LA-13474-C
Conference

Proceedings from the Conference on

High Speed Computing

LANL • LLNL

The Art of High Speed Computing

April 20–23, 1998

Salishan Lodge
Gleneden Beach, Oregon

Los Alamos
NATIONAL LABORATORY

Los Alamos National Laboratory is operated by the University of California for the United States Department of Energy under contract W-7405-ENG-36.

Photocomposition by Wendy Burditt, Group CIC-1

Special thanks to Orlinie Velasquez and Verna VanAken for coordinating this effort.

An Affirmative Action/Equal Opportunity Employer

This report was prepared as an account of work sponsored by an agency of the United States Government. Neither The Regents of the University of California, the United States Government nor any agency thereof, 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 product, process, or service by trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by The Regents of the University of California, the United States Government, or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of The Regents of the University of California, the United States Government, or any agency thereof. Los Alamos National Laboratory strongly supports academic freedom and a researcher's right to publish; as an institution, however, the Laboratory does not endorse the viewpoint of a publication or guarantee its technical correctness.

LA-13474-C
Conference

UC-705
Issued: August 1998

Proceedings from the Conference on High Speed Computing

The Art of High Speed Computing

April 20–23, 1998

Compiled by
Kathleen P. Hirons
Manuel Vigil
Ralph Carlson

Table of Contents

Conference Program vii

Abstract ix

Keynote Address 1
Steve Wallach, CenterPoint Venture Partners

Stockpile Stewardship and Management Program 25
Larry Ferderber, LLNL

Predictability, and the Challenge of Certifying a Stockpile Without Nuclear Testing 41
Ray Juzaitis, LANL

100 TeraFLOPs and Beyond, an Industry View into the Future 59
Panel: Moderator—John Morrison, LANL; Mark Seager, LLNL; Panelists—Tilak Agerwala, IBM; Greg Astfal, Hewlett Packard; Erik Hagersten, Sun Microsystems; Richard Kaufmann, Digital Equipment Corp.; Steve Oberlin, SGI/Cray.

ASCI Aliances Program.. 105
Ann Hayes, LANL

The Next Fifty Years of Computer Architecture 123
Burton Smith, Tera Computer Company

Full Wave Modeling of Signal and Power Interconnects for High Speed Digital Circuits 133
Gary Haussman, University of Colorado

Simulating the Physical-Biological Factors Affecting Abundance of Calanus finmarchicus in the Gulf of Maine 141
Wendy Gentleman, Dartmouth College

The Next Generation Internet 143
Bob Aiken, DOE

Petaflops Computing: Opportunites and Challenges 151
David Bailey, LBNL

President's Information Technology Advisory Committee (PITAC): A Mid-Term Report 165
David M. Cooper, LLNL

Processing-in-Memory: Past and Present 177
Ken Iobst, IDA/CCS

High Volume Technology for HPC Systems 185
Justin Rattner, Intel Corporation

High Performance Computing and the NCAR Procurement—Before, During, and After 209
Bill Buzbee, NCAR; Jim Hack and Steve Hammond, National Center for Atmospheric Research

Economies of Scale: HPC into the Next Millennium 223
Gary Smaby, Smaby Group

The Other Side of Computing 241
William Trimmer, Belle Mead Research, Inc.

Crystalline Computing 249
Norm Margolus, MIT

Quantum Computing 257
Emanual Knill, LANL

List of Attendees 277

Conference Program

Monday, April 20, 1998

Keynote Session:

Keynote Address: Billions and Billions
Steve Wallach, CenterPoint Venture Partners

Tuesday, April 21, 1998

Session 1: The Stockpile Stewardship and Management Program

Stockpile Stewardship and Management Program
Larry Ferderber, LLNL

Predictability, and the Challenge of Certifying a Stockpile Without Nuclear Testing
Ray Juzaitis, LANL

Session 2: The Challenge of 100 TeraFLOP Computing

100 TeraFLOPs and Beyond, an Industry View into the Future
Panel Discussion: Moderator—John Morrison, LANL and Mark Seager, LLNL; Panelists—Tilak Agerwala, IBM; Greg Astfalk, Hewlett-Packard; Erik Hagersten, Sun Microsystems; Richard Kaufmann, Digital Equipment Corp.; Steve Oberlin, SGI/Cray.

Session 3: ASCI Alliance

ASCI Alliances Program
Ann Hayes, LANL

Session 3.5: Hardware Design

The Next Fifty Years of Computer Architecture
Burton Smith, Tera Computer Company

Banquet

Adversarial Inspection in Iraq: 1991 and Thereafter
Jay Davis, LLNL

Wednesday, April 22, 1998

Session 5: Student Session

Full Wave Modeling of Signal and Power Interconnects for High Speed Digital Circuits
Gary Haussmann, University of Colorado

Simulating the Physical-Biological Factors Affecting Abundance of Calanus finmarchicus in the Gulf of Maine
Wendy Gentleman, Dartmouth College

Session 6: News You Can Use

The Next Generation Internet
Bob Aiken, DOE

Petaflops Computing: Opportunities and Challenges
David Bailey, LBNL

President's Information Technology Advisory Committee (PITAC): A Mid-Term Report
David M. Cooper, LLNL

Thursday, April 23, 1998

Session 7: Chip Technology for Large Scale Systems

Processing-in-Memory: Past and Present
Ken Iobst, IDA/CCS

High Volume Technology for HPC Systems
Justin Rattner, Intel

Session 8: Reality Check

High Performance Computing and the NCAR Procurement—Before, During, and After
Bill Buzbee, NCAR; Jim Hack and Steve Hammond, National Center for Atmospheric Research

Economies of Scale: HPC in the Next Millennium
Gary Smaby, Smaby Group

Session 9: Future

The Other Side of Computing
William Trimmer, Belle Mead Research, Inc.

Crystalline Computation
Norm Margolus, MIT

Quantum Computing
Emanuel Knill, LANL

Proceedings from the Conference on High Speed Computing

The Art of High Speed Computing

April 20–23, 1998

Compiled by
Kathleen P. Hirons
Manuel Vigil
Ralph Carlson

Abstract

This document provides a written compilation of the
presentations and viewgraphs from the
1998 Conference on High Speed Computing.
"The Art of High Speed Computing,"
held at Gleneden Beach, Oregon, on April 20 through 23, 1998.

BILLIONS & BILLIONS

STEVE WALLACH
CENTER POINT VENTURES
WALLACH@CENTERPOINTVP.COM

ASPECTS OF BILLIONS

- Raised to the power (giga, tera, peta, exa)
- The inverse (nano, pico, femto, atto)
- In the computer industry they are closely related. From a technology and investment perspective
- US government policy must be consistent with industry trends. (the ultimate venture capitalist)

2

PRESENTATION OUTLINE

- Fundamental Laws- Physics
- Trends in Telecommunications
- Trends in Semi-conductors
- Trends in Computer Architecture
- Draw some conclusions
- US Government Policy

3

FUNDAMENTAL LAWS

- C - Speed of light
- Power Consumption
- Propagation Delay

4

POWER CONSUMPTION

$$P \cup C * V^2 * F$$

- C= capacitance
- V= voltage
- F= frequency

5

PROPAGATION DELAY

- Lossless Line

  $$\text{Time} = \sqrt{LC}$$

- Lossly Line

  $$\text{Time} = L * \sqrt{\varepsilon_r / c_o}$$

  $$\varepsilon_r = \text{Dielectric Constant}$$

  $$c_o = \text{Speed of Light}$$

6

OTHER CONSTRAINTS

- Cost of Investment - I (billions)
- Size of Market - M (millions)
- L'Hospital's Rule of Profit
  - Profit = dM/dI
    - **as I approaches infinity**
    - **as M approaches K (sometimes 0)**
  - result is { 1 (success) | 0 (failure)}
- The government uses different rules

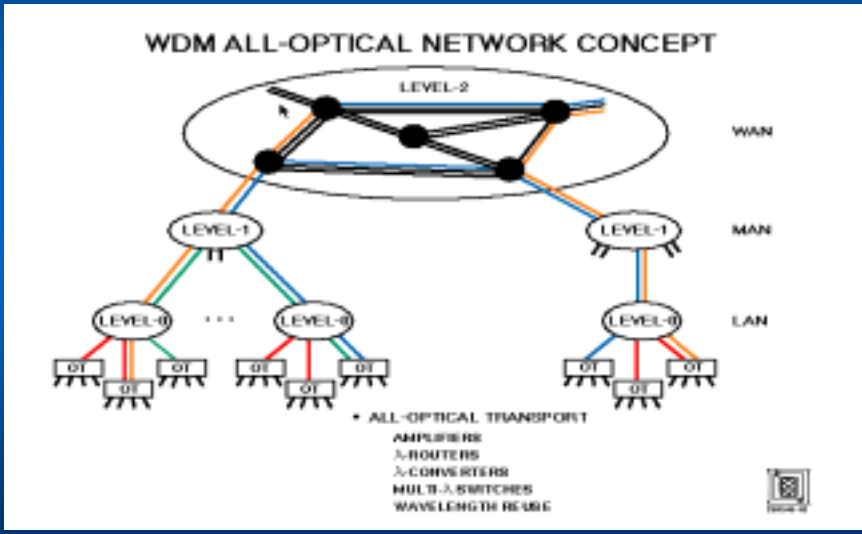
7

TELECOMMUNICATIONS

- Advances in ***PHOTONIC*** (mainly ***WDM***) technology.
- ***TERAHZ*** (THz) requirements
- All optical networks (***AON***)
- Effect on digital computer architecture
- The next supercomputer topology
  - **www.ll.mit.edu/aon/**
  - **Lemott, et. al., " low-cost WDM", Aug. 97, IEEE summer topicals, Montreal.**

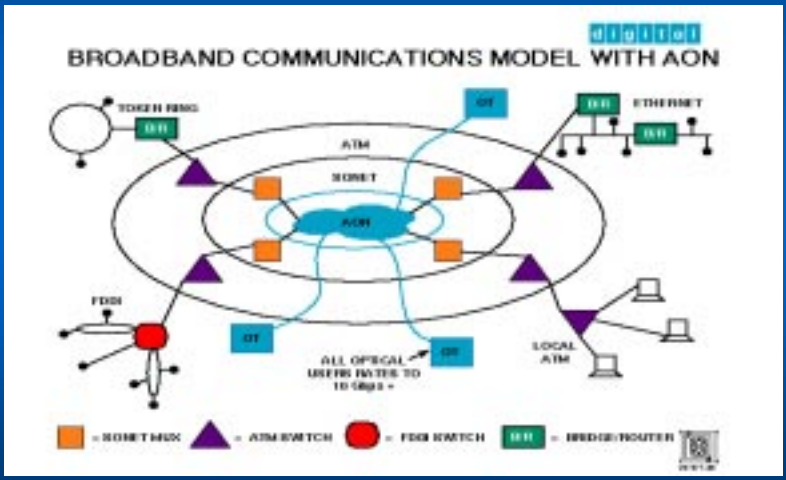
8

WDM ARCHITECTURE



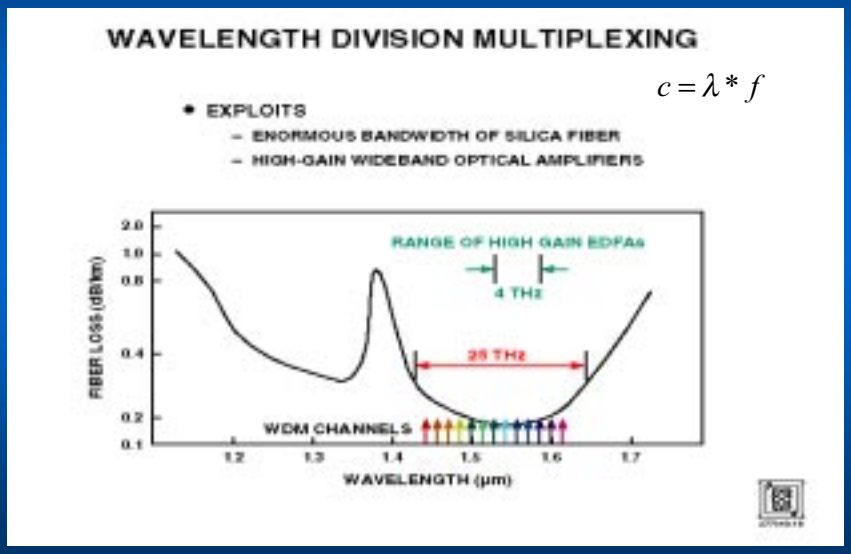
9

WDM ARCHITECTURE



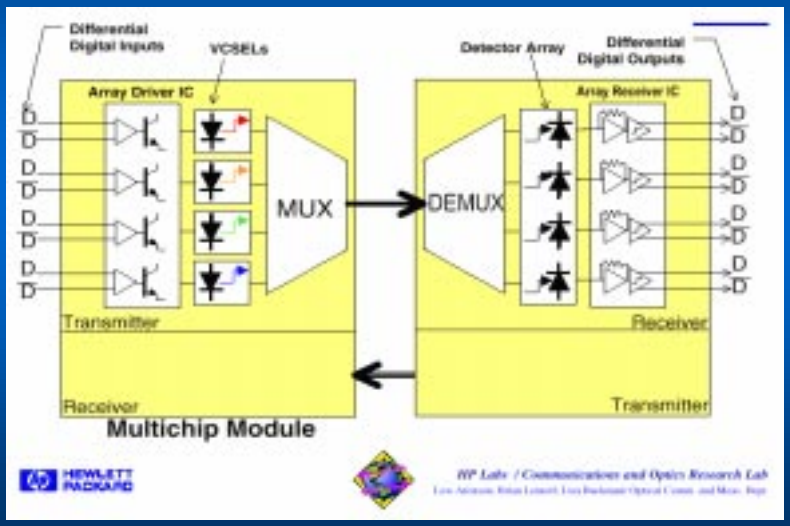
10

TECHNOLOGY



11

WDM IMPLEMENTATION



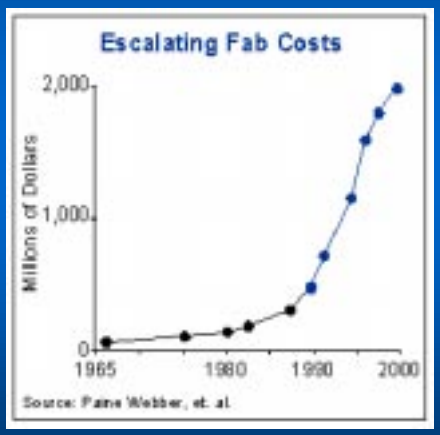
12

SEMI-CONDUCTORS

- Lets examine what is driving the ***I*** (investment) in our equation for success.
  - Information from 1997 SIA report (www.semichips.org)

13

THE COST OF "FABS"



- 2 billion and climbing
- One per continent?
- Put on the moon?
- Only million piece design can be made?

14

UNDERLYING REASONS

- 300 mm (12inch) wafers
- Billions to replace 8 inch fabs.
- Good news: keeps costs of chips down

15

UNDERLYING REASONS

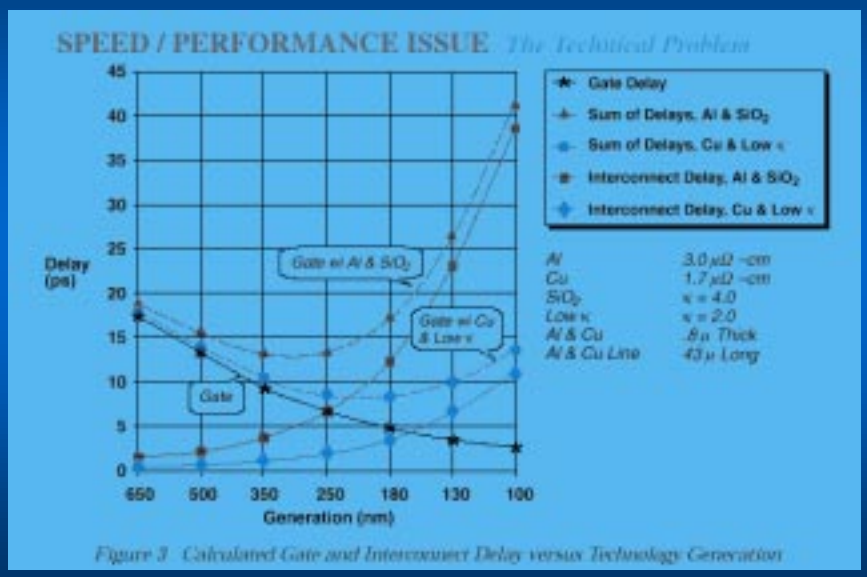


Figure 3 Calculated Gate and Interconnect Delay versus Technology Generation

16

HOW WE GET THERE

Table 63 Modeling & Simulation Technology Requirements (Continued)

Year of First Product Shipment Technology Generation	*1997 250 nm*	*1999 180 nm*	*2001 150 nm*	*2003 130 nm*	*2006 100 nm*	*2009 70 nm*	*2012 50 nm*
Numerical Methods							
Linear solvers—equations/minute	100k	150k	250k	250k	2.5M	5M	5M
Parallel speedup	—	4	6	9	16 ×	30 ×	50 ×
Grid reliability (pph)	300	180	120	90	26	14	7
MFLOPS* required	50	80	400	1000	4000	8000	8000
MC noise	NA	NA	NA	0.05	0.02	0.01	0.001
Simulation Environments							
Time needed for statistical sim.	10 weeks	6 weeks	4 weeks	2 weeks	2 weeks	1 week	1 week
Time needed for multi-tool initial problem setup	4 weeks	2 weeks	1 week	4 days	2 days	2 days	2 days
Correct data analyses per improvement cycle	0.1	1	1	1	2	4	10

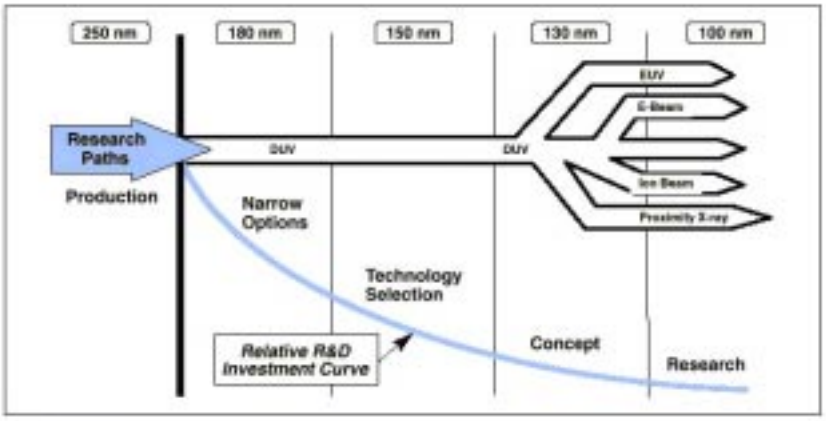
Solutions Exist *Solutions Being Pursued* *No Known Solution*

* *MFLOPS—million floating point operations per second*

† *Number of linear equations generated by discretizing an increasing number of PDEs over a typical device grid of 5000 nodes in 2-D and later 50000 nodes in 3-D*

17

HOW WE GET THERE



DUV—deep ultraviolet
EUV—extreme ultraviolet
E-beam—electron beam

Figure 1 Conceptual Illustration of Today's Research and Development Investments for Future Production Technologies

18

WHAT WE GET

Table 3 Performance of Packaged Chips

Year of First Product Shipment	*1997*	*1999*	*2001*	*2003*	*2006*	*2009*
Technology Generations Dense Lines (DRAM Half Pitch) (nm)	*250*	*180*	*150*	*130*	*100*	*70*
Isolated Lines (MPU Gates) (nm)	*200*	*140*	*120*	*100*	*70*	*50*
Number of Chip I/Os						
Chip-to-package (pads) high-performance	1450	2000	2400	3000	4000	5400
Chip-to-package (pads) cost-performance	800	975	1195	1460	1970	2655
Number of Package Pins/Balls						
ASIC (high-performance)	1100	1500	1800	2200	3000	4100
MPU/controller, cost-performance	600	810	900	1100	1500	2000
Cost-performance package cost (cents/pin)	1.40–2.80	1.25–2.50	1.15–2.30	1.05–2.05	0.90–1.75	0.75–1.50
Chip Frequency (MHz)						
On-chip local clock, high-performance	750	1250	1500	2100	3500	6000
On-chip, across-chip clock, high-performance	750	1200	1400	1600	2000	2500
On-chip, across-chip clock, cost-performance	400	600	700	800	1100	1400
On-chip, across-chip clock, high-performance ASIC	300	500	600	700	900	1200
Chip-to-board (off-chip) speed, high-performance (Reduced-width, multiplexed bus)	750	1200	1400	1600	2000	2500

WHAT WE GET

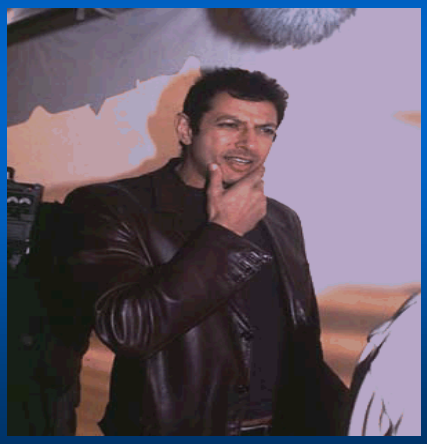
Table 24 Product Critical Level Lithography Requirements

Year of First Product Shipment Technology Generation	*1997 250 nm*	*1999 180 nm*	*2001 150 nm*	*2003 130 nm*	*2006 100 nm*	*2009 70 nm*	*2012 50 nm*
Product Application							
DRAM (bits)	256M	1G	—	4G	16G	64G	256G
MPU (logic transistors/cm^2)	4M	6M	10M	18M	39M	84M	180M
ASIC (usable transistors/cm^2)*	8M	14M	16M	24M	40M	64M	100M
*Minimum Feature Size (nm)***							
Isolated lines (MPU Gates)	200	140	120	100	70	50	35
Dense lines (DRAM Half Pitch)	250	180	150	130	100	70	50
Contacts	280	200	170	140	110	80	60
Development capability (minimum feature size, nm)	140	120	100	70	50	35	25
Gate CD control (nm, 3 sigma at post-etch)**	20	14	12	10	7	5	4
Product overlay (nm, mean + 3 sigma)**	85	65	55	45	35	25	20

20

HOW WE USE IT

- TRENDS IN COMPUTER ARCHITECTURE



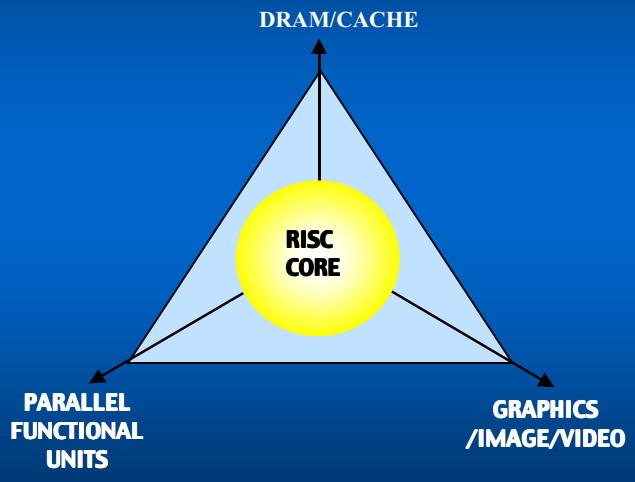
GENERAL VIEWS

2 TO 4 YEARS

10 YEARS (USING SIA STUDY)

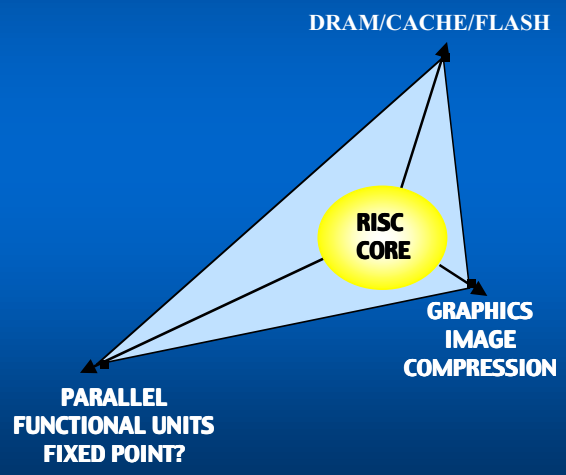
21

PC/PERSONAL WORKSTATION



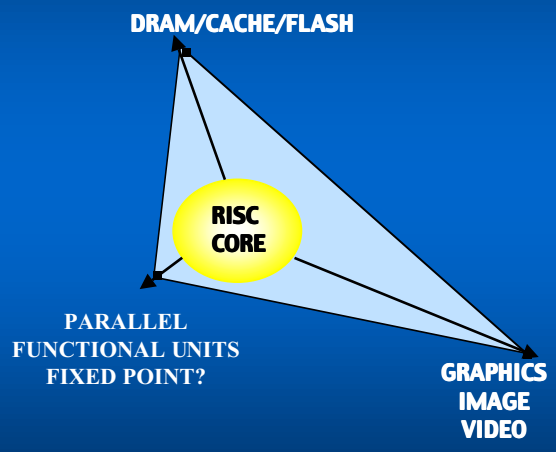
22

DIGITAL SIGNAL PROCESSOR



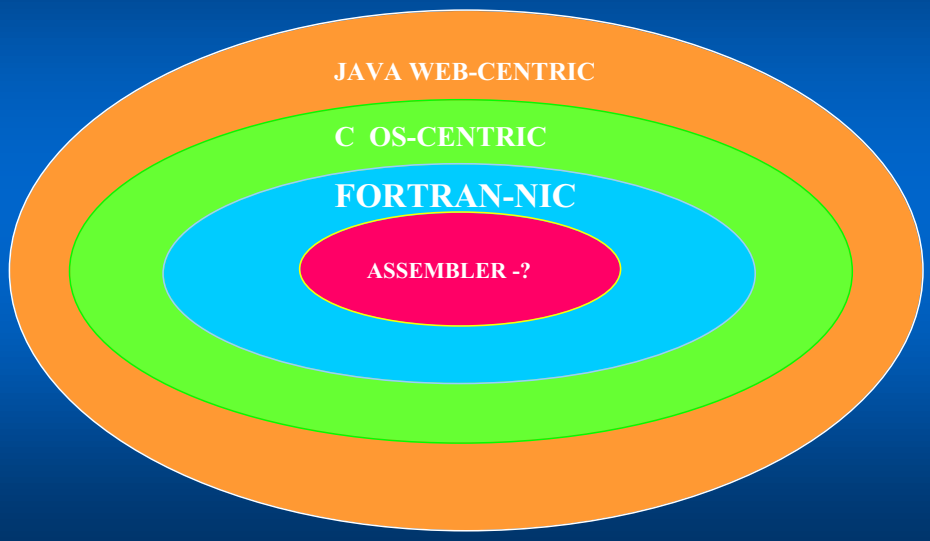
23

GRAPHICS/IMAGE PROCESSOR



24

APPLICATION STRUCTURE ?



25

JAVA Vs C++

FEATURE	JAVA	C++
Memory Management	Garbage collected	Explicit Memory Freeing
Multi-threading	YES (Mesa-style)	NO
Inheritance Model	Simpler (separate sub-typing)	Complex
Exception handling	Supported	Sporadic
Parametric type	Does Not	Has template
Type casts	Checked Thus easier to write protected subsystems	Unchecked (pointer ⟷ integer)

26

SO WHAT HAPPENS?

- Fundamentally the following architecture evolves:
  - ***PIM (processor in memory) or System-on-a-chip***
    - **more memory bandwidth**
    - **lower latency**
    - **consistent with PC pricing and technology curves**

27

WHICH APPROACH?

- SIMD
- MIMD
- MULTI-THREADED
- SUPERSCALAR
- VLIW

28

INTERCONNECT TYPE

- ***SOFTWARE***
  - Message Passing
  - Distributed Shared Memory (DSM)
  - Cache Only (COMA)
  - Object oriented
  - Emulated DSM (e.g.. Threadmarks)

29

INTERCONNECT TYPE

- ***HARDWARE***
  - Hierarchical - number of levels is a function of the number of cpu's.
  - Physical - combination of copper and photonic. Ultimately ***WDM*** will play an important role in external chip interconnects.

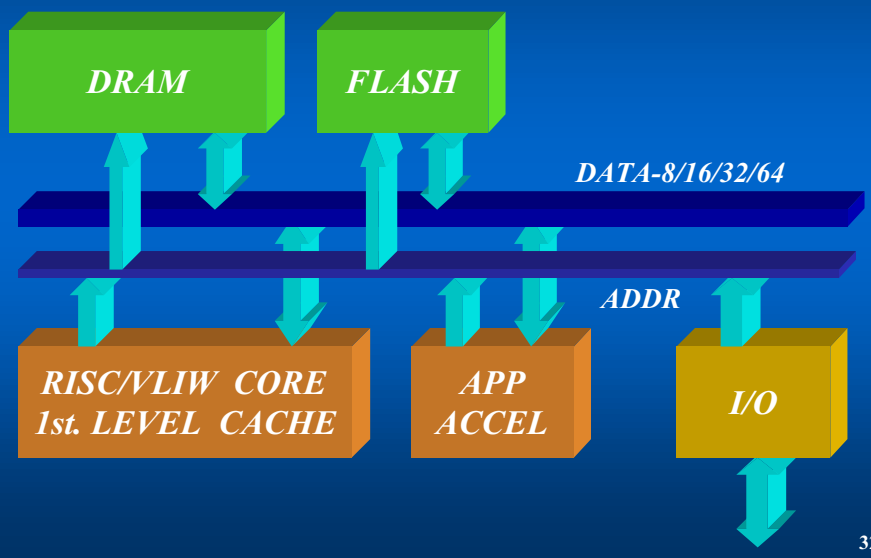
30

NEXT ARCHITECTURES

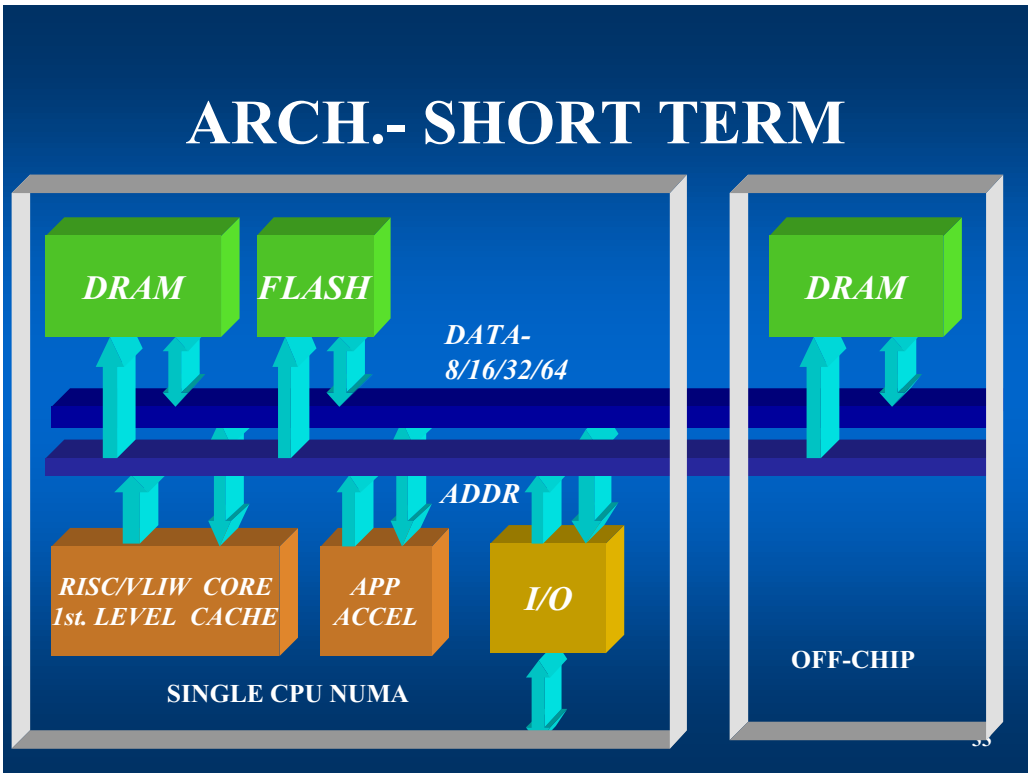
- Short Term - 2 To 4 years- low performance - System-on-a-chip (SOAC)
- Long Term - 10 years (using SIA study)
  - High Performance
  - Supercomputing
- US Gov't R&D Policy

31

ARCH.- SHORT TERM



32



ARCH. - LONG TERM- 2009

- **THE SIA STUDY TEACHES US:**
  - 64 gbits of dram - (8 gbytes)
  - 8 gbits of sram
  - 520 million MPU transistors
  - 70 nm lithography, 2.54 cm on-a-side
  - 6 ghz clock within vliw/risc core
  - 2.5 ghz across die
  - 2500 external signal pins

