



The NNSA ASCI Program: Advanced Simulation and Computing

Steve Louis

Lawrence Livermore National Laboratory
7000 East Avenue, Livermore, CA, 94550-0234
Phone: +1-925-422-1550 FAX: +1-925-423-8715
E-mail: stlouis@llnl.gov

Presented at the October 9, 2001 THIC Meeting
WestCoast Silverdale Hotel, Silverdale WA 98383-9191

UCRL-PRES-146034

This work was performed under the auspices of the U.S. Department of Energy by University of California
Lawrence Livermore National Laboratory under contract No. W-7405-Eng-48.

THIC Inc.

The Premier Advanced Recording Technology Forum





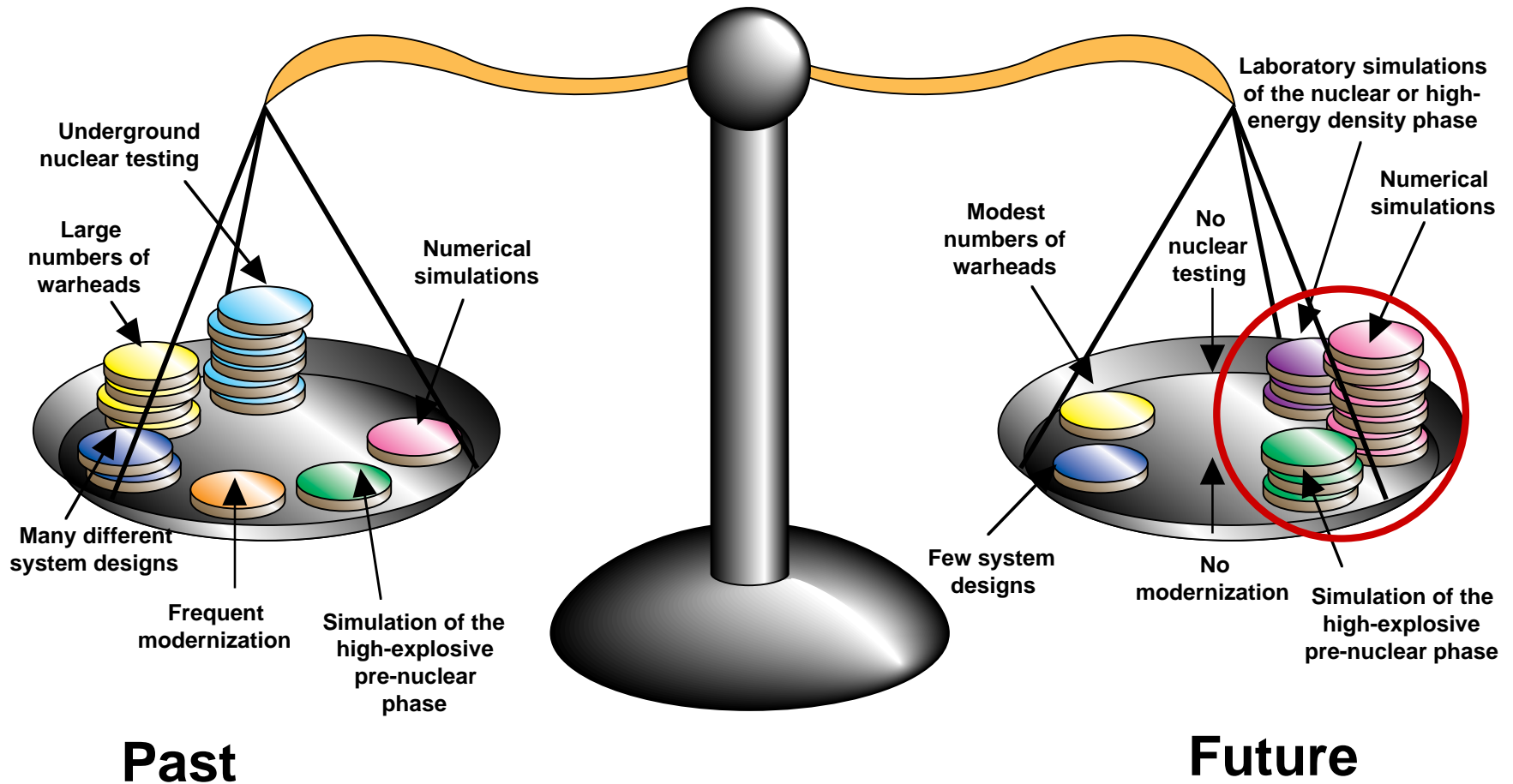
Outline



- **The NNSA Stockpile Stewardship Program**
- **Where We are Now: ASCI/LLNL Computing**
- **Challenges for Today and for the Future**

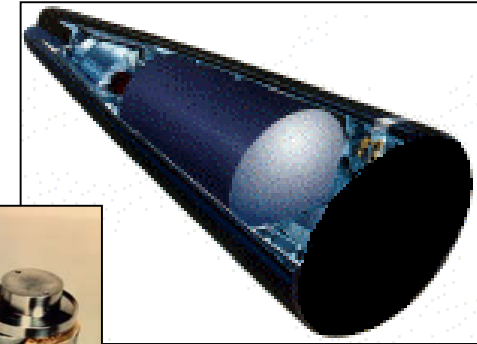
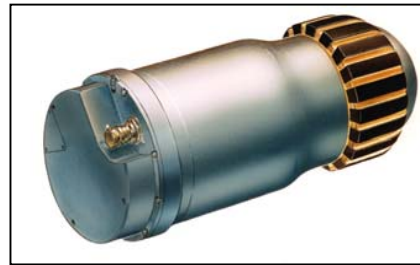
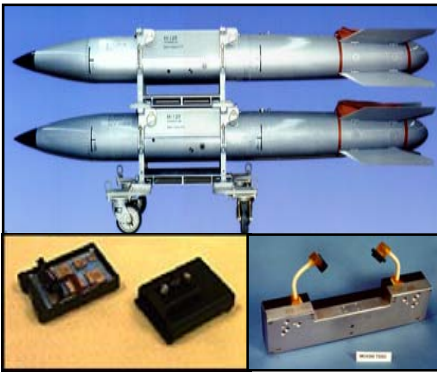


ASCI supports the DOE/NNSA Stockpile Stewardship Program





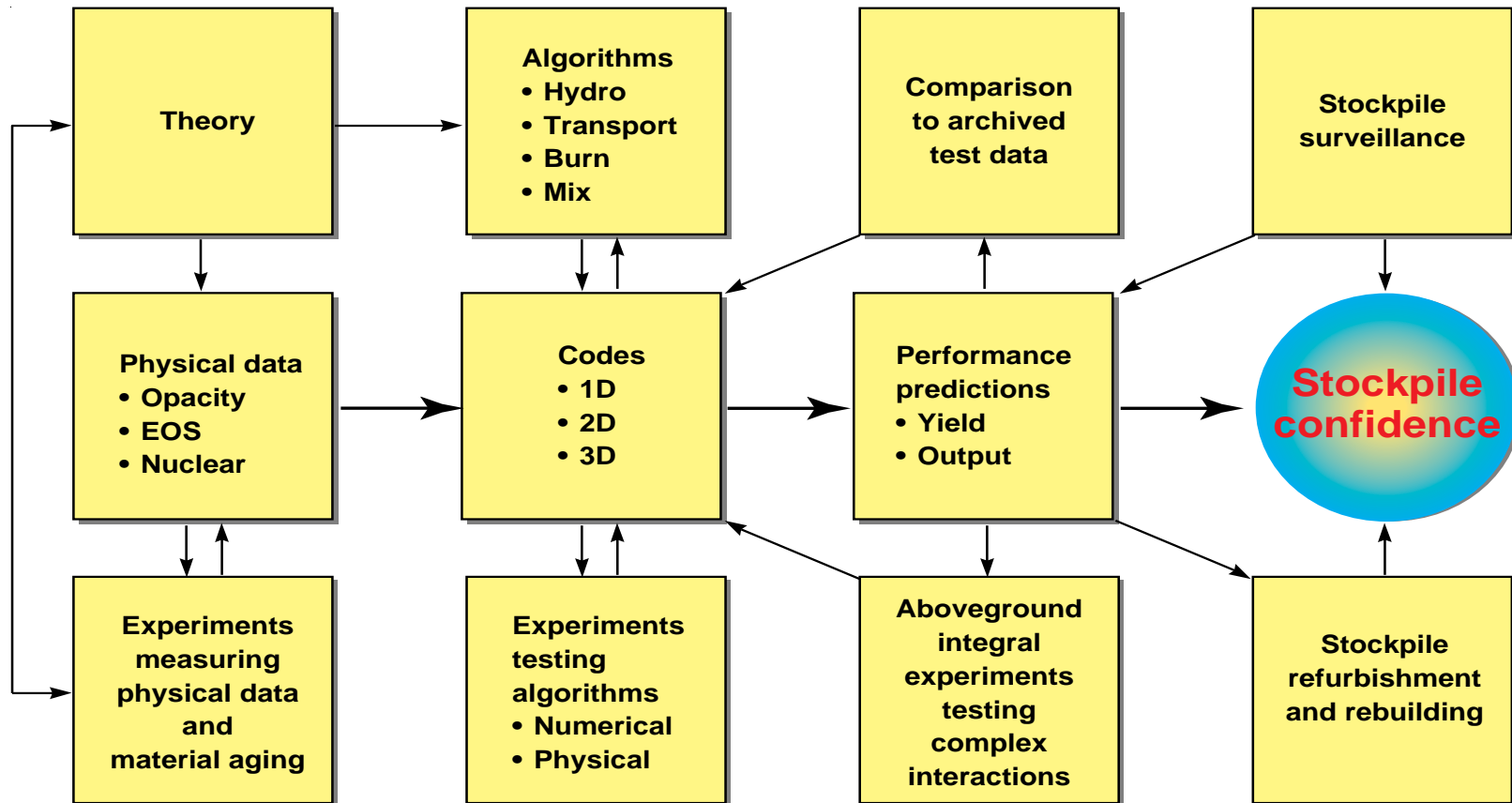
Challenge: Maintain stockpile confidence as changes occur



2004/2005 is a critical time period. Confidence in the stockpile is at risk.



