Shifting to multi/many core with attendant parallelism
  - Compute nodes with added hardware accelerators are introducing additional complexity of heterogeneous architectures
  - Processor cost is increasingly driven by pins and packaging, which means the memory wall is growing in proportion to the number of cores on a processor socket

- Development of large-scale Leadership-class supercomputers from commodity computer components will require collaboration
  - Leadership-class supercomputers cannot be built from only commodity components
  - Supercomputer architectures must be designed with an understanding of the applications they are intended to run

Software Trends

Science becoming more difficult on Leadership systems

Application trends

- Scaling limitations of present algorithms
- More complex multi-physics requires large memory footprint
- Need for automated fault tolerance, performance analysis, and verification
- Software strategies to mitigate memory latencies and hierarchy
- Need for innovative algorithms for multi-core, heterogeneous architectures
- Model coupling complexity for more realistic formulations

Emerging Applications

- Growing importance of data intensive applications
- Mining of experimental and simulation data

Production Quality Software Environments

- High-performance parallel I/O standards
- Future programming models
  - MPI/OpenMP replacements
  - Methodologies and tools required to exploit highly parallel architectures
    - performance analysis tools, libraries,…
  - Tools for refactoring application codes
  - Language improvements
- Componentization
  - verification; unit testing, …
- Scalable and distributed analysis software
- Math and application frameworks
- Benefits to partnerships in development of software environment
  - Substantial investment in software for current and future machines a priority

Technology transitions have been observed over time

Logistic change is characterized by an initial period of slow growth, followed by a period of exponential growth, then a point of inflection, and finally a period of asymptotic growth as the technology approaches a limit. This pattern of change was first observed in population studies [28], and it has since been found to be descriptive of change in a remarkably diverse set of circumstances, including technological evolution in general and the evolution of electronic and computer technologies in particular.

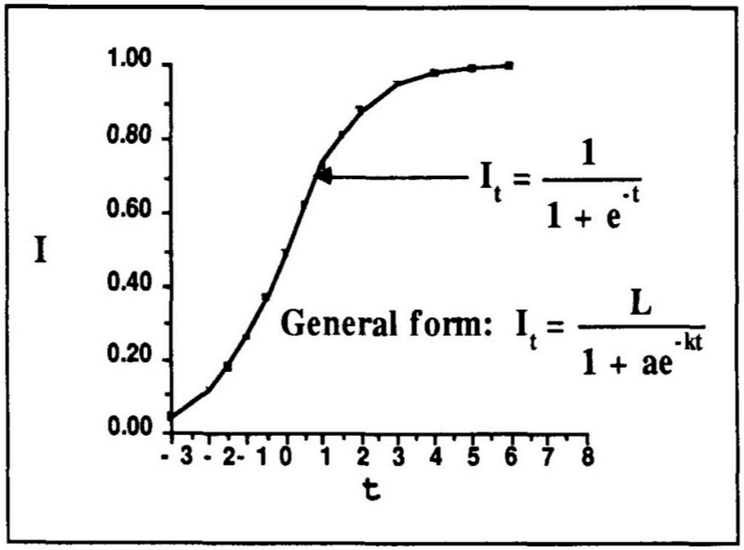


Figure 9. Logistic change.

Worlton (1988)