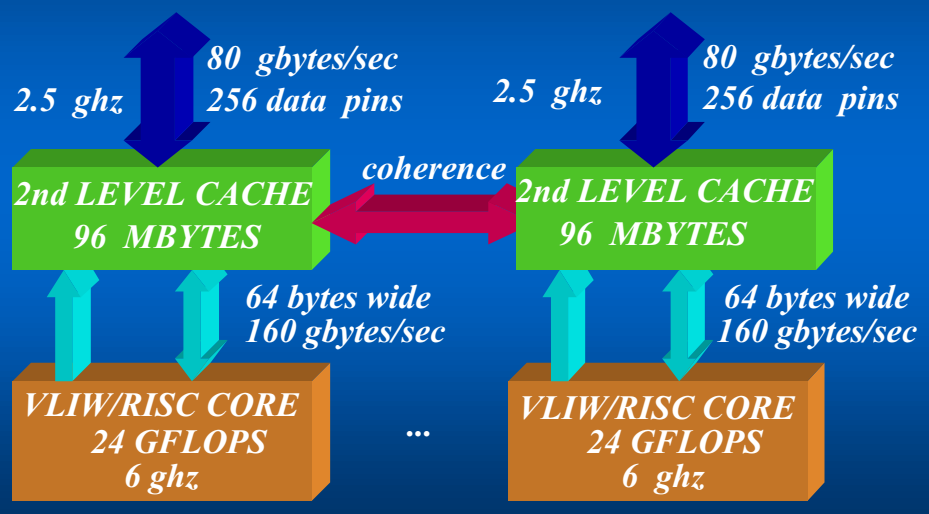
34

ARCH. - LONG TERM - 2009 DESIGN ASSUMPTIONS

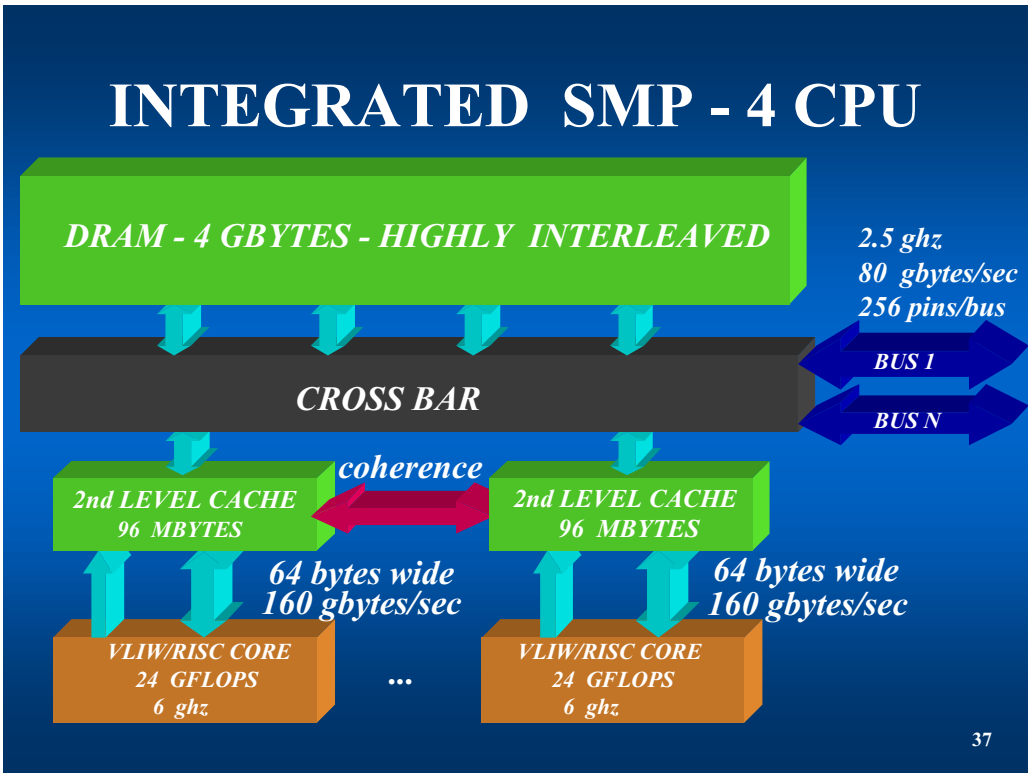
- 9 million transistors - vliw/risc core with first level cache.
- 2nd. Level cache - rule of thumb. 1/4 to 1/2 mbyte per 100 mflops peak.
- 96 mbyte 2nd. Level (6 Inst, 90 data)
- 170 watts
- .6 to .9 volts power supply

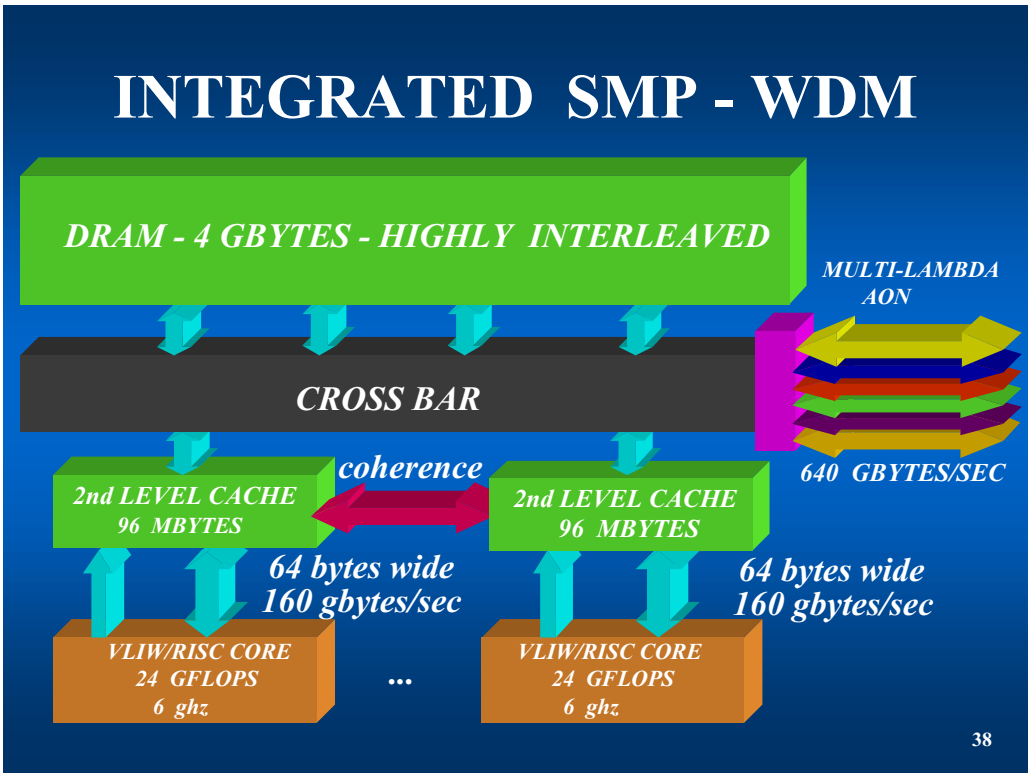
35

MAXIMUM PIN-USE EXTERNAL SMP- 6/8 CPU'S



36


INTEGRATED SMP - 4 CPU
DRAM - 4 GBYTES - HIGHLY INTERLEAVED
2.5 ghz
80 gbytes/sec
256 pins/bus
BUS 1
BUS N
CROSS BAR
coherence
2nd LEVEL CACHE
96 MBYTES
2nd LEVEL CACHE
96 MBYTES
64 bytes wide
160 gbytes/sec
64 bytes wide
160 gbytes/sec
VLIW/RISC CORE
24 GFLOPS
6 ghz
...
VLIW/RISC CORE
24 GFLOPS
6 ghz
37

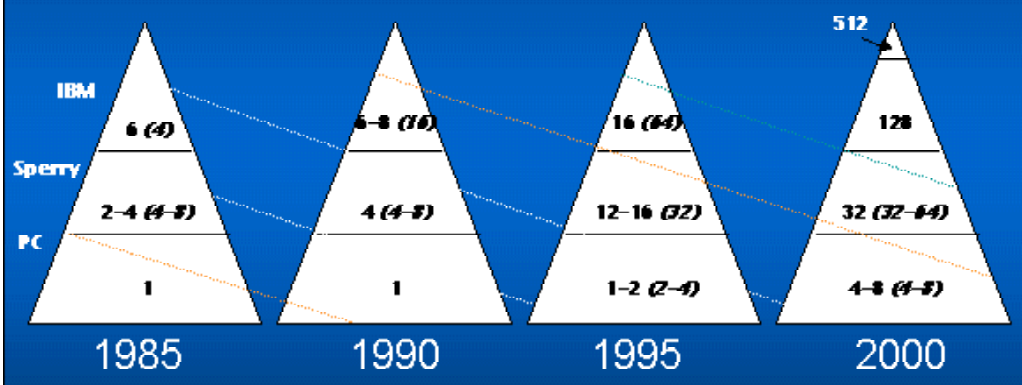

INTEGRATED SMP - WDM
DRAM - 4 GBYTES - HIGHLY INTERLEAVED
MULTI-LAMBDA
AON
CROSS BAR
640 GBYTES/SEC
coherence
2nd LEVEL CACHE
96 MBYTES
2nd LEVEL CACHE
96 MBYTES
64 bytes wide
160 gbytes/sec
64 bytes wide
160 gbytes/sec
VLIW/RISC CORE
24 GFLOPS
6 ghz
...
VLIW/RISC CORE
24 GFLOPS
6 ghz
38

US GOVERNMENT POLICY

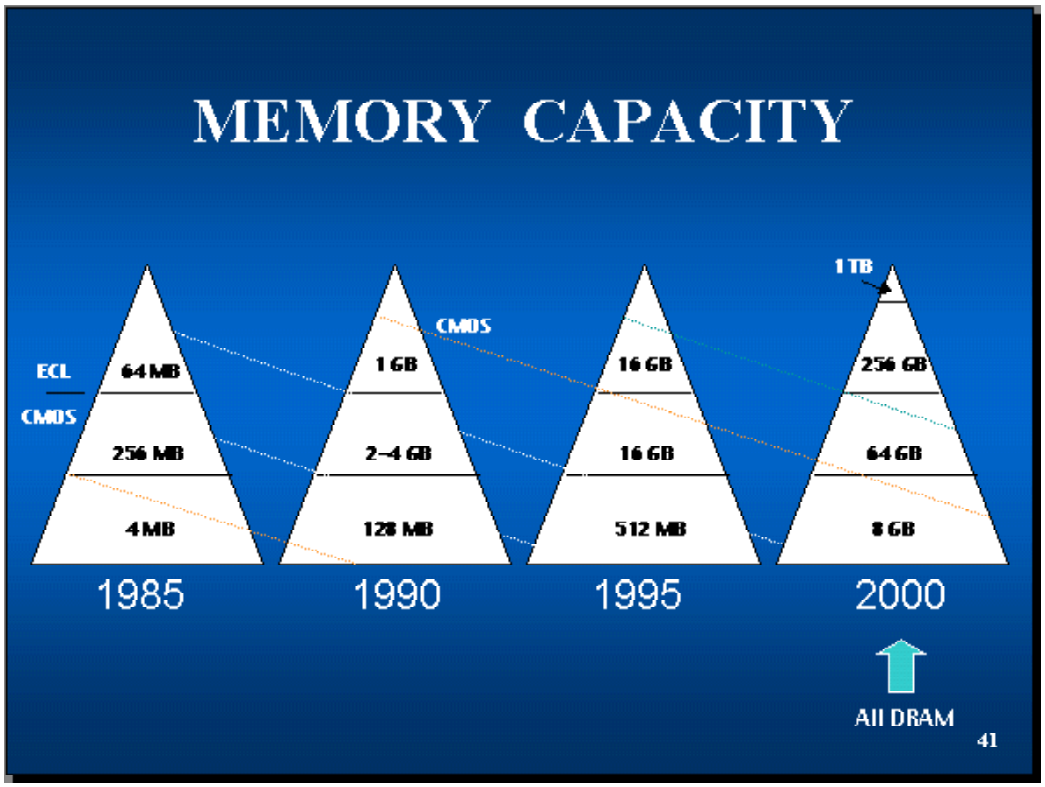
- Examine the Past
- Use Tops 500 - LINPACK
- Observe Venture Capital Investments
- What should happen in the future

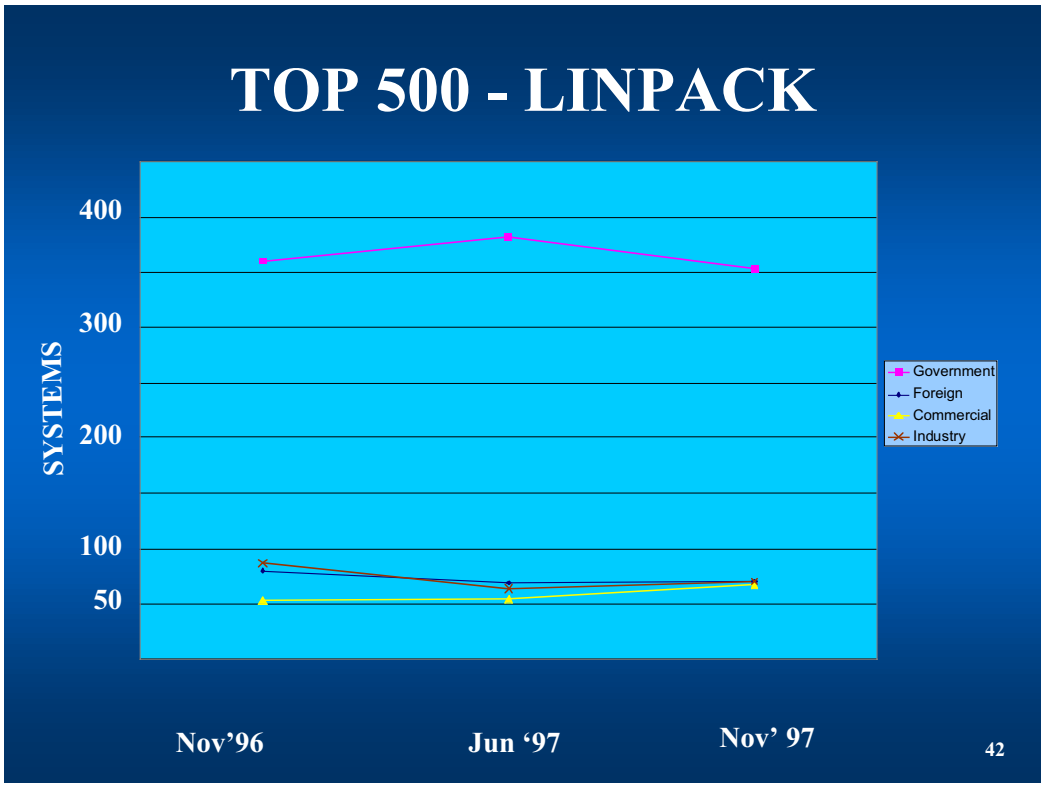
39

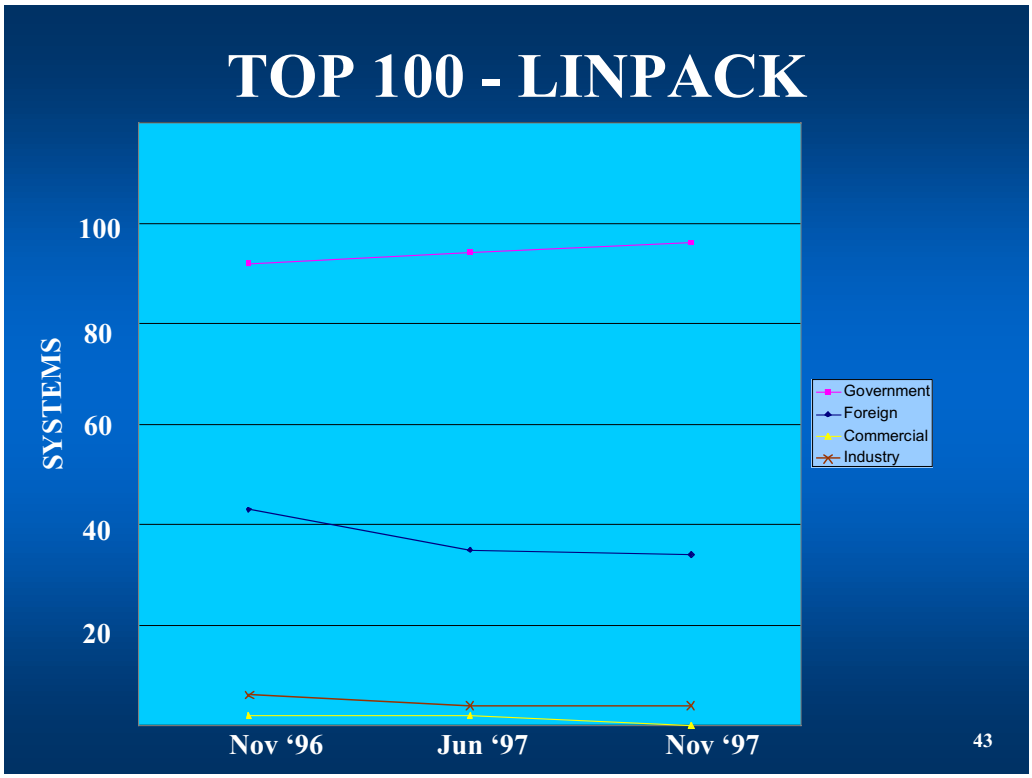
PROCESSOR SCALABILITY GENERAL PURPOSE-CPU

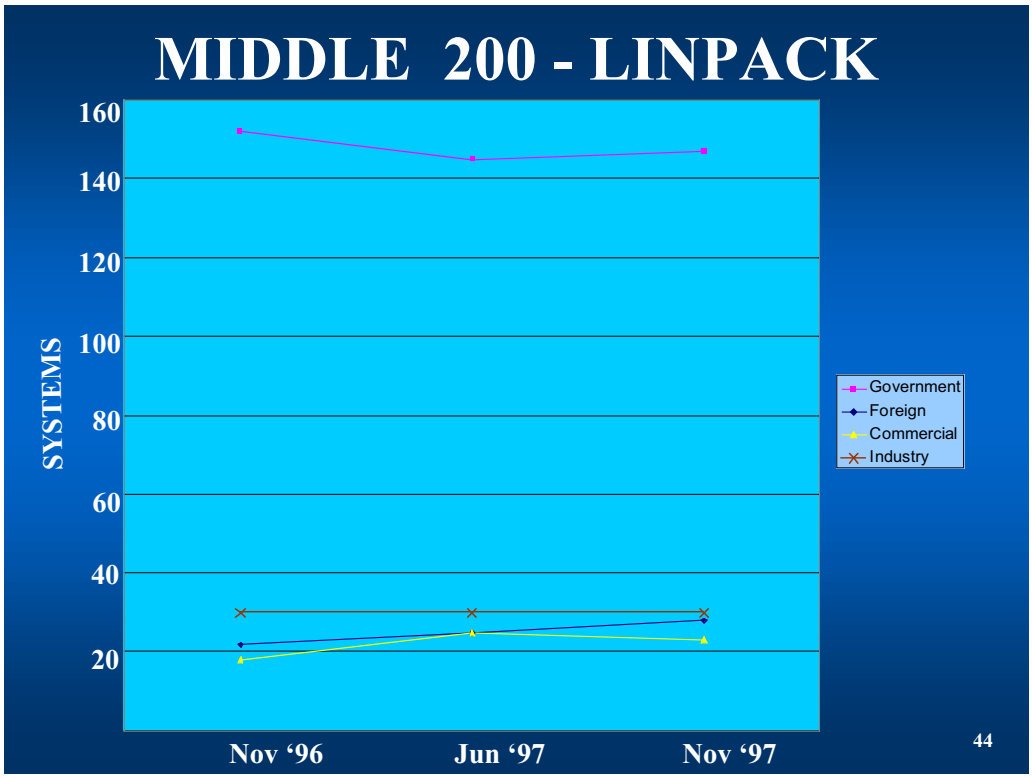


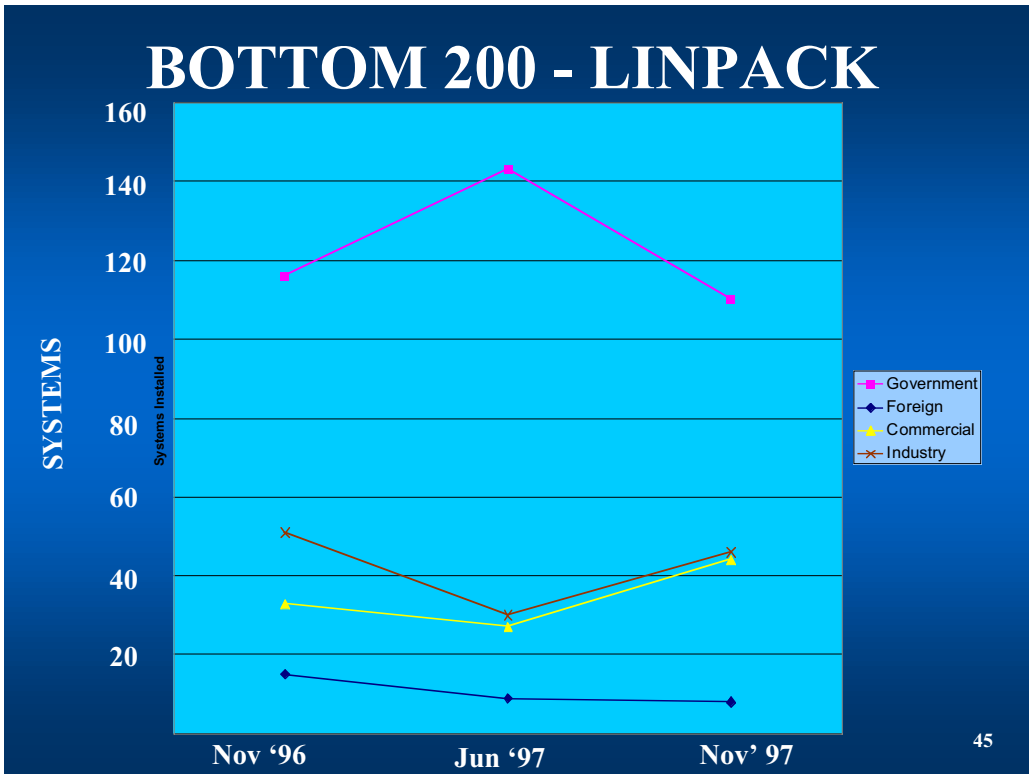
40

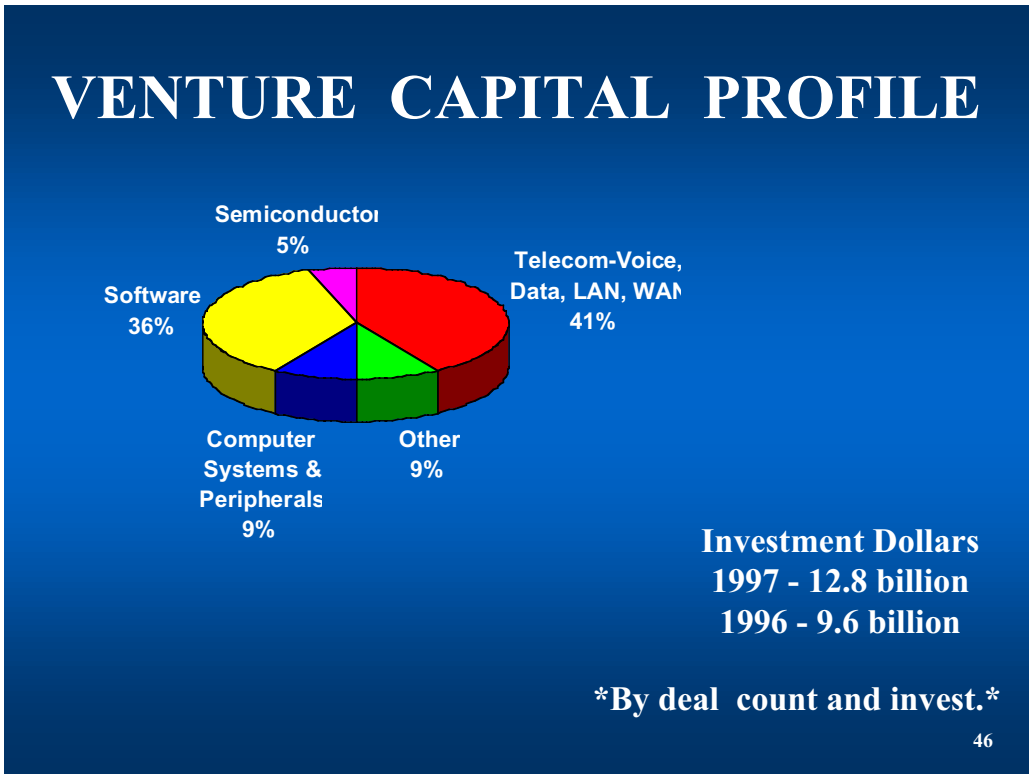

MEMORY CAPACITY
ECL
CMOS
64 MB
256 MB
4 MB
CMOS
1 GB
2-4 GB
128 MB
16 GB
16 GB
512 MB
1 TB
256 GB
64 GB
8 GB
1985
1990
1995
2000
All DRAM
41


TOP 500 - LINPACK
SYSTEMS
400
300
200
100
50
Government
Foreign
Commercial
Industry
Nov'96
Jun '97
Nov' 97
42


TOP 100 - LINPACK
SYSTEMS
100
80
60
40
20
Nov '96
Jun '97
Nov '97
Government
Foreign
Commercial
Industry
43


MIDDLE 200 - LINPACK
SYSTEMS
160
140
120
100
80
60
40
20
Nov '96
Jun '97
Nov '97
Government
Foreign
Commercial
Industry
44


BOTTOM 200 - LINPACK
SYSTEMS
Systems Installed
160
140
120
100
80
60
40
20
Nov '96
Jun '97
Nov' 97
Government
Foreign
Commercial
Industry
45


VENTURE CAPITAL PROFILE
Semiconductor
5%
Telecom-Voice,
Data, LAN, WAN
41%
Software
36%
Computer
Systems &
Peripherals
9%
Other
9%
Investment Dollars
1997 - 12.8 billion
1996 - 9.6 billion
By deal count and invest.
46

US GOVERNMENT POLICY

- Provide seed money - high risk/reward (darpa, nsf, dod, doe)
- Further national defense initiatives
- Begin the trickle down, technology xfer. What starts out as a US Gov't special becomes COTS after 1 or 2 generations
- Keep the US the most advanced and competitive in the world
- www.hpcc.gov/talks/petaflops-24june97

47

CONCLUDING

- **Convergence of telecommunications/computing**
  - **everything is digital**
  - **everything requires high bandwidth**
  - **voice is a digital packet (IP switching)**
  - **digital TV (a TV with a computer or a computer with a TV?)**
  - **overall system topology mirrors an AON**
- **Commodity Teraflop Computing**

Stockpile Stewardship Program (U)

1998 Conference on High Speed Computing
Gleneden Beach, Oregon
April 20-23, 1998

Lawrence J. Ferderber
Deputy Associate Director for National Security
Lawrence Livermore National Laboratory

Lawrence Livermore National Laboratory, P.O. Box 808, Livermore, CA 94551

NS-98-031.1

The President tasked DOE to help maintain the nuclear deterrent through the Stockpile Stewardship Program

"... I consider the maintenance of a safe and reliable nuclear stockpile to be a supreme national interest of the United States."

"I am assured by the Secretary of Energy and the Directors or our nuclear weapons labs that we can meet the challenge of maintaining our nuclear deterrent under a Comprehensive Test Ban Treaty through a Science-Based Stockpile Stewardship program without nuclear testing..."

"In order for this program to succeed, both the Administration and the Congress must provide sustained bipartisan support for the stockpile stewardship program over the next decade and beyond. I am committed to working with the Congress to ensure this support."

"As part of this arrangement, I am today directing the establishment of a new annual reporting and certification requirement that will ensure that our nuclear weapons remain safe and reliable under a comprehensive test ban."

– August 11, 1995

NS-98-031. 2

Today the stockpile is safe and reliable, but we already require a Stockpile Stewardship Program to keep it that way

- Today's stockpile has a good "pedigree" based on
  - Nuclear tests
  - An experienced workforce
  - State of the art design (then)
- But
  - The stockpile is aging beyond our experience
  - Refurbished components will be made by new processes, in new plants by new people
  - Our experienced workforce is retiring
  - We have no nuclear tests to verify the validity of our decisions
- We need a program that will:
  - Attract and train a new workforce
  - Be able to assess the effect of changes in the stockpile
  - Certify that refurbished components are functionally equivalent to the original ones

NS-98-031. 3

Like every other technological object, a nuclear weapon ages and sometimes we are surprised when we test it

- One-point safety
- Performance at cold temperatures
- Performance under aged conditions
- Marginal performance
- Degradation of various key materials
- Pit quality control
- Metal components cracking
- Yield-select problems
- HE degrading
- HE cracking
- Detonators corroding
- Detonator system redesign
- Metal components corroding

NS-98-031. 4

The Stockpile Stewardship Program responds to these challenges via a few fundamental principals

- We have experimental data from nuclear tests which indicate that details matter – remanufactured components sometimes behave anonymously
- Current experimental and computational capabilities are not sufficient to preclude that these anomalies will occur in the future
- Without nuclear testing, we must take the conservative approach in proving our fixes are real fixes which do not introduce new problems
- We must also develop a strategy to deal with the "unknown anomalies" (e.g. Challenger O-ring) ... including residual design flaws that have not yet manifested themselves
- The stockpile will continue to age and we will be required to deal with changes to almost every components

The SSP approach is not without risk

NS-98-031. 5

Simple remanufacture is not a credible solution for highly optimized and complex products like nuclear weapons

- This point is illustrated by the Polaris A3 motor rebuild
  - The U.S. production line was placed on standby in 1963
  - Procedures were carefully documented
  - Nineteen years later, in 1982, it was found that the "replica" rebuild of the rocket motors required extensive full scale testing to get it right (four of the flight tests failed)
  - A recall of retired personnel was necessary

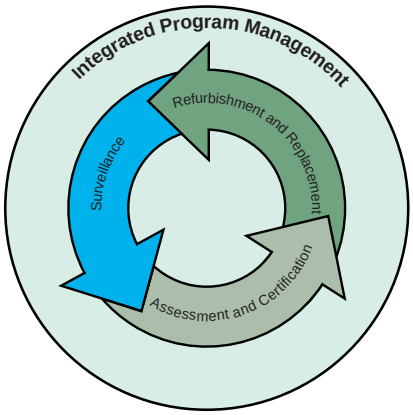
Replicating nuclear weapons would be more difficult (impossible) than replicating rocket motors

NS-98-031. 6

The SSP provides integrated capabilities to address DoD's near-term and longer-term issues

Four SSP Strategies:

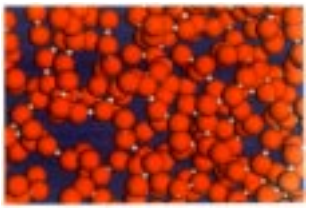
- Surveillance
  - to monitor, maintain and predict the condition of the stockpile
- Assessment & Certification
  - of the consequences of change
  - that modifications and maintenance do not degrade warhead safety and reliability
- Refurbishment
  - design and manufacture of refurbished components
- Tritium replacement



NS-98-031. 7

Our surveillance program is being expanded to meet the needs of an aging stockpile

- How do weapons age?
- What are the most likely issues?
- How will these issues affect performance and safety?
- When do components need to be refurbished?