Simulation plays central role to maintain stockpile confidence



But its value is critically dependent on the other elements of the integrated program



Nature of simulations changing with the loss of nuclear testing



With Nuclear Experiments



- Will it work as designed?
- Is the simulation good enough to risk cost of a nuclear experiment?

Without Nuclear Experiments

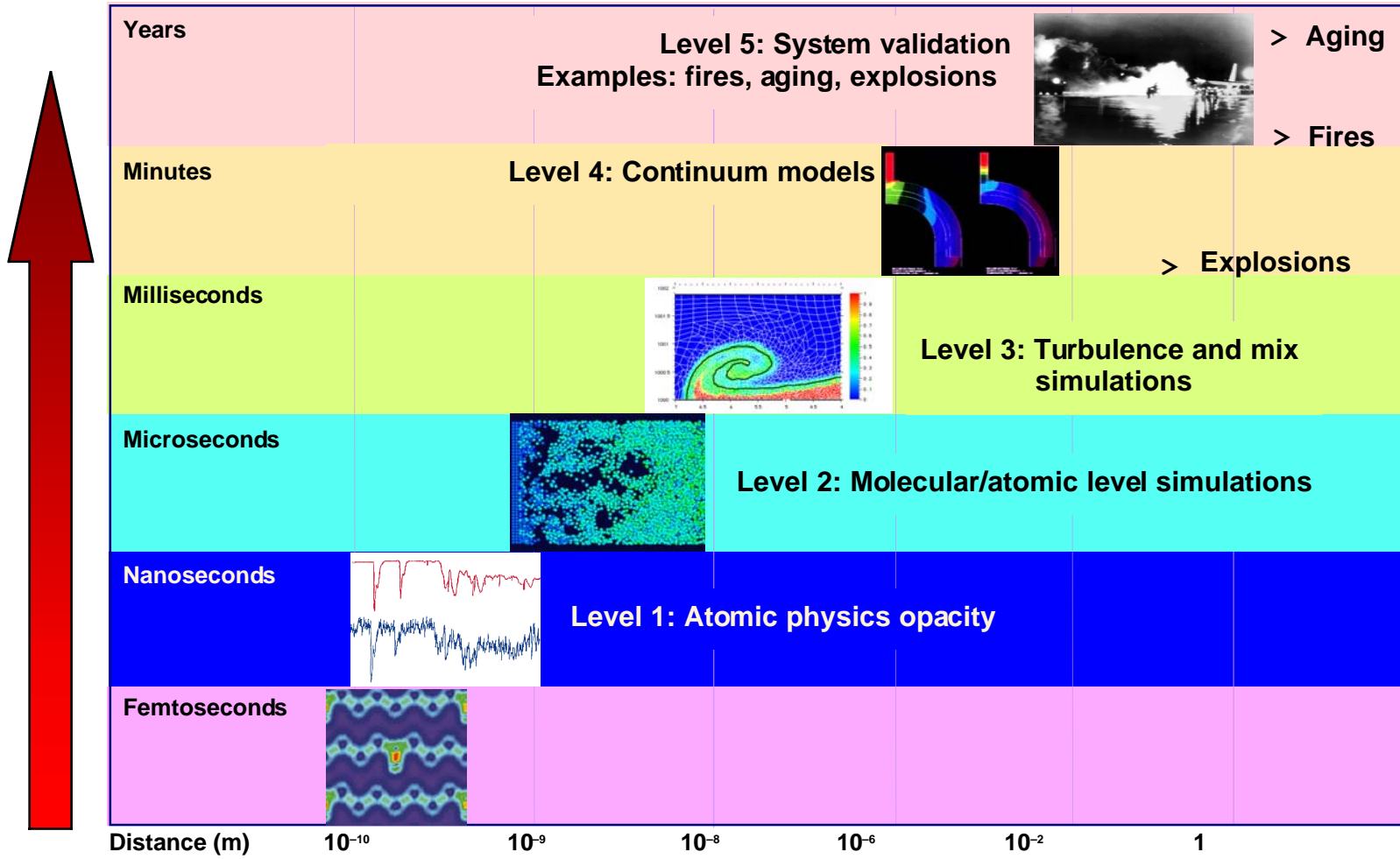


- Will it continue to work as it ages?
- Is the simulation adequate for making decisions affecting national security?

Supporting a stockpile of aging, highly optimized nuclear weapons demands advanced simulation capability

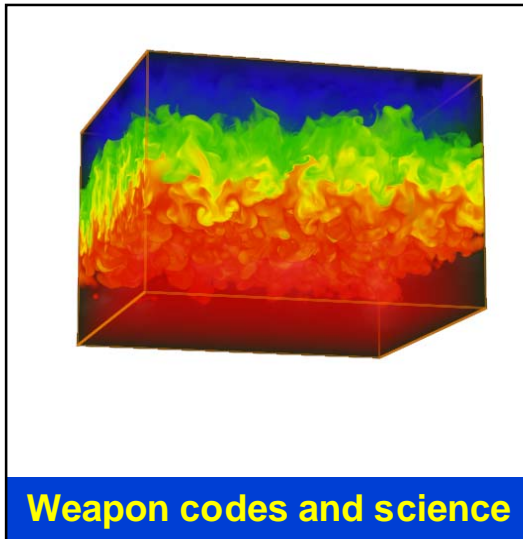


ASCI simulation time scales go from femtoseconds to years

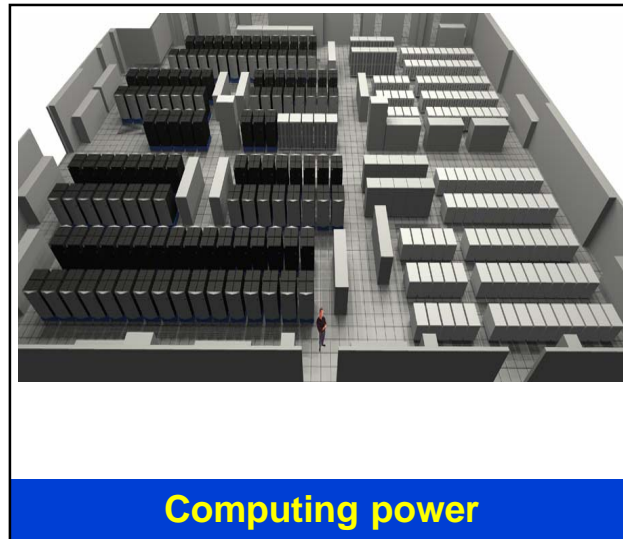




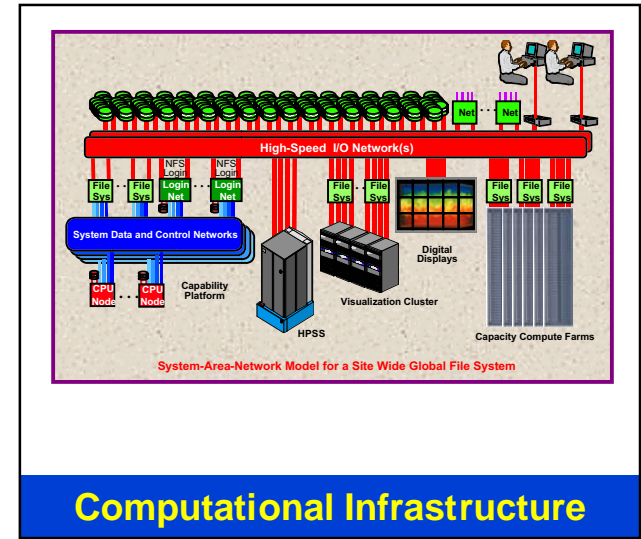
Computing capabilities require advances in several key areas