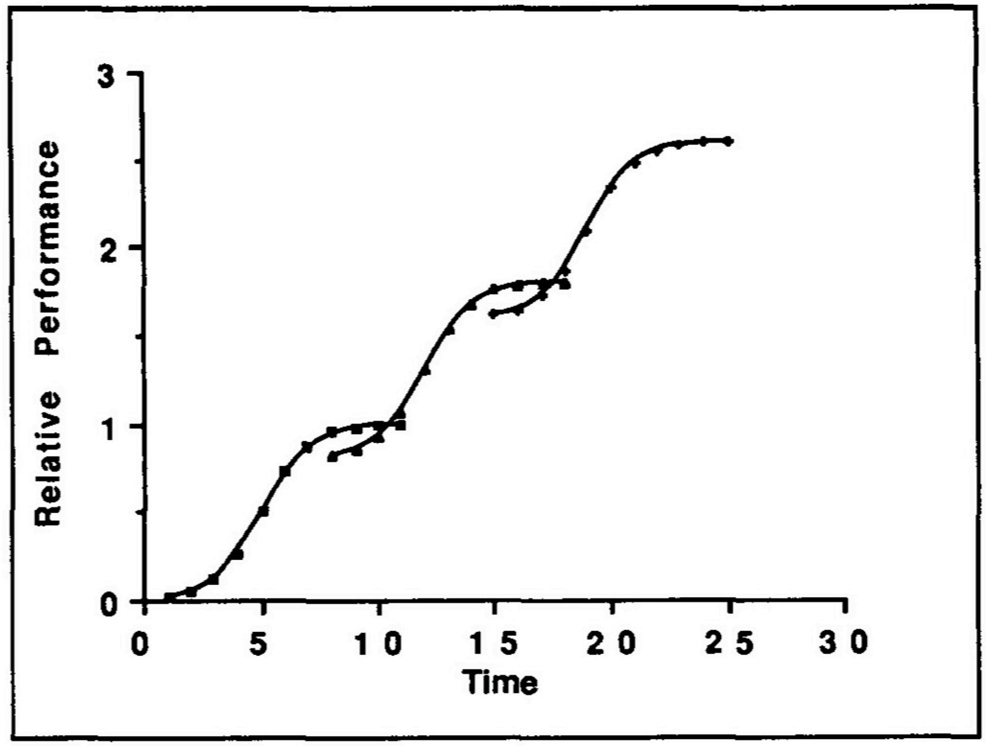
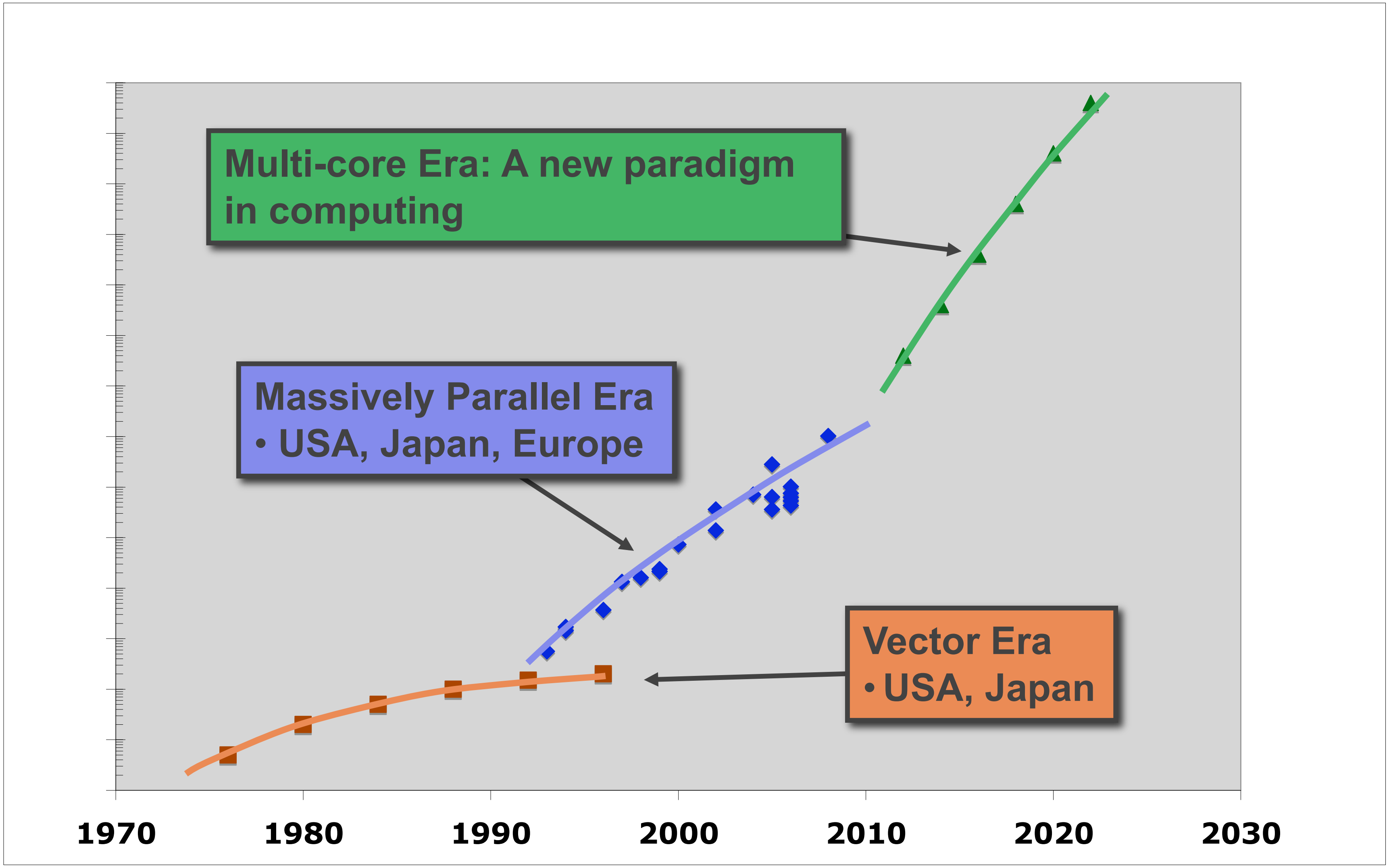


Figure 10. Piecewise-logistic patterns of change.