Interstitial helium

Assessment of disassembled components

Forensic surveillance techniques

NS-98-031. 8

The new complex must refurbish/replace components to counter age, performance, or safety degradation

Plutonium pits
Los Alamos, New Mexico

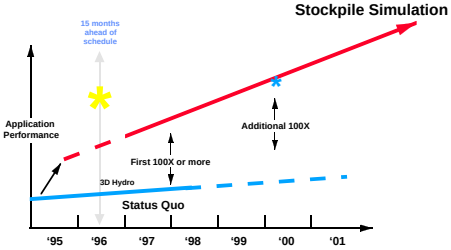
Assembly expertise

Integrated plutonium processing

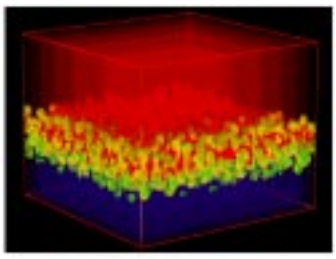
These new plants, people and processes must be certified to be functionally equivalent those originally used

NS-98-031. 9

SSP requires dramatic advances in computational capabilities



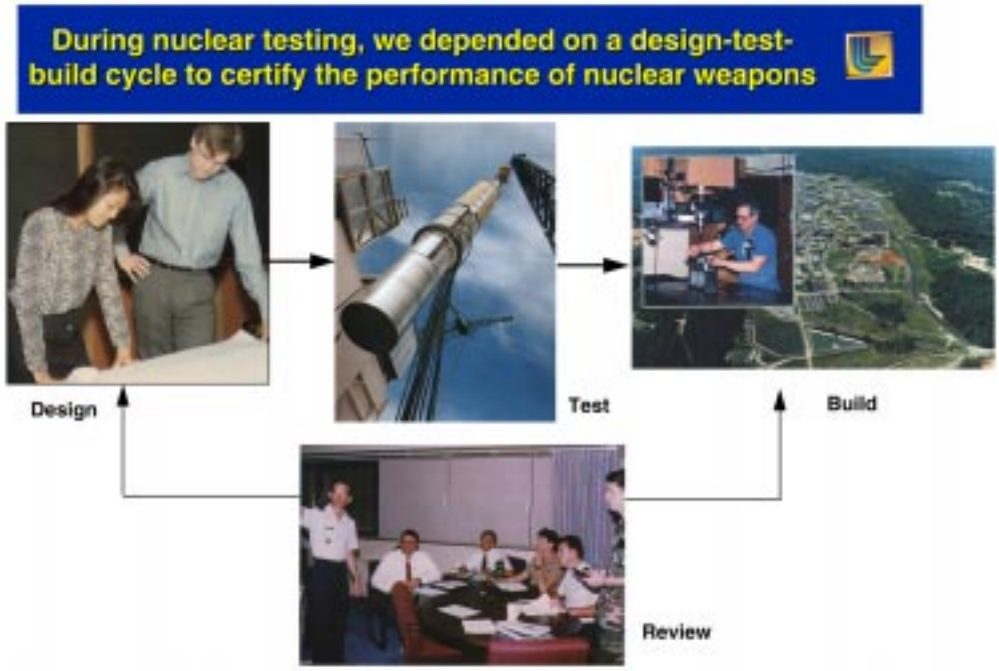
Accelerated Strategic Computing Initiative (ASCI)

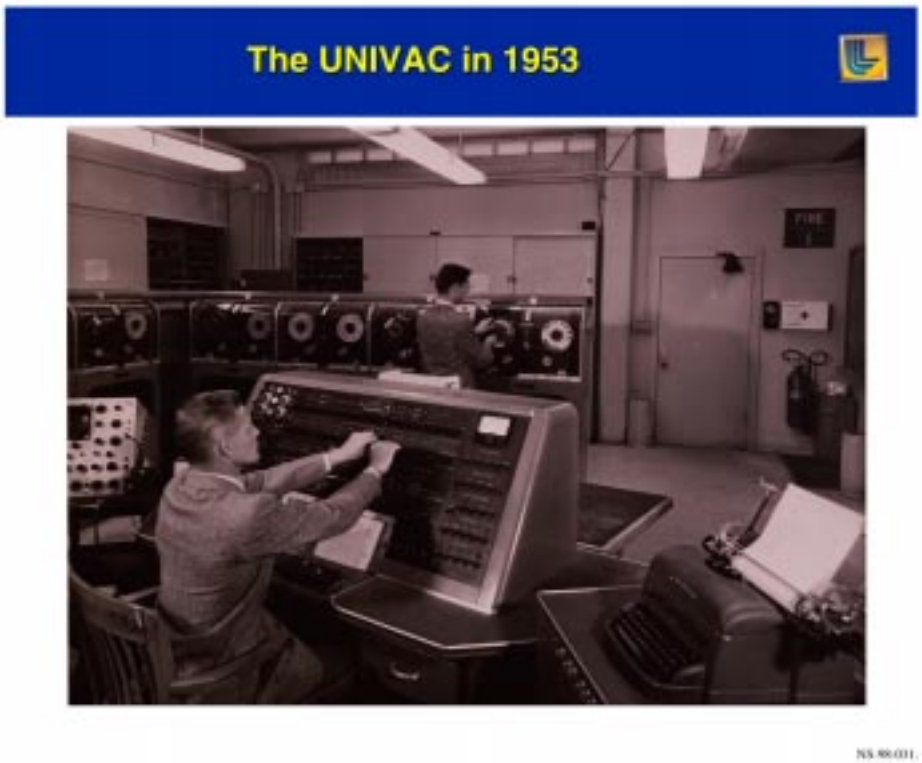


3D turbulent mix simulation

ASCI Blue Pacific SST, LLNL
3.3 TeraFlops
2.5 Terabytes

NS-98-031. 11


During nuclear testing, we depended on a design-test-build cycle to certify the performance of nuclear weapons
Design
Test
Build
Review
The computational capability was sufficient to provide reasonable assurance that the test would function properly (a cost issue)


The UNIVAC in 1953

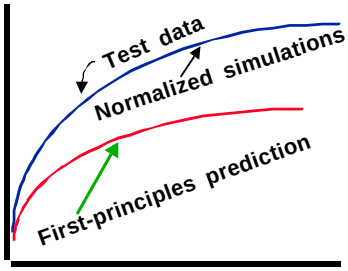
LLNL's first "nuclear" test was designed on the UNIVAC and slide rules

- LLNL was less than one year old
- The device was placed on a 300-foot tower and the physicists stood far away, observing with dark glasses
- Upon detonation, only a small cloud of dust appeared
- When the dust cleared, the tower was still standing

Nowhere to go but up ... and 50 years later, the ASCI program

NS-98-031. 1

During design-test-build, our simulation codes normalized complex phenomena against test data



- Computers lacked speed and memory to run full problems
- Some nonlinear physical processes not understood
- Nuclear test data provided normalizing factors to make simulations accurate
- Normalization factors differed from system to system

NS-98-031. 2

At one time, the bulk of LLNL's work was done on four CDC 7600s, and everyone wanted their own 7600

CDC 7600

Apple "7600"

Memory: 16 MB (expandable to 512 MB)
Cycle time: 8.3 ns (upgradeable to 5 ns)
Performance: Livermore Fortran Kernels (LFK)

Memory: 4 MB
Cycle time: 27.5 ns
Performance: LFK
Geometric Mean - 4.0 MFlops

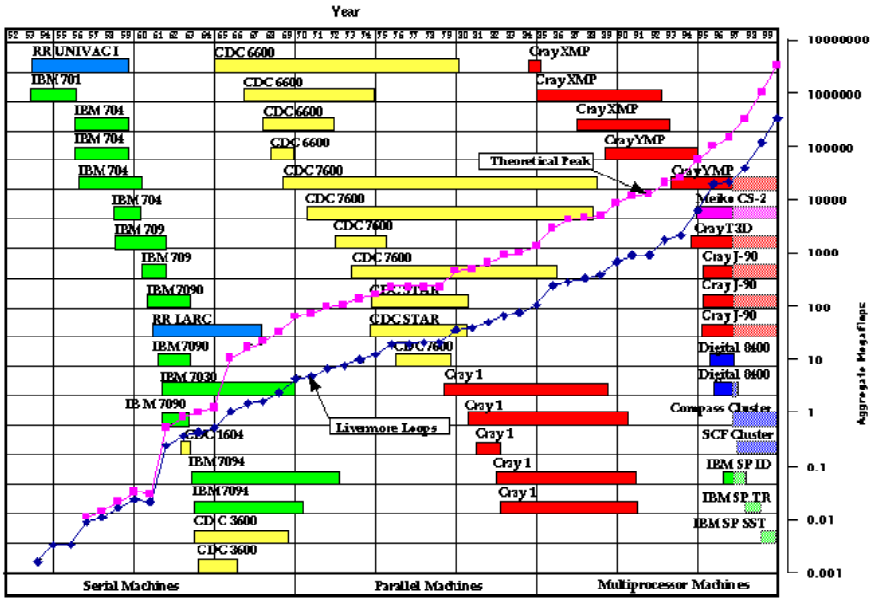
Cost:
$4,000,000 (in 1976 dollars)
$10,000,000 (in 1996 dollars)

			Mega-Flops/Sec
Maximum	Rate	=	122.28
Quartile	Q3	=	34.34
Average	Rate	=	27.97
GEOMETRIC	MEAN	=	21.12
Median	Q2	=	18.97
Harmonic	Mean	=	16.54
Quartile	Q1	=	12.84
Minimum	Rate	=	4.81

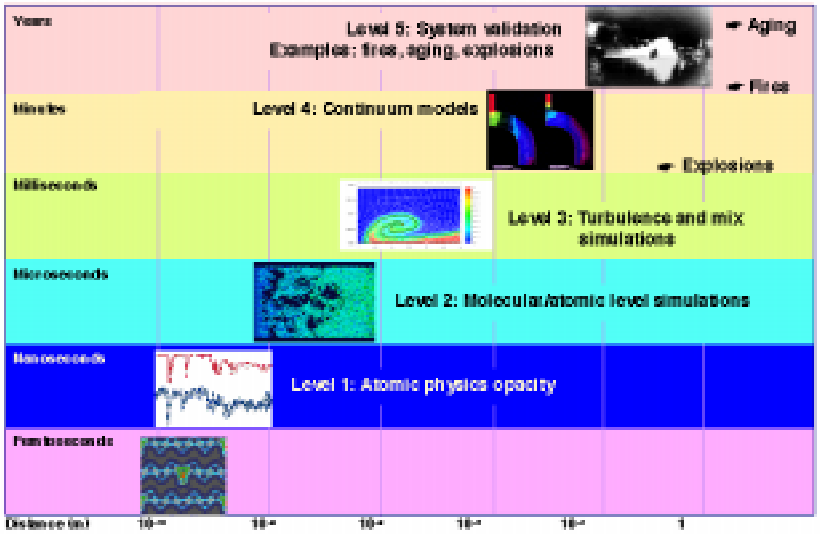
Cost:
$3,000 (in 1996 dollars)

NS-00-031. 15

LLNL Computer History



The computational needs of the SSP span many time and length scales

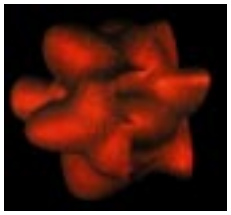


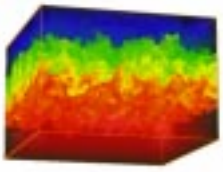
We need a hierarchy of models and modeling methods to enable predictive capability for all processes relevant to weapon performance

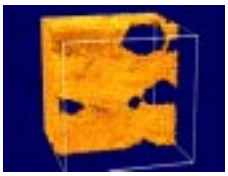
NS-031

The ASCI applications strategy emphasizes both integrated simulation codes and the sub-grid physics they rely on

Full-Scale Integrated Codes	Subgrid Models / Zonal Physics	
	Turbulence	Materials Models





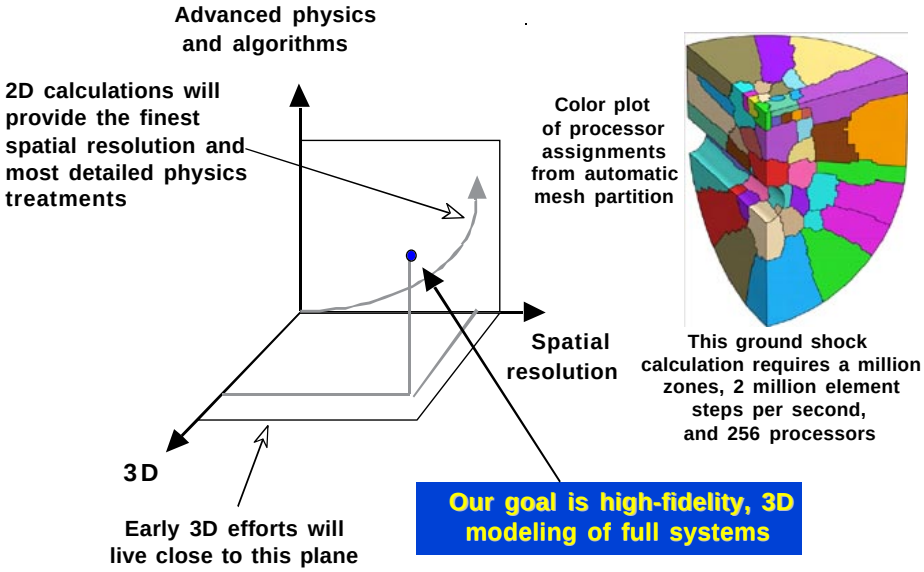


- 3D simulations
- High resolution
- Improved algorithms
  - Accuracy
  - Efficiency
  - Scalability
- Applied mathematics
- Mesh generation
- Visualization
- Validation

- Direct numerical simulations
- First principle approaches
- Predictive physics models
- Rigorous treatments of physical phenomena

NS-98-031. 1

Our full scale, integrated codes must support tradeoffs between dimensionality, resolution, and detailed physics

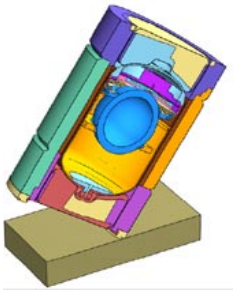


Color plot of processor assignments from automatic mesh partition

This ground shock calculation requires a million zones, 2 million element steps per second, and 256 processors

NS-98-031. 2

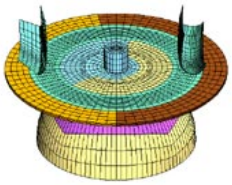
SSP structural analysis codes will develop mesh and boundary partitioning for a wide variety of integrated simulations



AT-400 shipping container drop test

50,000 elements
27 contact boundary conditions

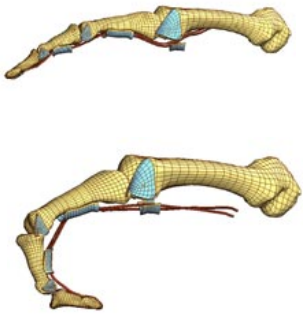
code: ParaDyn



Spin plate metal forming application

4,800 elements
4 contact boundaries

code: ParaDyn



Human index finger under flexion

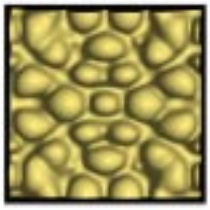
26,000 nodal points
21,000 8-node brick elements
15 sliding interfaces

code: Nike3D

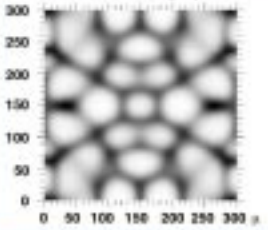
NS-98-031. 3

Direct 3D numerical simulation code HYDRA provides detailed comparisons between NIF predictions and experiments

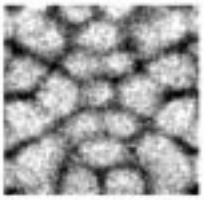
Isodensity contour of 0.5 g/cm^3 from HYDRA simulation



Simulated radiograph



Experimental radiograph



Nonlinear saturation of ablative Rayleigh-Taylor instability

Experimental measurements of 3D multimode surface perturbations on an ablatively driven foil are directly compared to simulations

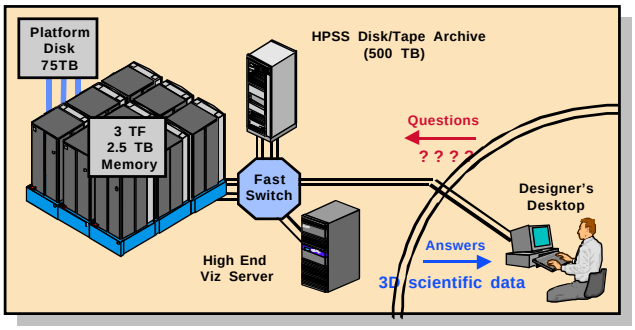
NS-98-031. 21

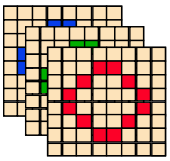
Analysis and visualization of terascale data sets places severe demands on all aspects of the computing environment

>> 1 billion zone physics problem

Physics Application

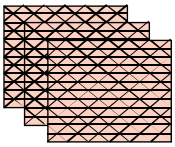
> > 100 GB visualization file every 3 minutes
(667 Mb/s)





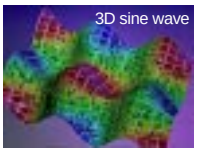
>> 3GB Iso-surface dataset extracted...
(20 MB/s)

Geometry Engine Software



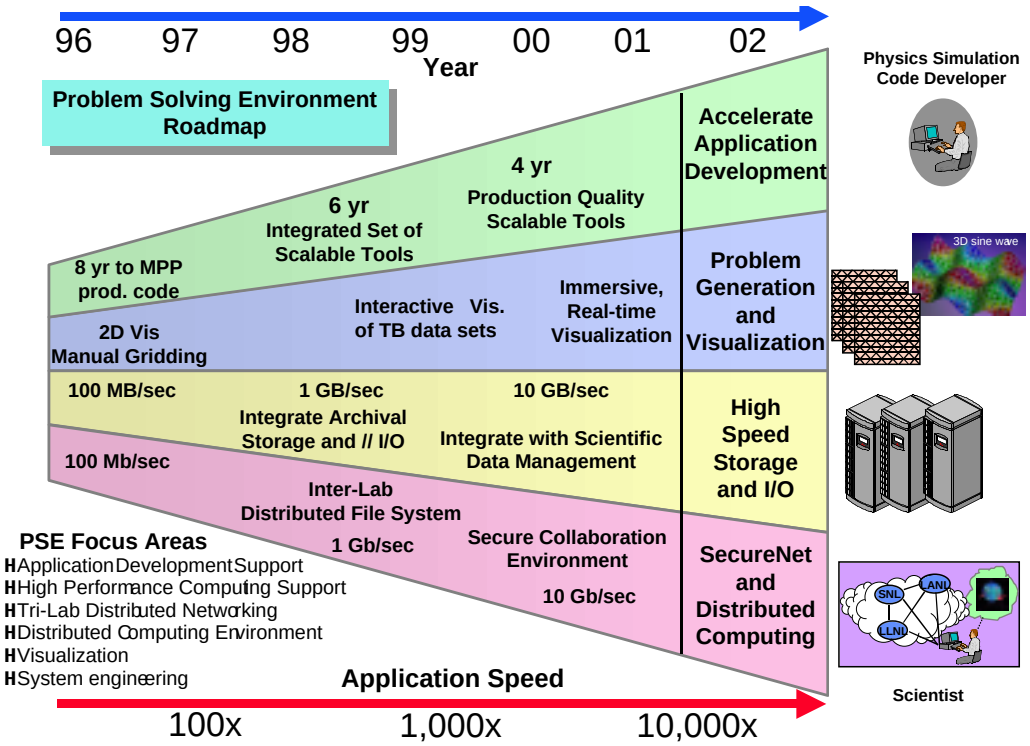
>> and 400 MB geometry data set generated
(2.7 MB/s)

Vis Server

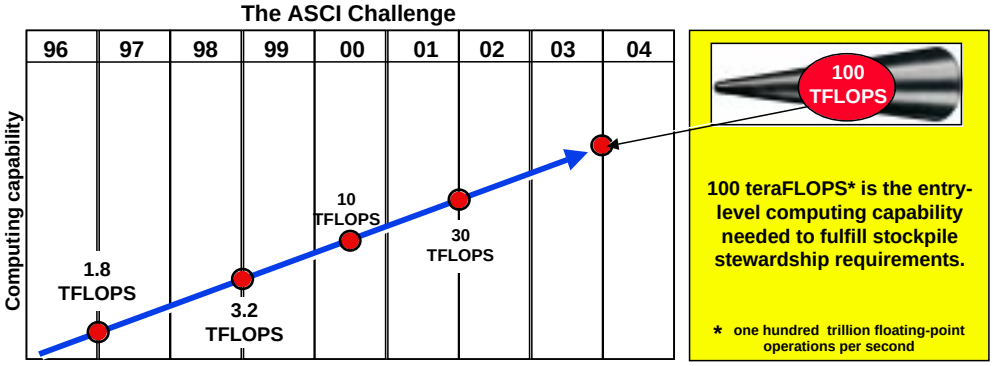


>> 3MB raster image to desktop, fly-around at 10 fps
(30 MB/s)

NS-98-031. 1



Meeting the ASCI challenge requires partnerships and collaborations among the three laboratories, industrial partners, and universities



Achieving the 100-teraFLOPS milestone will require carefully integrated efforts to develop unprecedented computer platforms, high-fidelity physics codes, and a world-class computing environment.

NS-98-031. 3

The ASCI machines are research partnerships with U.S. Industry

Intel
ASCI "Red" Computer
Sandia National Laboratories

SGI/Cray
ASCI "Blue Mountain" Computer
Los Alamos National Laboratory

ASCI "Blue Pacific" Computer
Lawrence Livermore National Laboratory
IBM

NS-98-031. 25

The SSP announced strategic academic alliances with five universities

- Stanford University
  - The Center for Integrated Turbulence Simulations
    - William C. Reynolds (wcr@thermo.stanford.edu)
- The University of Chicago
  - Astrophysical Thermonuclear Flashes
    - Robert Rosner (rrosner@oddjob.uchicago.edu)
- The University of Illinois an Champaigne, Urbana
  - Center for Simulation of Advanced Rockets
    - Michael T. Heath (m-heath@uiuc.edu)
- The University of Utah
  - Center for Simulation of Accidental Fires and Explosions
    - David W. Pershing (David.Pershing@dean.eng.utah.edu)
- The California Institute of Technology
  - Facility for Simulating the Dynamic Response of Materials
    - Daniel I. Meiron (dim@ama.caltech.edu)

NS-98-031. 1

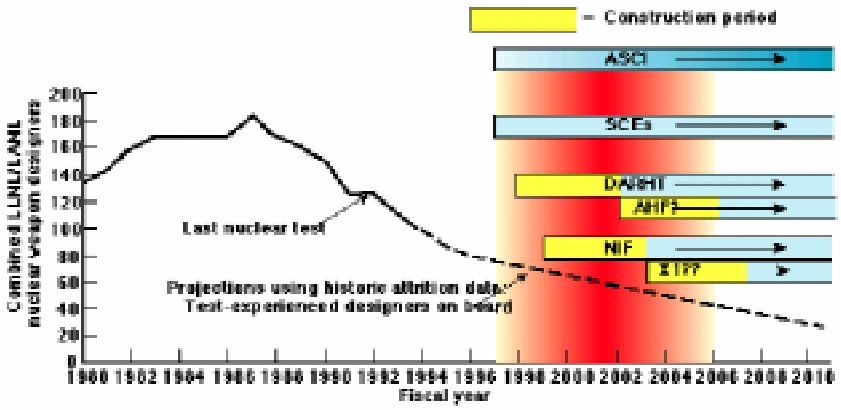
ASCI is an essential part of the rapidly evolving Stockpile Stewardship Program

- ASCI provides leading-edge, high-end simulation capabilities to meet weapon certification requirements
- ASCI integrates the resources of national laboratories, computer manufacturers, and academic institutions
  - national labs focus on application codes and related applied science
  - computer manufacturers develop technology and systems for 100 TeraFlops
  - Academic institutions research the basic science

The ASCI codes will need to be continually evaluated against experimental data in the relevant regimes

NS-98-031. 2

We estimate it will take ten years to fully implement the SSP investment



Time is critical because, in the transition, we will need to rely on the judgment of a diminishing number of nuclear-test trained weapon designers

NS-98-031. 3

Our future certification of the stockpile will rely on informed judgments

- Trained, knowledgeable people are required to assess and certify the stockpile
- A deeper understanding of the underlying science is required for practical weapon assessment capabilities
- New computational capabilities are needed to provide the integration formerly done with nuclear tests
- New experimental capabilities are required to provide detailed component level tests and validate the computation tools

Full implementation of SSP is required to sustain nuclear deterrence

NS-98-031. 29

There are many risks inherent in the SSP

- NASA did not accept the judgment of its engineers that the design was unacceptable and,
- As the problems grew in number and severity, NASA minimized them in management briefings and reports.
  - Reports of the Presidential Commission on the Challenger accident

"The contractor did not accept the implications of tests early in the program that the design had a serious and unanticipated flaw."

NS-98-031. 5

XD98-7044 4/20/98

Predictability and Stewardship

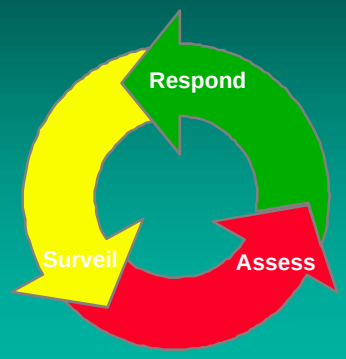
1998 Salishan Conference on High-Speed Computing
Gleneden Beach, Oregon
April 21, 1998

Raymond J. Juzaitis
Division Director, Applied Theoretical & Computational Physics (X)
Los Alamos National Laboratory

Applied Theoretical & Computational Physics Division Los Alamos

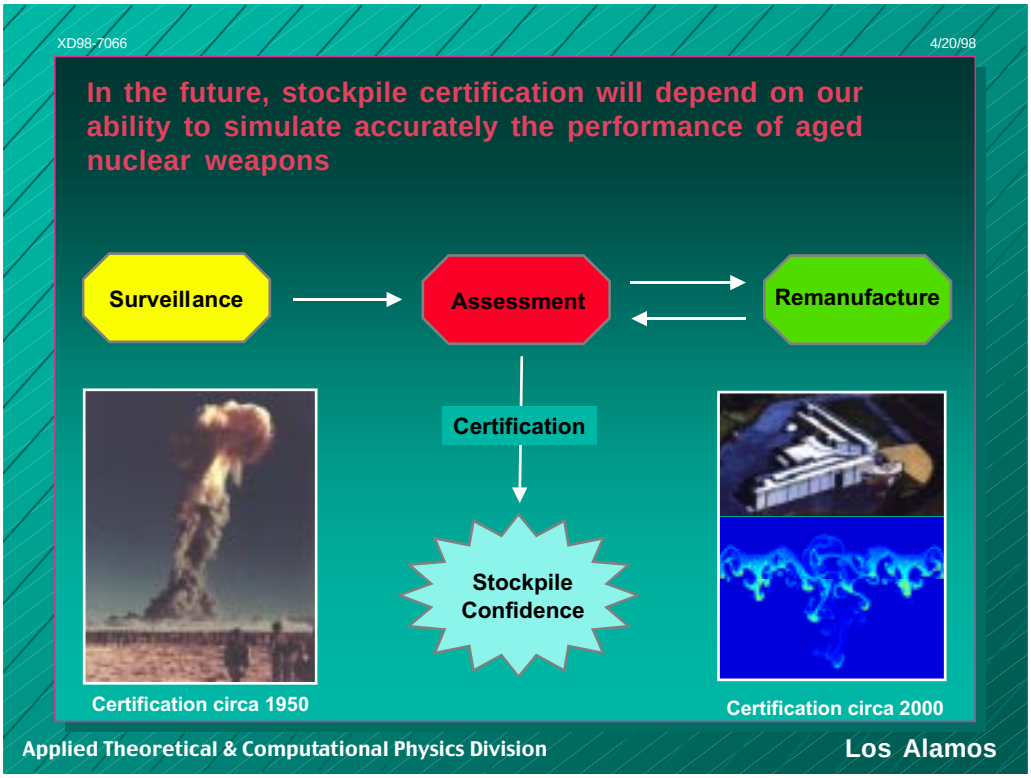
XD98-7065 4/20/98

Stockpile stewardship and management must be a continuous, fully integrated process



- An enhanced surveillance process, which is capable of diagnosing and predicting aging-related phenomena in stockpile weapons.
- A fundamental understanding of the consequences of the aging and manufacturing processes on weapons performance.
- The capability to repair and/or remanufacture and to revalidate stockpile weapons.

Applied Theoretical & Computational Physics Division Los Alamos

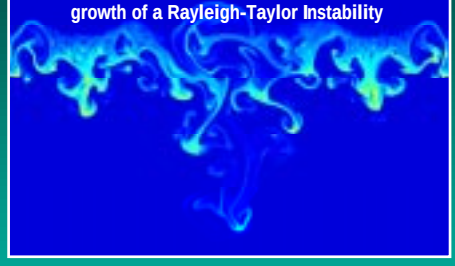

XD98-7066
4/20/98
In the future, stockpile certification will depend on our ability to simulate accurately the performance of aged nuclear weapons
Surveillance
Assessment
Remanufacture
Certification
Stockpile Confidence
Certification circa 1950
Certification circa 2000
Applied Theoretical & Computational Physics Division
Los Alamos

XD98-7067
4/20/98
Successful stockpile stewardship will require surveillance, assessment, and remanufacture
Surveillance—
regular destructive and nondestructive sampling of stockpile systems
Remanufacture—
refurbishment or replacement to meet safety, reliability, and performance requirements
Assessment—
investigation of surveillance observations and evaluation of refurbishment options
Applied Theoretical & Computational Physics Division
Los Alamos

XD98-7068 4/20/98

Without nuclear testing, computations will provide the only integrating mechanism to assure the performance and surety of the aging stockpile

Direct numerical simulation of the nonlinear growth of a Rayleigh-Taylor Instability



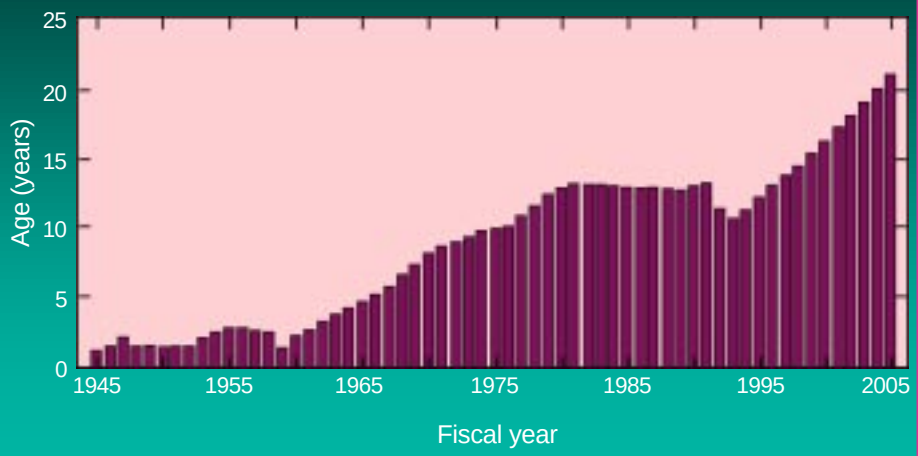
- Surety and aging questions are far more challenging than designing new weapons
- New numerical methods will be required to address the problems of the aging stockpile
- Greatly enhanced computational capabilities will be required to address the 3D problems which are sure to arise

- Major physics improvements must be incorporated in the weapon design codes
  - 3D hydrodynamics with fully coupled photon and particle transport and complete fission-fusion reaction networks
  - Improved models for HE performance, friction, spall, ejecta, fracture, tensile damage, mix, etc.
  - Improved data bases for material equations of state, opacity, etc.

Applied Theoretical & Computational Physics Division Los Alamos

XD97-6981 4/20/98

Average age of the stockpile* FY1945-2004



*Stockpile is the total stockpile, i.e., all capsule systems, gun-type weapons, and sealed-pit systems. Derived from stockpile data in *A History of the Nuclear Weapons Stockpile (U), FY1945-FY1991* and P&PD 95-0.

Applied Theoretical & Computational Physics Division Los Alamos

CIC-1/96-1296b 7/96 4/20/98

Plutonium aging is one focus of enhanced surveillance

Concern: Structural instability caused by helium bubble nucleation and ingrowth of uranium, neptunium, and americium

Impacts:
- swelling/density
- phase stability
- alloy segregation
- surface structure

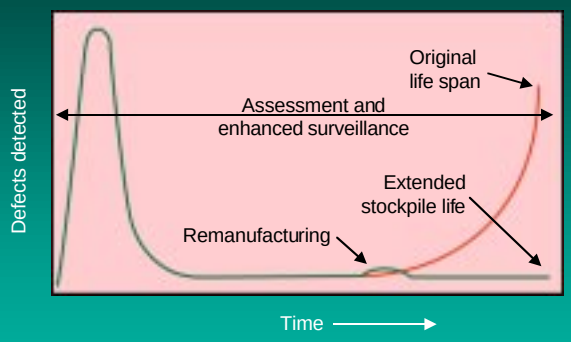
Tools:
- transmission electron microscopy
- extended x-ray absorption
- neutron diffraction/backscatter
- Auger electron spectroscopy

Helium bubbles in aged 20-yr old plutonium

Applied Theoretical & Computational Physics Division Los Alamos

XD97-6977 4/20/98

Life cycle of manufactured systems



The stewardship program will extend the useful life of U.S. stockpile warheads through enhanced surveillance, assessment, and remanufacturing

Applied Theoretical & Computational Physics Division Los Alamos

XD98-7069 4/20/98

The Accelerated Strategic Computing Initiative (ASCI) has four objectives

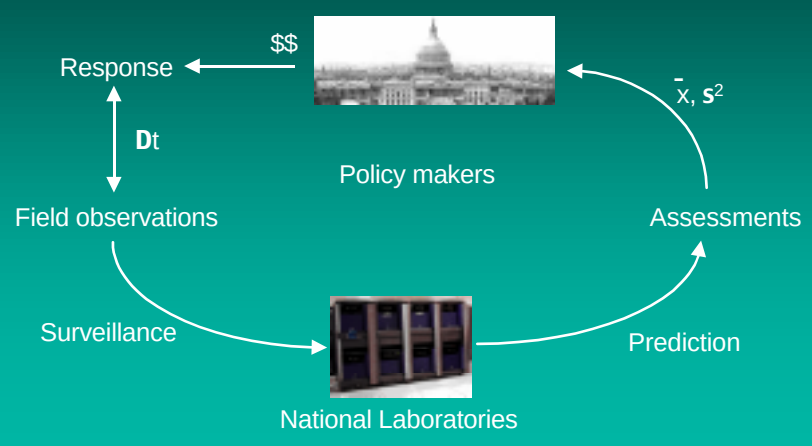
- ***Performance:*** create predictive simulations of nuclear weapon systems to analyze behavior and assess performance in an environment without nuclear testing.
- ***Safety:*** predict with high certainty the behavior of full weapon systems in complex accident scenarios.
- ***Reliability:*** achieve sufficient, validated predictive simulations to extend the lifetime of the stockpile, predict failure mechanisms, and reduce routine maintenance.
- ***Remanufacturing and Renewal:*** use virtual prototyping and modeling to understand how new production processes and materials impact performance, safety, reliability, and aging issues.