Weapon codes and science



Computing power

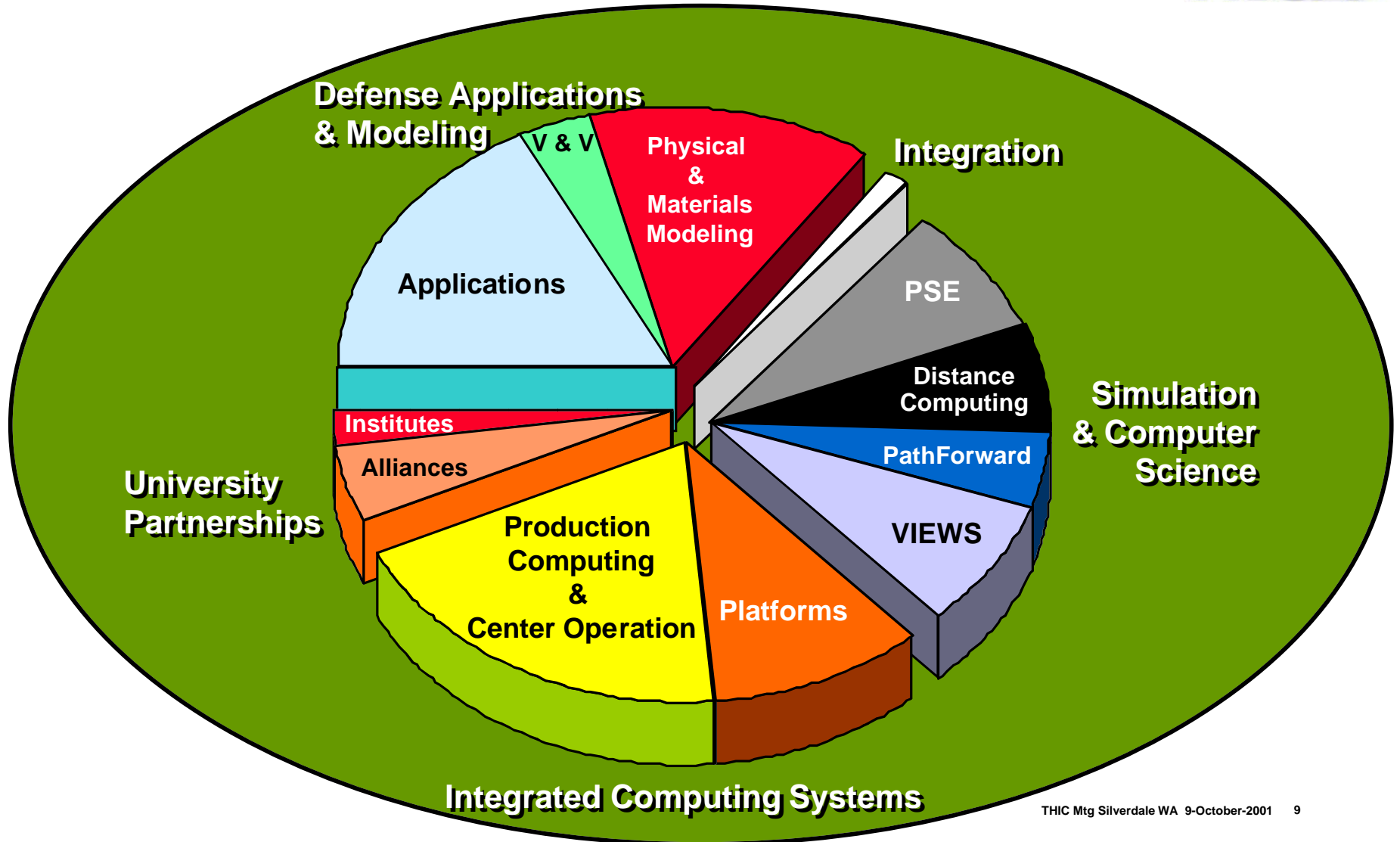


Computational Infrastructure

Stockpile stewardship pushes the limits in weapon simulation codes, computational power, and supporting infrastructures

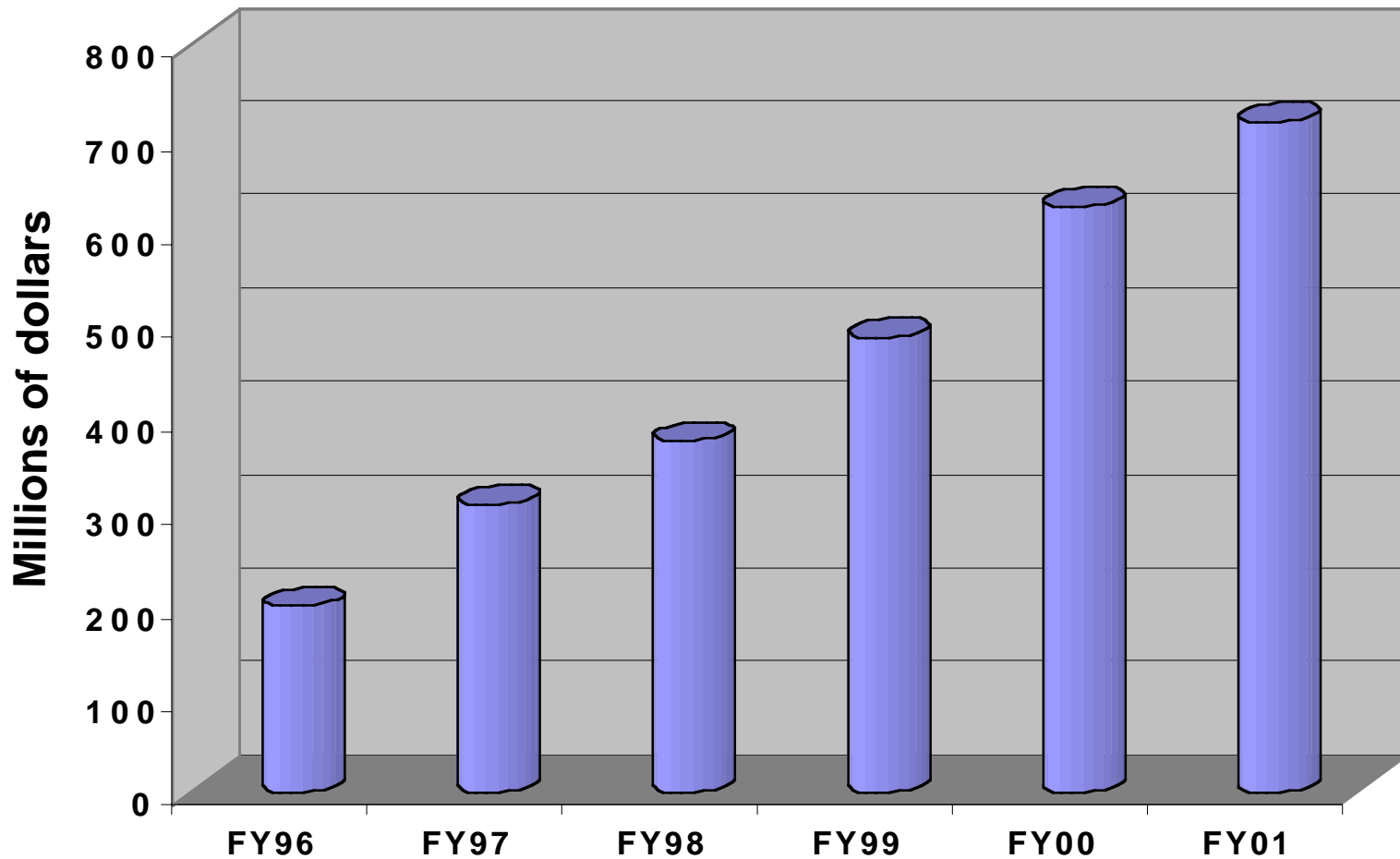


ASCI's elements have evolved to meet the SSP requirements





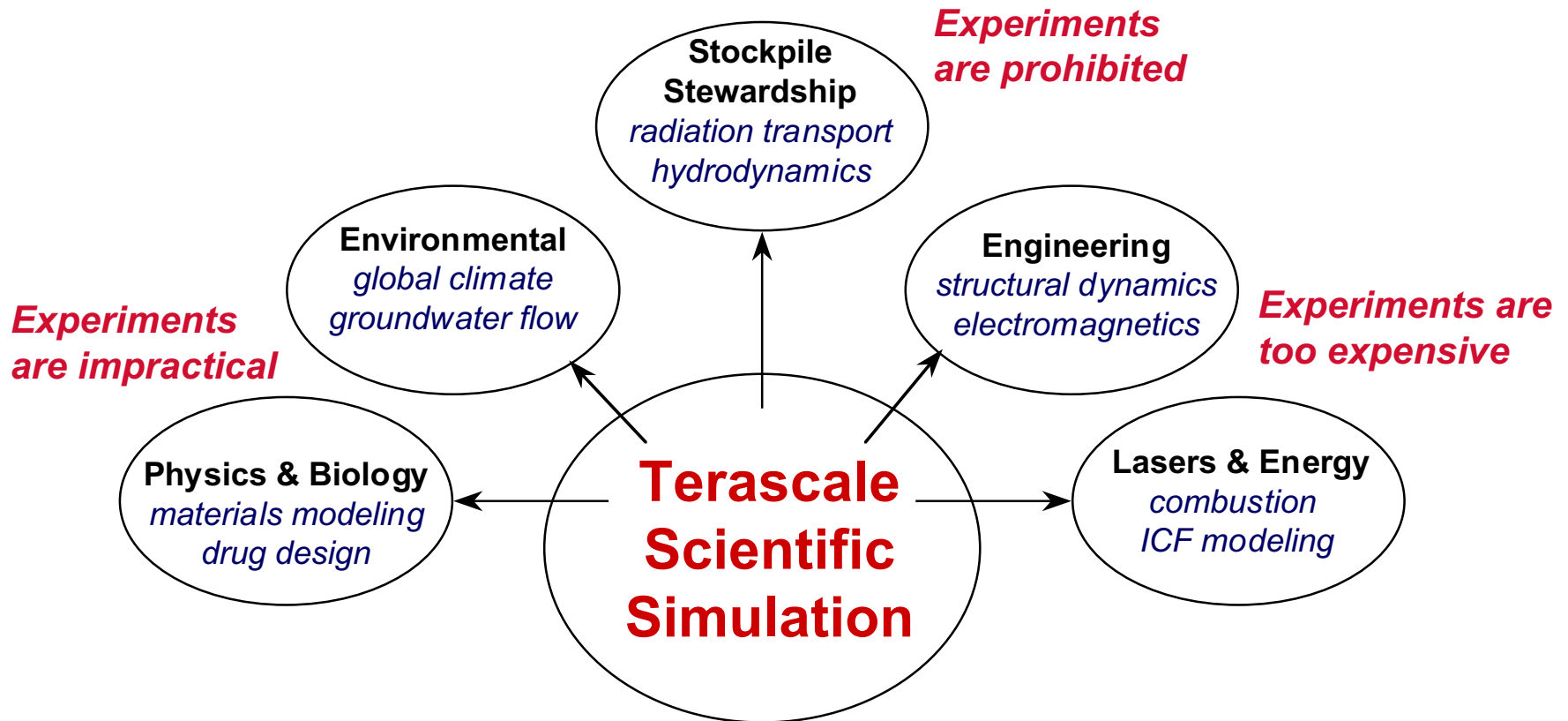
The ASCI budget is healthy and still growing