Examples of inflection points where technology changed

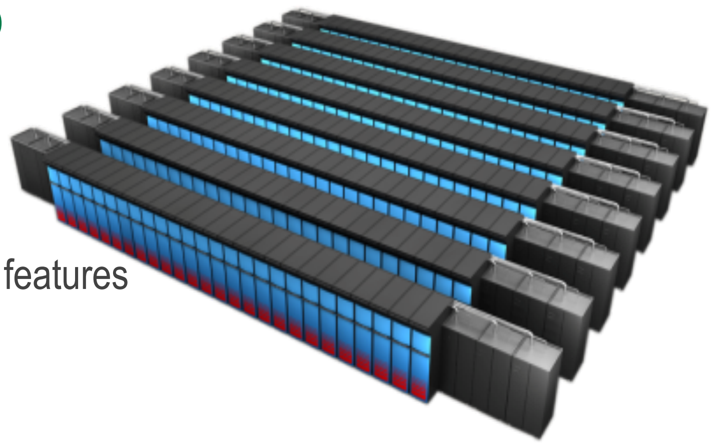


Processor Architecture Directions

- Requirements for achieving scale and sustained performance per $ or watt
  - Performance through parallelism, not frequency
    - fast serial threads coupled to many efficient parallel threads
  - Exploit on-chip locality
  - Deep, explicitly managed memory hierarchy
    - to better exploit locality, improve predictability, and reduce overhead
  - A microarchitecture to exploit parallelism at all levels of a code
    - distributed memory, shared memory, vector/SIMD, multithreaded

ORNL's "Titan" System Road Map

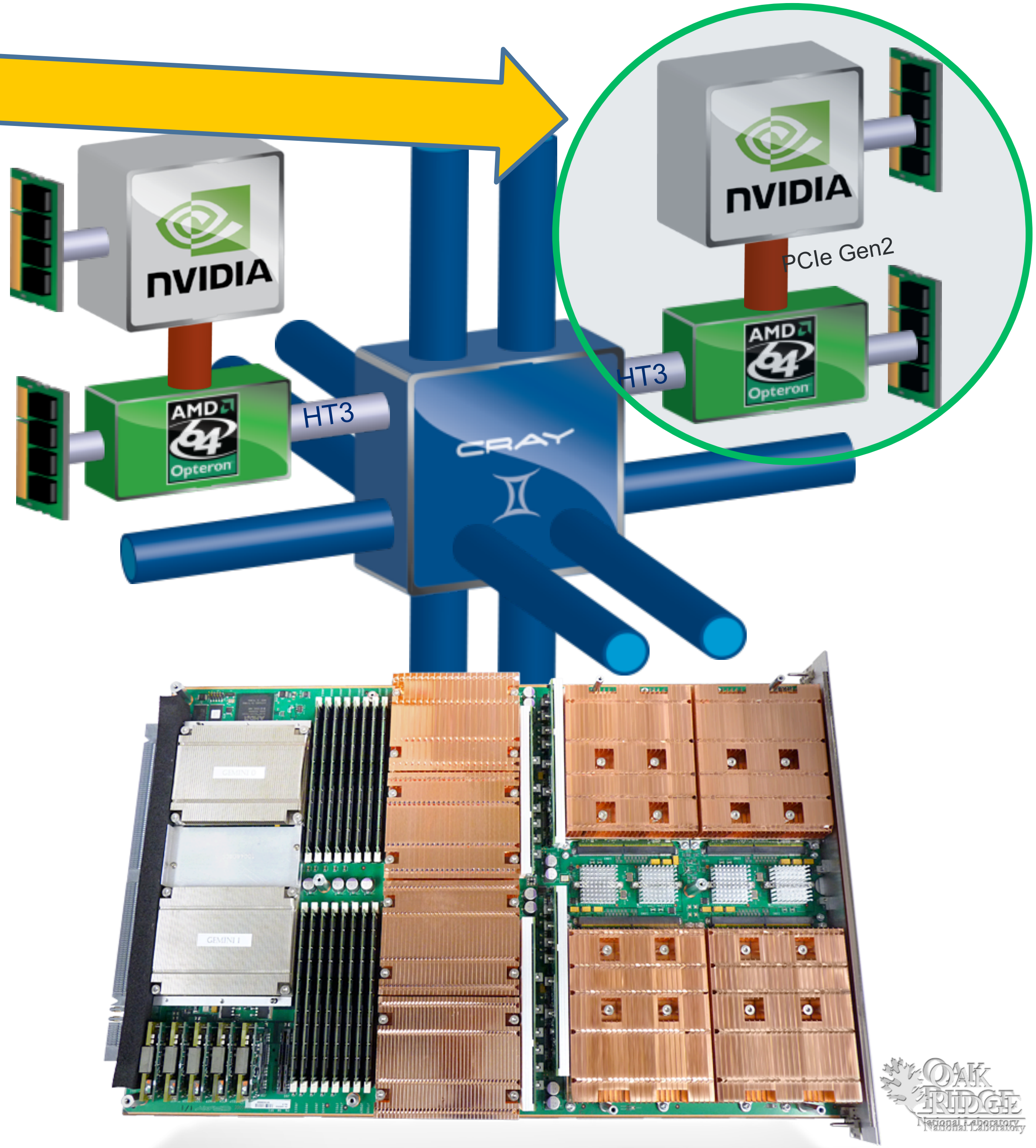
- Same number of cabinets, cabinet design, cooling and node count
- Operating system upgrade of today's Linux Operating System
- Gemini high-performance interconnect
  - 3-D Torus, globally addressable memory, advanced synchronization features
- AMD Opteron16-core processor (Interlagos)
- New accelerated node design using NVIDIA multi-core processors
- Larger memory - more than 2x more memory per node
- ~20 PF peak performance



