National Nuclear Security Administration

Advanced Simulation and Computing

Advanced Architectures

Dr. José L Muñoz

Acting Dir. Of Simulation and Computer Science

jose.munoz@nnsa.doe.gov

Technology Barriers to PetaFlops [1994-95]

- Systems software

- Memory Speed

- Aggregate I/O

- Inter-processor speed

- Processor speed

- Packaging

- Power Management

Same issues remain today 2001

Proceedings of Petaflops Frontier Workshop

Motivation for Advanced Architectures

- ASCI has a well defined, and is currently executing, a roadmap leading to 100Tflop machines by year 2005 (funding constrained)

  - beyond 100Tflop?

  - beyond 2005?

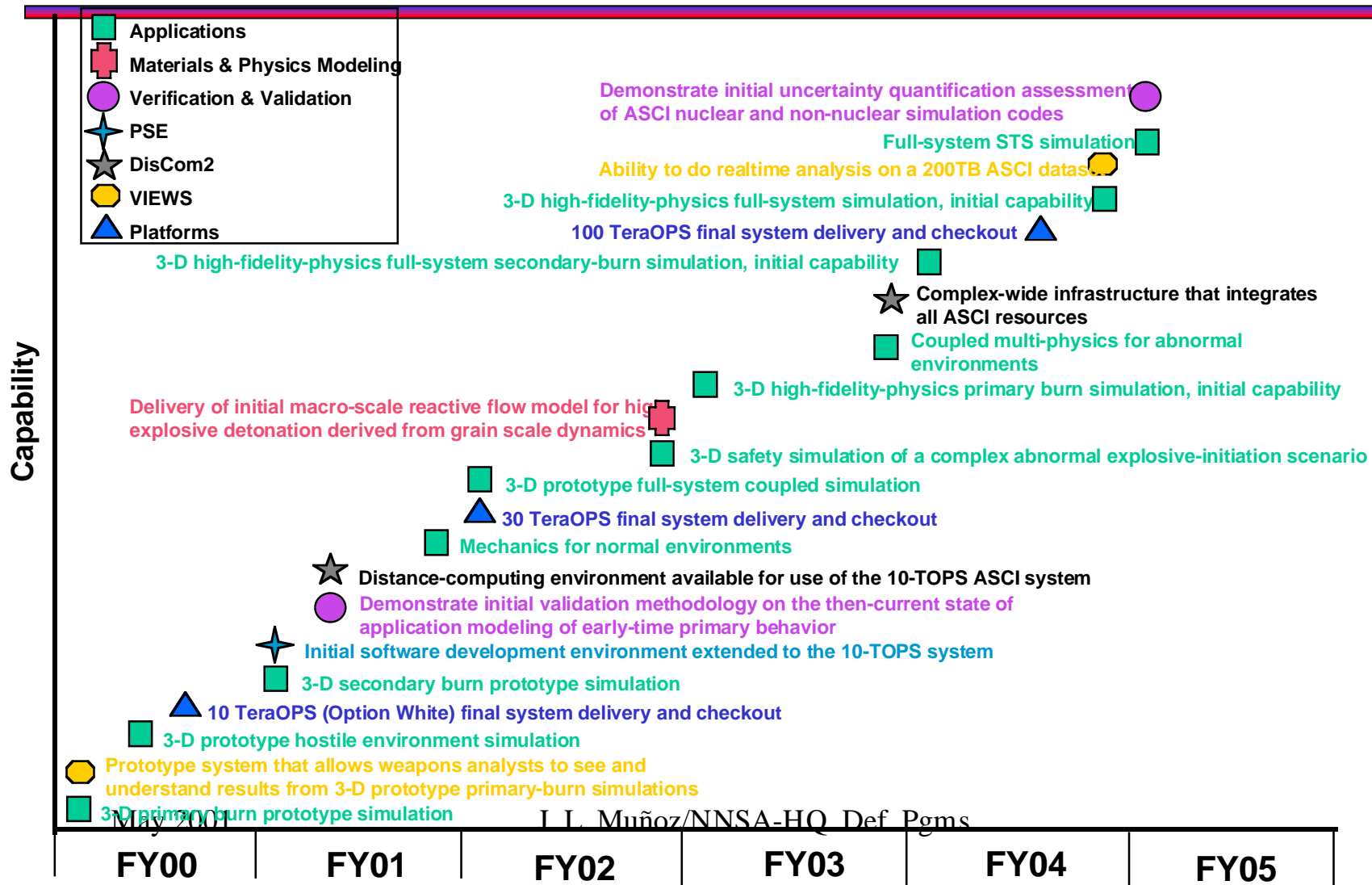 - alternative to building "super-computers" using commodity parts

Commodity Parts Model

- Given N X-Macho FLOP processors and "some kind" of processor interconnect, it is possible to build an NX Macho-Flop peak rating supercomputer:
  - how far can/should such a model be pushed?
  - Once we've reached that point (have we?) …. What alternatives do we have?