Excludes building construction



Simulation also plays a key role in virtually every LLNL program



Simulation emerging as a peer to theory and experiment in scientific discovery since it challenges these to refine quality and accuracy



Outline



- The NNSA Stockpile Stewardship Program
- **Where We are Now: ASCI White at LLNL**
- Challenges for Today and for the Future



Requirements for very large parallel computer systems



- **Many thousands of processors (up to ~20,000), with high reliability and high parallel efficiency**
- **Typical calculations will require a large fraction of the total machine for a hundred or more hours**
- **Some examples of problems that need extremely large-scale computing capabilities beyond ASCI**
 - **Micro and macro weather simulations**
 - **Global climate and ocean simulations**
 - **Material aging studies**
 - **Pharmaceutical design**
 - **Biology (brain function, circulatory systems, DNA)**



What can you get for \$250M?



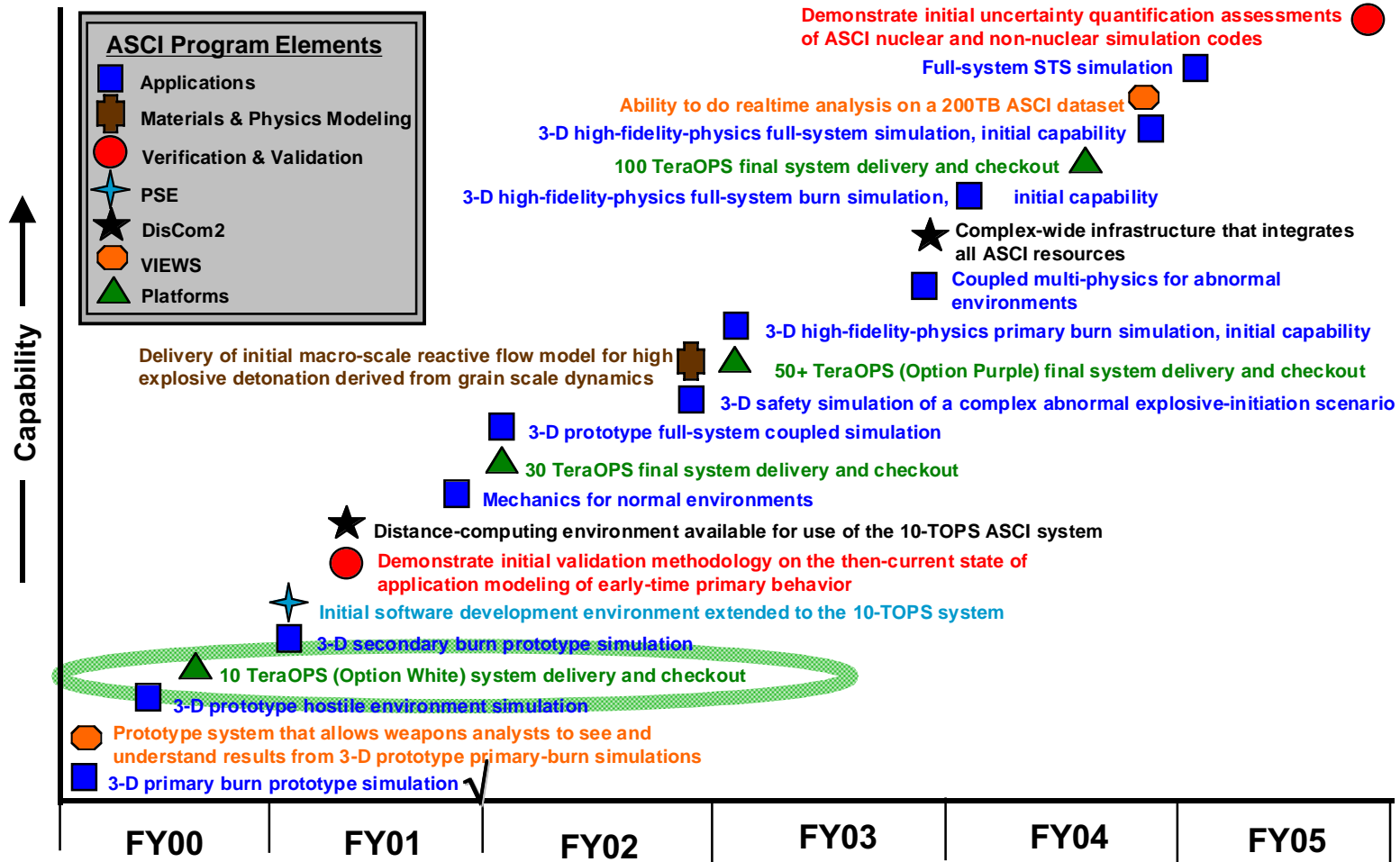
Shortstop for the NNSA Softball Team



100-TF National Security Computer

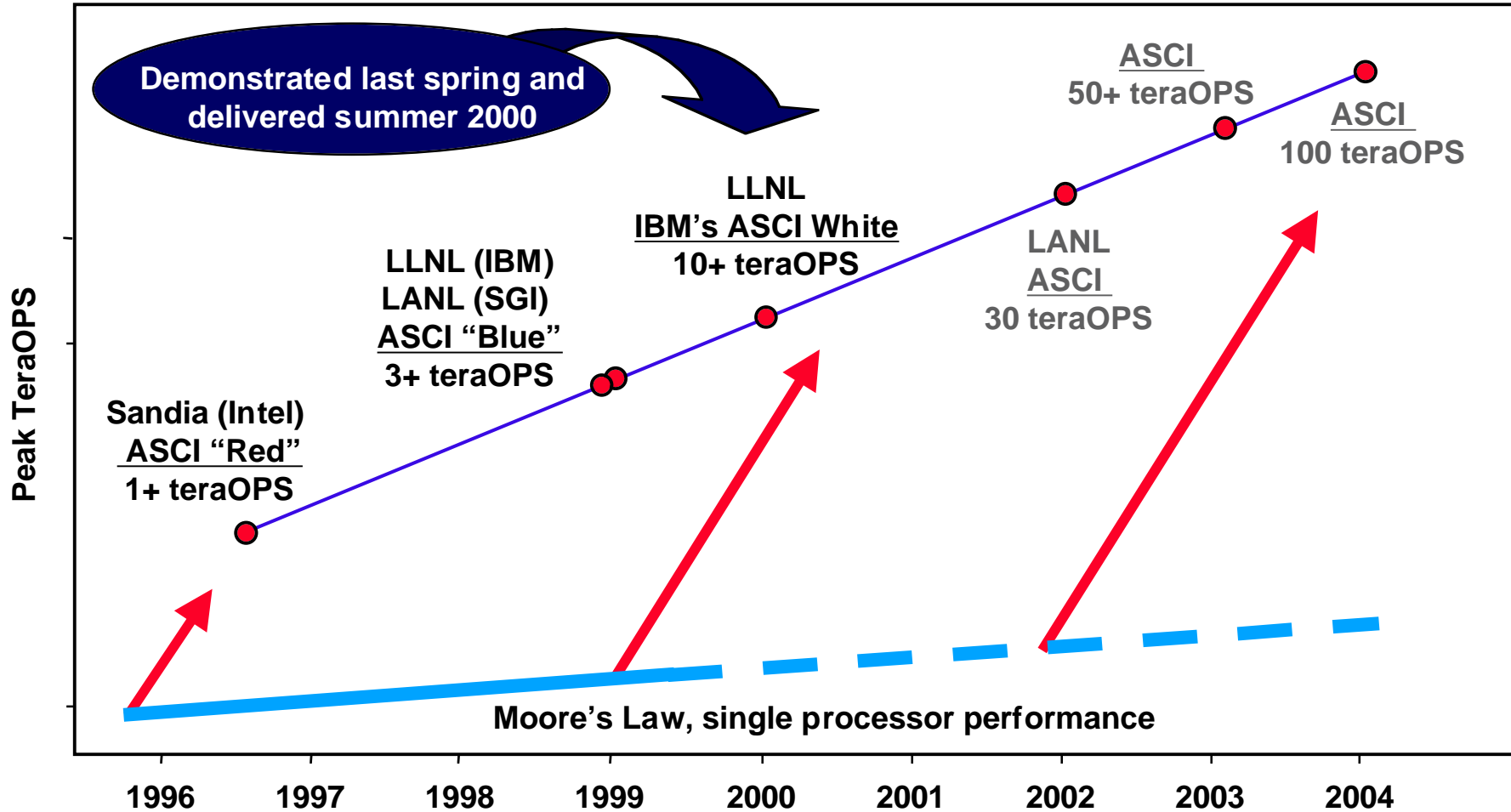


ASCI White delivery in FY00 was a key programmatic milestone





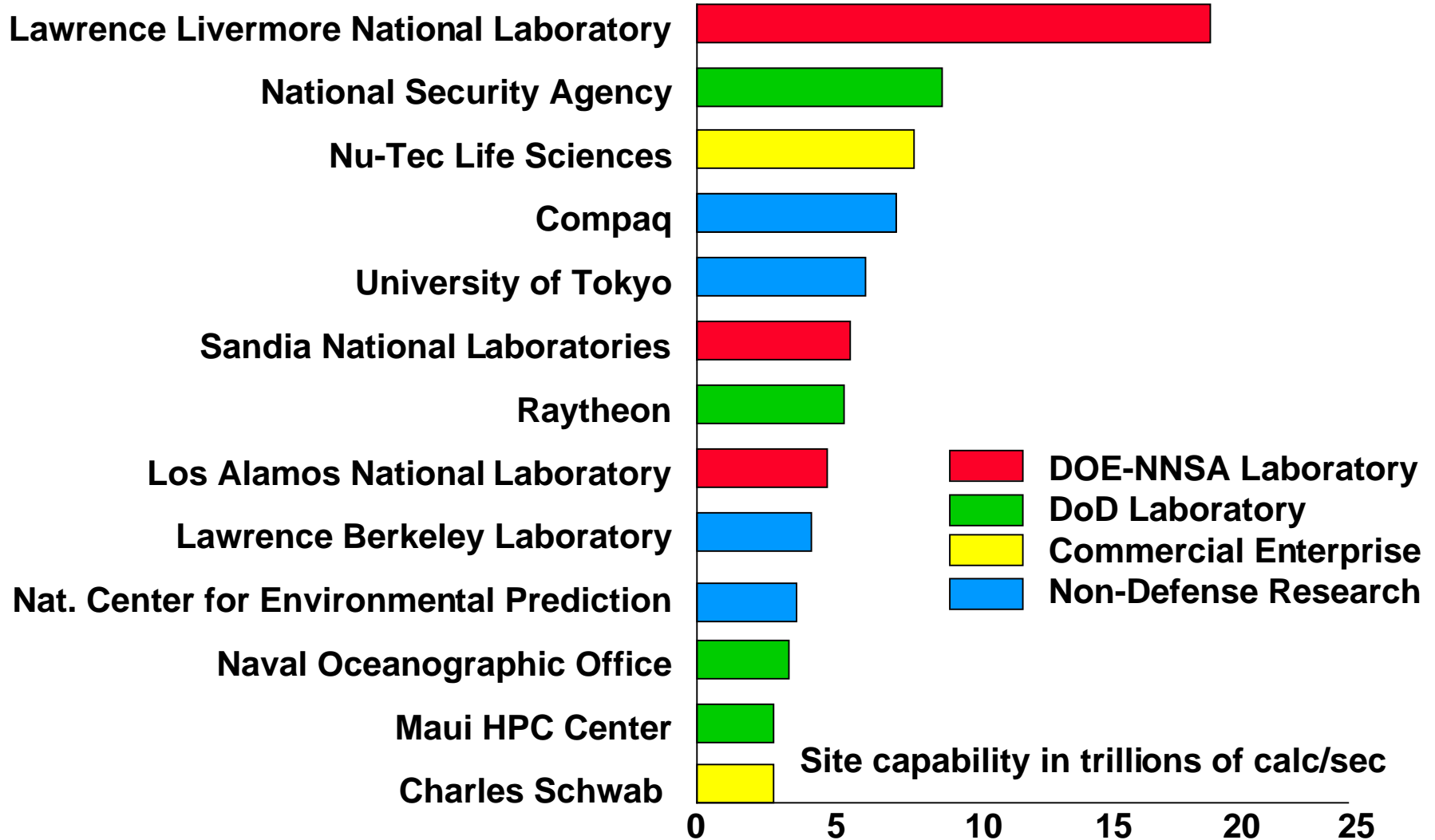
100 TeraOPS is the entry-level capability for SSP requirements