- ~1,000 Fermi accelerators now for application development
- >14,592 next generation accelerators in late 2012

Cray XK6 Compute Node

XK6 Compute Node Characteristics
AMD Opteron 6200 "Interlagos" 16 core processor @ 2.2GHz
Tesla M2090 "Fermi" @ 665 GF with 6GB GDDR5 memory
Host Memory 32GB 1600 MHz DDR3
Gemini High Speed Interconnect
Upgradeable to NVIDIA's next generation "Kepler" processor in 2012
Four compute nodes per XK6 blade. 24 blades per rack



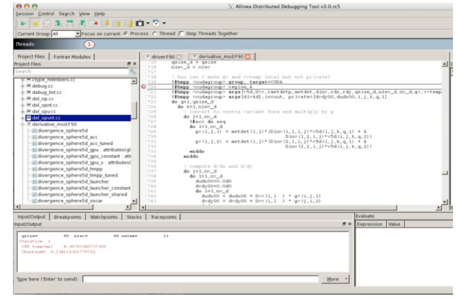
Why use an accelerator?

- Best way to get to a very powerful node
  - Our users tell us that they want fewer, much more powerful nodes
  - Titan nodes will be greater than 1.5 TeraFLOPS per node
- Power consumption per GF is much better than a conventional processor

Processor type	GigaFLOPS / Watt
Cray XE6 (Magny-Cours)	1
Titan (Projected)	6.3

- Explicitly managed memory hierarchy
  - Programmer places the data in the appropriate memory and manages to save energy

Programming Environment



Goals:

- Full functionality hybrid programming environment
  - Compilers, Debuggers, Performance Analysis tools, Mathematical Libraries
- Hardware agnostic programming model - *portable*
  - Describe execution parallelism and data layout: expose (hierarchical) parallelism
  - Standardization through OpenMP Architecture Review Board and other industry initiatives

Exploitation of node-level parallelism

- Recognition and exploitation of hierarchical parallelism
- Development of effective programming tools to facilitate rapid refactoring of applications
- Deployment of useful performance and debugging tools to speed refactoring

Scalable Debugging for Hybrid Systems

- collaboration with Allinea to develop a scalable hybrid aware debugger based on DDT