Aggressive Schedule

Capability (vertical axis label)

Legend:
- 🟩 Applications
- 🟥 Materials & Physics Modeling
- 🟣 Verification & Validation
- ✦ PSE
- ★ DisCom2
- ⬡ VIEWS
- 🔺 Platforms

🟣 Demonstrate initial uncertainty quantification assessment of ASCI nuclear and non-nuclear simulation codes

🟩 Full-system STS simulation

⬡ Ability to do realtime analysis on a 200TB ASCI dataset

🟩 3-D high-fidelity-physics full-system simulation, initial capability

🔺 100 TeraOPS final system delivery and checkout

🟩 3-D high-fidelity-physics full-system secondary-burn simulation, initial capability

★ Complex-wide infrastructure that integrates all ASCI resources

🟩 Coupled multi-physics for abnormal environments

🟩 3-D high-fidelity-physics primary burn simulation, initial capability

🟥 Delivery of initial macro-scale reactive flow model for high explosive detonation derived from grain scale dynamics

🟩 3-D safety simulation of a complex abnormal explosive-initiation scenario

🟩 3-D prototype full-system coupled simulation

🔺 30 TeraOPS final system delivery and checkout

🟩 Mechanics for normal environments

★ Distance-computing environment available for use of the 10-TOPS ASCI system

🟣 Demonstrate initial validation methodology on the then-current state of application modeling of early-time primary behavior

✦ Initial software development environment extended to the 10-TOPS system

🟩 3-D secondary burn prototype simulation

🔺 10 TeraOPS (Option White) final system delivery and checkout

🟩 3-D prototype hostile environment simulation

⬡ Prototype system that allows weapons analysts to see and understand results from 3-D prototype primary-burn simulations

🟩 3-D primary burn prototype simulation

FY00	FY01	FY02	FY03	FY04	FY05

Near-term Platform Strategy

- **FY00 and Beyond:**
  - LANL *Q,* ~30 teraOPS 2002
  - SNL, 20 teraOPS   2003
  - LLNL 60 teraOPS   2004
  - LANL 100 teraOPS 2005

- The sizes of these systems

is such that we are required

to build multi-1000 ft^2., multi-mega

watt facilities costing ~$100M
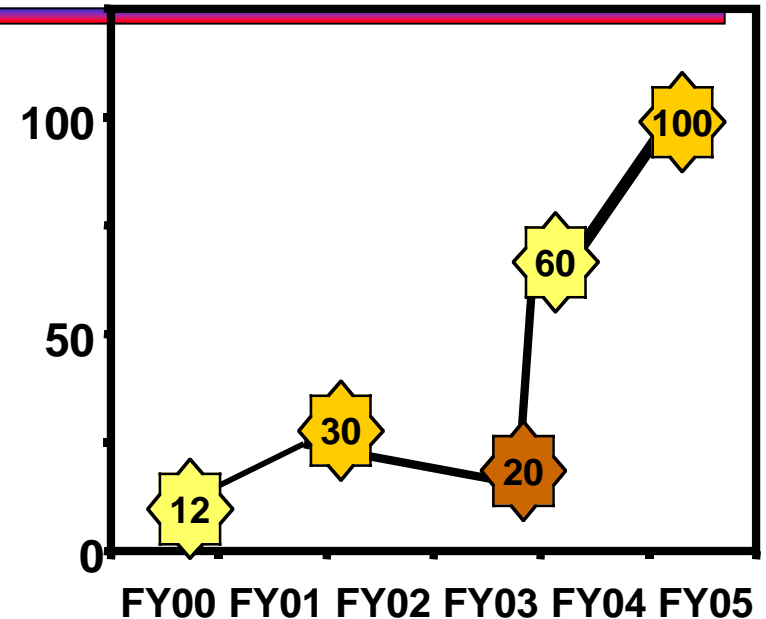
system capacities are budget limited