Our deadline is the year 2004

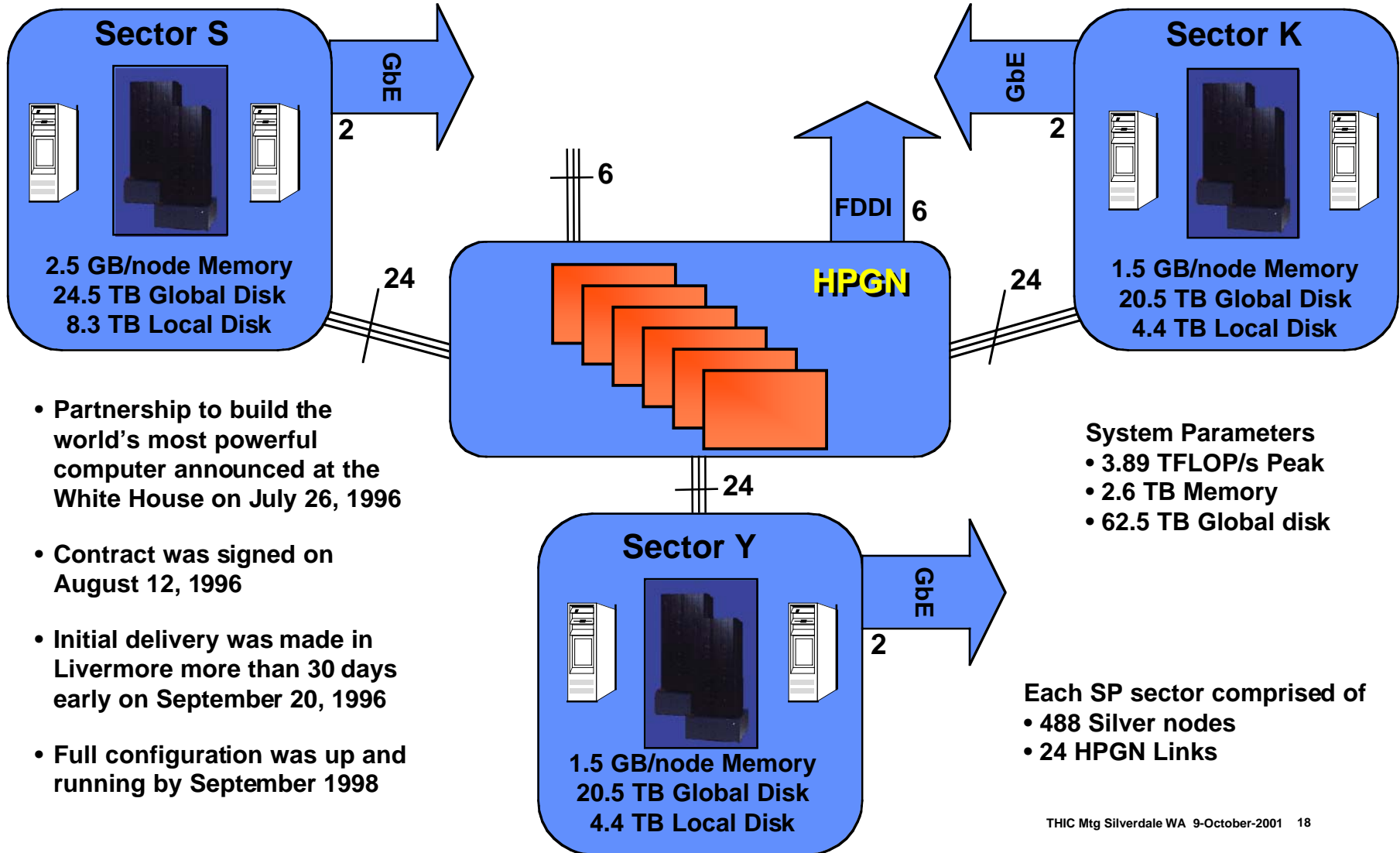


National security continues to require very high capability





Blue Pacific SST 3.9 TeraOP Hyper-Cluster Architecture

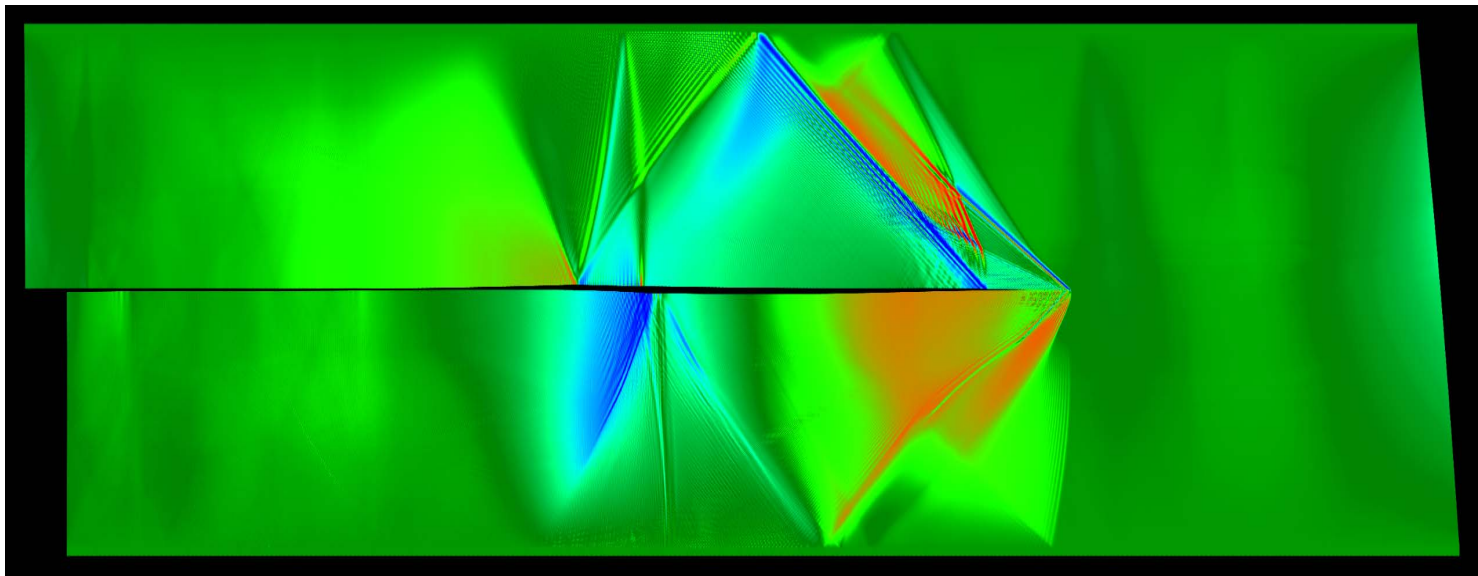




Smaller White system used for science runs Nov '00 - Feb '01



- Recent science runs on Frost have provided a unique opportunity to study material dynamics at the atomistic level with unprecedented problem sizes.
- **IBM Almaden, with LLNL materials scientists and visualization experts, successfully ran computations involving a billion atoms on 2000-5000 CPUs.**
- **Results: surprising discoveries that cracks can travel at supersonic speeds, and showing in unprecedented detail the complexities and structure of the molecular dislocation dynamics.**





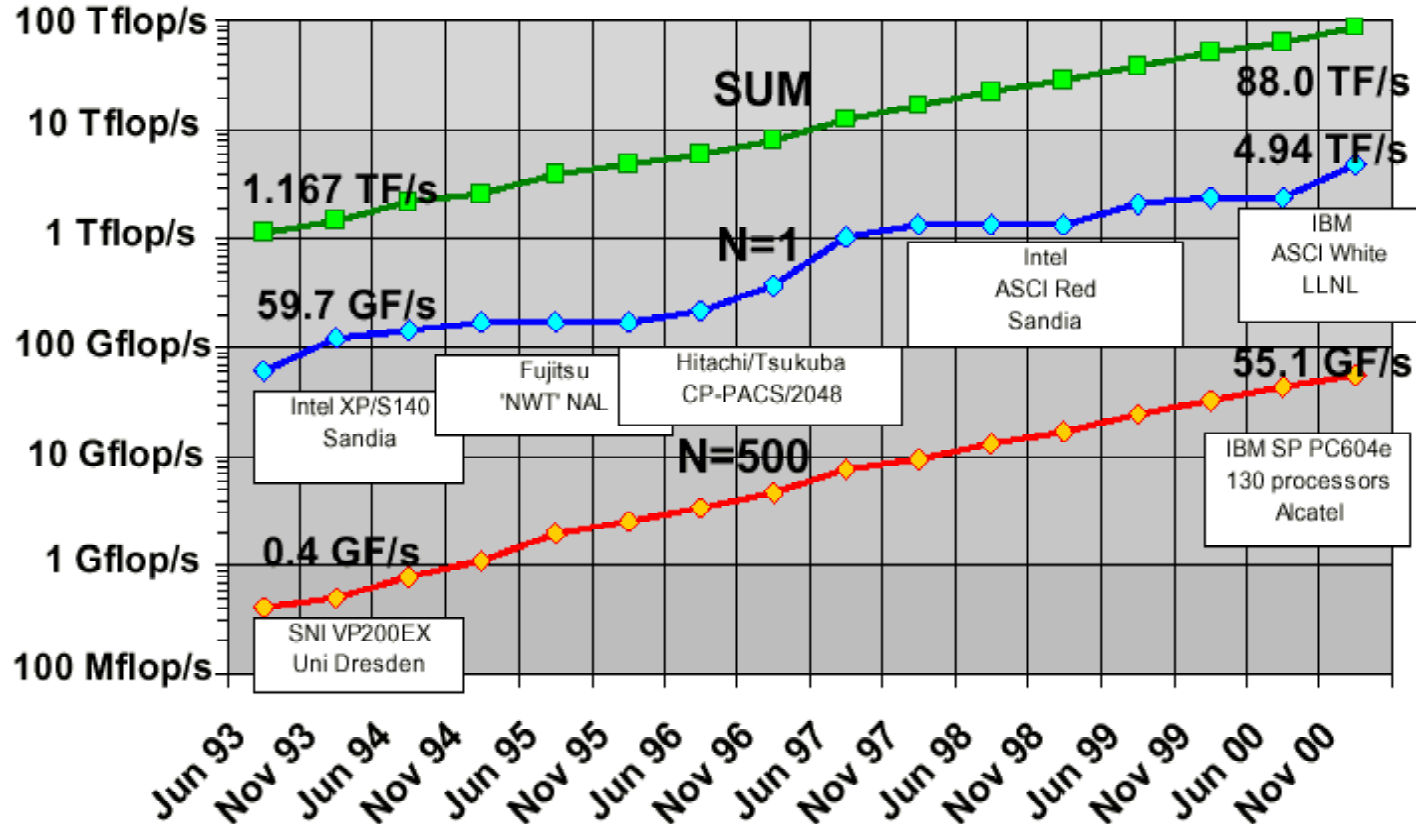
LLNL now operates the world's most powerful supercomputer



TOP500

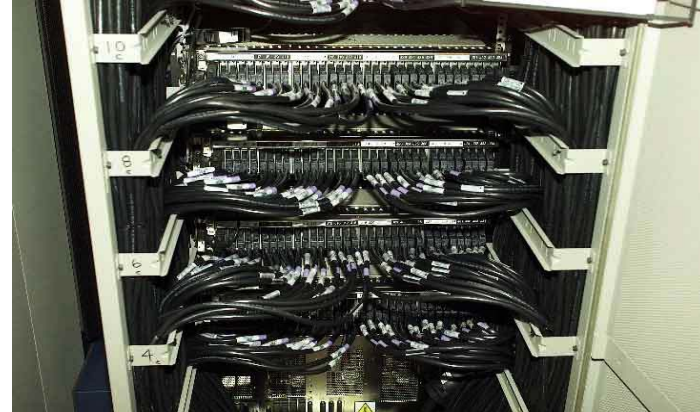
11/00

Performance Development





LLNL now operates the world's most powerful supercomputer