High-productivity Hybrid-programming Through Compiler Innovation

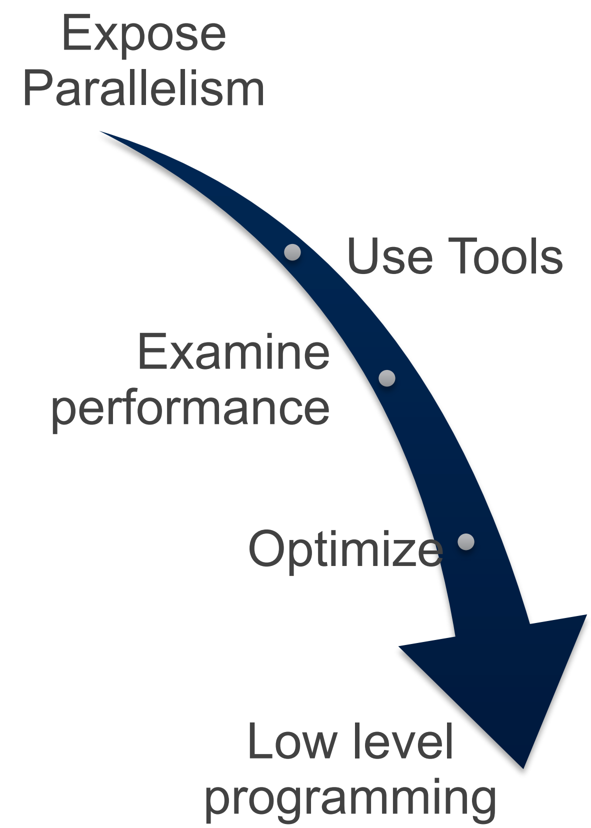
- collaboration with HMPP to develop directive based compiler technology in CAPS compiler
  - CAPS support for OpenACC set of directives; support for all common languages used at the OLCF, …

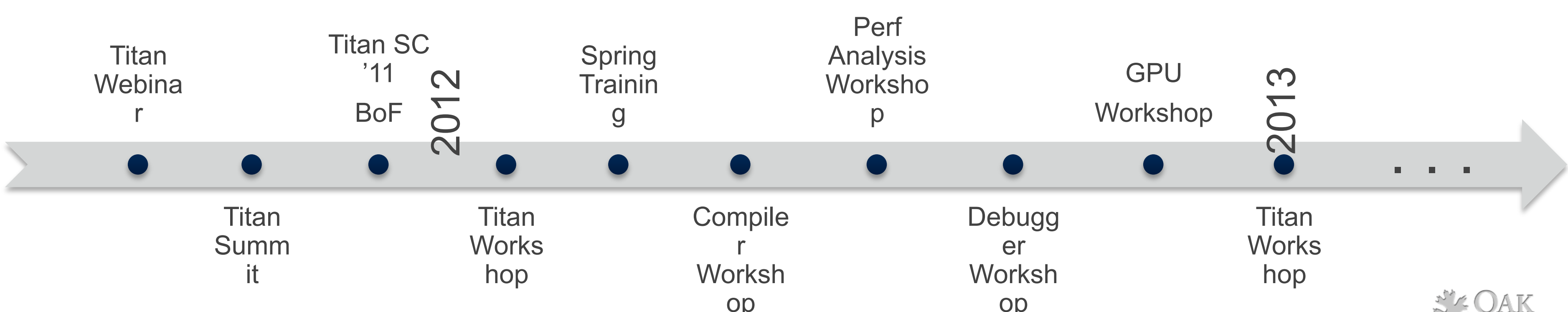
Scalable Performance Analysis for Hybrid Systems

- collaboration with Technische Universitat Dresden to add support for Hybrid (CPU/GPU) performance analysis in Vampir

Titan Training Program

- Goal: Enable break-through science through education
- Strategy: Provide conferences, workshops, tutorials, case studies, and lessons learned on tools and techniques for realizing hybrid architecture benefits. Provide content via traditional venues, online, and pre-recorded sessions.
- Objective: Users will be able to expose hierarchical parallelism, use compiler directive-based tools, analyze / optimize / debug codes, and use low-level programming techniques if required





Maintaining leadership: data infrastructure

- The OLCF Spider project was groundbreaking
  - Largest scale and highest performance parallel file system ever deployed
  - Success, which leveraged Lustre experience at LLNL, has served as a blueprint for HPC community
- Collaboration was key to success
  - Leveraged R&D across ORNL to architect, prototype, develop and deploy
  - Partnerships with Cray, DDN, and the Lustre development team were critical
- Titan environment demands new capabilities
  - Scalable metadata performance
  - Bandwidth scalability to over a Terabyte/sec
  - Scalable system resiliency for improved fault tolerance
- Challenges would not be met without our efforts
  - Leadership role in Lustre – Open Scalable File Systems Consortium
  - Continued partnerships both internally and externally to ensure solutions that meet our requirements are available and affordable

Isn’t this a risky strategy?