ASCI High-end Computing in 2001

- Application/architecture mapping and performance

  - ~30%, uniform grid

  - ~10%, block structured applications

  - ~1-3%, arbitrarily connected grids

- Processor speed/memory BW mismatch

- Programming models

Pathforward

- Near-term: 6 months - 2 years technology investments

- Much more Development, big "D", than research, little "r"

- commercial/industrial hardware and software vendors

- $25M/year over about 12 efforts

Advanced Architectures

Stockpile Stewardship Driven

- **Promote research and development**
  - **explore "alternative" high-performance computing architectures**
  - **not necessarily constrained by today's market forces**
  - **4+ years out**
  - **bigger "r", but not big-R**
- **Architectural approaches**
  - **directly addressing inadequacies in today's high-end systems that make it <u>very</u> difficult to achieve sustained high performance on complex applications.**

Advanced Architectures

- **Must** include system software

- **Well defined programming model**

- **Should address I/O and storage**

- **Work with other agencies such as DOE Office of Science, DARPA, NSF, NSA, ...**

Seeking complete system solution
"iron" alone is NOT SUFFICIENT

Goal is to address SSP needs… ASC is not a high-end computing program

Adv Arch Status

- Funding is only defined for FY01
  - < $4M
  - out-year funding must be identified
- Currently considering a proposal to fund in FY01
  - review panel has been convened
- Forming an ASCI "internal" team
  - process, "what", "when", "risks", "how"
  - budget (fund 3-4 efforts?)

Advanced Architectures

Back-Up

3 teraOPS ASCI platforms are operational at all Labs

ASCI Red
- SNL (Intel prime contractor)
- 3.15 teraOPS peak
- 2 processors per scalable unit
- UNIX operating system
- 9360 333 MHz processors
- 1.2 terabyte memory
- 12.5 terabyte storage

ASCI Blue-Mountain
- LANL (SGI prime contractor)
- 3.072 teraOPS peak
- 128 processors per scalable unit
- IRIX operating system
- 6144 250 MHz processors
- 1.5 terabyte memory
- 76 terabyte storage

ASCI Blue-Pacific
- LLNL (IBM prime contractor)
- 3.89 teraOPS peak
- 4 processors per scalable unit
- AIX operating system
- 5856 332 MHz processors
- 2.6 terabyte memory
- 52.5 terabyte storage

# Newer, more powerful platforms are in the pipeline

ASCI White

- LLNL (IBM prime contractor)
- 12.3 teraOPS peak
- 16 processors per scalable unit
- AIX operating system
- 8192 310 MHz processors
- 6 terabyte memory
- 160 terabyte storage

ASCI Q

- LANL (Compaq prime contractor)
- 30 teraOPS peak
- 32 processors per scalable unit
- Tru64 Unix operating system
- 11,968 1250 MHz processors
- 12 terabyte memory
- 600 terabyte storage
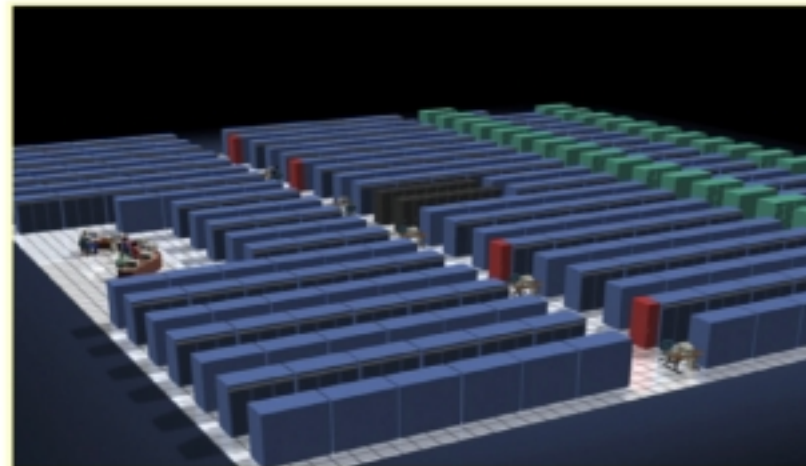
Pathforward Efforts

