CSTB • Computer Science and Telecommunications Board

The Future of Supercomputing Study

Marc Snir

April, 2004

THE NATIONAL ACADEMIES
Advisers to the Nation on Science, Engineering, and Medicine

About the Study

- Conducted by CSTB (within the NRC at the National Academies)
- Sponsored by DOE Office of Science and DOE Advanced Simulation and Computing
- Study Goal:
  - Supercomputing R&D in support of U.S. needs
  - Applications and implications for design
  - Market, national security and the role of U.S. govt
  - Options for progress/recommendations (Final Report)
- Emphasis on “one-machine-room” systems
- Interim (July 2003) and Final (end 2004) Reports

CSTB • Computer Science and Telecommunications Board

Study Committee

- **SUSAN L. GRAHAM**, University of California, Berkeley, *Co-chair*
- **MARC SNIR**, University of Illinois at Urbana-Champaign, *Co-chair*
- WILLIAM J. DALLY, Stanford University
- JAMES DEMMEL, University of California, Berkeley
- **JACK J. DONGARRA**, University of Tennessee, Knoxville
- KENNETH S. FLAMM, University of Texas at Austin
- MARY JANE IRWIN, Pennsylvania State University
- **CHARLES KOELBEL**, Rice University
- BUTLER W. LAMPSON, Microsoft Corporation
- ROBERT LUCAS, University of Southern California, Information Sciences Institute
- PAUL C. MESSINA, Argonne National Laboratory (part-time)
- JEFFREY PERLOFF, Department of Agricultural and Resource Economics, University of California, Berkeley
- WILLIAM H. PRESS, Los Alamos National Laboratory
- ALBERT J. SEMTNER, Oceanography Department, Naval Postgraduate School
- SCOTT STERN, Kellogg School of Management, Northwestern University
- SHANKAR SUBRAMANIAM, Departments of Bioengineering, Chemistry and Biochemistry, University of California, San Diego
- LAWRENCE C. TARBELL, JR., Technology Futures Office, Eagle Alliance
- STEVEN J. WALLACH, Chiaro Networks
- CSTB: Cynthia A. Patterson (Study Director, Margaret Huynh

What is in the Interim Report?

- Summary of earlier reports
- Current state of supercomputing
- Identification of issues
  - Evolution
  - Innovation
  - The role of government
- No specific findings or recommendations

Supercomputing Technologies: Microprocessor

32-bit

Better memory subsystem

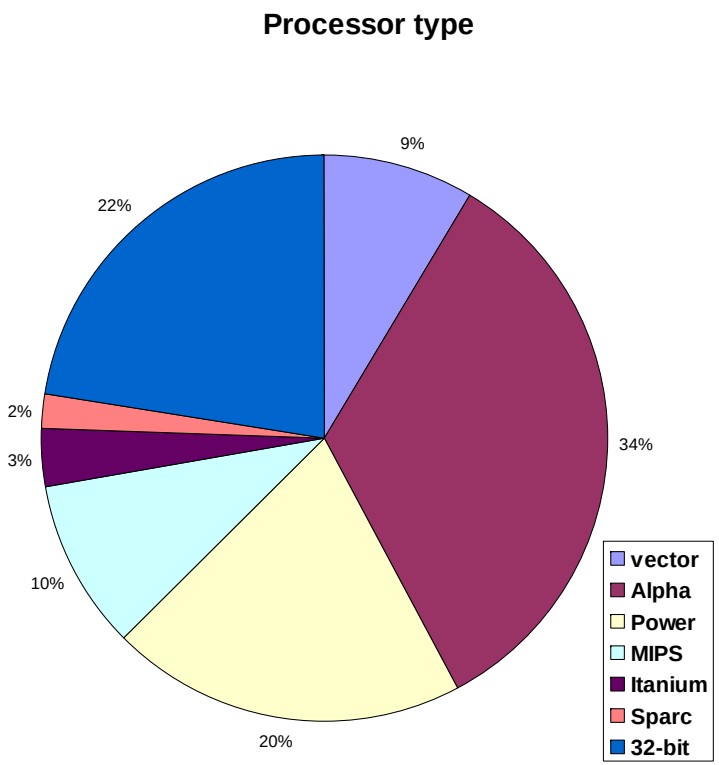
Volume (=price)