- **Hardware risk is low**

Component	Risk
System Scale	Same as Jaguar
Processor	Opteron follow-on
Accelerator	Fermi follow-on
Interconnect	Deployed in Cielo & Franklin at scale
Operating System	Incremental changes from Jaguar

- **Applications are where there is risk**

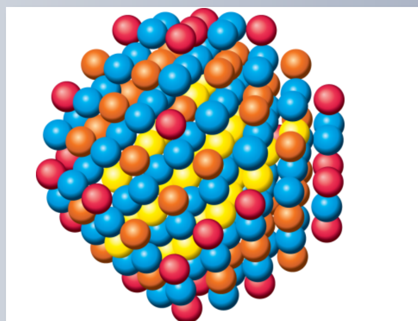
Component	Risk
Programmability	Rewrite in CUDA or something else?
Tools	Debuggers and development tools
Programmer reluctance to rewrite for GPUs	Fermi follow-on will provide a lot of capability

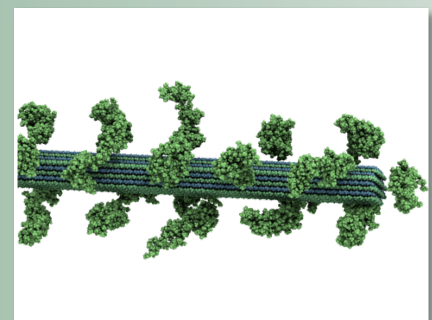
Titan: Early Science Applications

Driven by science mission, algorithms, data structures, programming models, libraries

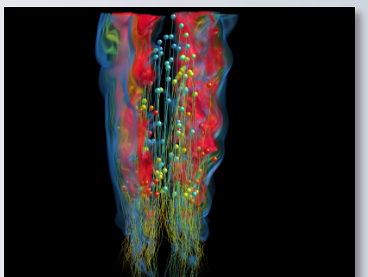
Facilitated by creation of the Center for Accelerated Application Readiness (CAAR)

WL-LSMS
Role of material disorder, statistics, and fluctuations in nanoscale materials and systems.



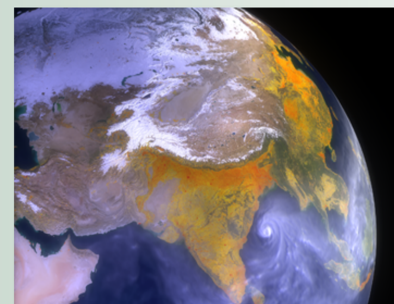


LAMMPS
A parallel particle simulator that can simulate soft materials (biomolecules, polymers), solid-state materials (metals, semiconductors) and coarse-grained or mesoscopic systems

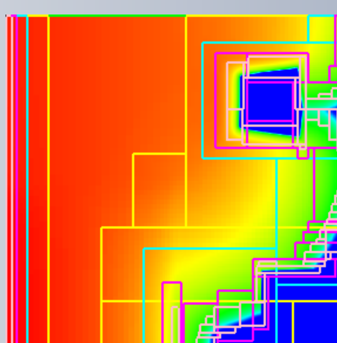


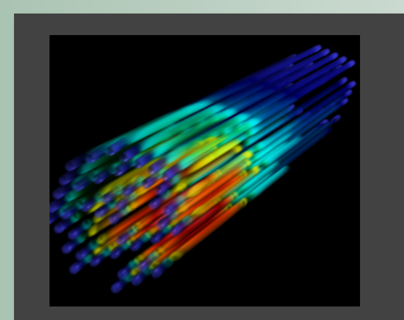
S3D
How are going to efficiently burn next generation diesel/bio fuels?
.

CAM / HOMME
Answer questions about specific climate change adaptation and mitigation scenarios; realistically represent features like precipitation patterns/statistics and tropical storms



NRDF
Radiation transport – important in astrophysics, laser fusion, combustion, atmospheric dynamics, and medical imaging – computed on AMR grids.





Denovo
Unprecedented high-fidelity radiation transport calculations that can be used in a variety of nuclear energy and technology applications.

How Effective are GPUs on Scalable Applications?