These objectives will be realized by implementing the five ASCI strategies

Applied Theoretical & Computational Physics Division Los Alamos

XD98-7026 4/20/98

Policy community demands predictions of high-order accuracy to effectively address societal problems in a timely and resource-efficient way



Applied Theoretical & Computational Physics Division Los Alamos

XD98-7029 4/20/98

A broad class of nationally-important problems requires predicting response of complex systems outside the envelope of controlled experiments and direct reliable observation

- Science-based stewardship of the nuclear weapons stockpile
- Global climate predictions
- Nuclear reactor technology
- Virtual testing aerospace, auto, military technologies
- Natural disaster forecasting
- National infrastructure security

Applied Theoretical & Computational Physics Division Los Alamos

XD98-7064 4/20/98

What does the policy community require to build confidence in "predictive" assessments?

- Rigor in technical work, formality in the process
- Technical integrity
- Some plan or approach to software quality assurance (SQA)
- Peer review
- Quantitative metrics for uncertainty

Applied Theoretical & Computational Physics Division Los Alamos

XD98-7030 4/20/98

When "predictability" is the issue, quantification of error is just as important as solution itself

- Errors almost always imply some acceptance of "cost," in the response to initiating circumstances
- Confidence-building in the policy community
- Theoretical foundations are not always developed
- Increasing number of computed cycles is not always the answer

Applied Theoretical & Computational Physics Division Los Alamos

XD98-7023 4/20/98

Euler's taxonomy of "certainty"

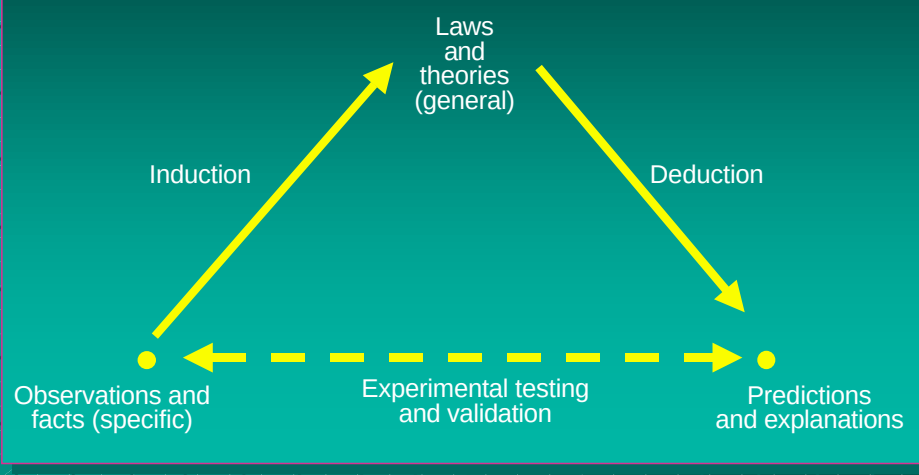
- Perceptual certainty (direct experience)
- Demonstrative certainty (deduction, tools of logic)
- Moral certainty (knowledge "by faith")

Demonstrative certainty is attained through the process of scientific inquiry

Applied Theoretical & Computational Physics Division Los Alamos

XD98-7024 4/20/98

The process of scientific inquiry is driven by our need to make sense of the things we see, so that we can predict the future course of events



Applied Theoretical & Computational Physics Division Los Alamos

XD98-7028 4/20/98

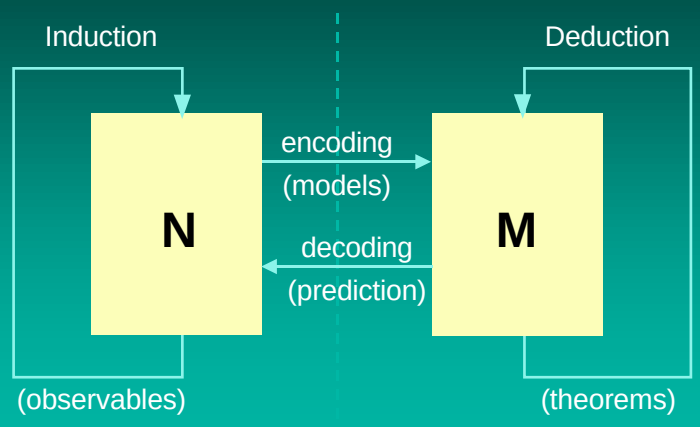
The nature of scientific prediction ...

- Governed by “rules,” “laws” tested by time
- Explicit schemes (clear and detailed rules, applied by “anyone”
- Publicly available, can be tested independently
- Can postulate unobservable “stats” or events as part of the explanation mechanism
- “Causality” does not imply “predictability,” e.g., celestial mechanics, earthquake prediction

Applied Theoretical & Computational Physics Division Los Alamos

XD98-7025 4/20/98

Mathematics (modeling) gives a systematic, reliable procedure for generating new truths from old (rules of logical inference)



Applied Theoretical & Computational Physics Division Los Alamos

XD98-7022 4/20/98

Sources of uncertainty in numerical modeling

- Incomplete understanding of governing natural laws (unknown physics)
- Fundamental data limitations (initial and boundary specifications, physics data, and parameters)
- “Bugs” in computer codes
- Stochastic problem representations
- Multiscale approximating and fidelity across scales (representation of unresolved scales)
- Temporal evolution of dynamic systems

Applied Theoretical & Computational Physics Division Los Alamos

XD98-7020 4/20/98

Strategic computing and simulation for stockpile stewardship and management…meeting "green book" requirements

Strategic Computing and Simulation Program

- **Accelerated Strategic Computing Initiative (ASCI)**
  - ASCI provides the leading edge, high-end simulation capabilities needed to meet weapons assessment and certification requirements without nuclear testing
- **Distance Computing and Distributed Computing for Weapon Simulation (DisCom²)**
  - DisCom² will develop and provide the technology needed to deploy an integrated environment that permits DP labs and plants to access computing (from desktops to TFlops) across thousands of miles
- **Stockpile Computing (SC)**
  - Conduct computing operations, models development, and code maintenance to support execution of the SSMP
- **Numeric Environment for Weapon Simulation (NEWS)**
  - A local computational environment for large numbers of designers to use high-end simulation capabilities to simultaneously address large numbers of stockpile issues
- **Validation an Verification (V²)**
  - Provide the tools, methodologies, and data to ensure high-end simulation capabilities reflect and predict the real world

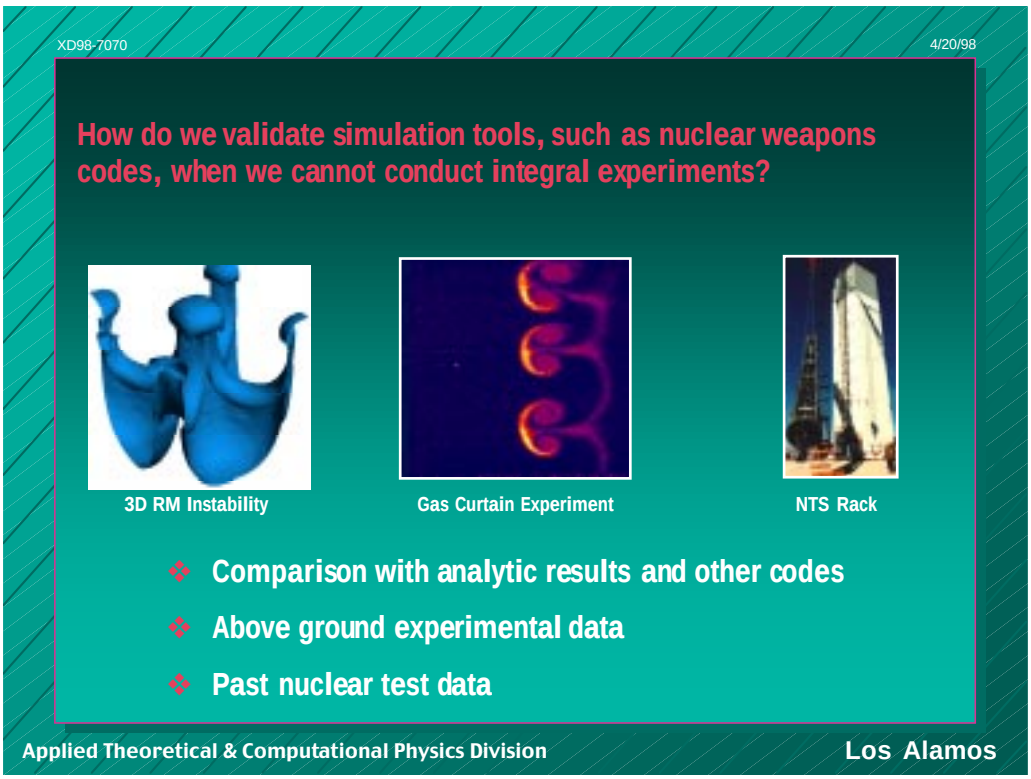
Applied Theoretical & Computational Physics Division Los Alamos

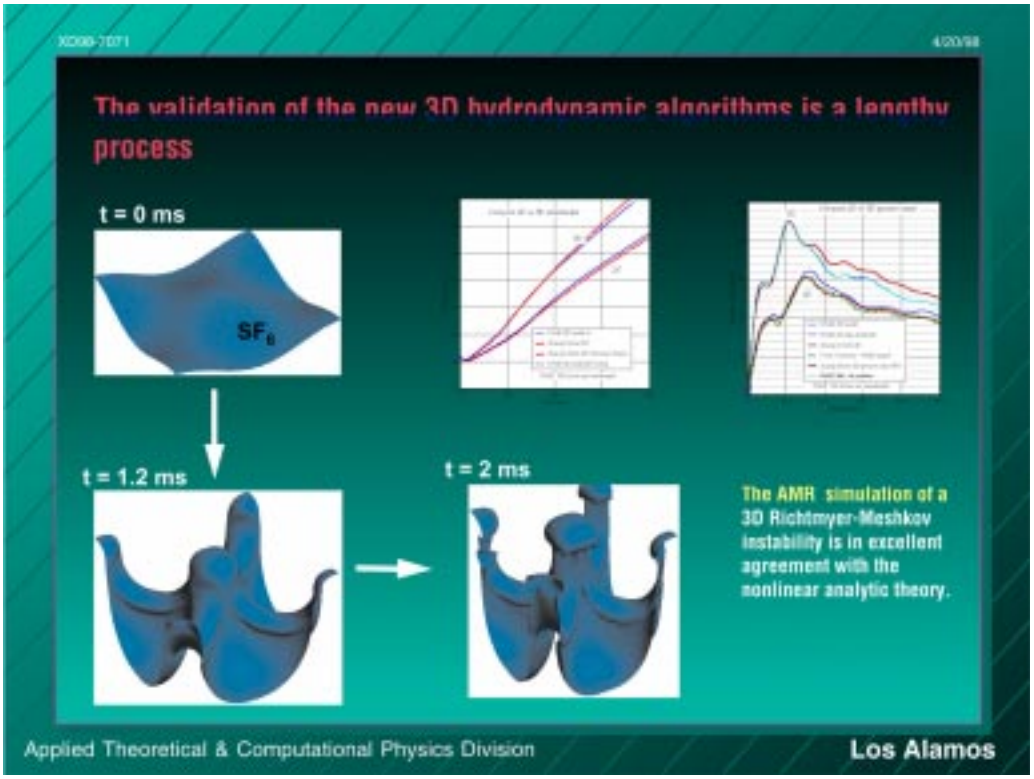
XD98-7045 4/20/98

Tools and techniques for verification

- Applying "best practices" and standards in software engineering
- Effective debuggers
- Configuration management
- Development of appropriate test suites and ????
- Use of analytical benchmarks
- Peer review, software review process

Applied Theoretical & Computational Physics Division Los Alamos

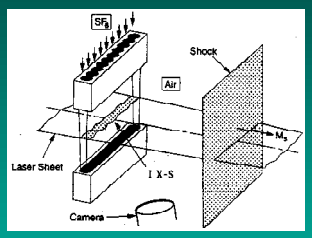

XD98-7070
4/20/98
How do we validate simulation tools, such as nuclear weapons codes, when we cannot conduct integral experiments?
3D RM Instability
Gas Curtain Experiment
NTS Rack
Comparison with analytic results and other codes
Above ground experimental data
Past nuclear test data
Applied Theoretical & Computational Physics Division
Los Alamos


process
t = 0 ms
SF6
t = 1.2 ms
t = 2 ms
The AMR simulation of a 3D Richtmyer-Meshkov instability is in excellent agreement with the nonlinear analytic theory.
Applied Theoretical & Computational Physics Division
Los Alamos

XD98-7072 4/20/98

Laboratory experiments can be used to validate the algorithms for direct numerical simulation of hydrodynamic instabilities

- Los Alamos DX-13 shock tube data:
  - Gas curtain experiments.
  - High quality data for onset of turbulence
  - Mach 1.2 shock induced instability
- Experimental innovations include:
  - Laminar gas jet to produce interfaces
  - Laser Rayleigh scattering sheet to observe cross-section with high resolution



Direct numerical simulation of laboratory data provides confidence in the hydrodynamics that is "necessary but not sufficient"

Applied Theoretical & Computational Physics Division Los Alamos

XD98-7073 4/20/98

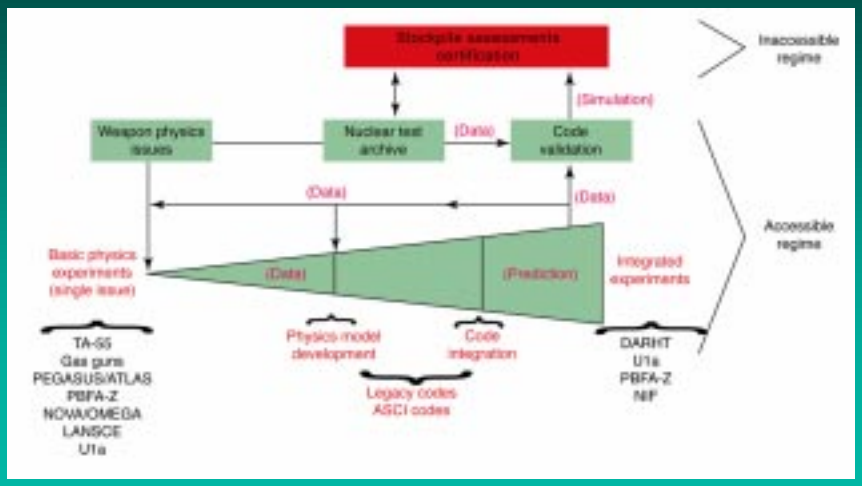
The development of a Richtmyer-Meshkov instability is very sensitive to the initial conditions

- Density plots
- The Benjamin and Budzinski data are shown on the left
- The RAGE Adaptive Mesh Refinement Eulerian simulations are shown on the right
- The RAGE calculations used initial conditions taken directly from the experimental data

Applied Theoretical & Computational Physics Division Los Alamos

XD97-6982 4/20/98

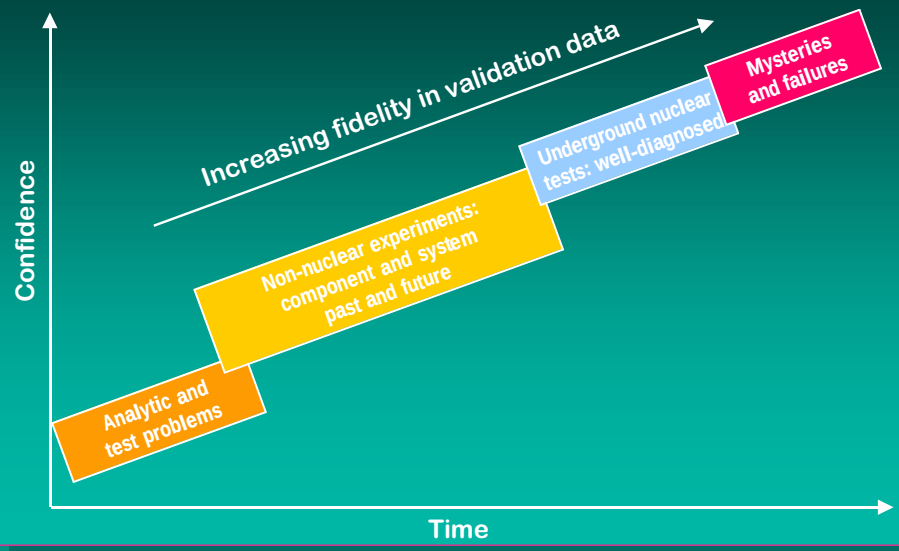
Stockpile assessments must be grounded in rigorous adherence to scientific method



Applied Theoretical & Computational Physics Division Los Alamos

XD98-7021 4/20/98

Validation curve: nuclear performance/safety and hostile environment codes



Applied Theoretical & Computational Physics Division Los Alamos

XD98-7048 4/20/98

Data sources for validation

- Laboratory-scale and above-ground experiments in the most complete range of thermophysical conditions allowable
- Archive of underground nuclear test data
- Stockpile surveillance data

Establish, with increasing confidence, that a code is predictive. Can you confidently extrapolate beyond the range of experimental data a code was validated with?

Applied Theoretical & Computational Physics Division Los Alamos

XD98-7074 4/20/98

Probabilistic risk assessment (PRA) has been employed in major technology sectors to predict behavior of complex systems (e.g., quantify risk)

- 1974 Reactor Safety Study (WASH-1400)
- Major application in nuclear reactor safety analysis (core meltdown, containment releases, consequences)
- Uncertainty analysis employs
  - system-level *fault trees*
  - accident sequence *event trees*
  - appropriate failure data
- Subjective judgements may be used when data is sparse or unavailable (expert elicitation)
- Weapons safety applications—probabilistic weapon response models
- Methodology aids in identification of major contributors to uncertainty; provides insight for resource investment
- May not be a tractable tool in a highly coupled nonlinear dynamic system

Applied Theoretical & Computational Physics Division Los Alamos

XD98-7050 4/20/98

SBSS assessment uncertainties should be quantified within a full probabilistic framework

- Successful prediction must account for uncertainty
- Two-part approach has strong potential for success
  - stochastic PDEs (SPDEs) for forward prediction
  - Markov Chain Monte Carlo (MCMC) and other modern statistical inference methods for inverse prediction

Applied Theoretical & Computational Physics Division Los Alamos

XD98-7049 4/20/98

Predictive science can be regarded as essentially a two-step procedure *(Glimm, Sharp, et.al.)*

- Forward problem (forward prediction)
  - given model - equations, initial data, parameters
  - predict behavior
- Inverse problem (inverse prediction)
  - given observations
  - validate forward solution
  - improve (predict) model
- How can we do this in presence of uncertainties?

Applied Theoretical & Computational Physics Division Los Alamos

XD98-7052 4/20/98

Role of stochastic PDEs

- SPDEs propagate uncertainty in input to determine uncertainty in output
- Stochastic PDEs combine
  - determinism of physical law (PDEs)
  - modeling of uncertainty (statistical analysis)

Applied Theoretical & Computational Physics Division Los Alamos

XD98-7054 4/20/98

MCMC is based on three ideas

- Bayesian inference
- Monte Carlo simulation
- Markov chain stochastic processes

Applied Theoretical & Computational Physics Division Los Alamos

XD98-7047 4/20/98

Bayesian inference

- A "model," *m*, specifies
  - all equations, initial data, parameters used to describe system
- A *prior* distribution is a probability *p(m)*, which represents our current knowledge of *m, including assessment of uncertainties*
- Further observations *O*
  - add information and constrain model
- Constraints expressed as probabilistic consistency
  - probability of *O* given $m = p(O|m)$
- *Posterior* distribution $p(m|O)$ is updated knowledge of *m*, again including uncertainties

Applied Theoretical & Computational Physics Division Los Alamos

XD98-7053 4/20/98

Markov Chain Monte Carlo (MCMC)—what is it for?

- Statistical inference deals with the reality of incomplete information
- It is a tool to answer crucial question
  - How does adding or removing information affect uncertainty in the answer?
  - What is the uncertainty due to missing physics?
  - How do you quantify effect of solution errors?
- MCMC is a method to carry out Bayesian statistical inference in complex systems

Applied Theoretical & Computational Physics Division Los Alamos

XD98-7055 4/20/98

Blending responsible conservatism with a commitment to the methodical development of true predictive capability

- Establish modern baseline for each enduring stockpile weapon system
- Draw a “box” in parameter regime about each system, based on quantified uncertainty
- Don’t certify “outside the box”
- Enhance simulation fidelity (SBSS, ASCI)
- Use verification, validation, and stochastically-based “predictivity tools” to demonstrate an expanded range of design parameter space (go to second bullet)

Applied Theoretical & Computational Physics Division Los Alamos

XD98-7056 4/20/98

In summary

- The fundamental science-based stockpile stewardship paradigm joins a larger class of complex simulations which directly impact the policy/technical interface
- Quantified uncertainties in predictions are just as important as the assessments themselves—they are required for ensuring long-term confidence of the policymakers
- Methodical verification and validation of our ASCI codes is time-urgent and critical, but not sufficient for establishing their absolute “predictability”
- Parallel investments must be made in developing the appropriate theoretical foundations for the quantification of uncertainty in a probabilistic framework

Applied Theoretical & Computational Physics Division Los Alamos

RS/6000

100 TFLOPS and Beyond

Tilak Agerwala
Director, Server Architecture & System Strategy
IBM Corporation

(C) Copyright 1998 IBM Corp.

RS/6000

Special notices

All statements regarding IBM's future direction and intent are subject to change or withdrawal without notice, and represent goals and objectives only. Contact your IBM local Branch Office or IBM Authorized Reseller for the full text of a specific Statement of General Direction.

The information contained in this document represents the current views of IBM on the issues discussed as of the date of publication. IBM cannot guarantee the accuracy of any information presented after the date of publication.

The following terms are trademarks of International Business Machines Corporation in the United States and/or other countries: AIX, PowerPC, RS/6000, SP.

(C) Copyright 1998 IBM Corp Tilak Agerwala
Salishan

RS/6000 THE PATH TO 10 TFLOPS

IBM

1997	1998	1999	2000

Technology Refresh

- 336 PPC 604 SMP nodes
- 32 Power2 nodes
- 900+ GFLOPS
- 178 GB Memory
- 2.6 TB disk

ASCI Blue

- 512-way system
  - PPC G3 SMP nodes
- 3.2 TFLOPS
- 2.5 TB Memory
- 75 TB disk

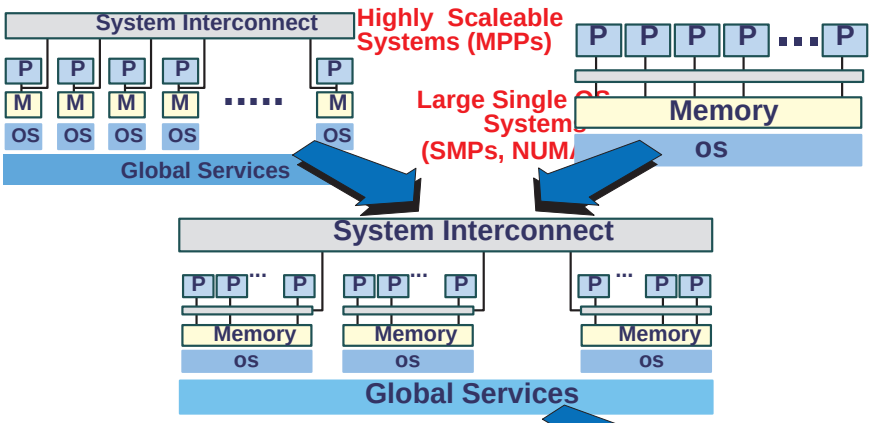
ASCI White

- 512-way system
  - Enhanced PPC G3 SMP node
- 10 TFLOPS
- 4 TB Memory
- 150 TB disk

(C) Copyright 1998 IBM Corp Tilak Agerwala Salishan

RS/6000 Converging to a Common Architecture

IBM



Architectures converging to

- Multiple partitioned sets of resources
- Managed by own copy of operating system
- With globalized resources and services as required

Evolving to ...

- Greater flexibility and
- Global Shared Memory across OS images

(C) Copyright 1998 IBM Corp Tilak Agerwala Salishan

RS/6000

Future Architecture Evolution

IBM

- Key emphasis on system availability
  - Failure isolation
  - Recovery services
  - Hot pluggable at all levels
- Scaleable and flexible system management
- Scaleable I/O
- Programmer's task can be greatly facilitated by support of all dominant programming models consistently up and down the hierarchy
- Increasing latencies, deepening hierarchy
  - Locality of data will be critical

(C) Copyright 1998 IBM Corp Tilak Agerwala
Salishan

RS/6000

100 TFLOPS AND BEYOND ...

IBM

1999-00
~ 4-10 TFLOPS
~ 1-1.3 GF/proc
~1-1.3 GF/chip
8 - 16 procs/ node
8-20 GF/node
512 nodes
150 frames
Standard air cooling
1 - 1.5 MW

2001-02
30+ TFLOPS
~2X GF/proc
~2X GF/chip
2X procs/ node
3-4X GF/node
512 nodes
150 frames
Standard air cooling
1 - 1.5 MW

2003-2004
100+ TFLOPS
~2+ X GF/proc
~4+ X GF/chip
2X procs/ node
4+ X GF/node
512 nodes
200 frames
Aggressive cooling ?
~ 1.5- 2 MW

2008-2010
~ 1 PetaFLOPS
~2 X GF/proc
~16X GF/chip
4X proc/node
12X GF/node
512+ nodes
250 frames
Aggressive cooling
~ 3 - 4 MW

(C) Copyright 1998 IBM Corp Tilak Agerwala
Salishan

RS/6000

Getting to 100 TFLOPS ...

IBM

- Applications must scale up to circumvent Amdahl's law
- Increased need for tuning
  - Parallelism level ~ 60-70 X today
    - (~30-35K threads or processes)
    - Staged increase over 6 years
  - Increased latencies, constrained bandwidth
  - Increased levels in hierarchy
- Consistent programming model across the total hierarchy will greatly facilitate tuning

- Can complexity of parallel architectures and algorithms be hidden from users without sacrificing performance ?
- How do you debug and optimize across 30-35K threads/processes ?
- Managing 512 nodes and 30-35K processors
  - Streamlined management approach
  - AI based management systems
- 10s of PBytes storage will require
  - Fully integrated, application-transparent storage hierarchy
  - Alternative to disks ?
    - Density, reliability, speed, cost
- Hot-Plug at reasonable granularity
- Fault tolerant structures needed ?

Compared to today,

Memory latency : ~ 10 X worse in processor cycles

Memory bandwidth balance (B/F) : degrades by factor of 5-10

Much of BW degradation will come with the introduction of SMP nodes

Bulk cache will maintain effective latency to 50-100 cycles, and bandwidth balance (B/F) to around 1 to 2

(C) Copyright 1998 IBM Corp Tilak Agerwala Salishan

RS/6000

Summary

IBM

- Ultra-Large systems must be based on leading edge mainstream technology of the day
- Such system with conventional architecture is probably feasible given current technology trends projections
  - But will market pressures move technology in different direction ?
- Price/performance trends rather than technical feasibility will drive when such systems emerge
- Without breakthroughs, stress points will emerge throughout the memory/storage hierarchy
  - Considerable application tuning at all levels
- Aggressive packaging with advanced, scaleable cooling necessary for manageable size
- Semiconductor manufacturing breakthroughs required to hit projected performance levels

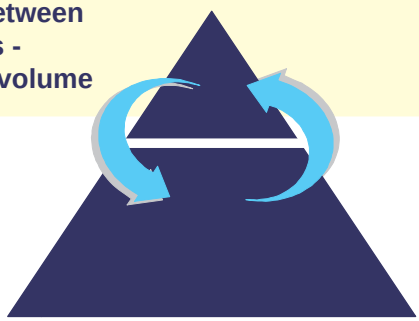
(C) Copyright 1998 IBM Corp Tilak Agerwala Salishan

RS/6000

Ultra-Large Scale Systems

IBM

- Market dynamics will not support substantial development that is unique to ultra-large systems
- Development of ultra-large systems will be accelerated if mainstream applications can effectively use smaller clones of these systems.
- Whatever new technology is required for ultra-large systems must ultimately be applicable to mainstream commercial applications
- Designed well, there is a synergy between ultra-large and mainstream systems - between the high-end and the high-volume



(C) Copyright 1998 IBM Corp Tilak Agerwala
Salishan

Scaling to Very Large Systems

Greg Astfalk

Hewlett-Packard Company
3000 Waterview Pkwy.
Richardson, TX 75080–1400 USA
astfalk@rsn.hp.com

Prologue

- We are interested in how technology will, or will not, permit us to scale up to systems with performance in excess of 10 (peak) Tflops.
- The bias we show is to be "practical" and consider what is possible in the next ~ 5 years.
- Miracles and blue-sky technologies are not considered here.

DRAM density

- No surprise that this continues to grow at a phenomenal rate.
- It is not a good indicator of processor development.
- Memory manufacturers may depart from the 4× increases every 3 years.
- The density growth rate is outstripping the demand and this reduces the volume sold; poor economic formula.
- An important point: the density is increasing rapidly but the access time is NOT decreasing rapidly.

Processor frequency

- Processor frequency is following a milder ramp than DRAM density.
- Approximately 20% every 3 years.
- 1.6 GHz in 2003, which is approximately a doubling in the next 5 years.

Processor frequency

- Processor frequency is following a milder ramp than DRAM density.
- Approximately 20% every 3 years.
- 1.6 GHz in 2003, which is approximately a doubling in the next 5 years.

Feature size

- It is not evident that the supporting technologies, primarily lithography, will be available to go beyond 0.13 μ.
- Feature size is not the primary driver for increasing processor frequency.

Being a computer vendor

- Rule #1: be profitable to stay in business to design the next system.
- You can only spend approximately 10% of your revenue on R&D.
- You must have a new system every 2–3 years.
- You don't fab you own silicon (hopefully).

Economics

- A vendor needs to have at least 2 design teams in place for the system and the processor.
- Without overlapping development the vendor can not have a stream of products to market on time.
- All of this tends to indicate that;
  - technical-only business can't succeed,
  - commodity processors must be used.
- Executive summary:
  - This is a very expensive business!

How to get to 10 Tflops

- Consider some possible paths to a 10 Tflops system.
- We could use a:
  - very, very, very impressive uniprocessor.
  - very, very impressive SMP.
  - very impressive ccNUMA.
  - large SMP or ccNUMA cluster.
  - very, very impressive PC cluster.

Processor speed

- Lets consider how to get a single processor to 10 Tflops.
- Assuming we have 8-way issue the frequency then needs to be 1,250 GHz.
- Today we are not even at 1 GHz.
- In 5 years 1,250 GHz is 780× the SIA projection.
- This constitutes a miracle and therefore it won't happen.

Using PCs

- Lets consider building the 10 Tflops system from PCs.
- This is Beowoulf on steroids!
- Consider that each PC is 0.5 cubic feet and 200 (peak) Mflops.
- A mere 50,000 processors that occupy 25,000 cubic feet or $60 \times 60 \times 8$ feet (with no voids).

Using PCs (cont'd)

- Considering the MTTBSOD† for a PC you find that the entire system will be almost never be available.
- Any credible interconnect that could handle the communication volume would be as expensive as the processors.

†—Mean Time To The Blue Screen Of Death.

Using a SMP

- Lets consider a somewhat more contemporary approach; a SMP.
- SIA projects a 1600 MHz processor in 2003.
- Assume 4-way issue to give us a 6.4 (peak) Gflop processor
- Thus we need 1,563 processors in the SMP.

Using SMPs (cont'd)

- To get the true SMP, with flat memory we need to use either a;
  - flat crossbar for low latency

or a

  - multi-stage crossbar for uniformly bad latency.
- If we use a flat, crossbar we would need 3,126 ports.
  - Assume 150–200 wires per bus to yield a $\sim$ 500k pin-out ASIC for the crossbar.
- This is a miracle that won't happen.

NUMA SMPs

- If we relax the requirement for flat memory then the size of the "SMP" can be increased.
- We can accomplish this now by using cascading crossbars.
- Each foot of separation between the components will induce a 4 nanosecond delay ($\sim$ 2 nanosoconds per foot)
- Each ASIC in the path adds $\sim$ 25–50 nanoseconds.
- We are left with a NUMA node (not strictly a SMP).
- The industry is trending to this architecture.
- What about the o/s for >1000 processors?

ccNUMA (cont'd)

- Assume we have 6.4 (peak) Gflop processors.
- Relaxing the restriction of flat memory then we can do a "NUMA SMP" node of $\sim$ 64 processors.
- Now we are only faced with putting $\sim$ 25 nodes together in some fashion.

Clustering

- The most viable approach to scaling to large systems is through the use of clustering.
- The differentiating attributes of clustering are;
  - the presence of multiple, autonomous operating systems,
  - loss of a single, global name-space.
- Software layers can provide some semblance of SSI for the user and the administrator.

Why clusters?

- They are affordable and feasible!
- All vendors will be producing, for "volume" sale, SMPs or ccNUMA systems.
- Using the SMP or ccNUMA system as the building block allows for reasonable product development.
- Does not require any "extrordinary" technology (i.e., o/s, hardware, interconnect, etc.)
- There is some negative onus placed on the end-user.

Clustering: user perspective

- The loss of the global name-space.
- The presence of a deep memory hierarchy.
- Explicit messaging among some parts of the hierarchy but not among others (for efficiency).

Cluster interconnect

- It must avoid the protocol stack and the o/s.
- Scheduled transfer looks to have genuine promise.
- HIPPI6400 appears to be a viable approach with a number of distinct benefits:
  - 800 Mbyte/sec bi-directional bandwidth,
  - scheduled transfer,
  - switched implementation possible.

Latency management

- Processors will not likely catch up to the point that they can cover the latency to even the nearest memory.
- Latency management will be a focal point of design in both hardware and software.
- The memory hierarchy is deep and going to get deeper.