64-bit

Higher memory bandwidth

custom

Memory latency hiding (e.g., vector)



Supercomputing Technologies: Switch & Interface

LAN technology (Ethernet, GEthernet)

Volume (=price)

performance, function

Custom switch on standard IO bus (Quadrics, Myriacom…)
Higher bandwidth, lower latency, protocol offload

Custom switch on proprietary direct memory attach interface (Cray, IBM…)
Higher bandwidth, lower latency, direct memory access

Supercomputers: 3 Main Types

- **All custom** (9%)
  - Within TOP500, synonymous with vector
    - NEC Earth Simulator (#1), Cray X1

- Best processor performance even for codes that are not "cache friendly"
- Good communication performance
- Simplest programming model
- Most expensive

- **Commodity microprocessor, custom interface & switch** (50%)
    - IBM ASCI White (#4), SGI Origin, Sun Fire, Cray T3E

- Good communication performance
- Good scalability

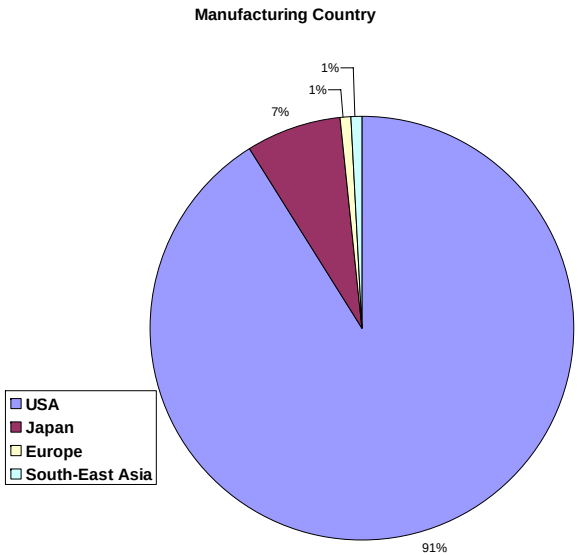
- **Commodity microprocessor, standard interface** (40%)
  - Most use custom (non LAN) switch
  - Half use 32-bit processors today
    - HP ASCI Q (#2), LLNL Linux Network (#3)

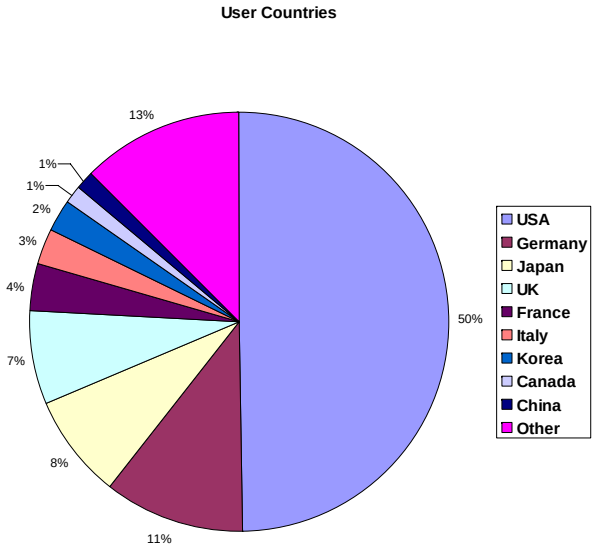
- Best price/performance (for codes that work well with caches and are latency tolerant)
- More complex programming model

Each type has its own niche!