March 2012 Performance Snapshot on TitanDev (Fermi accelerator)

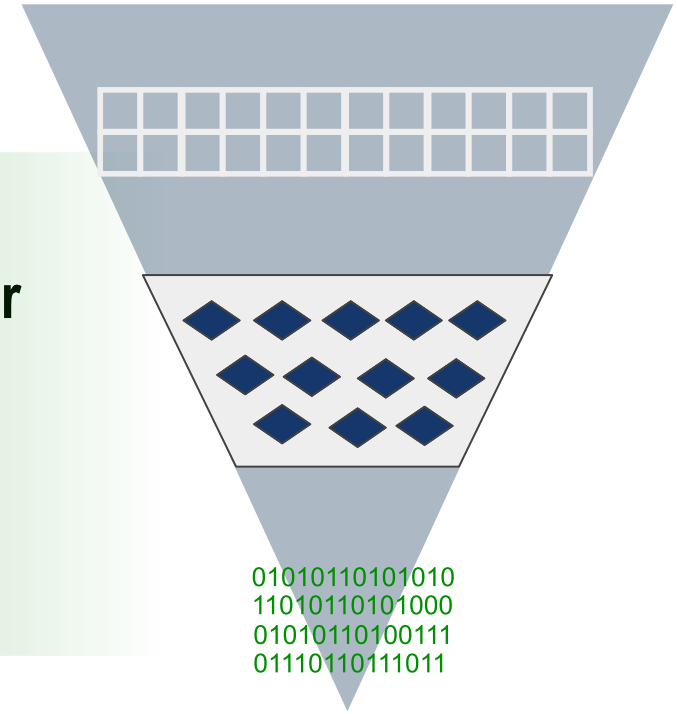
Application	XK6 (w/ GPU) vs. XK6 (w/o GPU) Performance Ratio Titan Dev : Jaguar	XK6 (w/ GPU) vs. XE6 Performance Ratio Titan Dev : Monte Rosa
S3D Turbulent combustion	1.5	1.4
Denovo 3D neutron transport for nuclear reactors	3.5	3.3
LAMMPS High-performance molecular dynamics	6.5	3.2
WL-LSMS Statistical mechanics of magnetic materials	3.1	1.6
CAM-SE Community atmosphere model	2.6	1.5
NAMD High-performance molecular dynamics	2.6	1.4
Chroma High-energy nuclear physics	8.8	6.1
QMCPACK Electronic structure of materials	3.8	3.0
SPECFEM-3D Seismology	4.7	2.5
GTC Plasma physics for fusion-energy	2.5	1.6
CP2K Chemical physics	2.8	1.5

Cray XK6: Fermi GPU plus Interlagos CPU
Cray XE6: Dual Interlagos and no GPU

Gordon Bell Winner

Hierarchical Parallelism

- MPI parallelism between nodes (or PGAS)
- On-node, SMP-like parallelism via threads (or subcommunicators, or…)
- Vector parallelism
  - SSE/AVX on CPUs
  - GPU threaded parallelism