Level	clocks
register	1
primary (L1) cache	2–3
secondary (L2) cache	6–12
tertiary (L3) cache	14–20
near local memory	60–100
far local memory	100–250
remote shared memory	$\mathcal{O}(10^2)$
message-passing	$\mathcal{O}(10^3)$–$\mathcal{O}(10^4)$

100TFLOPS by 2004

Evolution OR Revolution?

Commodity OR Special Magic?

What is Sun doing here anyhow?

Erik Hagersten

Chief Architect, High-End Server Engineering, SMCC

Server Market [%]

Application	Market Share 1995	Market Share 2000
Database	32	39
FileServer	32	25
Eng+Sci	16	14
Print Serv	9	7
Media Email	6	8
Others	5	7

(Source: Dataquest)

Server Market [%]

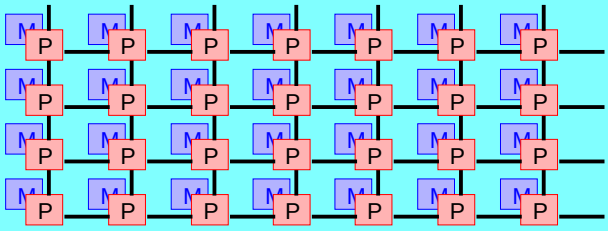
Server size	Technical/ Scientific	Commercial
<$10k	1	19
<$50k	5	25
<$250k	5	24
<$1M	2	9
>$1M	3	8

Sun's Path to Scalable HPC systems

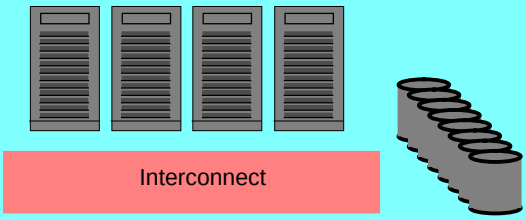
(Source: Dataquest)

HPC Server Architectures

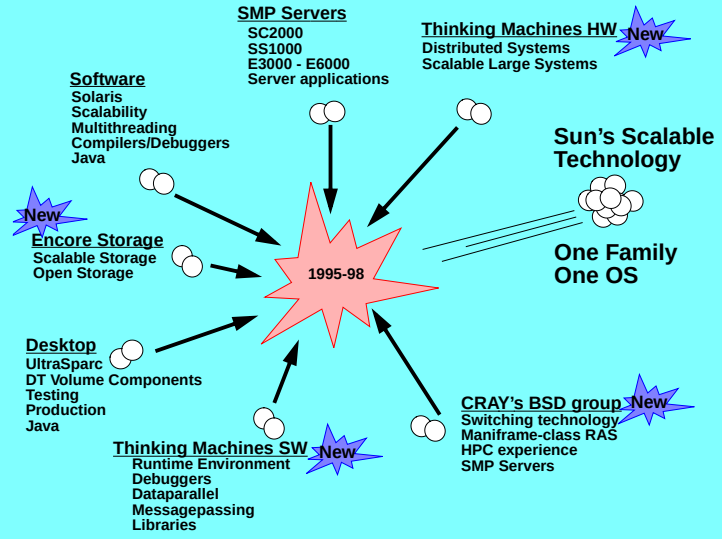
Yesterday: Boutique Shoping



Today: Commercial Building Blocks



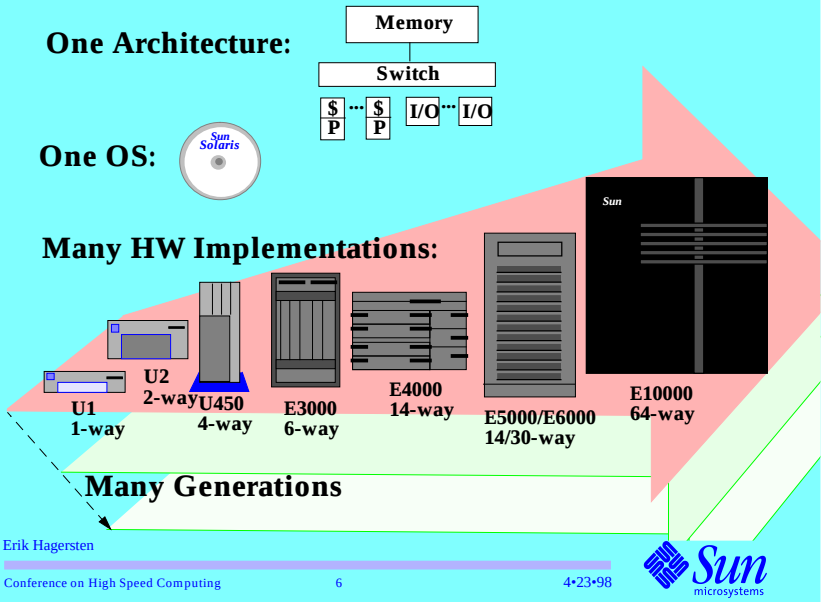
Sun's Technology Fusion



Erik Hagersten

Conference on High Speed Computing 5 4•23•98

Sun's Family of Enterprise Servers

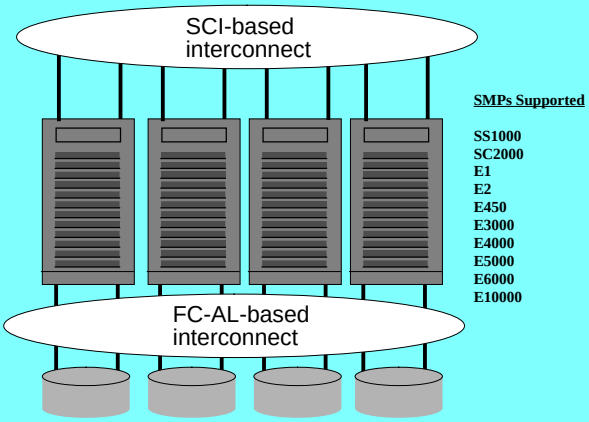


Erik Hagersten

Conference on High Speed Computing 6 4•23•98

Sun HA Clusters & Enterprise PDB

- Scaling beyond a box: Clustering
- The scalable SMP technology is the base
- High-availability, fail-over and further scalability



SMPs Supported

SS1000
SC2000
E1
E2
E450
E3000
E4000
E5000
E6000
E10000

Erik Hagersten

Conference on High Speed Computing 7 4•23•98

HPC achievements

- **Former TMC-HW, TMC-SW, and Cray-BSD ==> Sun 1996**
  - Experience from CM-1, CM-2, and CM-5.
  - Parallel Programming Environment and Debuggers (PRISM)
  - Switching Technology
  - Scalable Operatingsystem
  - Technology Leveraging
  - Large Stable Company
- **"The Network is the Computer" works for HPC as well**
  - Fat SMP nodes
  - Leveraging Commercial High-volume Technology
  - RAS: Reliability, Availability, Serviceability
  - Few nodes ==> Easy to Manage, Explore Locality
  - Scalable Network Technology
- **LINPACK**
  - 34.2 GFLOPS: 64-way SMP 336 MHz USPARC II (80% efficiency)
  - 100.4 GFLOPS: 4 x 64-way SMP 250 MHz USPARC II
- **Top 500 List**
  - 1996: 0 entries
  - 1997: 87 entries (#2 vendor)
  - 1998: "more than half" [Erik's guess]
- **Awarded ASCI PathForward R&D Contract 1998**

Erik Hagersten

Conference on High Speed Computing 8 4•23•98

100TFLOPS by 2004: extrapolation by looking 6 years back

- CPU speed 2x every 18 months [~Moore's]
- SMP bandwidth 2x every 13-14 months
- Sun's TPC-C number doubles every 9 months (this will drive the "node performance" for HPC)

How Many CPUs?

- 1998: 670MFlops (now)
- 2004 (6 years later): 2^4 * 670MFlops = 10 GFlops /CPU
- Need ~10,000 CPUs

How Large Nodes (i.e., "SMPs")?

- 1997: 64 CPUs/SMP
- 2004: 2^{5-6} / 2^4 * 64 = 128 - 256 CPUs per node

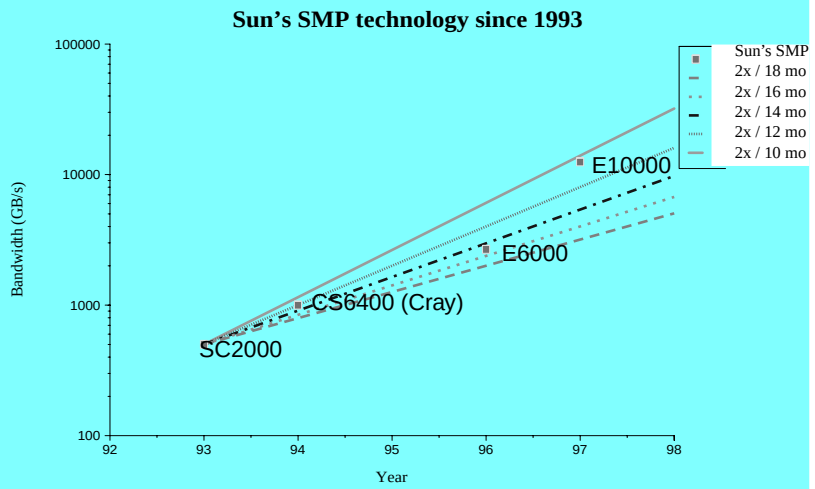
How Many Nodes?

- 10,000 / 128 = ~ 80 nodes
- 10,000 / 256 = ~ 40 nodes

Erik Hagersten

Conference on High Speed Computing 9 4•23•98

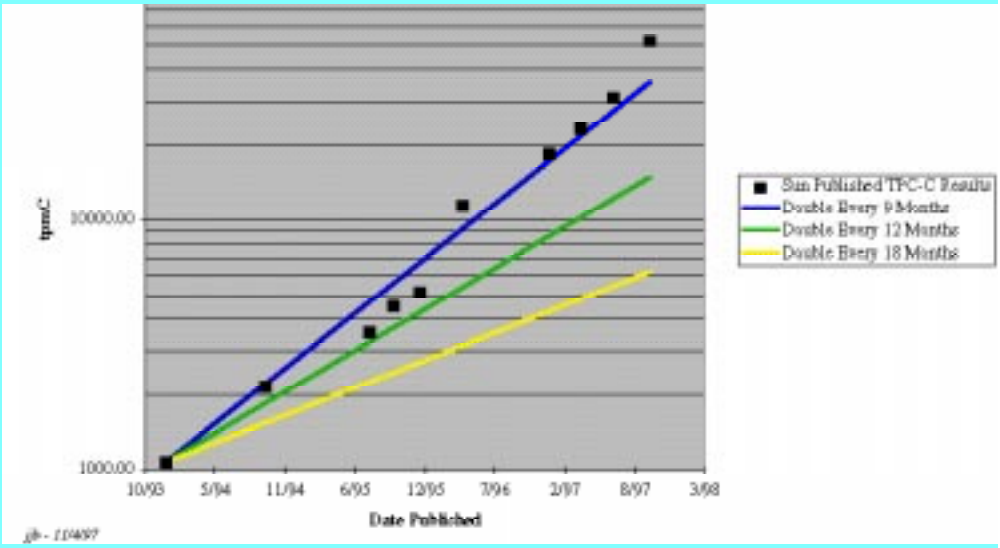
Who said that SMPs can's scale?



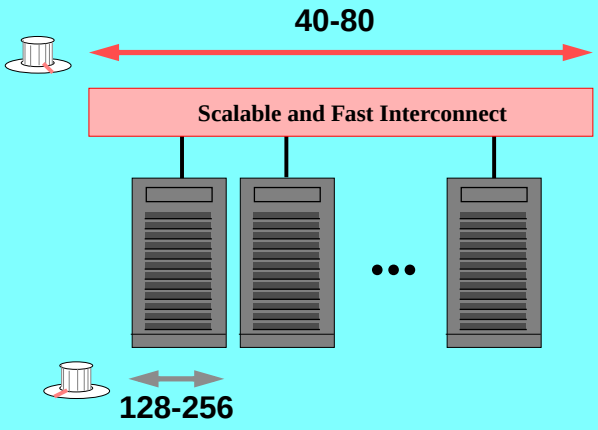
Erik Hagersten

Conference on High Speed Computing 10 4•23•98

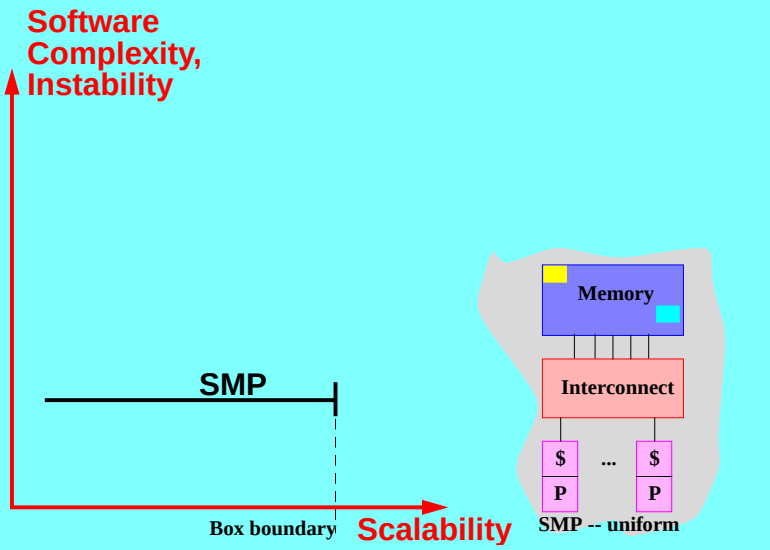
Sun's Trend for Database Applications (Published TPC-C Results)



Two-Knob Architecture: 100 Tflops 2004



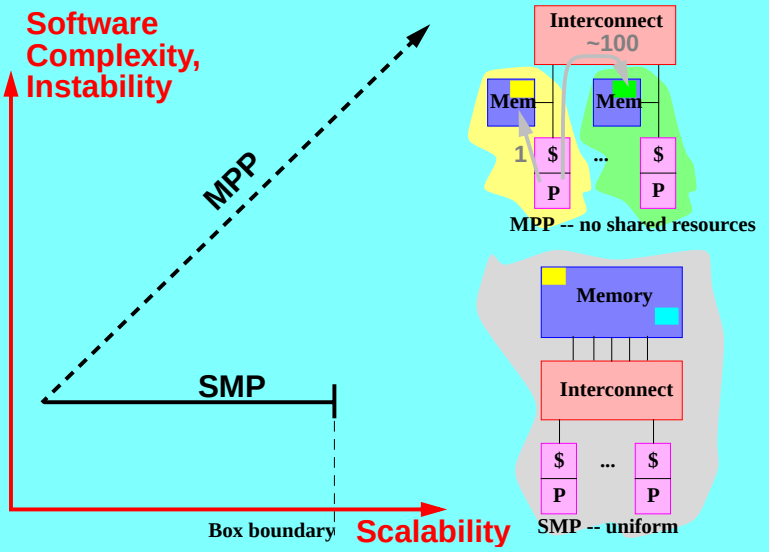
Options for Scalable Architectures



Erik Hagersten

Conference on High Speed Computing 13 4•23•98

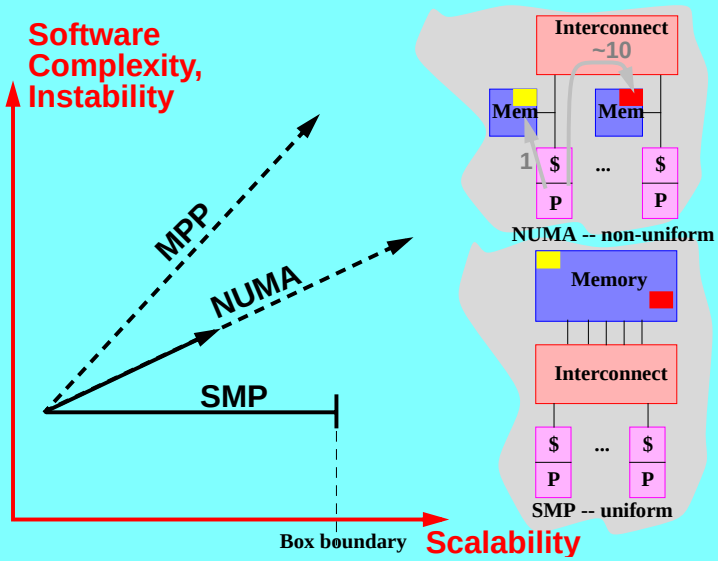
Options for Scalable Architectures



Erik Hagersten

Conference on High Speed Computing 14 4•23•98

Options for Shared Memory Architectures



Erik Hagersten

Conference on High Speed Computing 15 4•23•98

Which one is more scalable?

TPC-D 300GB (Informix)

Machine	Config	Qppd	QthD	$/QphD
Sun Ultra 6000	24-way	3,270	1,477	1,553
Sequent NUMA-Q	32-way	2,667	1,028	3,464

TPC-C

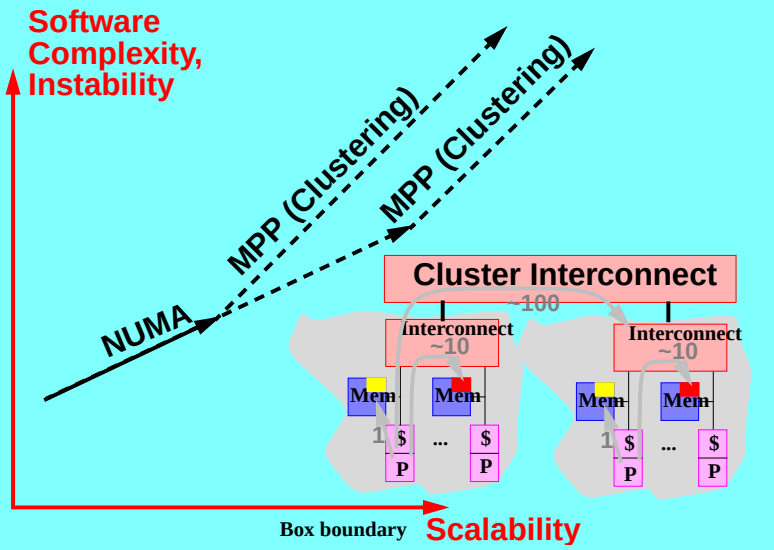
Machine	Config	tpmC	$/tpmC
Sun Ultra 6000	24-way	31,147	108
SGI Origin	28-way	25,309	139

==> SMPs is cost-effective for the “mid-range”

Erik Hagersten

Conference on High Speed Computing 16 4•23•98

Combining NUMA and Cluster is too Complex

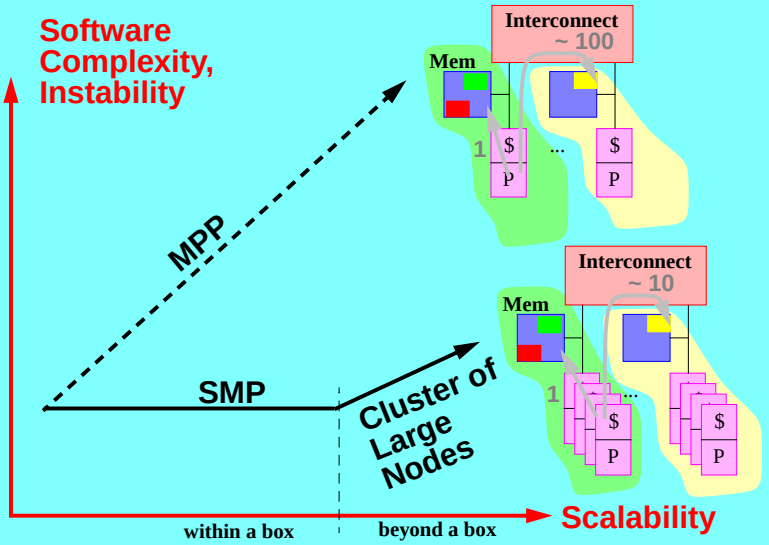


Erik Hagersten

Conference on High Speed Computing 17 4•23•98

Sun microsystems

Few Large Nodes + Fast Interconnect = Simplest Solution



Erik Hagersten

Conference on High Speed Computing 18 4•23•98

Sun microsystems

Clusters of large nodes work well for DSS Databases

All the published numbers for the largest TPC-D database

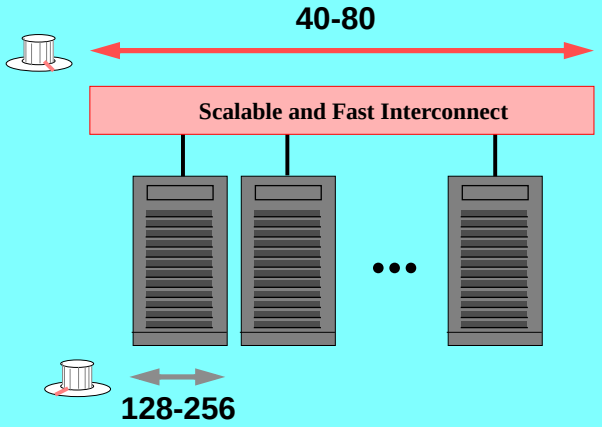
TPC-D 1TB

Machine	Config	Qppd "power"	QthD "throughput"	$/QphD "price/perf."	Date Available
Sun Ultra E6000	4 x 24 way	12,931	5,850	1,353	6/15/98
NCR WorldMark	32 x 4 way	12,149	3,912	2,103	3/15/98
Sun Ultra E10000	64-way	8,870	3,612	1,508	6/01/98
IBM RS/6000	32 x 8 way	7,633	5,155	2,095	3/31/98

Erik Hagersten

Conference on High Speed Computing 19 4•23•98

Two-Knob Architecture: 100 Tflops 2004



- The commercial market will promote SMPs as the high-volume nodes
- NUMAs are not suitable for database applications
-

Erik Hagersten

Conference on High Speed Computing 20 4•23•98

digital™

EV6, EV7, EV8, ...

Richard Kaufmann
Technical Director,
High Performance Computing
Richard.Kaufmann@ digital.com

digital™

DIGITAL/Intel agreement
The facts

- Intel and DIGITAL announced broad-based business relationship
- 10 year patent cross-license agreement

Intel will...

…manufacture Alpha chips for DIGITAL based on the accelerated long term Alpha roadmap

…buy Hudson FAB - $700M (book value)

DIGITAL will...

… continue to own and develop Alpha, maintaining all design engineering resources

… source Alpha from Intel (plus Samsung and potentially others)

digital™

Alpha semiconductor technology roadmap

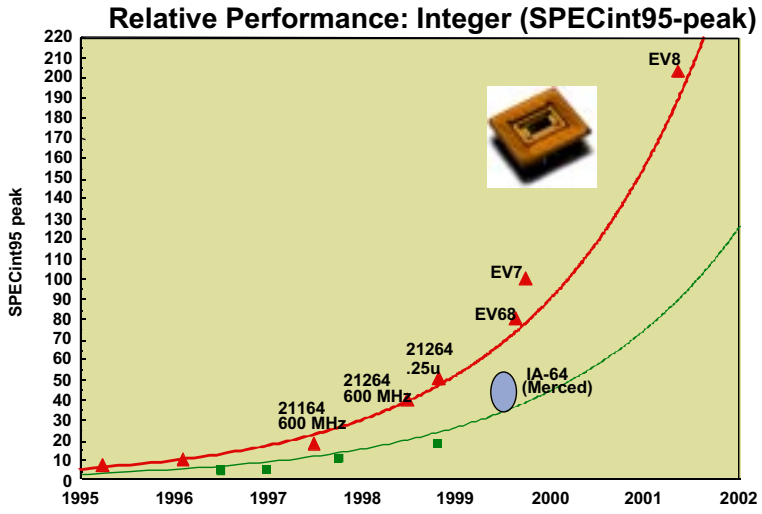
Wafer Size	FAB-4 & FAB-5" 6"			FAB-6 8"			Intel FAB		
	CMOS 3	CMOS 4	CMOS 5	CMOS 6		CMOS 7	CMOS 8		
Flagship Chip	1989	1991	1995	1996	May'98	Dec'98	Dec'99	Dec2000	Dec2001
	Mariah	NVAX/EV-4	EV-5	EV-56	EV-6	EV-67	EV-68	EV-7	EV-8
Min. Feature	1.0 um	.75 um	.5 um	.35 um	.35 um	.25 um	.18 um	.18 um	.18 um
Clock Rate - MHz	63	200	300	400-600	500-800	600-1200	600-1200	600-1200	750-1200+
Transistors	0.8M	1.5M	9.3M	9.3M	15M	15M	15M	100+M	250M

Benefits of using Intel semiconductor manufacturing:

- Significantly lower chip costs
- Significantly lower manufacturing investment
- Much earlier access to 0.18u CMOS
  - Allows EV7 to include larger on-chip caches
  - Accelerates future Alpha roadmap enhancements

digital™

Alpha processor updated roadmap



digital™

New agreement with Samsung announced…..

- DIGITAL:
  - grants Samsung an Alpha architecture license
  - will manage the architectural process to ensure architectural consistency, and will continue to develop future Alpha designs
- Samsung:
  - gains access to all Alpha microprocessor intellectual property, including patents and future implementations
  - can make its own versions of the Alpha processor for specific markets
  - creates a subsidiary dedicated to the marketing and sales of Alpha microprocessors

digital™

Alpha 21264 (EV6) highlights

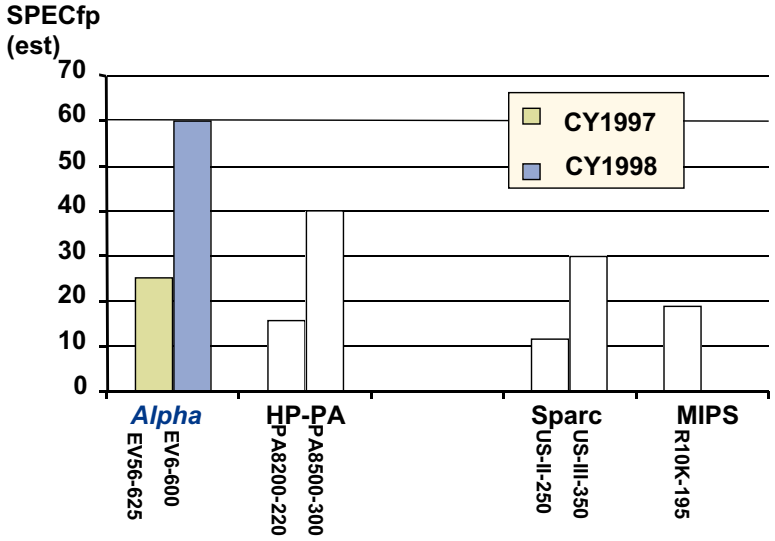
Performance

- **500-600 MHz operating frequency**
- **40 SPECint95 (est)**
- **60 SPECfp95 (est)**
- **MPEGII encode in real-time**
- **Spectacular memory bandwidth**
  - **4+ GB/s L2 cache bandwidth**
  - **2+ GB/s peak memory bandwidth**
  - **1600MB/s McCalpin STREAM**
  - **900 MFLOPs Linpack 1000x1000**

Features

- **Four-way instruction map**
  - **Dynamic execution**
  - **Register renaming**
  - **Out-of-order execution**
  - **Quad integer pipelines**
  - **Dual floating-point pipelines**
- **Motion video instructions**
- **Square root & divide**
- **64 KB I-cache -- 2-set**
- **64 KB D-cache -- 2-set**
- **CMOS6 -- 0.35u, 2.0V, 6LM**

SPECfp95 performance comparison

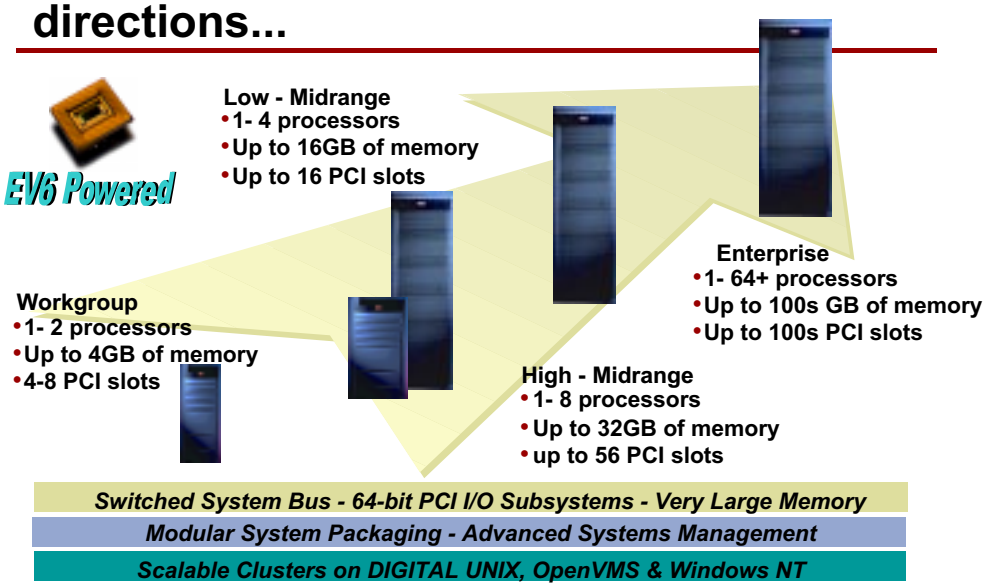


The system is the silicon... an evolution

- Reflects new philosophy of coupling chip design to system design
  - Optimize for the system
- Evolution from EV68
- EV7 - on-chip switches
  - Fewer, more powerful chips, another order of magnitude memory bandwidth
- EV8 - Symmetric multithread approach (SMT)
  - More efficient multi-execution units

digital™

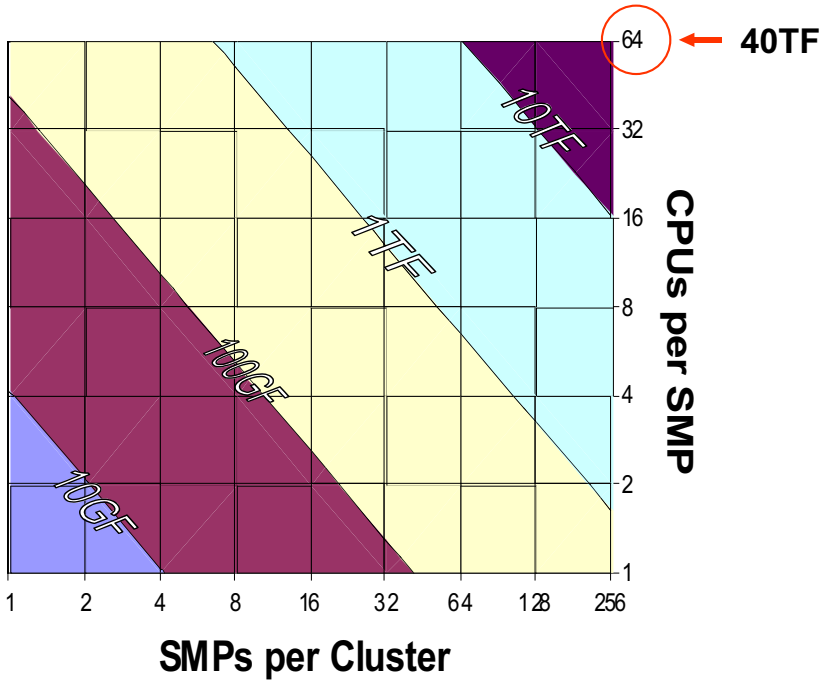
Alpha-based server directions...



digital™

Digital Plans a Very Strong Bid on ASCI Turquoise

- Alpha will remain the most capable CPU
- The bid will be based on EV7
  - EV7 will continue the trend towards increased FLOPS/CPU, and will again provide more memory bandwidth per FLOP
- The system architecture is flexible enough to allow aggregation through large SMPs and/or large node count clusters
- "In the same box" upgrade from our EV6 server
  - Same footprint transition from 10TF to 40TF systems



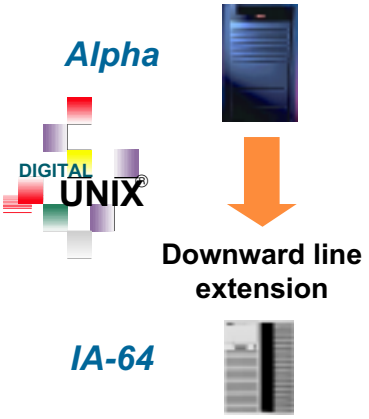
digital™

EV8 the basis for 100TF

- Chip in 2001
- Easily "on the curve"
- We can either deliver 100TF in a smaller footprint, or a few years early!

digital™

Strengthening DIGITAL UNIX



- Design center for DIGITAL UNIX will remain Alpha
- For customers, this means NO migrations -- can stay with DIGITAL UNIX on Alpha
- For ISVs, this means no porting required
- Provides broad base for DIGITAL UNIX to sustain investment in enterprise-class features

digital™

Enhanced DIGITAL's Windows NT

AlphaPowered™

MICROSOFT WINDOWS NT

intel inside

- On x86 and Alpha *today* - IA-32, IA-64 and Alpha *tomorrow*
- Alpha is *Microsoft's* development platform for 64-bit Windows NT
- Windows NT 5.0 with *64-bit VLM* addressing on Alpha in 1998 for high-end computing, plus FX32!
- Common Programming Model for 64-bit Windows NT on IA-64 and Alpha = *simple recompile* for applications between IA-64 and Alpha

digital™

digital™

Digital Equipment Corporation
© 1998
www.digital.com/info/hpc

To 100 TFLOPS
(*and beyond*)

Steve Oberlin
VP, Hardware, SGI/Cray
Salishan 98
4/21/98

© Cray Research and Silicon Graphics - Preliminary and Proprietary Information

Max System (1998)

- Postulate: ASCI Blue ~= Largest "practical" systems today
- 3 TFLOPS peak (~1 TFLOPS sustained on sPPM)
- 1.5 TB memory
- 75 TB rotating storage
- .5 Tb/s interconnect bisection BW

© Cray Research and Silicon Graphics - Preliminary and Proprietary Information

Scaling Drivers

- **Density:** Scales at "Moore's Law" rate: 2x every ~18 months.
- **Connectivity:** Scales at *slower* rate: 2x every ~26-30 months.