- **Hardware**
  - ◇ Compaq
  - ◇ IBM
  - ◇ Sun
  - ◇ SGI
  - ◇ ABBA
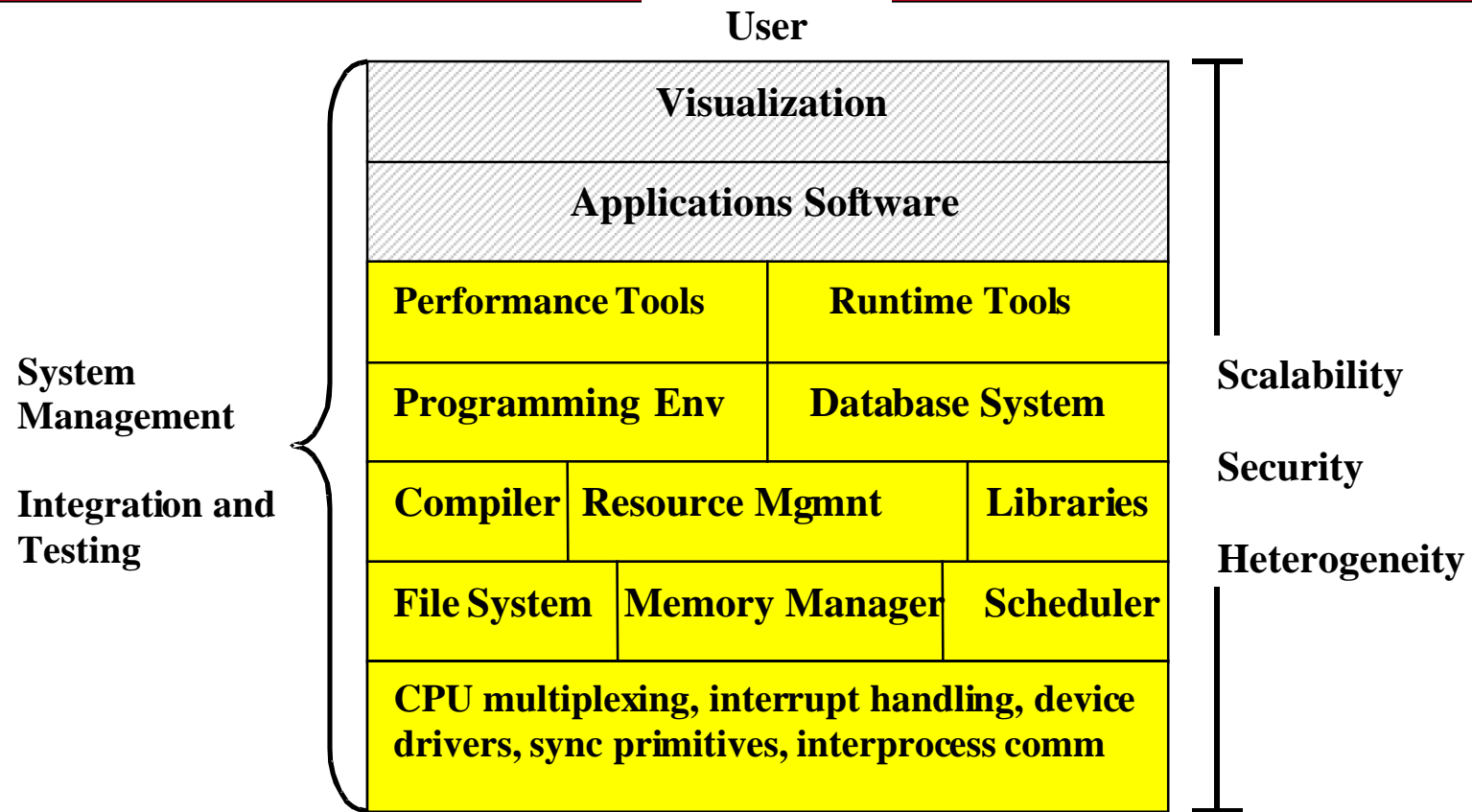 - ◇ LOTS
  - ◇ Storage Tek
  - ◇ "Scalable rendering

- **Software**
  - ◇ Etnus
  - ◇ KAI/Pallas
  - ◇ MSTI

  □ **End in FY01**

Systems Software

User

Visualization		
Applications Software		
Performance Tools	Runtime Tools	
Programming Env	Database System	
Compiler	Resource Mgmnt	Libraries
File System	Memory Manager	Scheduler
CPU multiplexing, interrupt handling, device drivers, sync primitives, interprocess comm		

System Management

Integration and Testing

Scalability

Security

Heterogeneity

Adapted from Distributed Operating System & Algorithms, Chow/Johnson