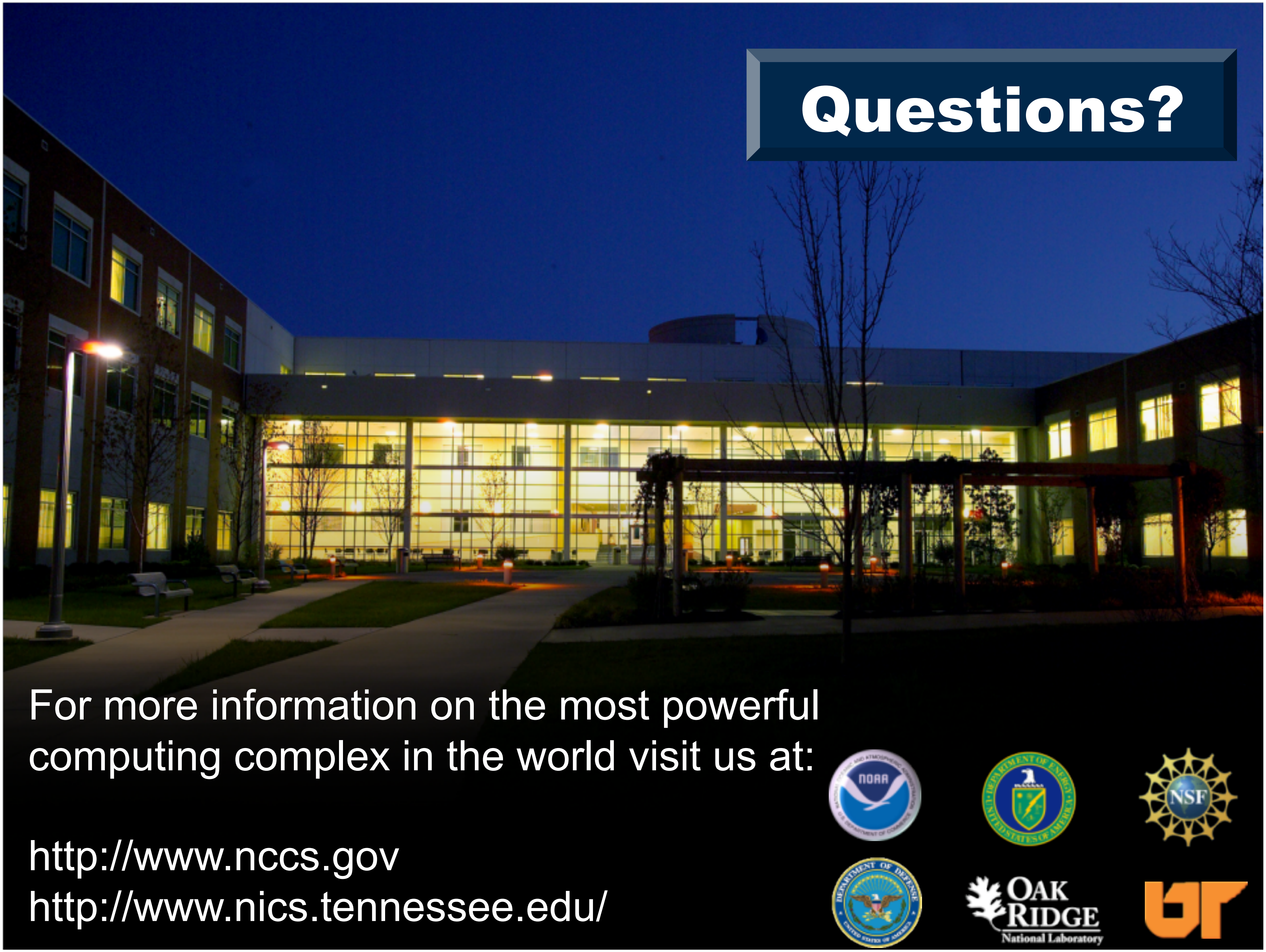
- Exposure of unrealized parallelism is essential to exploit all near-future architectures.
- Uncovering unrealized parallelism and improving data locality improves the performance of even CPU-only code.

Some Lessons Learned

- Exposure of unrealized parallelism
  - Figuring out where is often straightforward
  - Making changes to exploit it is hard work (made easier by better tools)
  - Developers can quickly learn, e.g., CUDA and put it to effective use
  - A directives-based approach offers a straightforward path to portable performance
- For those codes that already make effective use of scientific libraries, the possibility of continued use is important.
  - HW-aware choices
  - Help (or, at least, no hindrance) to overlapping computation with device communication
- Ensuring that changes are communicated back and remain in the production "trunk" is every bit as important as we initially thought.
  - Other development work taking place on all CAAR codes could quickly make acceleration changes obsolete/broken otherwise

Summary

- **Climate science is at multiple inflection points**
  - new, growing, and demanding expectations from simulation tools
  - multiscale/multiphysics simulation challenges
  - new directions in HPC computer architectures/environments
- **How do we survive this?**
  - continued investment in basic scientific understanding
  - investments in scalable computational algorithms
  - investments in disciplined software development environments
  - creation of multi-disciplinary project-like teams


Questions?
For more information on the most powerful computing complex in the world visit us at:
http://www.nccs.gov
http://www.nics.tennessee.edu/
NOAA
NSF
OAK RIDGE
National Laboratory

Summary

- Partnering has demonstrated value in navigating architectural transition
  - highly integrated engagement with user community has led to early success
  - CAAR application effort already demonstrating advantages of hybrid architectures
  - user assistance and outreach will help codify best practices and inform the broader community via educational opportunities

- Important investments in and collaborations with technology providers
  - Scalable Debugging for Hybrid Systems
    - collaboration with Allinea to develop a scalable hybrid aware debugger based on DDT
  - High-productivity Hybrid-programming Through Compiler Innovation
    - collaboration with HMPP to develop directive based compiler technology in CAPS compiler
      - CAPS support for OpenACC set of directives; support for all common languages used at the OLCF
  - Scalable Performance Analysis for Hybrid Systems
    - collaboration with Technische Universitat Dresden to add support for Hybrid (CPU/GPU) performance analysis in Vampir