© Cray Research and Silicon Graphics - Preliminary and Proprietary Information

DRAM Memory Chip Performance

1979: Standard DRAM
- 16 Kbit
- 1-bit wide interface
- 5 Mb/s uniform access BW
- 2 Mb/s random access BW

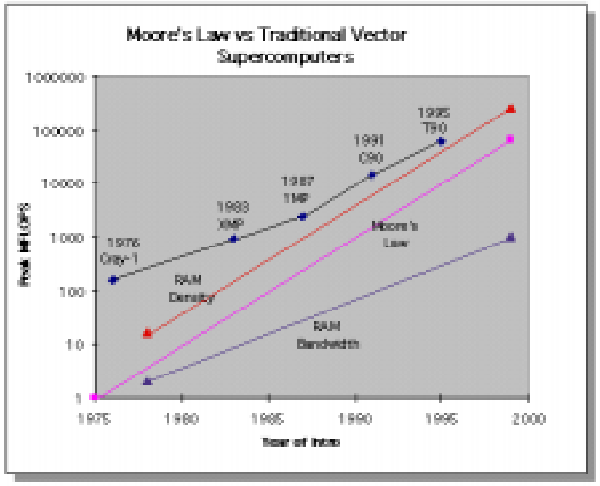
→

1999: 200 MHz SDRAM
- 256 Mbit
- 16 bit wide interface
- 3200 Mb/s uniform access BW
- 1000 Mb/s random access BW

✿ 1979 -> 1999:
- ✿ 16000X density increase
- ✿ 640X uniform access BW increase
- ✿ 500X random access BW increase
- ✿ 25X *less* per-bit memory BW

© Cray Research and Silicon Graphics - Preliminary and Proprietary Information

Connectivity-Limited Performance



© Cray Research and Silicon Graphics - Preliminary and Proprietary Information

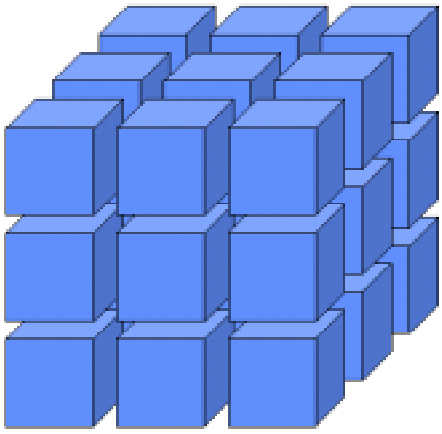
Overcoming Connectivity Limits

- Co-locate data and compute engines as much as possible
  - Distributed memory architecture and programming models
- Filter repetitive accesses to same or nearby data
  - Caches
- Increase node capability (volume) to minimize communications across packaging boundaries (surface)
  - SMP nodes

© Cray Research and Silicon Graphics - Preliminary and Proprietary Information

3D Scaling Effects

- Volume (FLOPS, memory) scales $O(N^3)$
- Surface (communication) scales $O(N^2)$
- Density/connectivity scaling disparity over 6 years: ~2.3x
- Scaling SMP node from 8 -> 128 processors reduces communications/FLOPS by ~2.5x



© Cray Research and Silicon Graphics - Preliminary and Proprietary Information

Complication: Stride-Independence

- Scientific computing requires good performance on variety of memory access patterns
  - Large stride
  - Scatter/gather
- Stresses system design in several dimensions:
  - Address generation bandwidth
  - Network single word transport efficiency
  - Data port efficiency on single words
  - Latency hiding (more on this later)

© Cray Research and Silicon Graphics - Preliminary and Proprietary Information

Example: NSA "Tabletoy"

System	Peak Performance (FLOPS)	PeakMemory Bandwidth	Time (seconds)	Normalized Speed
Triton T4	7.2 GFLOPS	100 GB/s	1.39	90.1
C90	8 GFLOPS	128 GB/s	2.84	44.1
Jedi J932	6.4 GFLOPS	38.4 GB/s	11.42	11
Origin 2000 32 PEs	12.5 GFLOPS	12.8 GB/s	125.26	1

Single processor performance, table size = 256 MB

© Cray Research and Silicon Graphics - Preliminary and Proprietary Information

Stride-Independence Strategies

- Direct architecture support
  - T90
  - J90
- Augmented communications
  - T3E
- Localize working set (cache blocking)
  - O2000
- Avoid global communications
  - Eschew implicit solvers

© Cray Research and Silicon Graphics - Preliminary and Proprietary Information

Attacking Latency

- Avoid it.
  - Caches
  - Large register sets
- Minimize it.
  - Fast components
  - Distributed memory
  - Single-system image (SSI)
  - Minimal diameter interconnect network
  - Small physical size (more on this later)
- Hide it.
  - Large number of outstanding memory references
  - Cache prefetching
  - Large register set

© Cray Research and Silicon Graphics - Preliminary and Proprietary Information

Synchronization

- Necessary every time you communicate…
- Different algorithms may call for different types of synchronization
  - Gang (barrier)
  - Producer-consumer
- Synchronization latency is (usually) dead time.
- Scalable synchronization strategies:
  - Hardware support
  - Pipeline when you can

© Cray Research and Silicon Graphics - Preliminary and Proprietary Information

More Scaling Issues

- **Reliability:**
  - Inversely proportional to how much HW/SW is tied together.
- **Structure:**
  - Physical configuration can limit bandwidth, increase latency
  - Scaled configuration compatibility with existing infrastructure

© Cray Research and Silicon Graphics - Preliminary and Proprietary Information

Reliability Issues

- More connections, more components, and more complexity mean increased probability of failure.
- Increasing scale decreases MTBF
  - Components that "work" for small configs are inadequate at large scales.
- Hardware MTBF *falling* each generation as scale increases.

© Cray Research and Silicon Graphics - Preliminary and Proprietary Information

Reliability Strategies

- **Increase component size, not numbers.**
  - Example: Muffin fans vs. big blowers
- **Increase quality, design margins**
- **Design for resilience, redundancy:**
  - Extra compute nodes (e.g. T3E)
  - Redundant components (e.g. T90 memory chips)
  - Multiple paths (e.g. O2000, T3E)
  - Concurrent maintenance
  - *Requires unified HW, SW, Service, Sales strategy to work.*

© Cray Research and Silicon Graphics - Preliminary and Proprietary Information

System Structure Issues

- **Making it smaller (faster) increases thermal density.**
  - CMOS: Power is frequency dependent.
  - VLSI: More on a chip
  - MCMs: More on a substrate
  - High-I/O counts: Driving lots of transmission lines takes power.
- **Complications:**
  - Higher temperature = lower speed, lower reliability
  - High thermal gradients pushing technology limits
  - Exotic cooling drives up cost

© Cray Research and Silicon Graphics - Preliminary and Proprietary Information

Structural Strategies

- Only make dense what you must.
  - Minimize latency-sensitive paths: e.g. processors <-> caches
  - Maximize wiring bisection for maximum bandwidth
  - Don't bother with dense DRAM packaging.
- Use liquid where you must.
- Scale structurally
- Fit into your natural environment (high end systems go into computer rooms)

© Cray Research and Silicon Graphics - Preliminary and Proprietary Information

Balance

- Attention to detail: Everything is important:
  - Processor
  - Memory
  - Interconnect
  - I/O
  - ...
  - *No "weak links" allowed*

© Cray Research and Silicon Graphics - Preliminary and Proprietary Information

Max "Practical" System (2004)

- 2004 - 1998 = 6 years
- Density improvement: ~= 16x at "standard" Moore's Law rate
- Connectivity improvement: ~= 3.5x
- 48 TFLOPS peak, ~16 TFLOPS Sustained
- 72 TB memory
- 1.2 PB rotating storage
- 3.5 Tb/s interconnect bisection BW

© Cray Research and Silicon Graphics - Preliminary and Proprietary Information

Issues

- 100 TFLOPS in 2004 will cost ~2x ASCI Blue
- *Will it do what you want?*
  - Drive industry with benchmarks that really represent your codes.

© Cray Research and Silicon Graphics - Preliminary and Proprietary Information

ASCI Academic Strategic Alliances Program (ASAP) Level 1 University Centers

The Conference on High Speed Computing

Ann Hayes
LANL
4/21/98

6/5/98 1

Outline

- **What/why is ASCI ASAP?**
- **Selection process for Alliance Centers**
- **Discussion of each Center**
  - » **overarching application**
  - » **major thrust areas**
  - » **some first year achievements**

6/5/98 2

ASCI ASAP Strategy

- **Motivation**
  - » **ASCI simulation and computing problems are sufficiently difficult and complex that Labs cannot solve them alone,**
  - » **develop a broad consensus that simulation is an appropriate means of ensuring confidence in the safety, performance and reliability of the stockpile,**
  - » **help train the next generation of stockpile stewards.**

6/5/98 3

Alliances Structure

- **Three Levels in Program**
  - » **Strategic Alliances (Level 1)**
  - » **Strategic Investigations (Level 2)**
  - » **Individual Collaborations (Level 3)**

6/5/98 4

Level 1 - Strategic Alliances

- large, 5-10 yr. relationships between Labs and universities to achieve broad milestones
- support confidence in simulation
- multidisciplinary, coordinated effort with one *open* overarching application for each Center
- 5 Centers established
- ramp up funding to $4M/yr/Center
- no experiments will be funded through Alliances funding

5/5/98 5

Center Review Criteria (per RFP)

- *applicability*, including relevance to ASCI goals and objectives
- *feasibility*, including roadmap and simulation validation
- *capabilities* (personnel and facilities)
- *cost* /budget

6/5/98 6

Tri-Laboratory Oversight/Management

- **Trilab Sponsor Team (TST)**
  - » **each Center will have a TST**
  - » **comprised of one application and one CS representative/Lab, contract representative, Alliances Strategy Team member**
  - » **assure ASCI apps and CS/CM goals are being met in Center**
  - » **serve as liaison between Labs and Centers to foster useful collaborations**

6/5/98 7

Tri-Laboratory Oversight/Management, cont'd.

- **all Centers allocated 2% of each Lab ASCI computing system + usage at Pittsburgh Supercomputer Center**
- **Computer Resource Team (CRT)**
  - » **comprised of 1 person/Lab, 1/Center**
  - » **establish uniform access mechanisms to machines**
  - » **develop resource allocation plan**
  - » **provide technical aid in connection, usage issues; network support; software questions**

6/5/98 8

Center Characteristics

- single unifying application
- multidisciplinary
- multiple space and time scales
- tera-scale computing requirements
- at least 40% of effort devoted to advances in computational simulation
- validation component
- relevant to ASCI programmatic goals

6/5/98 9

Five Universities Designated ASC Academic Strategic Alliance Cente

The University of Chicago
Center on Astrophysical
Thermonuclear Flashes

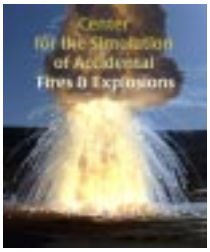
Center for Simulation of Advanced Rockets

University of Illinois at Urbana-Champaign

California Institute of Technology

A Facility for Simulating the Dynamic Response of Materials

Stanford University
Center for Integrated Turbulence Simulations (CITS)



Academic Strategic Alliance Program University Involvement

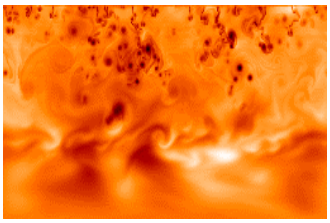
	CalTech	Stanford	Utah	Illinois	Chicago	Total
Faculty	8	26	22	42	13	**111**
Post-Docs/ Research Staff	19	8	17	17	8	**69**
Graduate Students	12	10	14	29	6	**71**
Professional Staff	3	4	5	3	2	**17**
Total	**42**	**48**	**58**	**91**	**29**	**268**

6/5/98 11

University of Chicago Center on Astrophysical Thermonuclear Flashes

Goal

Solve the long-standing problem of thermonuclear flashes (nuclear burning) on the surfaces of compact stars such as neutron stars and white dwarfs, and in the interior of white dwarfs (i.e., Type I supernova).



6/5/98 12

University of Chicago Center on Astrophysical Thermonuclear Flashes
Major Research Interests

- **astrophysics - develop physics modules for FLASH code**
  - » **ignition problem**
  - » **“flame” propagation problem**
  - » **deflagration-detonation transition problem**

6/5/98 13

University of Chicago Center on Astrophysical Thermonuclear Flashes
Computer Science Objectives

- **Implementation of one-sided MPI-2 operations**
  - – **can allow new numerical algorithms**
- **Integration of U. of C., ANL, ASCI sites via Globus**
- **ASTRO3D code ported to MPI and run on all ASCI machines**
- **Multiresolution volume rendering of ASTRO3D datasets**
- **Exploration of standard data formats**

6/5/98 14

University of Chicago Center on Astrophysical Thermonuclear Flashes
ANL Computer Science Role

FLASH code:

- simulate matter accumulation on the surface of compact stars, nuclear ignition of the accumulated (and stellar) matter, and the subsequent evolution of the star's interior, surface, and exterior. (Based on U. of Chicago's existing 3D multipurpose code ASTRO3D for modeling the above phenomena).

6/5/98 15

University of Chicago Center on Astrophysical Thermonuclear Flashes
First Year Accomplishments

- **4 new computational sciences introduced into the U. of Chicago curriculum**
- **major portions of MPI-I/O implemented on all ASCI machines**
- **ASTRO3D output in CAVE and visualization of 256^3 datasets using new volume-rendering algorithms**
- **ASTRO3D ported to all ASCI machines and run on small number of processors-larger runs anticipated soon**

6/5/98 16

University of Utah Center for Simulation of Accidental Fires and Explosions (CSAFE)

Goal

Develop technical capability to simulate accidental fires and explosions involving hydrocarbons, structures, containers, and high energy materials, within context of handling and storage of highly flammable materials

- will not include any atomic devices or detonation on impact or destruction after impact (emphasis is on fires)

- focus of this Center is on engineering

6/5/98 17

University of Utah Center for Simulation of Accidental Fires and Explosions (CSAFE)
Major Research Interests

- **C-SAFE product for the simulation of accidentalfundamental chemistry and physics fully integrated into problem solving environment**
  - **non-linear solvers, optimization, computational steering, and visualization**
  - **verified accuracy through comparison with detailed experimental results**
- **Selected sample simulations with comparison to actual data for targeted experiments**

6/5/98 18

University of Utah Center for Simulation of Accidental Fires and Explosions (CSAFE)

Computer Science Objectives

- **SCIRun - environment using Overture interactions**
- **high end visualization using multiple Infinite Reality pipes**
- **Overture evaluation for application development**

SCIRun Problem Solving Environment

6/5/98 19

University of Utah Center for Simulation of Accidental Fires and Explosions (CSAFE)

First Year Accomplishments

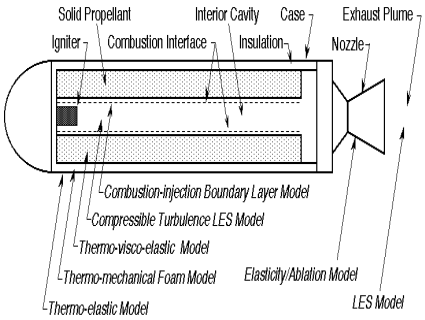
- **work progressing on getting SCIRun onto computational scientists' desks, improving its robustness and incorporating performance tools into it**
- **initial validation experiments identified and required experimental facilities being designed. Matched experiments will be created in U. of U. pilot scale combustion lab and at Thiokol Corp. - all will perform detailed measurements for fires involving transient burning of hydrocarbons in conjunction with a HE device**
- **80% of key faculty /student positions have been filled**

6/5/98 20

U. of Illinois Center for Simulation of Advanced Rockets

Goal

Focus on detailed whole-system simulation of solid propellant rockets under both normal and abnormal operating conditions



6/5/98 21

U. of Illinois Center for Simulation of Advanced Rockets

Major Research Interests

- structural and solid mechanics
- fluid flow simulation
- combustion and energetic materials
- molecular and materials simulation

6/5/98 22

U. of Illinois Center for Simulation of Advanced Rockets

Computer Science Objectives

- joint development of threads libraries for application with array classes
- particle classes for aluminum particles in rocket fuels
- Overture geometry and AMR capabilities
- develop grid-based methods for linear solvers of large linear systems
- system integration code to link components

6/5/98 23

U. of Illinois Center for Simulation of Advanced Rockets

First Year Accomplishments

- Gen 0 code designed and being developed (prototype code to validate ideas for Gen 1 and Gen 2)
- incorporation of automated performance optimization into Panda parallel I/O library
- scalable I/O instrumentation and characterization, with VR based visualization, applied to ASCI codes (e.g., DYNA3D, ALE3D)
- extension of ParaGraph performance visualization tool to support full functionality of MPI
- development of mesh generation, partitioning tools with dynamic load balancing and grid-based methods for partitioning and solution of sparse linear systems)

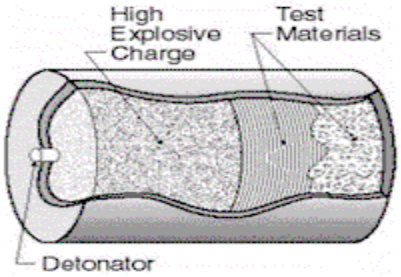
6/5/98 24

CalTech Virtual Facility for Simulation of Dynamic Response in Materials

- **Goals**
  - » **facilitate computation of a variety of experiments with various materials in which strong shock and detonation waves impinge on targets**
  - » **compute subsequent dynamic responses of target materials**
  - » **validate computations against experimental data**

6/5/98 25

CalTech Virtual Facility for Simulation of Dynamic Response in Materials
Facility Prototype



Virtual Facility

Virtual shock physics facility where 3D response of a variety of target materials can be computed for wide ranges of loadings, such as produced by detonation of energetic materials

6/5/98 26

CalTech Virtual Facility for Simulation of Dynamic Response in Materials
Major Research Interests

Will perform work related to the three stages required to conduct a virtual experiment:

- detonation of high explosives

- interaction of shock waves with materials

- shock-induced compressible turbulence and mixing

6/5/98 27

CalTech Virtual Facility for Simulation of Dynamic Response in Materials
Computer Science Objectives

- **optimization of I/O rates to match computational speeds (SIO)**
- **PSE development to facilitate use of virtual test facility across distance computing resources**
- **high performance PC clusters**
- **Globus metacomputing toolkit provides fundamental infrastructure on distributed applications**
- **AMRITA initial coordination and composition framework for CFD applications**

6/5/98 28

CalTech Virtual Facility for Simulation of Dynamic Response in Materials

First Year Accomplishments

- **high explosives**
  - » **reduced model for hydrogen-oxygen combustion has been formulated**
  - » **detonation initiation prediction comparison between LANL engineering model with AMRITA and Caltech calculations**
- **solid dynamics**
  - » **3D finite element AMR capability using tetrahedra being applied to shock wave computation in solid materials**
  - » **Equation of State tables for metal Tantalum obtained-plan to use for impact calculations**

6/5/98 29

CalTech Virtual Facility for Simulation of Dynamic Response in Materials

First Year Accomplishments, cont'd.

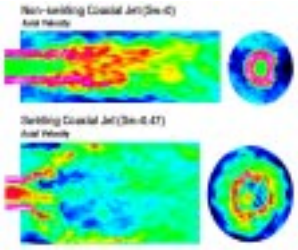
- **materials properties**
  - » **S-threads package used to parallelize a molecular dynamics (MD) code used by Mat. Prop. group-speedup results look good**
  - » **atomistic impact calculations with up to 10^6 atoms performed using Caltech MPSim MD code**
- **compressible turbulence**
  - » **Rayleigh-Taylor mixing of miscible fluids**
  - » **3D shock-contact interactions calculated on PSC T3E-used in 3D shock/vortex and shock/contact interactions**

6/5/98 30

Stanford Center for Integrated Turbulence Simulations (CITS)

Goal

Develop new turbulence models and associated numerical simulation methodologies to enable new paradigm for design of advanced systems where turbulence plays a controlling role. Use this to drive Center's development of parallel scientific computing, software, systems, and architectures. Major application is aircraft gas turbine engine.



6/5/98 31

Stanford Center for Integrated Turbulence Simulations (CITS)
Major Research Interests

- **turbomachinery**
  - » **RANS turbulence modeling - develop models for discrete-hole turbine blade film cooling; develop models for transition and relaminarization in turbines and compressors**
- **combustor**
  - » **gas-phase combustion**
  - » **spray modeling**
- **integrated full engine flow simulation**

6/5/98 32

Stanford Center for Integrated Turbulence Simulations (CITS)
Computer Science Objectives

- **SimOS - simulator for multiple CPU architectures to study machine performance**
- **SUIF compiler and associated development environment (goal is scalability)**
- **computer architectures**
  - » **characterization**
  - » **alternative design**
  - » **high performance interconnects**
- **numerical linear algebra algorithms**

6/5/98 33

Stanford Center for Integrated Turbulence Simulations (CITS)
First Year Accomplishments

- SUIF Explorer used to parallelize 13,000 line program yielding greater than 4x improvement
- strained homogeneous turbulence in a rotating frame being run on Sandia ASCI computer. Producing key archival data needed for turbulence model development
- new fully conservative unstructured mesh algorithm (2D, 3D) developed for Navier-Stokes eqn.; demonstrated in 2D
- scaling laws derived for optimal loan distribution in parallel simulation of particle-field (gas turbine spray combustor) eqn.
- completed first “implementation-oriented” study of scalability of coherence protocols.

6/5/98 34

Issues for the Future

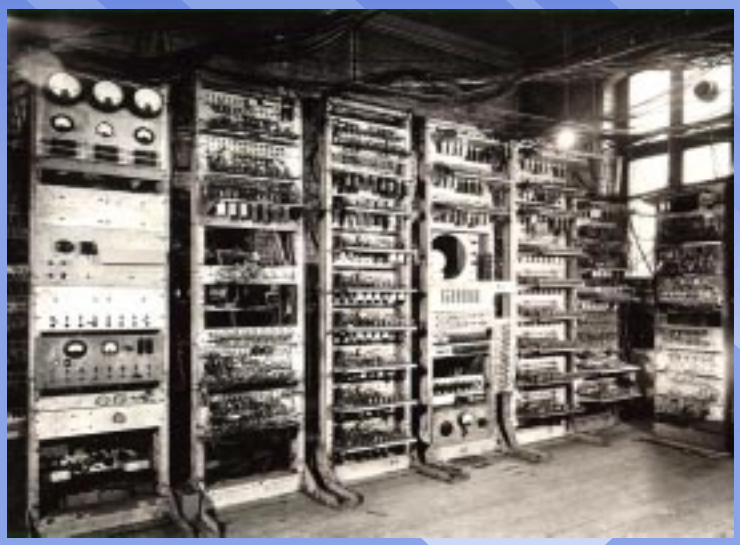
- achieve balance between university/Lab collaboration efforts and allowing Centers to fulfill the goals set forth in overarching applications
- successful validation of simulations

6/5/98 35

The Next 50 Years of Computer Architecture

Burton Smith
Tera Computer Company

50 Years Ago



What Is Driving Computer Architecture?

- Requirements of the marketplace
- Architectural understanding
- Technology imperatives and opportunities
- Recurring and non-recurring costs
- Fashion

Marketplace Requirements in 1998

- Support the "killer apps"— word processing, spreadsheets, etc.
- Deliver high performance for multimedia applications, e.g. games
- Implement decent SMP and NUMA multiprocessor support
- Simplify hardware and software configuration management
- Reduce power consumption
- The trends:
  - » PCs are driving the market; penetration is over 50%
  - » PC prices are dropping precipitously

This list reads like a recipe for stagnation.

Architectural Understanding in 1998

- Prevalent wisdom calls for:
  - » concurrent operation issue to multiple functional units
  - » a deep hierarchy of instruction and data caches
  - » branch prediction with respectable accuracy
  - » some degree of speculative execution
- Latency tolerance is also important, but it is unclear what mix of prefetching, vector pipelining, and multithreading is best
- Massive parallelism is decidedly not general purpose parallelism, but instruction-level parallelism decidedly is

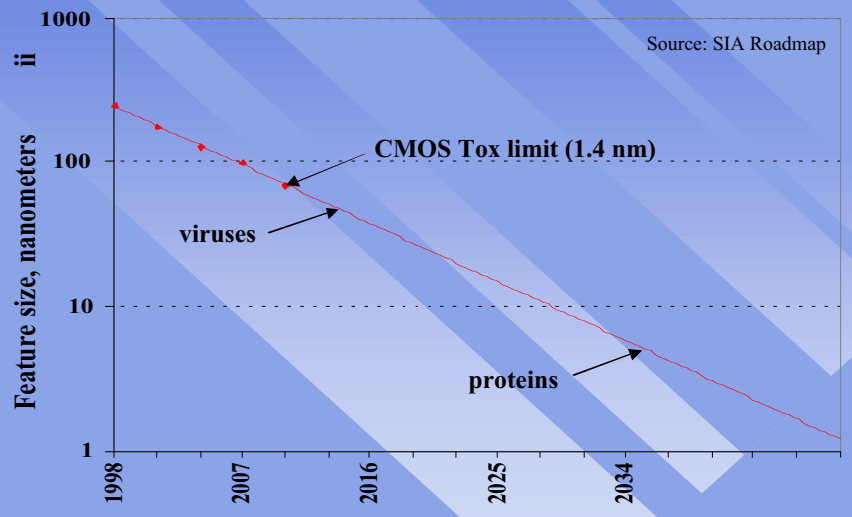
Architectural Trends in 1998

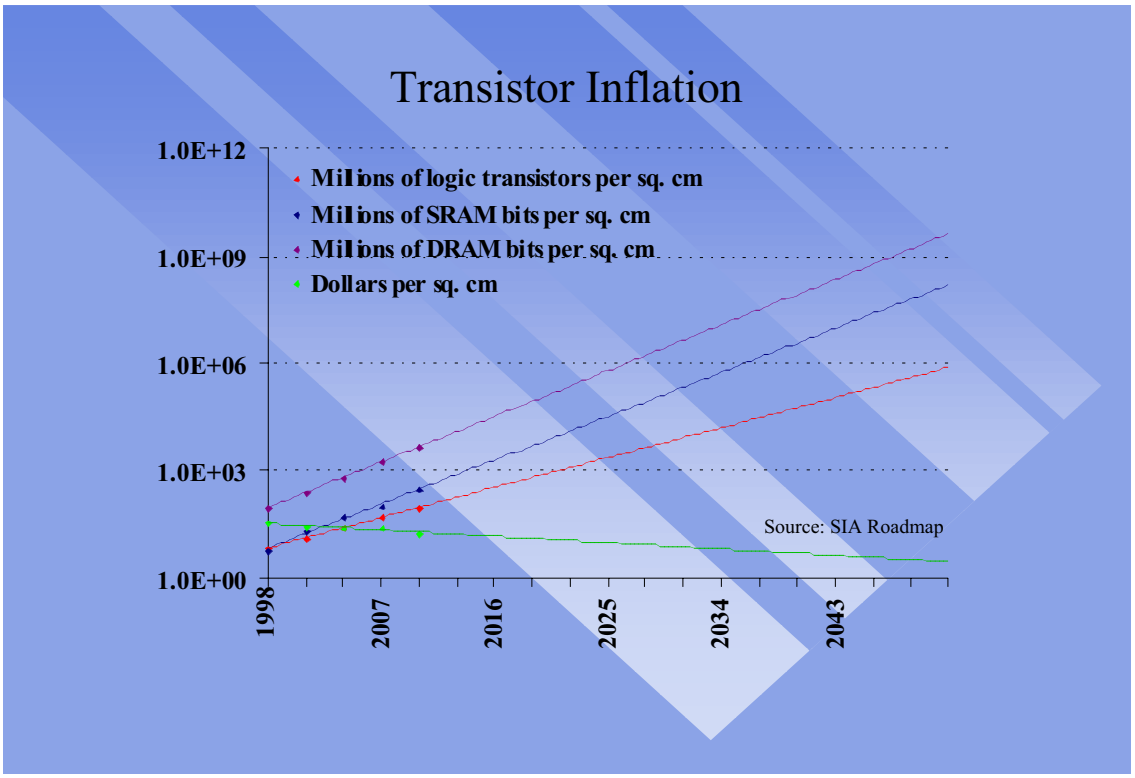
- Memory latency is a narrowing bottleneck, even for uniprocessors
- Instruction-level parallelism (ILP) is increasingly relied upon:
  - » to feed more and more functional units
  - » to feed deeper and deeper pipelines
- System responsiveness fails to improve in spite of faster clock rates
  - » because (# registers) $\propto$ (ILP used) $\geq$ (pipe depth) $\propto$ (clock rate)
- Meanwhile, clients and servers are diverging architecturally

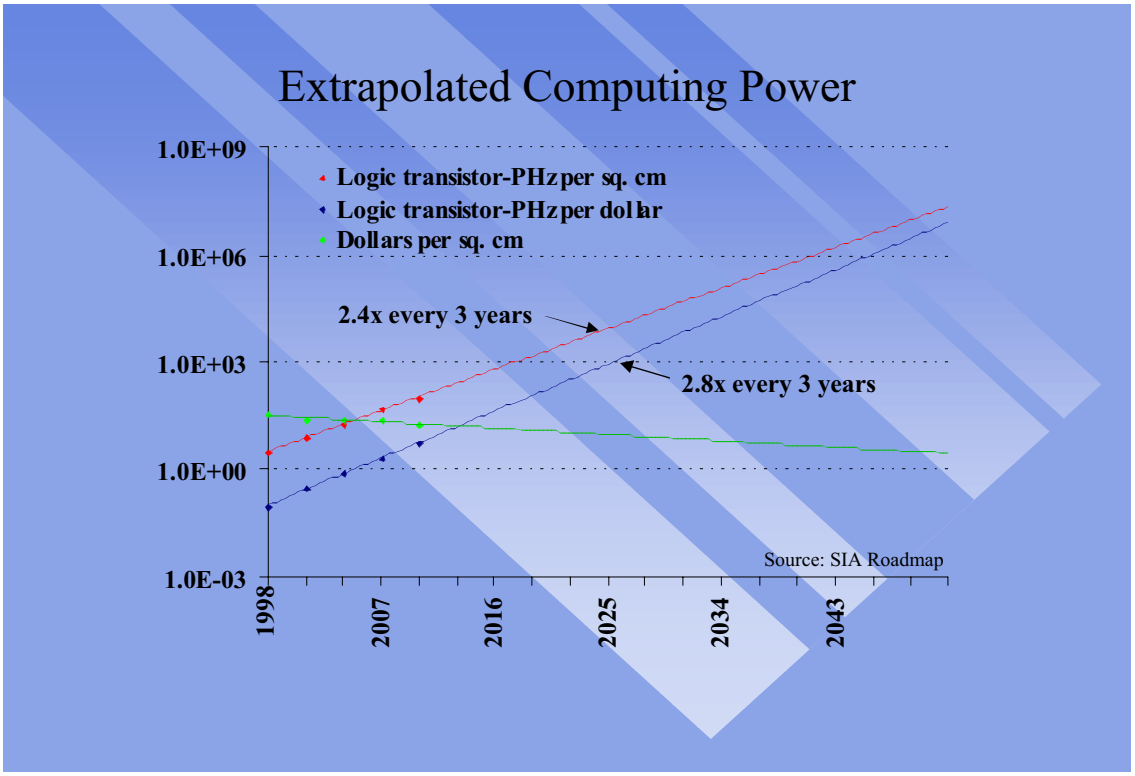
Clients and Servers are Diverging

Attribute	Clients	Servers
Display	megapixel, 24 bit color	none
Graphics capabilities	Instructions, coprocessor	none
Audio	sound card, spkrs, mike	none
Other interfaces	keyboard, mouse, camera	none
Disks	one	RAID
Operating system	Windows NT	“UNIX”
Multiprocessing	heterogeneous	homogeneous
Parallelism type	functional	data

Extrapolated Feature Size




Transistor Inflation
1.0E+12
1.0E+09
1.0E+06
1.0E+03
1.0E+00
Millions of logic transistors per sq. cm
Millions of SRAM bits per sq. cm
Millions of DRAM bits per sq. cm
Dollars per sq. cm
Source: SIA Roadmap
1998
2007
2016
2025
2034
2043


Extrapolated Computing Power
1.0E+09
1.0E+06
1.0E+03
1.0E+00
1.0E-03
Logic transistor-PHz per sq. cm
Logic transistor-PHz per dollar
Dollars per sq. cm
2.4x every 3 years
2.8x every 3 years
Source: SIA Roadmap
1998
2007
2016
2025
2034
2043

A Wild Guess at Technology in 2048

- Single-electron transistors a few nanometers on a side
- Trillions of devices per die
- Frequencies above 10 GHz
- Wafer-scale multiprocessors
- Optical interconnect, both on-wafer and off
  - » Implemented micromechanically
- Some circuit construction using biotechnology
- Quantum coprocessor support

Consequences of Continued Transistor Inflation

- Ubiquitous, large-scale parallel computing
  - » coprocessors for graphics, speech, vision, chess, …to a point
  - » homogeneous multiprocessors of ever-increasing scale
- Transistor-based companies must refocus periodically — or die
  - » No market can outgrow Moore's Law for long
  - » Scalability and software bloat can prolong a market's growth
- The parallel programming problem looms large
  - » How should parallel programs be written?
  - » How can we get a measure of independence from system scale?
  - » How can architecture help address these issues?