The Glass is Half Full

- Supercomputers are used to do important research
- Most supercomputers are manufactured and used in the US





The Glass is Half Empty: Supercomputer software is in bad shape

- Supercomputers are too hard to program
- Software developers have inadequate environments and tools
- Legacy software is difficult to adapt and evolve to newer platforms and more ambitious problems

Why is there a software problem?

- Inadequate investment
  - unlikely that the current model (platform vendors + open source) can provide standardized, high-quality programming environments
  - PC software is often unsuitable
- Few 3rd party supercomputing software vendors
  - few vendor-supported application codes
  - few vendor-supported portable tools (Etnus – Totalview)
- Lack of standards (IO, tools)
- Lack of perseverance: e.g., High Performance Fortran

CSTB • Computer Science and Telecommunications Board

Start from Scratch or Evolve?

Both

THE NATIONAL ACADEMIES
Advisers to the Nation on Science, Engineering, and Medicine

The Case for Evolution

- Current platforms do important work (ASCI)
  - Need both capability and capacity
- No near-term silver bullet in the offing
  - Technology pipeline is fairly empty
- It's never one size fits all
  - Relative ranking of architectures is problem/time/cost dependent; expect three main species to be around for a while
- Technology evolves over decades
  - X1 inherits from 25 years of Cray and T3D/E
  - Clusters inherit from 20+ years of MPP/NOW/COW/Beowulf research

The Case for Evolution (2)

- Commodity based clusters will continue to play a role
  - Cost/performance advantage (for suitable applications)
  - Scalable technology – compatible with technology used in academic research groups and departments
- Software and application work done on today's machines will be adapted to tomorrow's machines
  - Need massive parallelism to scale up on any architecture
- Legacy codes have continuing utility
  - Often need to run on old-style architecture
- Uncertainty and inconsistent policies are expensive
  - Companies disinvest, R&D teams disappear, researchers move to greener pastures

The Case for Sustained Research Investment

- Field is not mature: base technology continues fast evolution
  - Non-uniform scaling causes major dislocations (e.g., processor vs. memory speeds)
  - Supercomputers are early victims of non-uniform scaling
  - Solutions require both hardware and software innovation
- New applications challenges abound
  - Scaling and coupling
  - Massive amounts of data

Breadth and Continuity are Essential

- Continuous, steady investments at all stages of technology pipeline
  - basic research, technology demonstration, product development
- Continuous, steady investment in all major communities
  - academia, national labs, vendors …
- Mix of small science (individual projects) and large science (collaborative teams)
- Avoid linear view (successive elimination) and maximize flow of ideas and people across projects and concepts

Examples of Research Directions

- Architecture
  - Better memory architecture (higher bandwidth, latency hiding)
  - Better support for higher-level, portable, inherently parallel virtual architecture
- Software
  - Programming environments and tools where parallelism is innate
  - Programming environments and tools that match HPC code development process
    - Code often developed from mathematical formulation of physical problem
    - Code developed by small team of domain experts
  - OS that manages the entire system as one entity and that supports parallel applications as first class citizens

Examples of Research Directions (2)

- Applications and algorithms
  - Scaling of some existing disciplinary methods
  - New algorithms and formulations
  - Interdisciplinary challenges (e.g., coupling)
  - Shift from analysis to synthesis and optimization
  - Very large and complex data sets (e.g., biology)
  - Adaptation to changing machine models

The Role of Government in Supercomputing

- Government is main user of supercomputers
  - directly (e.g. defense and national security)
  - indirectly (e.g. innovation in drug design, via Medicare/Medicaid)

➡ Government must ensure that supercomputing technology evolves at a rate and in a direction that serves government missions

The Role of Government in Supercomputing (2)

- Supercomputers are essential to national security
  - Cryptanalysis, weapon design, battlefield-related calculations, homeland security…
- Government must ensure strong supercomputing technology base in the U.S.
  - Guarantee availability to the U.S.
  - Provide technological advantage to the U.S.
  - Ensure that systems can be trusted

The Role of Government in Supercomputing (3)

- Market-based incentives are insufficient because innovators do not capture the full value of their invention
  - supercomputer vendors have not captured economic benefit of technologies that flowed to mainstream computing
- ➡ Government needs to support development of supercomputing technology and supercomputer use in support of science

Interim Report Summary

- Supercomputing is important for the future
- Need balance between customization and commodities
- Need balance between evolution and innovation
- Need continuity and sustained investment
- Government role is essential

Personal Observations

- Cannot discuss content of final report
- Can offer some personal views that may or may not find their way in the final report.