- **ASCI White peak speed of 12.3 TeraOPS (trillion operations per second)**
- **ASCI White weighs 106 tons, covers 10,000 square feet of floor space**
- **Contains 8,192 P3 375 MHz processors in 512 shared memory nodes**
- **Latest IBM technology: silicon-on-insulator with copper interconnects**
- **8 TB of memory, 36 TB local disk, 110 TB global (GPFS) disk**
- **5.12 GB/s local I/O and 12.8 GB/s global I/O bandwidths**
- **4 login nodes, 3 system nodes, 16 GPFS server nodes, 32 VIEWS nodes**
- **457 batch/compute nodes (each with 16 GB memory)**

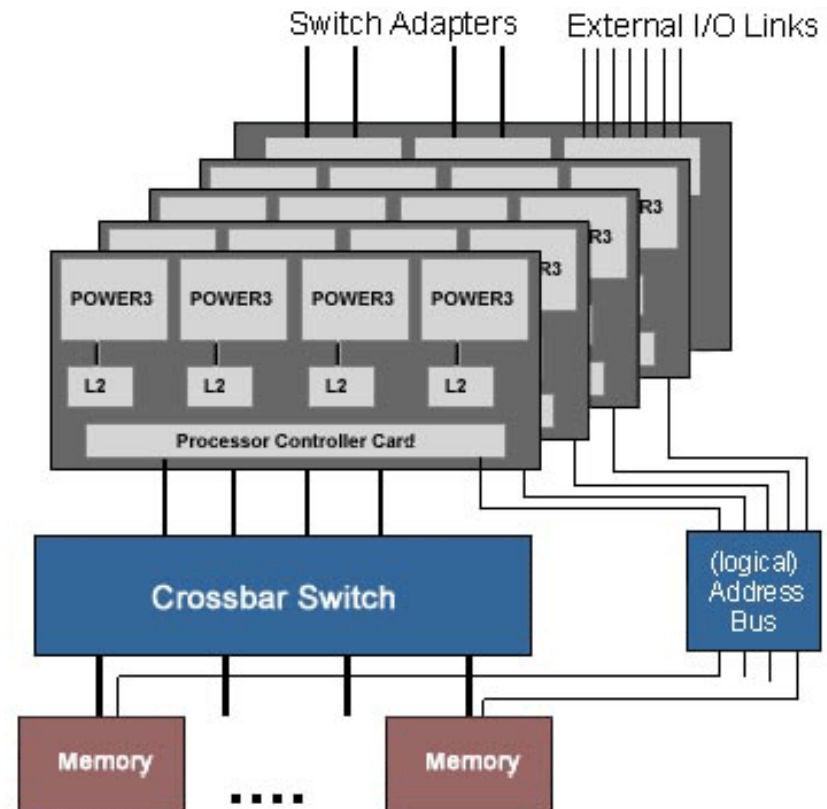


ASCI White IBM Nighthawk-2 Node Specifications



Number of CPUs per Node	16
CPU Clock Speed	375 MHz
Node Peak Perf.	~24 GigaOP/s
Memory per node	16 GB
Local Disk per node	72 GB

POWER3 processors are super-scalar pipelined 64-bit RISC chips with two floating-point units and three integer units. They are capable of executing up to eight instructions per clock cycle and up to four floating-point operations per cycle.