The Move to Multithreading

- Multithreading solves three important architectural problems:
  - » it deals with memory latency in a scale-invariant way
  - » it reconciles fine-grain synchronization with high performance
  - » it enables virtualization of the multiprocessor resource
- In a few years, nearly all processors will be multithreaded
  - » multiprocessors will become easier to use "just in time"
- As fallout, there will be major changes in:
  - » cache architecture
  - » branch prediction
  - » speculative execution

Example: Possible Changes in Cache Architecture

- Using caches to reduce bandwidth (moreso than latency)
- Compiler cache management
  - » deciding whether to cache a value at all
  - » what cache sector size to use
  - » whether the value needs to be written back or not
  - » lazy "epochal" invalidation
- Transactional cache
  - » for management of speculative state
  - » for atomic updates involving multiple variables

The Bandwidth Market

- In a conservative system, concurrency = latency times bandwidth
- Multithreading can supply lots of memory system concurrency
  - » even to the point of implementing uniform shared memory
- Bandwidth (not latency) limits practical system size
- The issue is the cost of a given amount of bandwidth
  - » How much bandwidth is enough?
  - » The answer pretty clearly depends on the application
- Systems will be differentiated by their bisection bandwidths
  - » System purchasers will buy the class of bandwidth they need
  - » System vendors will make sure their bandwidth scales properly

Reducing the Cost of Bandwidth

- Use on-wafer wires whenever possible
- Use optics for long-range interconnect
  - » Leverage communication technology whenever possible
- Use direct interconnection network topologies
  - » Indirect networks waste connections
- Use low-diameter graphs of moderate degree
  - » Bisection bandwidth per node $\propto$ degree /average distance

Trying to keep up with transistor inflation is not easy.

Largest Known Graph Size with Diameter ≈ Degree

Degree	Diameter = Degree	Diameter = Degree+1
3	20	38
4	95	364
5	532	2,734
6	7,817	13,056
7	35,154	93,744
8	234,360	820,260
9	3,019,632	15,686,400

Source: Bermond, Delorme, and Quisquater, JPDC **3** (1986), p. 433

Market Needs in 2048

- Support the "killer apps" — Polyglot, LA1986, etc.
- Improve exosurrogate nanoanimation
- Provide decent interfaces for VFSE transimagery
- Interconnection networks for larger scale GL_0 systems
- Simplify hardware and software configuration management
- Reduce power consumption
- The trends:
  - » HRs are driving the market; penetration is over 50%
  - » HR prices are dropping precipitously

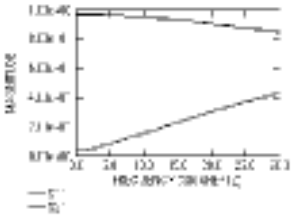
This list still reads like a recipe for stagnation.

Full Wave Modeling of Signal and Power Interconnects

Gary Haussmann
Melinda Piket May
University of Colorado Boulder

Kevin Thomas
Cray Research/Silicon Graphics

Modeling, Simulation,



Visualization, and Analysis

Presentation Outine

Background

- *FDTD Simulation Method*
- *LC Interface*
- *Analysis Capabilities*

Models and Simulations

- *Microstrip/SPICE Simulation*
- *CHAMP Signal Connector*
- *Photonic Bandgap Filters*
- *Power Distribution Array*

LC
An Electromagnetic Modeling Tool

- Integrated 3D model editing, simulation, visualization, and analysis features

- Based on FD−TD simulation

  Finite−Difference Time−Domain

- Developed for electrical interconnect modeling

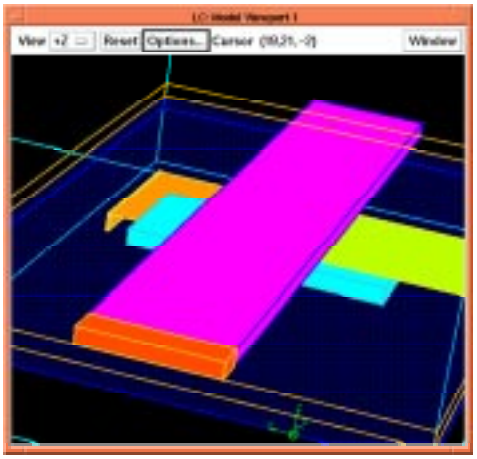
  Some newer features for radiation problems

Model Viewport

- View Model dialog
- Several may be used at the same time
- Pointer modes are used to manipulate the model

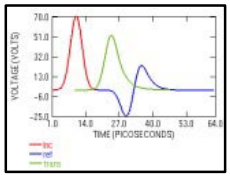
  Rotate
  Translate
  Magnify

- Plane probe visualization

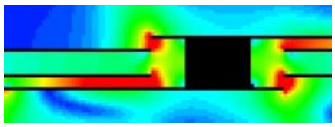


Plot Probes Dialog

- Probes may be plotted together



- Plane probes can be played back



- Probe data must be saved during the simulation

Plot Pulses Dialog

- Pulses can be plotted together
- Probe quantities can be plotted
  - Voltage
  - Current
  - Charge
  - Magnetic Flux
- Derived quantities can be plotted
  - Impedance
  - Power
  - Capacitance
  - Inductance
- Time domain or frequency domain

Define Sweeps Dialog

- Radiated power plots (RCS)
- Look angle sweep

  Phi angle from the positive X−axis in the X−Y plane (azimuth)

  Theta angle from the positive Z−axis (altitude)
- Frequency sweep
- Frequencies must agree with the Far Field Parameters Frame in the Define Model Parameters Dialog
- Must be created before the simulation is run

Other Features

- Direct editing of the FD−TD mesh
- SPICE circuit interface
- Algebraic manipulation of probe data

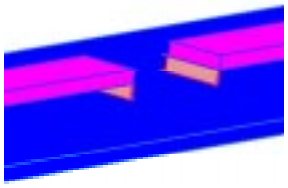
  Calculations
  Plot Calculator
- Other calculations derived from pulses

  S−Parameters
  Step Pulses dialog
- CAD data readers

  Gerber photoplotter
  NC drill files
  AutoCAD

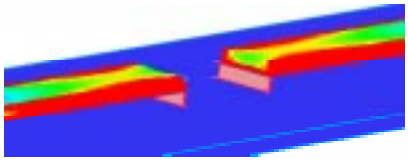
Multiport SPICE Circuit Interface

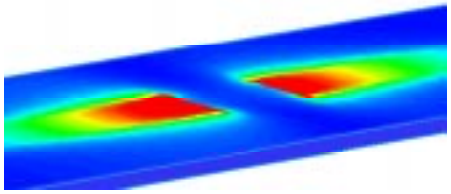
Microstrips Joined With Impedance−Matched SPICE Circuit



SPICE interface ports at the ends of the microstrips

Surface currents on the microstrips



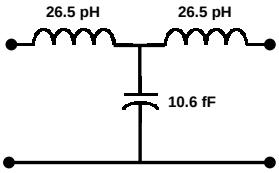


Surface currents on the ground plane

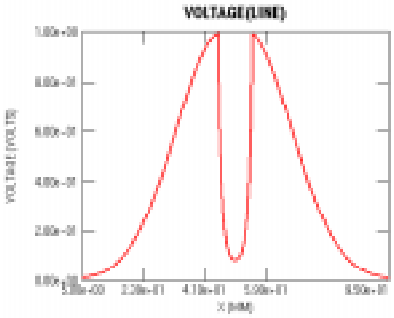
April 1998 Electromagnetic Analysis With LC http://www.cray.com/lc/

Multiport SPICE Circuit Interface

Microstrips Joined With Impedance−Matched SPICE Circuit



Impedance−matched SPICE circuit

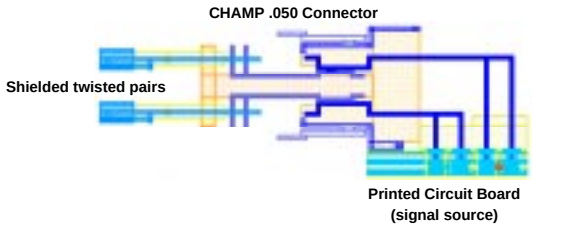


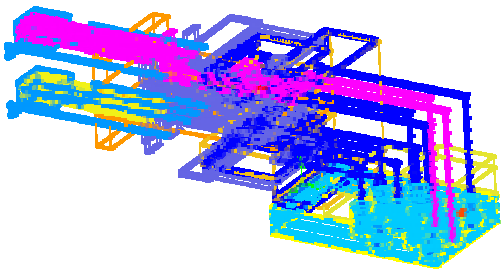
Voltage beneath the microstrips

Currents on the microstrips

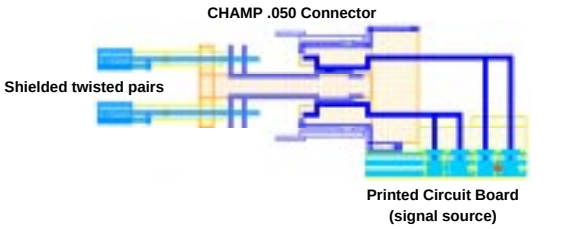
April 1998 Electromagnetic Analysis With LC http://www.cray.com/lc/

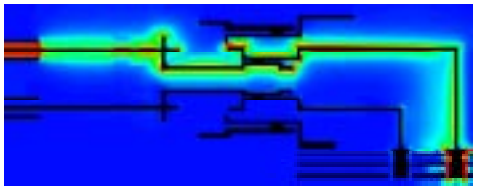
Electrical Connector Modeling For Gigabyte SCI Interface





Electrical Connector Modeling For Gigabyte SCI Interface



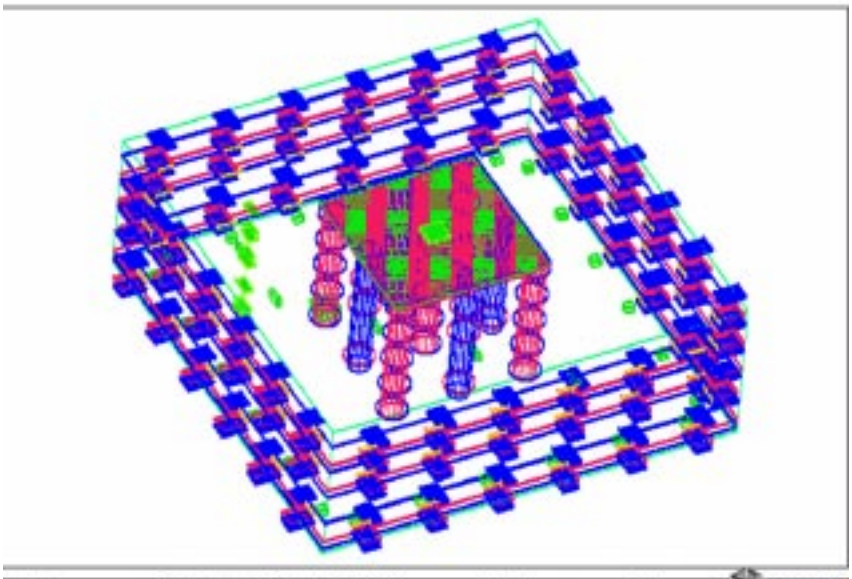


Electric field intensity after 400 picoseconds

April 1998 Electromagnetic Analysis With LC http://www.cray.com/lc/

Power Distribution Analysis

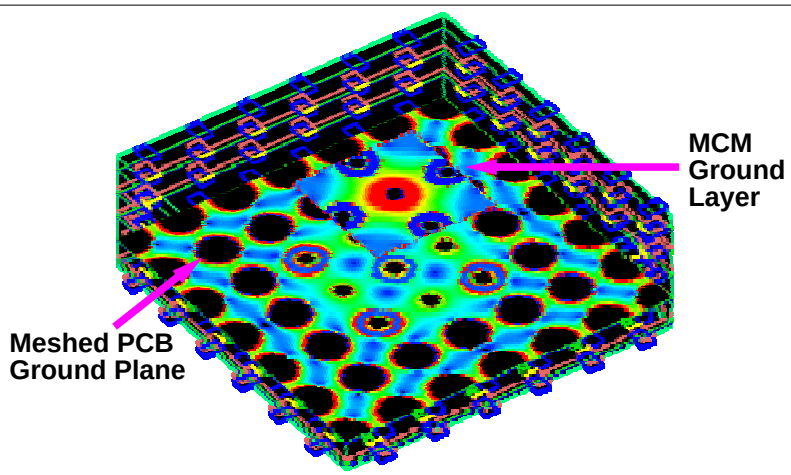
Printed Circuit Board With Multi–Chip Module Load



April 1998 Electromagnetic Analysis With LC http://www.cray.com/lc/

Power Distribution Analysis

Printed Circuit Board With Multi–Chip Module Load

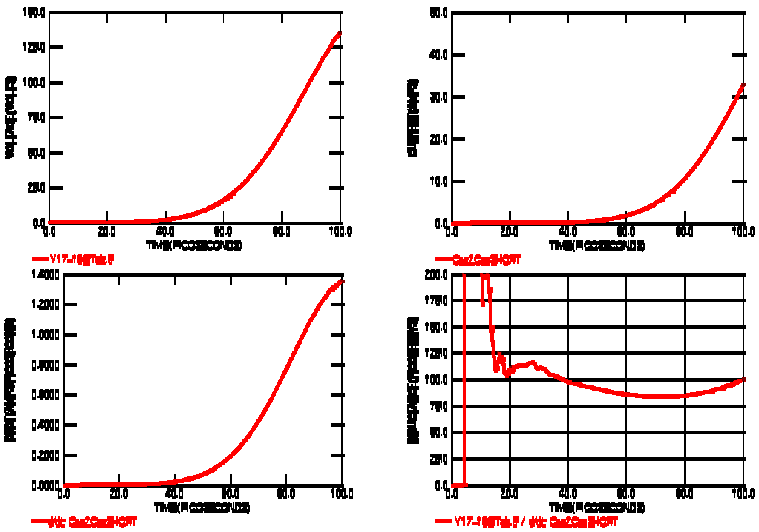


Current density in the PCB (source) and MCM (load) ground layers

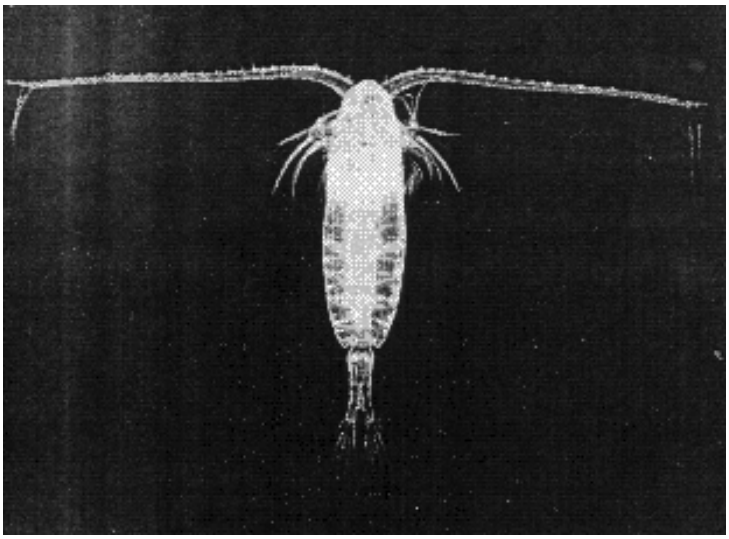
April 1998 Electromagnetic Analysis With LC http://www.cray.com/lc/

Power Distribution Analysis

Printed Circuit Board With Multi−Chip Module Load



Physical-Biological Simulations of the Abundance and Distribution of *Calanus finmarchicus* in the Gulf of Maine-Georges Bank Region



For more information regarding this talk, please contact
Wendy Gentleman
HB 8000
Thayer School of Engineering
Dartmouth College
Hanover, NH 03755
Phone (603) 646-2685
Fax (603) 646-3856
Email wendy.gentleman@dartmouth.edu

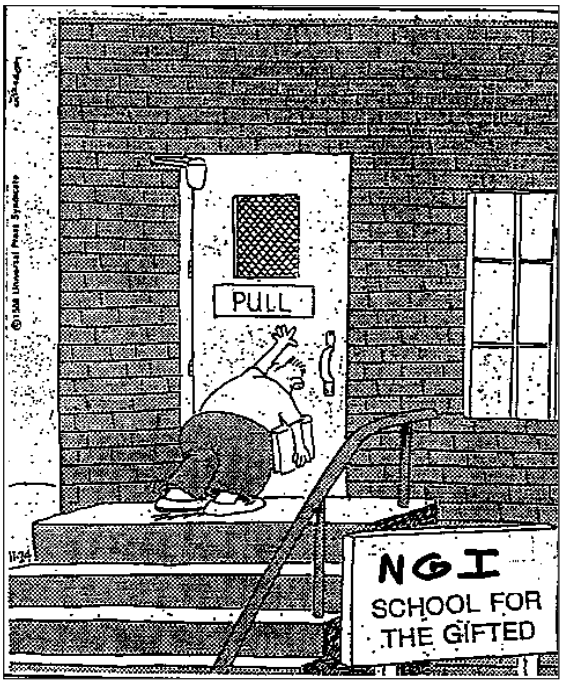
Next Generation Internet - NGI

Bob

Aiken

DOE

aiken@er.doe.gov



NGI Goals

- Goal 1 : Experimental R&D for Advanced Network Technology (DARPA lead)
  - Network growth and engineering
  - Quality of Service (QOS)
  - Security
- Goal 2: NGI Testbeds (NSF, NASA, DOE [99] lead)
  - High performance connectivity (100 megabits/sec +)
  - Ultra high speed networks (1 gigabit/sec +)
- Goal 3: Revolutionary Applications (Domain driven)

Experimental Research for Advanced Network Technologies

- Network Growth and Engineering
  - planning and simulation
  - monitoring, control, analysis, and display
  - integration
  - data delivery
  - managing lead user infrastructure

Experimental Research for Advanced Network Technologies

- Quality of Service (QOS)
  - baseline QOS architecture
  - drill down technologies
  - next generation QOS technologies
  - admission control, accounting,

Experimental Research for Advanced Network Technologies

- Security
  - cryptographic technologies and applications (e.g. PKI, inter-networks)
  - security criteria, test methods, tests

NGI Testbeds

- High performance connectivity
  - end-to-end 100 megabits/sec and up
  - 100+ Universities, Labs, and Federal Centers
  - support goals 2 and 3
  - gigapops and interconnect / peering points
- Ultra high speed networks
  - end-to-end gigabits/sec and up
  - 1000x network management tools / capabilities
  - (de) aggregation of high speed tributaries

Revolutionary Applications

Discipline Affinity Groups:

- Medicine
- Crises Management
- Basic Sciences
- Education
- Environment
- Manufacturing
- Federal Services

Technology Affinity Groups:

- Distributed Computing
- Remote Operation
- Digital Libraries
- Collaboratories
- Privacy / Security

NGI BUDGET

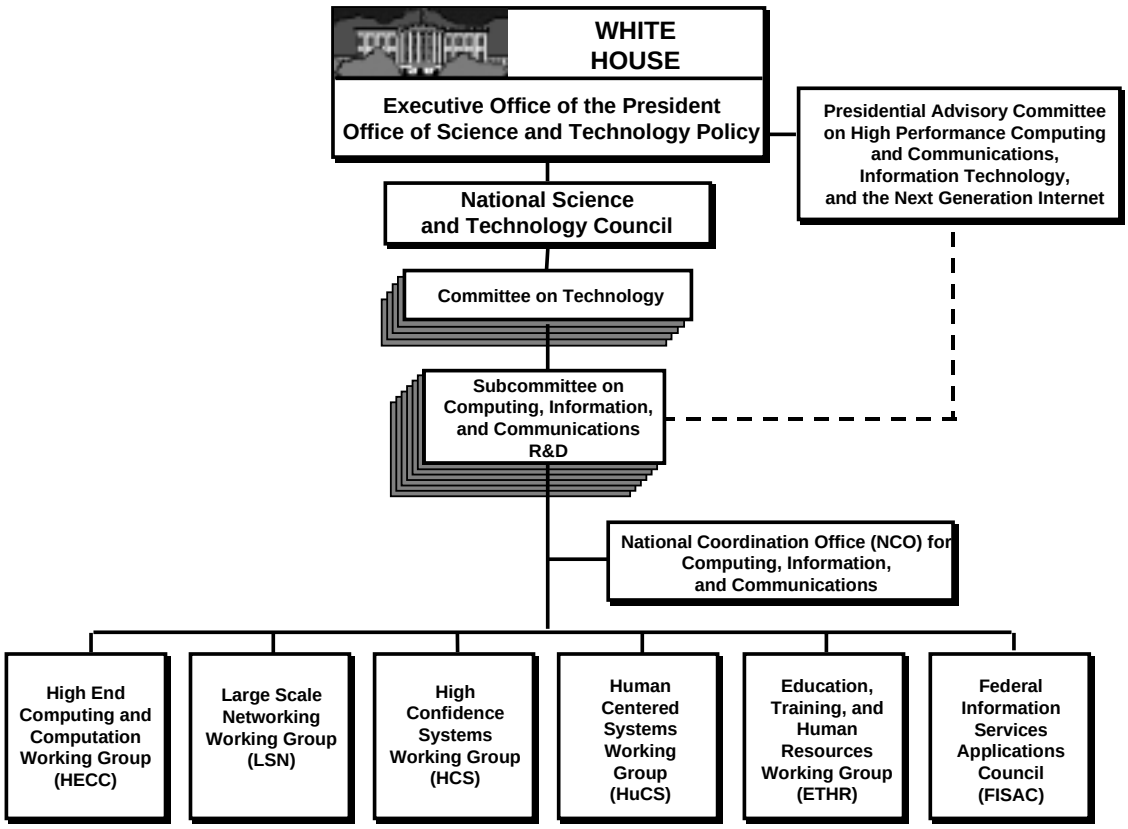
FY 1998 NGI Budget, $ in millions

Goal	DoD/DARPA	NSF	NASA	NIST	NLM/NIH	Total
Goal 1: Experimental Research	20	5	2	3		**30**
Goal 2: Next Generation Network Testbed	20	10	3			**33**
Goal 3: Revolutionary Applications	2	8	5	2	5	**22**
Total	**42**	**23**	**10**	**5**	**5**	**85**

Figure 1. NGI FY 1998 Funding by Goal

Proposed FY 1999 NGI Budget, $ in millions

DoD/DARPA	DoE	NSF	NASA	NIST	NLM/NIH	Total
40	**22**	**25**	**10**	**5**	**5**	**107**



HISTORY

- Internet 2 (summer of 1996)
- NGI concept (fall of 1996)
- First Interagency NGI mtg (1 / 97)
- Concept paper (4 / 97)
- NGI workshop (5 / 97)
- Presidential Advisory Committee (6 / 97)
- Implementation Plan Draft (7 / 97)
- DOE - no NGI (9 / 97)
- NSF must use NSI $ (9 / 97)
- DOE NGI workshop (1 / 98)
- NEW Implementation Plan (2 / 98)

NGI Issues

- Infrastructure vs Research
- Rural (haves and have nots)
- production and network research (e.g. MORPHNET)
- Internet is HOT - i.e. everyone thinks THEY are charting the future
- Industry vs Academia / Research Institutions
- I2 vs NGI
- DOE is Application Agency

For More Information - URLs

Next Generation Internet
- **http://www.ngi.gov**

Internet 2
- **http://www.internet2.edu**

NASA Research and Education Network
- **http://www.nren.nasa.gov**

DOE
- **http://www.es.net**
- http://www.anl.gov/ECT/Public/research/morphnet.html

DARPA
- **http://www.ito.darpa.mil/ResearchAreas.html**

NSF's Connections
- **http://www.vbns.net**

DOE FY 1999 NGI PROGRAM

FY 1999 President's Request: $22M

A new way of interacting with university partners
($5M)

Enhancements to accomplish critical DOE missions
($7M)

Core with a long history of significant accomplishments
($10M)

DOE FY 1999 NGI PROGRAM

Core with a long history of significant accomplishments
($10M)

Goal 1: Technologies
Goal 2: Testbeds
Goal 3: Applications

DOE2000 National Collaboratory Pilots

DOE2000 National Collaboratory Tools (i.e. CBQ, Multicast, Secure Access,…)

Network R&D: Management,Monitoring, High Speed Interfaces

Maintain Internetwork IP Infrastructure

ESnet ATM Experiments

DOE FY 1999 NGI PROGRAM

Enhancements to accomplish critical DOE missions
($7M)

Goal 1: Technologies
Goal 2: Testbeds
Goal 3: Applications

High Performance System Services (e.g. OS and OS bypass) for DOE Applications

Network R&D: Network Aware Intelligent Middleware

Multifunction OC-12 ESnet Testbed

Application driven WDM testbed

DOE FY 1999 NGI PROGRAM

A new way of interacting with university partners
($ 5 M)

Goal 1: Technologies
Goal 2: Testbeds
Goal 3: Applications

University involvement in a DOE critical mission application

University deployment and use of DOE2000 National Collaboratory Tools

Network R&D: Network peering and QOS management tools

Multifunction Campus Networks

Multifunction Interconnects

Challenges of Future High-End Computing

David H. Bailey
NERSC, Lawrence Berkeley Laboratory
Mail Stop 50B-2239
Berkeley, CA 94720
dhb@nersc.gov

A Petaflops Computer System

- 1 Pflop/s (10^{15} flop/s) in sustained computing power.
- Between 10,000 and 1,000,000 processors.
- Between 10 Tbyte and 1 Pbyte main memory (1 Pbyte =100 times the capacity of the U. C. Berkeley library).
- Between 1 Pbyte and 100 Pbyte on-line mass storage.
- Between 100 Pbyte and 10 Ebyte archival storage.
- Commensurate I/O bandwidth, etc.
- If built today, would cost $50 billion and consume 1,000 Mwatts of electric power.
- May be feasible and “affordable” by the year 2010 or sooner.

Petaflops Computers: Who Needs Them?

Expert predictions:

- (c. 1950) Thomas J. Watson: only about six computers are needed worldwide.
- (c. 1977) Seymour Cray: there are only about 100 potential customers worldwide for a Cray-1.
- (c. 1980) IBM study: only about 50 Cray-class computers will be sold per year.

Present reality:

- Some private homes now have six Cray-1-class computers!

Applications for Petaflops Systems

- Nuclear weapons stewardship.
- Cryptology and digital signal processing.
- Satellite data processing.
- Climate and environmental modeling.
- Design of advanced aircraft and spacecraft.
- Design of practical fusion energy systems.
- Pattern matching in DNA sequences.
- 3-D protein molecule simulations.
- Global-scale economic modeling.
- Virtual reality design tools for molecular nanotechnology.

Plus numerous novel applications that can only be dimly envisioned now.

SIA Semiconductor Technology Roadmap

Characteristic	1998	2001	2004	2007
Feature size (micron)	0.25	0.18	0.13	0.10
DRAM size (MByte)	32	128	512	2048
RISC processor clock (MHz)	500	600	800	1000
Density (Mtran/cm^2)	9	18	35	67
Cost per transistor (millicents)	0.5	0.2	0.1	0.05

Observations:

- Moore's Law of increasing density will continue until at least 2007.
- Clock rates of RISC processors and DRAM memories are **not** expected to be much faster than today's rates.

Conclusion: Future high-end systems will feature tens of thousands of processors, with deeply hierarchical memories.

Designs for a Petaflops System

Commodity technology design:

- 125,000 nodes, each with an 8-way symmetric multiprocessor.
- Clock rate = one GHz; each processor sustains one Gflop/s.
- Multi-stage switched network.

Hybrid technology, multi-threaded (HTMT) design:

- 10,000 nodes, each with one superconducting RSFQ processor.
- Clock rate = 100 GHz; each processor sustains 100 Gflop/s.
- Multi-threaded processor design handles a large number of outstanding memory references.
- Multi-level memory hierarchy (CRAM, SRAM, DRAM, etc.).
- Optical interconnection network.

Little's Law of Queuing Theory

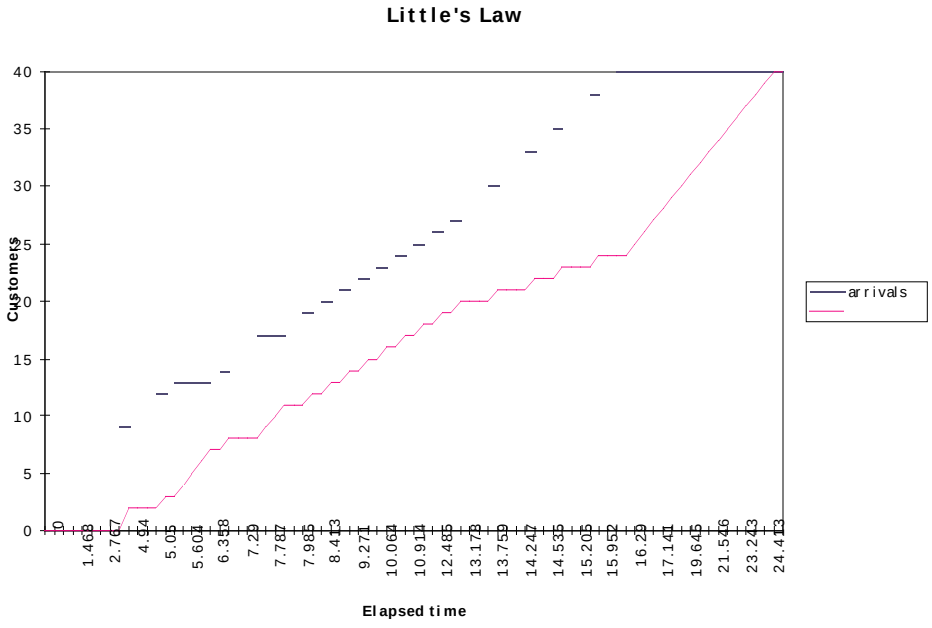
Little's Law:

Average number of waiting customers =

average arrival rate x average wait time per customer.

Proof:

Define f(t) = cumulative number of arrived customers, and g(t) = cumulative number of departed customers. Assume f(0) = g(0) = 0, and f(T) = g(T) = N. Consider the region between f(t) and g(t). By Fubini's theorem of measure theory, one can evaluate this area by integration along either axis. Thus Q T = D N, where Q is average length of queue, and D is average delay per customer. In other words, Q = (N/T) D.



Little's Law of High Performance Computing

Assume:

- Single processor-memory system.
- Computation deals with data in local main memory.
- Pipeline between main memory and processor is fully utilized.

Then by Little's Law, the number of words in transit between CPU and memory (i.e. length of vector pipe, size of cache lines, etc.)

= memory latency x bandwidth.

This observation generalizes to multiprocessor systems:

concurrency = latency x bandwidth,

where "concurrency" is aggregate system concurrency, and "bandwidth" is aggregate system memory bandwidth.

This form of Little's Law was first noted by Burton Smith of Tera.

Little's Law and Petaflops Computing

Assume:

- DRAM memory latency = 50 ns.
- There is a 1-1 ratio between memory bandwidth (word/s) and sustained performance (flop/s).
- Cache and/or processor system can maintain sufficient outstanding memory references to cover latency.

Commodity design:

Clock rate = 1 GHz, so latency = 50 CP. Then system concurrency = 125,000 x 8 x 50 = 5×10^7.

HTMT design:

Clock rate = 100 GHz, so latency = 5,000 CP. Then system concurrency = 10,000 x 5,000 = 5×10^7.

By Little's Law, system concurrency = $5 \times 10^{-8} \times 10^{15} = 5 \times 10^7$ in each case.

Amdahl's Law and Petaflops Computing

Assume:

- Commodity petaflops system -- one million CPUs, each of which can sustain one Gflop/s.
- 90% of operations can fully utilize one million CPUs.
- 10% can only utilize 1,000 or fewer processors.

Then by Amdahl's Law,

Sustained performance $< 1 / [0.9/10^{15} + 0.1/10^{12}]$

$= 9.91 \times 10^{12}$ flop/s,

which is only about 1% of the system's presumed achievable performance.

Concurrency and Petaflops Computing

Conclusion: No matter what type of processor technology is used, applications on petaflops computer systems must exhibit roughly 100 million way concurrency at virtually every step of the computation, or else performance will be disappointing.

- This assumes that most computations access data from local DRAM memory, with limited cache re-use (typical of many applications).
- If substantial long-distance communication is required, the concurrency requirement may be even higher!

Key question: Can applications for future systems be structured to exhibit these enormous levels of concurrency?

Latency and Data Locality

System	Latency Sec.	Latency Clocks
SGI O2, local DRAM	70 ns	30
SGI Origin, remote DRAM	250ns	100
IBM SP2, remote node	40 us	3,000
HTMT system, local DRAM	50 ns	5,000
HTMT system, remote memory	200 ns	20,000
SGI cluster, remote memory	3 ms	300,000

Algorithms and Data Locality

- Can we quantify the inherent data locality of key algorithms?
- Do there exist "hierarchical" variants of key algorithms?
- Do there exist "latency tolerant" variants of key algorithms?
- Can bandwidth-intensive algorithms be substituted for latency-sensitive algorithms?
- Can Little's Law be "beaten" by formulating algorithms that access data lower in the memory hierarchy? If so, then systems such as HTMT can be used effectively.