Need Better Supercomputing Advocacy

- Supercomputing has to be justified by the existence of problems
  - that are of major importance
  - that can solved via simulation
  - where the impediment to solution is machine performance
  - where time to solution is critical
- E.g., ASC or NSA story
- Few good (new) stories coming from science community, and none from industry
- Need Nobel prize winners to lobby for supercomputers, in addition to the HPC aficionados!

Need Mechanisms to Support Large, Persistent Investments in Software

- A major (ASCI) project code takes $100M, 8 years and an expert, stable, closely knit team of 20-30 people to develop; the team has to continue upgrading, porting, maintaining for 10-30 years [Post, Kendall].
  - DOE labs are paying the tab and providing the environment.
- Same (more?) is needed to develop and maintain a good portable compiler or a good portable parallel file system.
  - who is paying the tab?
  - who is providing the stable environment and access to platforms?
- Who will support community codes for 20 years?

Possible Mechanisms

- Well funded ISV (big enough and specialized enough not to be swallowed by Intel)
- Publicly funded applied parallel software institute (> 100 HC, > $30M/y)
- Teams within national labs.

An HPC Ecosystem

- A set of technologies that are interdependent and mutually reinforce each other
- A set of companies developing these technologies in strong interaction; a set of interdependent products
- A set of people that hold the expertise for these technologies and communicate frequently

Examples

clusters
switch (Myriacom, Quadrix)
cluster OS & middleware
message passing libraries
apps

vector
vector machines
vectorizing compilers
vectorized apps

scalable shared memory
scalable SMPs (SGI?)
scalable or cellular OS
OpenMP and shared memory libs
shared memory apps

The Ecology of Ecosystems

- Once created ecosystem is kept stable because of
  - networking effects
  - different life cycles of different technologies
- Need a critical mass to be viable
  - to have good coverage of all key technologies
  - to avoid extinction due to catastrophic events
  - to ensure cross-fertilization
- Creating a new ecosystem requires major investment (or major new opportunity)
  - barriers have grown as technology investments have grown
  - technology disruptions provide the opportunity

Observations on HPC Ecosystems

- Each ecosystem has its “raison d’être”
- Few ecosystems (any?) have critical mass
  - same problems rediscovered again and again
  - catastrophes can (almost) wipe out an ecosystem
- Only one ecosystem (cluster) is economically viable without heavy govt. investment
  - even “cluster” requires significant investments to continue scaling up and for increased productivity

Ecosystem Maintenance

- How many can we afford?
  - 1? 2? 3?
- How do we nurture ecosystems?
  - long-term, stable investments
  - communities, consortia
- How do we provide opportunities for the creation of new ecosystems?
  - large-scale prototypes, large community efforts

Portability: Growing the Ecosystem

- Ecosystem largely defined by "performance profile" of platforms
  - ratio compute speed to memory bw/ latency to global bw/latency…
- A programming model is implicitly designed to fit a performance profile
- Portability across broader range of performance characteristics would allow to merge ecosystems
- But portability across performance profiles is **hard**
  - Necessary, to some extend, because of differential growth of hardware technologies
- Tradeoff: better productivity within narrower range of platforms vs. better portability

CSTB • Computer Science and Telecommunications Board

Final Plea

- Report content will be finalized in coming 2-3 months
- A good report may not help but a bad report will certainly hurt; please send comments, suggestions, information to me or to Sue Graham

THE NATIONAL ACADEMIES
Advisers to the Nation on Science, Engineering, and Medicine