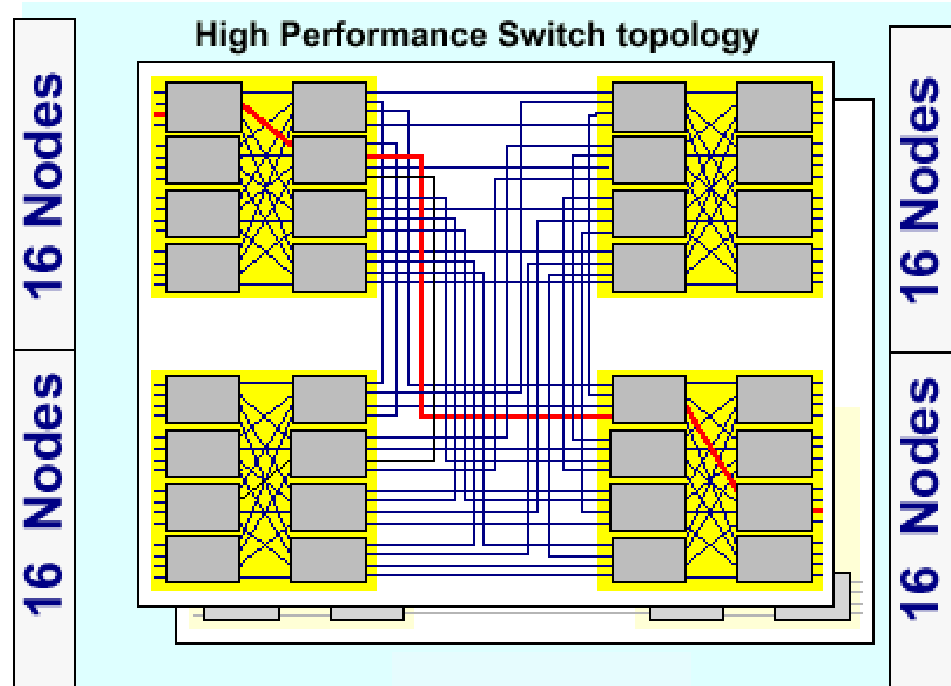




ASCI White IBM SP Switch2 high-performance technology



- Increased communication bandwidth with reduced message latencies
- Bi-sectional B/W over 4 Tb/s
- Compared to previous technologies, provides:
 - Faster interfaces, data paths, and microprocessor
 - Reduced microcode workloads using hardware assisted datagram reassembly
 - Microprocessor bus operations concurrent with data movement



- Switch adapters controlled by node software and on-card microcode
- 500 MB/s data rates each direction per adapter with message retry



Simulations managed from data generation to data assessment



- **High performance simulation requires balanced systems**
 - Supercomputers
 - GigE networks/switches
 - Local and archival storage
 - Data analysis/visualization
 - Algorithm development
 - Programming techniques
 - Facilities
 - floor space
 - power
 - cooling



State-of-the-art visualization facilities in same buildings that house code physicists/analysts



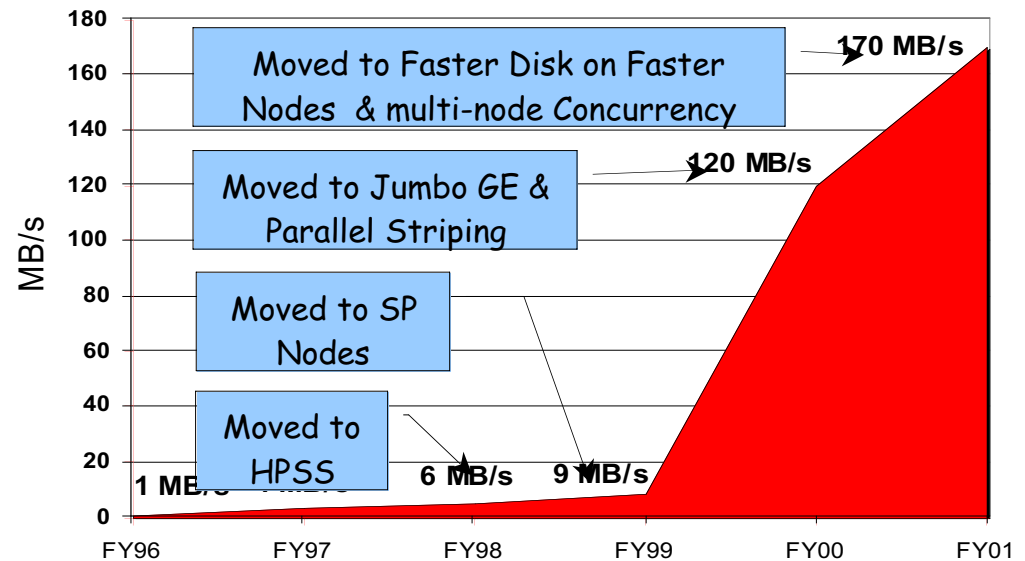
HPSS archival storage recent performance improvements



Accomplishments

- **A 20x performance increase in 15 months** (faster nets and disks)
- PSE Milepost demonstrated **170 MB/s aggregate throughput White-to-HPSS**
- Large single file transfer rates of up to **80MB/s White-to-HPSS**
- Large single file transfer rates of up to **150MB/s White-to-SGI**

Aggregate Throughput to Storage



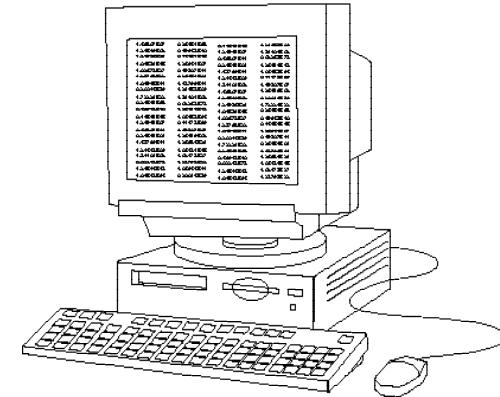
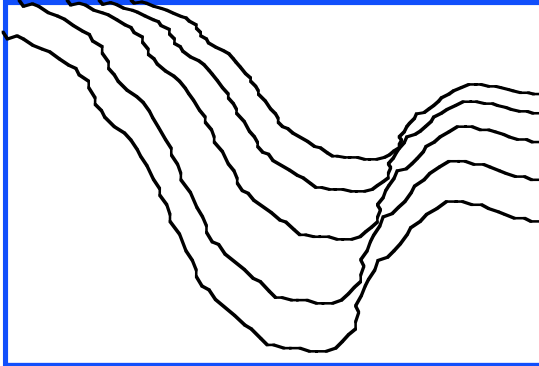
Challenges

- **Yearly doubling of throughput is needed for next machine**

At 170 MB/s, 2TB of data moves to storage in less than 4 hours. A year and a half ago it took two and a half days to move the same amount of data

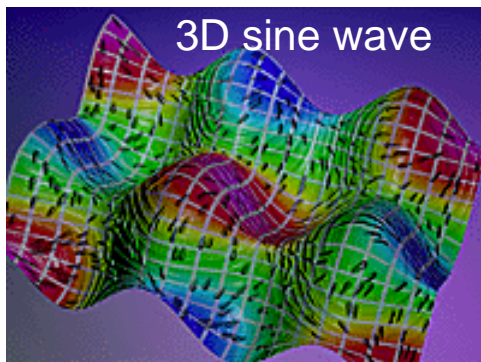


How does a scientist cope with terascale, 3D simulation data?



XY plots and 2D graphics cannot accurately represent 3D simulation data

“With 3D data sets, I can’t look at all the numbers anymore” ... LLNL scientist



New methods are needed to analyze high resolution, 3D simulation data.

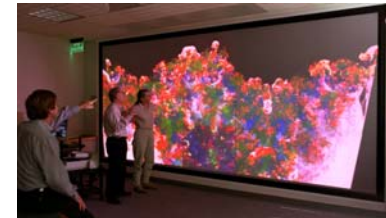
Terascale data sets will not fit on monitors.



Current ASCI PowerWall Capabilities at LLNL



- **B-132 Assessment Theater**
 - 5x3 tiled 20M pixel display
 - Usage: Two to three times per week for presentations to dignitaries, some data analysis, other demos
- **B-451 Video Cube PowerWall**
 - 3x2 modular 8M pixel display
 - Usage: Extensive both DNT and other
- **B-111 Visualization Work Center**
 - 4x2 tiled 10M pixel display
 - Usage: Milepost review in January, some for data analysis, several demos. Will grow more when resource management system deployed for instant access
- **B-451 Vis Development Lab PowerWall**
 - 2x2 tiled 20M pixel display
 - Usage: Demos (less since VideoCube), development





Outline



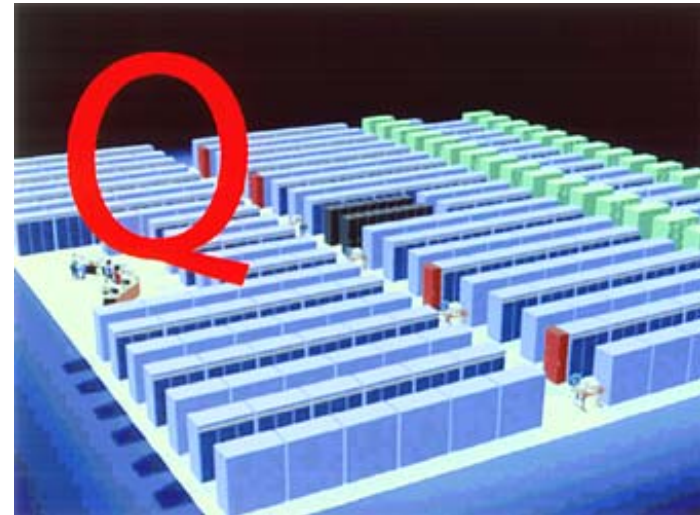
- The NNSA Stockpile Stewardship Program
- Where We are Now: ASCI White at LLNL
- Challenges for Today and for the Future



ASCI 30 TeraOPS “Q” System



- ~30 TeraOPS
- ~12,000 processors
- ~12 TB of memory
- ~600 TB usable disk storage
- Multi-rail high-speed switch
- ID System is first delivery
- FS-P1 - Final System Phase 1
- FS-BU - Final System Build-Up
- FS - Final System 30 TeraOPS





Strategic Computing Complex for siting the ASCI Q at LANL



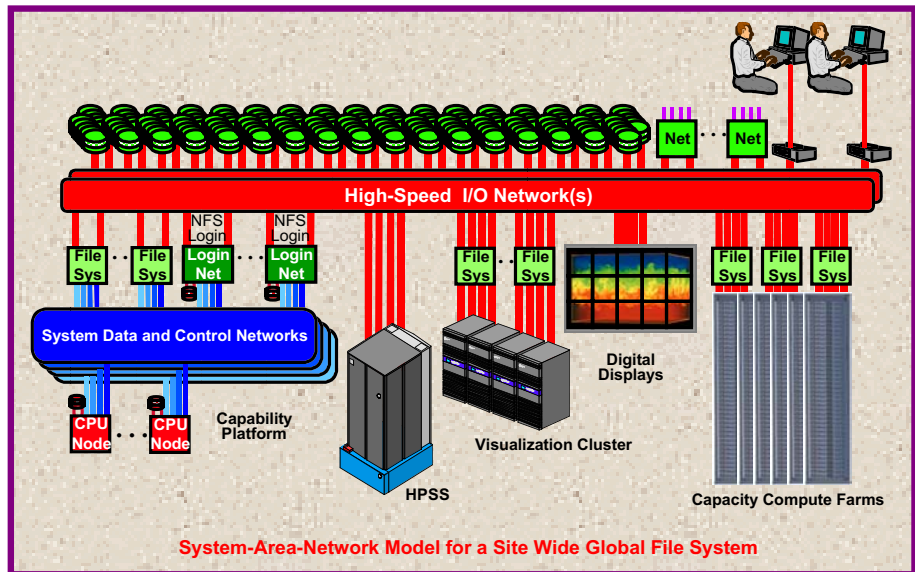
303,000 sq. ft. 43,500 sq. ft. unobstructed computer room
1 PowerWall Theater, 4 Collaboration rooms, 2 Immersive Rooms
Design Simulation Laboratories (200 classified, 100 unclassified)
200 seat auditorium



A 50+ TeraOPS procurement strategy for viz and filesystems



- **New ideas for visualization capabilities**
 - Include visualization requirements with platform procurement
 - Separately priced options with target of <10% platform budget
 - Framework for multiple solutions, bridge existing environment
 - Fast access to raw data and visualization files
 - Special network to commodity rendering resources
- **New ideas for networking, I/O and file systems**
 - Site-wide shared global file system
 - Possible open source development
 - 100+ GB/s delivered I/O xfer rates
 - External InfiniBand or 10Gb Enet
 - Parallel FTP transfers for speed





Warning: There is a major ASCI speed bump ahead



Some Challenges

- **ASCI will achieve its objectives.....**
 - 100 TF by 2004 + full systems code
 - Exceeding Moore's Law for 7-10 years
 - Energized other agencies and Nation
- **ASCI soon at its limit for acceleration**
 - 4x non-ASCI sites not sustainable
 - If the H/W doesn't break, the S/W complexity will kill you...
- **Major problems looming ...**
 - Cost, power, space
 - SW scalability, usability, reliability
 - Interconnect
 - Distance-to-memory issues
 - Availability



Quote from PITAC HPC summary

"Suppliers of high-end systems suffer from unusual market pressures..."



If ASCI acceleration stops, how to take simulation to next step?



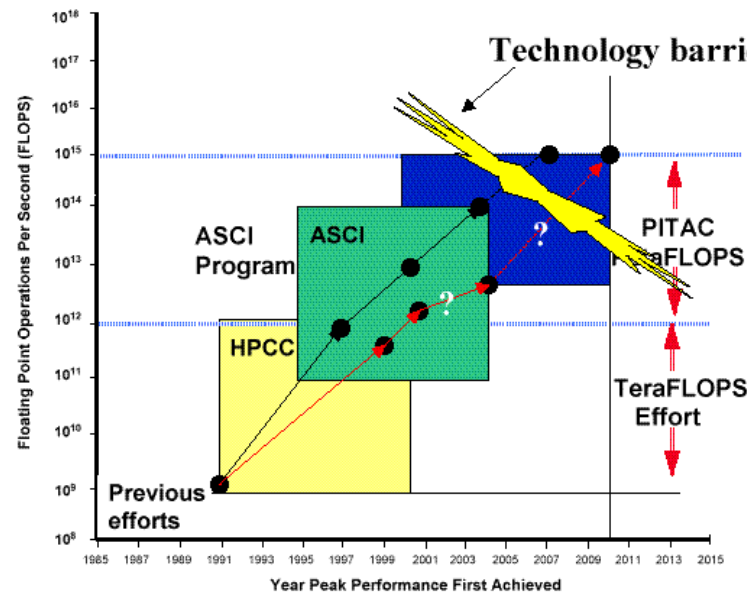
PITAC Recommendations

(President's IT Advisory Committee)

- \$\$ for R&D on innovative computing technologies
- \$\$ for software research
- \$\$ for Petaflops on some applications by 2010
- \$\$ to fund the most powerful high-end systems
- Can this be leveraged into a broad national program?

P
E
T
A
F
L
O
P
S

What Barriers to PetaFLOPS?

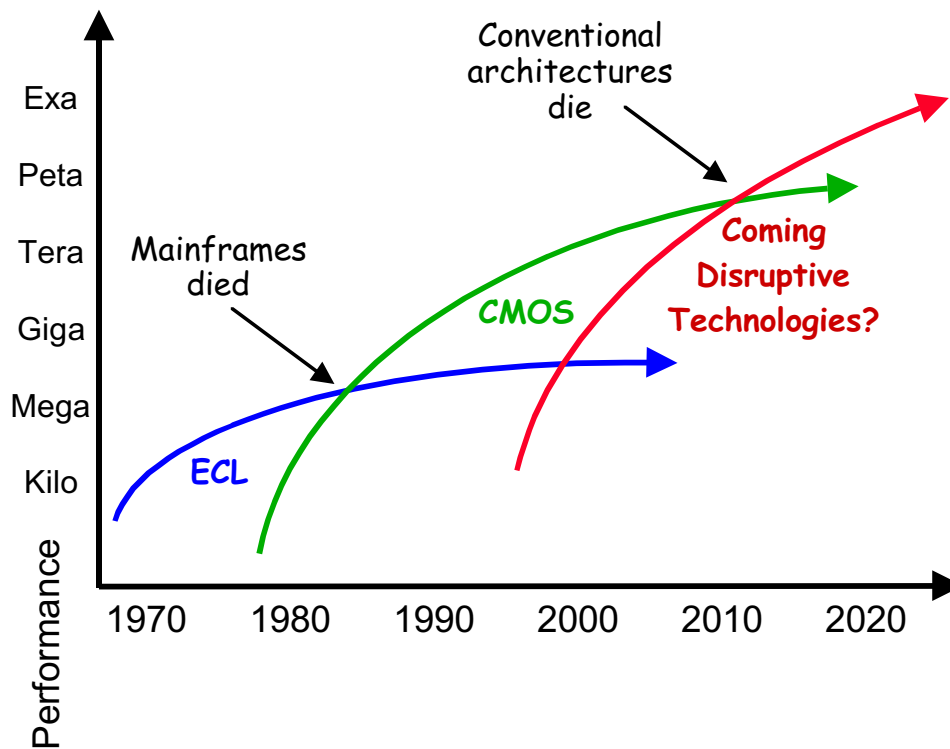




Disruptive technologies



Evolutionary Improvements
require R&D as well



- Disruptive technology tends to start small with much faster growth rate
- R&D today can impact 2010-2020 timeframe and accelerate a transition from PetaFLOP to ExaFLOP computations
- This transition will require fundamental research and development in:
 - Processor technology
 - Memory
 - Computer architecture
 - Operating systems
 - Programming environments
 - Scientific applications
 - Storage including MEMS and holographic devices



Summary and Conclusions



Unprecedented hardware and software simulation capabilities
have been built through the ASCI Program at LLNL, LANL, SNL

Advanced simulation capabilities have several major elements
(all of which must be present for effective use)

- Advanced codes, skilled scientists
- Advanced computing platforms and visualization

Simulation has become an integral part of science and technology
programs at the national labs

“We are changing the nature of scientific discovery”

ASCI-style acceleration is inevitably going to slow down

- R&D for evolutionary and disruptive technologies is the hope
- Partnerships with industry and academia critical to success

“The opportunities go to those who understand the trends”



DISCLAIMER

This document was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor the University of California nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial products, process, or service by trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or the University of California. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or the University of California, and shall not be used for advertising or product endorsement purposes.

This work was performed under the auspices of the U.S. Department of Energy by University of California Lawrence Livermore National Laboratory under contract No. W-7405-Eng-48.