A Hierarchical, Latency Tolerant Algorithm for Large 1-D FFTs

- Regard input data of length n = p q as a p x q complex matrix, distributed so that each node contains a block of columns.
- Transpose to q x p matrix.
- Perform q-point FFTs on each of the p columns.
- Multiply resulting matrix by exp (-2 pi i j k / n), where j and k are row and column indices of matrix.
- Transpose to p x q matrix.
- Perform p-point FFTs on each of the q columns.
- Transpose to a q x p matrix.

Features:

- Computational steps are embarrassingly parallel -- no communication.
- Transpose operations can be done as latency tolerant block transfers.
- This scheme can be recursively employed for each level of hierarchy.

Numerical Scalability

For the solvers used in most of today's codes, condition numbers of the linear systems increase linearly or quadratically with grid resolution.

The number of iterations required for convergence is directly proportional to the condition number.

Conclusions:

- Solvers used in most of today's applications are not numerically scalable.
- Research in novel techniques now being studied in the academic world, especially domain decomposition and multigrid, may yield fundamentally more efficient methods.

System Performance Modeling

Studies must be made of future computer system and network designs, years before they are constructed.

Scalability assessments must be made of future algorithms and applications, years before they are implemented on real computers.

Approach:

- Detailed cost models derived from analysis of codes.
- Statistical fits to analytic models.
- Detailed system and algorithm simulations, using discrete event simulation programs.

Performance Model of the NAS LU Benchmark

Total run time T per iteration is given by

$$T = 485\, F\, N^3\, 2^{-2K} + 320\, B\, N^2\, 2^{-K} + 4\, L + [1 + 2\,(2^K - 1) / (N - 2)]$$
$$\{2\,(N - 2)\,[279\, F\, N^2\, 2^{-2K} + 80\, B\, N\, 2^{-K} + L)] + 953\, F\, (N - 2)\, N^2\, 2^{-2k}\}$$

where

L = node-to-node latency (assumed not to degrade with large K)

B = node-to-node bandwidth (assumed not to degrade with large K)

F = floating point rate

N = grid size

$P = 2^{2K}$ = number of processors

Acknowledgment: Maurice Yarrow and Rob Van der Wijngaart

Hardware and Architecture Issues

- Commodity technology or advanced technology?
- How can the huge projected power consumption and heat dissipation requirements of future systems be brought under control?
- Conventional RISC or multi-threaded processors?
- Distributed memory or distributed shared memory?
- How many levels of memory hierarchy?
- How will cache coherence be handled?
- What design will best manage latency and hierarchical memories?

How Much Main Memory?

- 5-10 years ago: One word (8 byte) per sustained flop/s.
- Today: One byte per sustained flop/s.
- 5-10 years from now: 1/8 byte per sustained flop/s may be adequate.

3/4 rule: For many 3-D computational physics problems, main memory scales as d^3, while computational cost scales as d^4, where d is linear dimension.

However:

- Advances in algorithms, such as domain decomposition and multigrid, may overturn the 3/4 rule.
- Some data-intensive applications will still require one byte per flop/s or more.

Programming Languages and Models

MPI, PVM, etc.

- Difficult to learn, use and debug.
- Not a natural model for any notable body of applications.
- Inappropriate for distributed shared memory (DSM) systems.
- The software layer may be an impediment to performance.

HPF, HPC, etc.

- Performance significantly lags behind MPI for most applications.
- Inappropriate for a number of emerging applications, which feature large numbers of asynchronous tasks.

Java, SISAL, Linda, etc.

- Each has its advocates, but none has yet proved its superiority for a large class of highly parallel scientific applications.

Towards a Petaflops Language

- High-level features for application scientists.
- Low-level features for performance programmers.
- Handles both data and task parallelism, and both synchronous and asynchronous tasks.
- Efficient for systems with up to 1,000,000 processors.
- Appropriate for parallel clusters of distributed shared memory nodes.
- Permits both automatic and explicit data communication.
- Designed with a hierarchical memory system in mind.
- Permits the memory hierarchy to be explicitly controlled by performance programmers.

System Software

- How can tens or hundreds of thousands of processors, running possibly thousands of separate user jobs, be managed?
- How can hardware and software faults be detected and rectified?
- How can run-time performance phenomena be monitored?
- How should the mass storage system be organized?
- How can real-time visualization be supported?

Exotic techniques, such as expert systems and neural nets, may be needed to manage future systems.

Faith, Hope and Charity

Until recently, the high performance computing field was sustained by

- Faith in highly parallel computing technology.
- Hope that current faults will be rectified in the next generation.
- Charity of the federal government.

Results:

- Numerous firms have gone out of business.
- Government funding has been cut.
- Many scientists and lab managers have become cynical.

Where do we go from here?

Time to Get Quantitative

- Quantitative assessments of architecture scalability.
- Quantitative measurements of latency and bandwidth.
- Quantitative analyses of multi-level memory hierarchies.
- Quantitative analyses of algorithm and application scalability.
- Quantitative assessments of programming languages.
- Quantitative assessments of system software and tools.

Let the analyses begin!

Presentation to the Conference on High Speed Computing

David M. Cooper
April 22, 1998

High End Subcommittee Status Report

Co-Chairs: David Cooper and Susan Graham
Members: Vinton Cerf
Robert Ewald
James N. Gray
Daniel Hillis
Ken Kennedy
Larry Smarr
Joe Thompson
Steven Wallach
Irving Wladawsky-Berger

High End Subcommittee Scope

The subcommittee will focus on the technology (hardware and software), methodology, and infrastructure to support present and future applications that are highly computer-intensive, data-intensive, visually-intensive, interaction-intensive and/or computer communication-intensive.

The applications span the traditional and scientific engineering arenas, as well as the emerging societal and commercial issues such as education, entertainment, and economics.

High End Subcommittee Activities (1)

- Next Generation Internet Initiative (NGI)
  - Several members were on the NGI Subcommittee
  - Interactive meetings with government and industry reps
  - Reviewed NGI plan
  - Participated in NGI workshop
  - Formulated findings and recommendations
  - Submitted recommendations to President in June 1997

High End Subcommittee Activities (2)

- National Security High End Computing Integrated Process Team (IPT)
  - Unclassified report presented to subcommittee
  - Unclassified discussions with agency representatives
  - Preparing report on recommendations to OSTP

High End Subcommittee Activities (3)

- High Performance Computing & Communications Program (HPCC)
  - » Issues being addressed
    - Progress in implementing HPCC program
    - Need to revise the HPCC program
    - Balance among HPCC program components
    - Impact of HPCC R&D on maintaining US leadership in advanced computing and communications technologies and their applications

High End Subcommittee Activities (3 Con't)

- High Performance Computing & Communications Program (HPCC)
  - »Input: Numerous presentations and discussions on HPCC program
    - –Agency and industry funded personnel
    - –Non-funded industry personnel
    - –Town meeting at SC'97
  - »Output: Preparing a draft response addressing the issues

High End Subcommittee Activities (4)

- High End Computing and Computation Program (HECC)
  - » Reviewed October 1997 draft of Program Strategy
  - » Provided recommended changes
  - » Currently reviewing February 1998 draft revised Program Strategy
  - » Input to be provided by April 1998

High End Subcommittee Activities (5)

- Recommendations for FY00 budget
  - »Working with members of Broad Based Subcommittee to develop an single integrated report
    - – Integration of high end computing, high performance networking, management of massive amounts of data, coupling of scientific instruments, ...
    - – Importance of basic research
    - – Software challenges

Broad-Based Subcommittee

Co-Chairs: David Nagel and Ted Shortliffe

Members:
Eric Benhamou
Ching-Chih Chen
Steve Dorfman
Dave Dorman
Dave Farber
Sherrilynne Fuller
Hector Garcia-Molina
Bill Joy
John Miller
Raj Reddy
Les Vadasz
Andy Viterbi

The Vision That Guided Discussions

Our view of the kinds of capabilities that we expect to be made available by the nation's information infrastructure in the decades ahead:

A network that supports:

- 20ms-200ms latency
- High-quality video
- Every household
- All industry, including all small businesses
- All centers of learning
- Reliability, availability, maintenance comparable to the current telephone industry
- Connections to a global network
- Secure, authorized subnets
- Constant connectivity
- Mobility
- Standard interfaces

Presidential Advisory Committee on HPCC, IT, and NGI April 1998 ADC-572 Page 11

The Vision That Guided Discussions

Our view of the kinds of capabilities that we expect to be made available by the nation's information infrastructure in the decades ahead:

A network that supports:

- Rapid access to information in 1 minute or less
- Rapid access to other people
- Persistence of information, with archival functions in place
- Intelligent agents
- Electronic commerce, including an ability to purchase items anywhere in the world with management of security, currency exchange, and the like
- Assured quality of content
- Quality education and health care

Presidential Advisory Committee on HPCC, IT, and NGI April 1998 ADC-572 Page 12

Charge to the Broad-Based Subcommittee

- Anticipate the evolution of broad-based computing and communications over the next two decades with regard to the following issues:
- Scalability of the Internet
  - Representative anticipated applications
  - Middleware and software technologies
  - Anticipated end-user devices and connection methods
  - The financial model envelope that captures the trade-offs
  - Predicted societal impacts

Presidential Advisory Committee on HPCC, IT, and NGI April 1998 ADC-572 Page 3

Charge to the Broad-Based Subcommittee

Report to the full committee, based on economic considerations and results of information gathering, with regard to the following questions:

What are the research and development imperatives?

» Which will be adequately pursued in the commercial sector?

» Which require government support and coordination for timely impact?

» How can the government best achieve its unique facilitating roles?
  - research funding?
  - cross-agency coordination?
  - technology sharing?
  - support for applications?
  - regulation/legislation?
  - others?

Presidential Advisory Committee on HPCC, IT, and NGI April 1998 ADC-572 Page 4

Working Group #1: Scalability

- Steve Dorfman and Raj Reddy
- Scalability of the Internet in areas such as robustness, quality of service, security, guaranteed bandwidth, management of latency, and implementation of end-to-end bandwidth (broad-based infrastructure).
- White paper internally developed

Working Group #2: Applications

- Sherrilynne Fuller and Andy Viterbi
- Representative anticipated applications, with a special emphasis on those from the fields of communications and commerce that will in part define the infrastructure, including dependability, privacy, security, and useability, on which other applications are likely to build (in areas such as education, science, health care, the environment)

Working Group #2: Applications

- **White papers:**
  - » **Health care: "Telemedicine and the National Information Infrastructure: Issues and Opportunities"; Daniel R. Masys, University of California, San Diego**
  - » **Education: "Technology as a Tool for Revitalizing America's K-12 Education System"; Kathie Blankenship, Smart Valley, Inc.**
  - » **Digital libraries: "White Paper on Digital Libraries"; Cliff Lynch, Coalition on Networked Information**

Working Group #3: Devices

- **Les Vadasz and Ted Shortliffe**
- **Anticipated end-user devices and connection methods that will be used for personal and professional computing and communications as well as connectivity to the Internet.**
- **White paper:**
  - » **"White Paper on End-User Devices"; Michael Hawley, MIT Media Lab**

Working Group #4: Middleware

- Hector Garcia-Molina, Jim Gray, and Dave Nagel
- Middleware and software technologies for data management and mining, information retrieval and analysis, programming support, system survivability and interoperability, as well as for managing large scal networks and computer systems.
- White paper:
  - » "The Cooperative Computing Initiative"; Michael L. Brodie, GTE Laboratories, Inc.

Working Group #5: Economics

- Eric Benhamou and Bill Joy
- The financial model envelope that captures the trade-offs between broad-based affordability, financial returns to service providers and equipment vendors, and varying levels of service quality.
- White paper :
  - » "Economic Factors"; Scott O. Bradner, Harvard University

Working Group #6: Societal Issues

- Ching-Chih Chen and Dave Farber
- Predicted societal impacts, including identification of those which may have negative consequences and for which preventive measures can be taken.
- White paper :
  - » "The NGI as an Effective Support for Professional Work and Social Life"; Rob Kling, Indiana University

Presidential Advisory Committee on HPCC, IT, and NGI April 1998 ADC-572 Page 21

Draft Interim Report Outline

Major Themes:

- Increase the Federal investment in basic research in computing, information, and communications
- Create the knowledge and technology base for the Next Generation Information Infrastructure
- Enhance the research base and broaden the application of high-end computing
- Attack fundamental problems in software
- Implement programs to remedy the shortfall in the research and development workforce
- Develop a strategic vision for information technology R&D and a plan for integrating investment across the relevant Federal agencies

Presidential Advisory Committee on HPCC, IT, and NGI April 1998 ADC-572 Page 22

Processing-in-Memory: Past and Present

Ken Iobst
IDA Center for Computing Sciences
(Formally Supercomputing Research Center)
Bowie, Maryland

What is Processing-in-Memory (PIM)?

- A computer architecture that associates individual columns of a RAM with SIMD processors
- An architecture that issues instructions to its memory using simple assignment as a broadcast mechanism
- NOT a Memory-in-Processor (MIP) nor Cache-in-Processor (CIP) type architecture

What is the History of PIM?

- 1988
  - PASSWORK, PETASYS and PIM is envisioned
- 1989
  - PETASYS research notebook is signed and dated for PIM patent disclosure
  - PIM chip discussions are initiated with Motorola, TI, LSI Logic and VLSI Technology

History of PIM (Continued)

- 1990
  - PIM ALU is fully integrated into PASSWORK 3.4
  - PETASYS/TERASYS Project is formed to investigate PIM
  - PASSWORK, PETASYS and PIM is formally presented at the Frontiers of Supercomputing II: A National Reassessment
  - TERASYS workstation is designed using a SUN4 front end, Sbus interface and 16 petaboards (32 PIM chips per board and 64 processors per chip)
  - PIM chip specifications are finalized
  - Discussions are started with SSI to integrate PIM into the SS-1

History of PIM (Continued)

- 1991
  - US PIM Patent #5,396,641 is filed
  - TERASYS Critical Design Review is completed
  - PIM IEEE floating point is emulated in PASSWORK 3.6
  - Foreign PIM patents are filed in Japan, Great Britain and Germany
- 1992
  - First PIM chip is successfully tested
  - “Petasys Supercomputer” HPCC white paper is written
  - Discussions are started with CCC to integrated PIM into the CRAY-3

History of PIM (Continued)

- 1993
  - SSI goes bankrupt
  - 4K processor TERASYS workstation is successfully tested
  - Formal proposal is submitted by CCC to the NSA to build a Super Massively Parallel Processor Computer System
  - Cray-3 PIM hardware design is reviewed by Seymour Cray
  - (D)ARPA Joint Program Solicitation No: 93-29 of the Technology Reinvestment Project (TRP) between CCC, NSC and NSA for the "Design and Manufacture of a Prototype Super Massively Parallel Processor Computer System" is rejected

History of PIM (Continued)

- 1994
  - PIM context switch software for the Cray-3 is simulated on PASSWORK then executed on TERASYS
  - HPCC awards contract of 9.2 M$ to CCC, NSC and the SRC to design and develop a Prototype Super Scalable System (SSS)
  - First PIM module is successfully tested on the CRAY-3
  - CCC receives initial delivery of 5000 bare PIM die from NSC

History of PIM (Continued)

- 1995
  - Panel "SIMD Machines: Do they have a Significant Future?" at Frontiers of Massively Parallel Computation concludes YES!
  - SSS is successfully tested on the CRAY-3 with 256K PIM processors
  - CCC files for court protection under Chapter 11
  - United States PIM Patent #5,396,641 is awarded
  - "Processing in Memory: The Terasys Massively Parallel PIM Array" is published in IEEE Computer Magazine
  - Great Britain PIM Patent #GB2252185 is awarded

What were the Salient Features of the CRAY-3 SSS?

- SSS was a CRAY-3 system with 2 vector processing CPU's, 1 main memory "Octant" and up to 4 PIM memory octants (1M PIM processors). CPU's were clocked at 2 nsec and PIM memories at 32 nsec.
- SSS had 256M words of main memory (64 banks) and 8M "words" of PIM memory (256 banks)
- Main memory to PIM memory bandwidth was 64 Gbs
- PIM processor density was 3450 processors per inch^3 (not including power supply)
- PIM technology was 0.8μ CMOS SRAM organized as 2K rows by 64 columns/ processors

Amazing Things that can be done from a Single Instruction Stream

- Finding the largest value across all processors in time proportional to the length of the value NOT the number of processors
- Solving multiple linear systems with different pivots
- Performing faster more accurate floating point computations using a variable precision format, NOT normalizing after every computation and using CORDIC arithmetic where possible
- Performing integer addition in some processors and subtraction in others as well division in some processors and square root in others

What did we learn?

- You can control a process-in-memory with standard high level languages that issue SIMD instructions over a conventional memory "data" path
- There is virtually no need for an instruction cache because most SIMD instructions come from "data" registers
- Conditional global OR is important for associative operations whereby two SIMD instructions are issued and only one is actually broadcast/executed

What did we learn? (Continued)

- There are big payoffs for simulating and emulating a system before you build it
- In the case of TERASYS complete applications were actually run using PASSWORK before the first PIM chip was ever built
- Long term federally funded research and development can be successfully transferred to industry if that industry is still in business when the research matures

What did we learn? (Continued)

- There are big payoffs for simulating and emulating a system before you build it
- In the case of TERASYS complete applications were actually run using PASSWORK before the first PIM chip was ever built
- Long term federally funded research and development can be successfully transferred to industry if that industry is still in business when the research matures

What would we do differently next time?

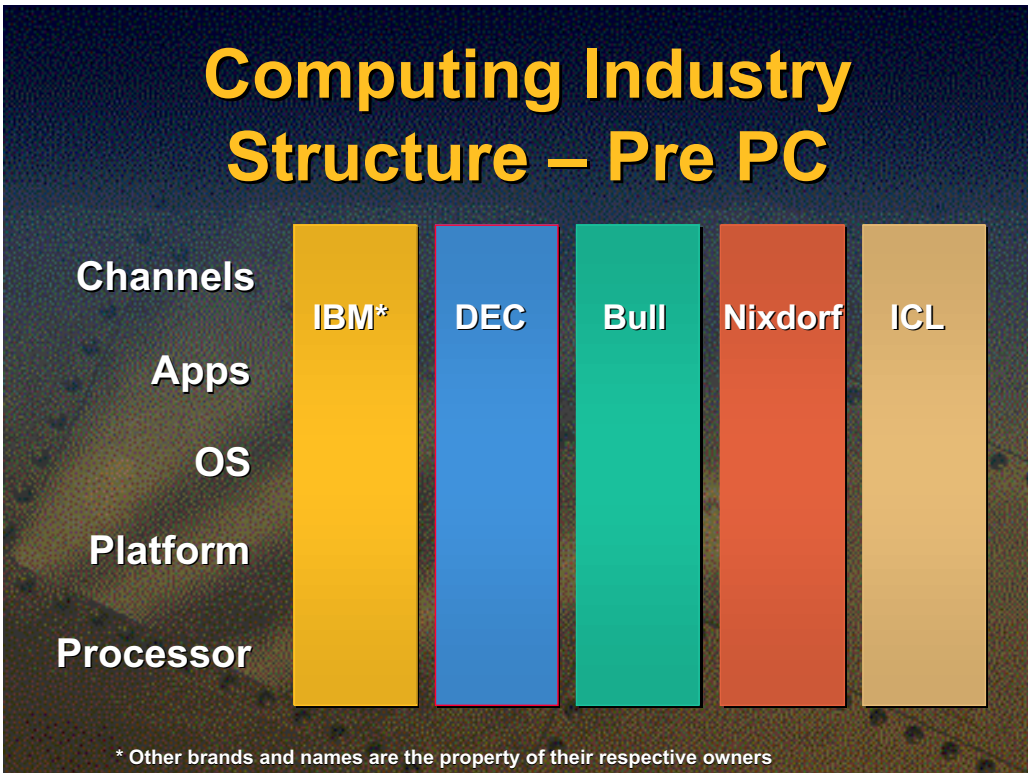
- Design PIM as a COTS based technology
- Design a better high level language to take full advantage of a user's intuitive concept of parallelism, i.e. the computer word
- Let all data objects be of variable power of two bit length
- De-emphasize SIMD and more fully integrate data parallel associative into the high level language
- Support both fine and coarse grain processes-in-memory at a high level

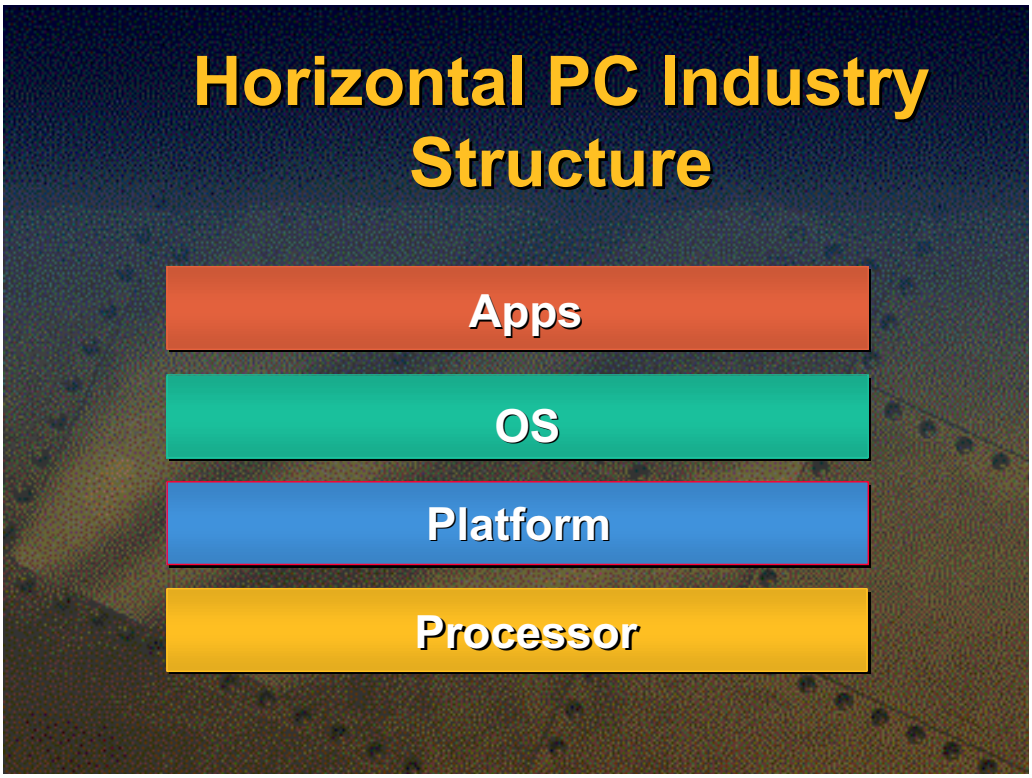
Summary

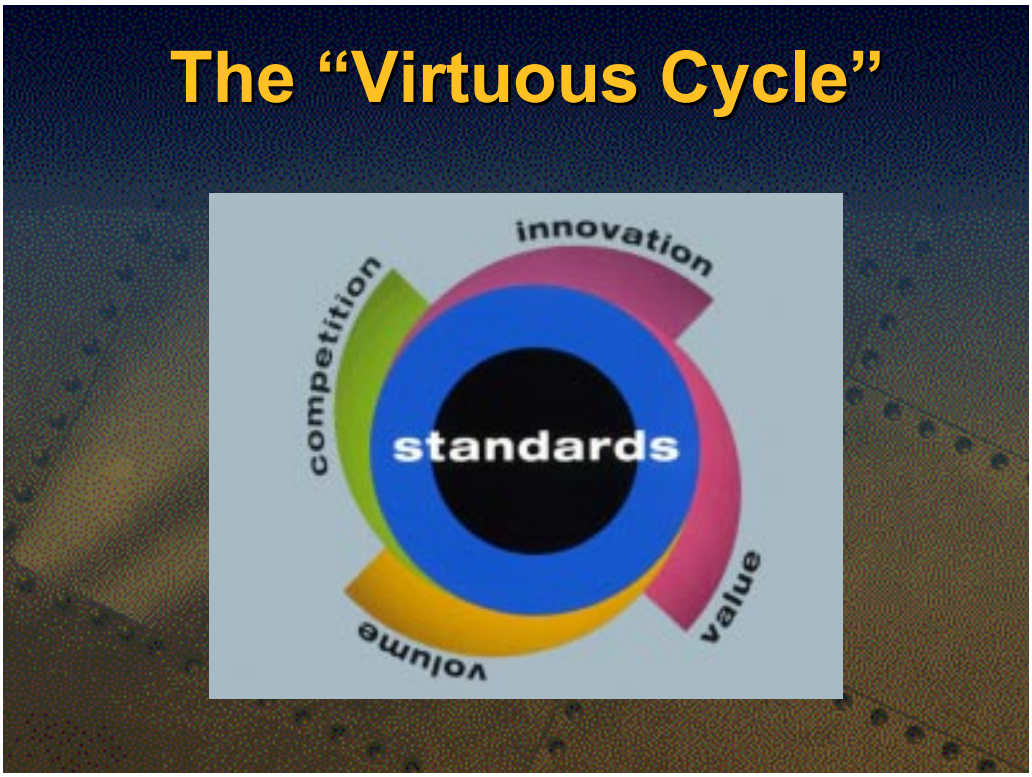
- PIM is memory centric not processor centric, very scalable and a perfect candidate for COTS implementation
- PIM programming paradigm is SIMD not MIMD with processors mapped on to memory columns of a 2-D array and row addresses/instructions being broadcast to all columns/processors
- Choice of SRAM rather than DRAM for PIM speeds computation by 10X
- Computers built with PIM technology are variable word length machines that support efficient high speed computation of vertical objects of any length and horizontal objects of power of two length

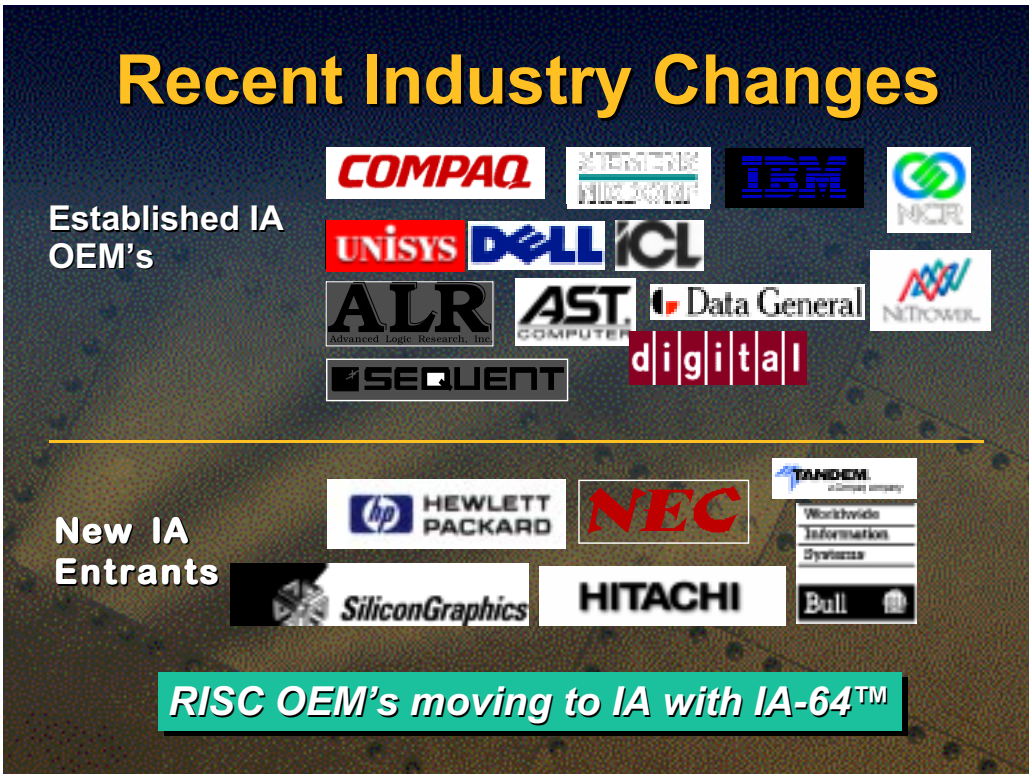
intel®

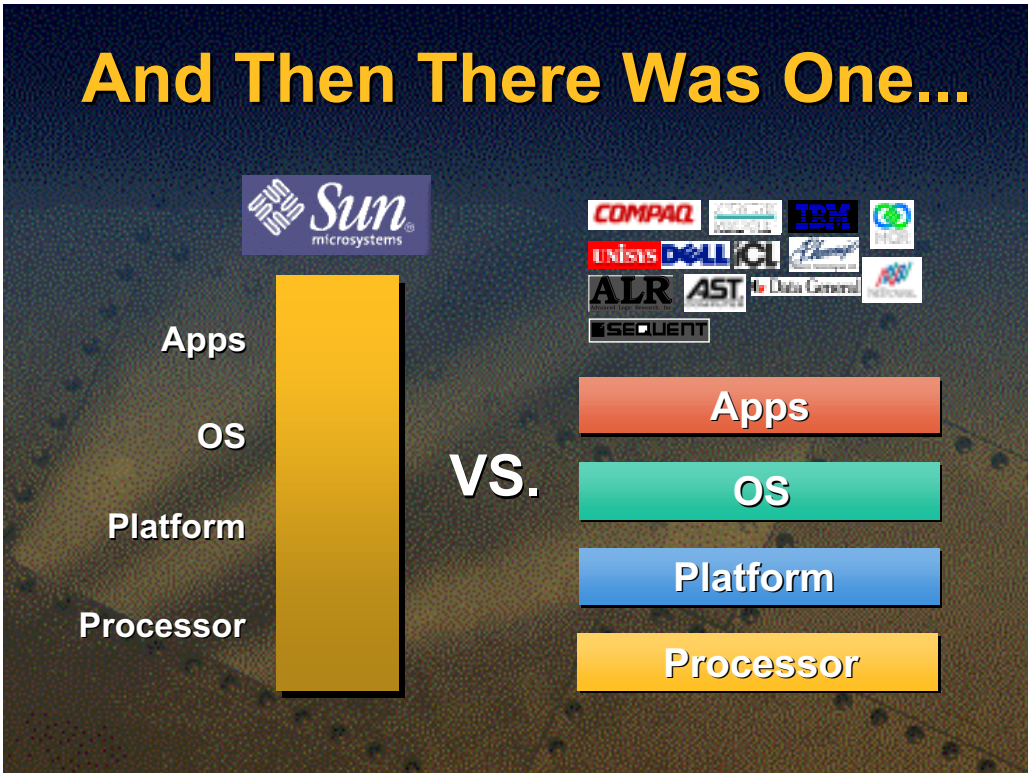
High Volume Technology
for HPC Systems
Justin Rattner
Director and Fellow
Server Architecture Lab
Enterprise Server Group
© Intel Corporation 1998

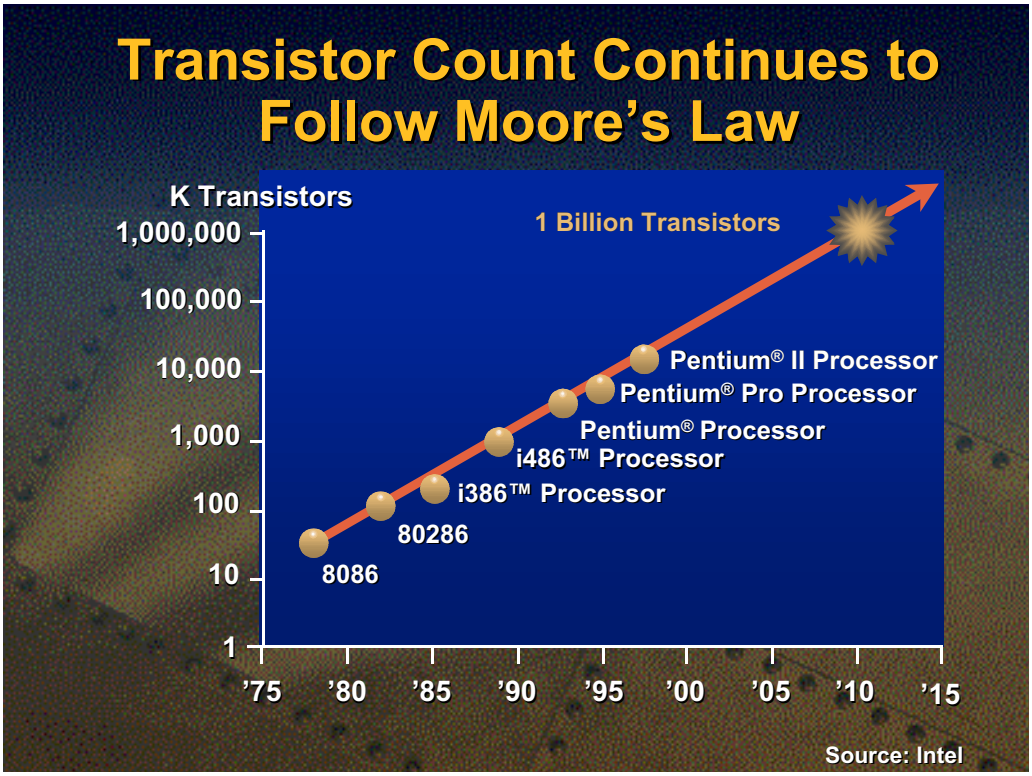

Computing Industry Structure – Pre PC
Channels
Apps
OS
Platform
Processor
IBM*
DEC
Bull
Nixdorf
ICL
* Other brands and names are the property of their respective owners

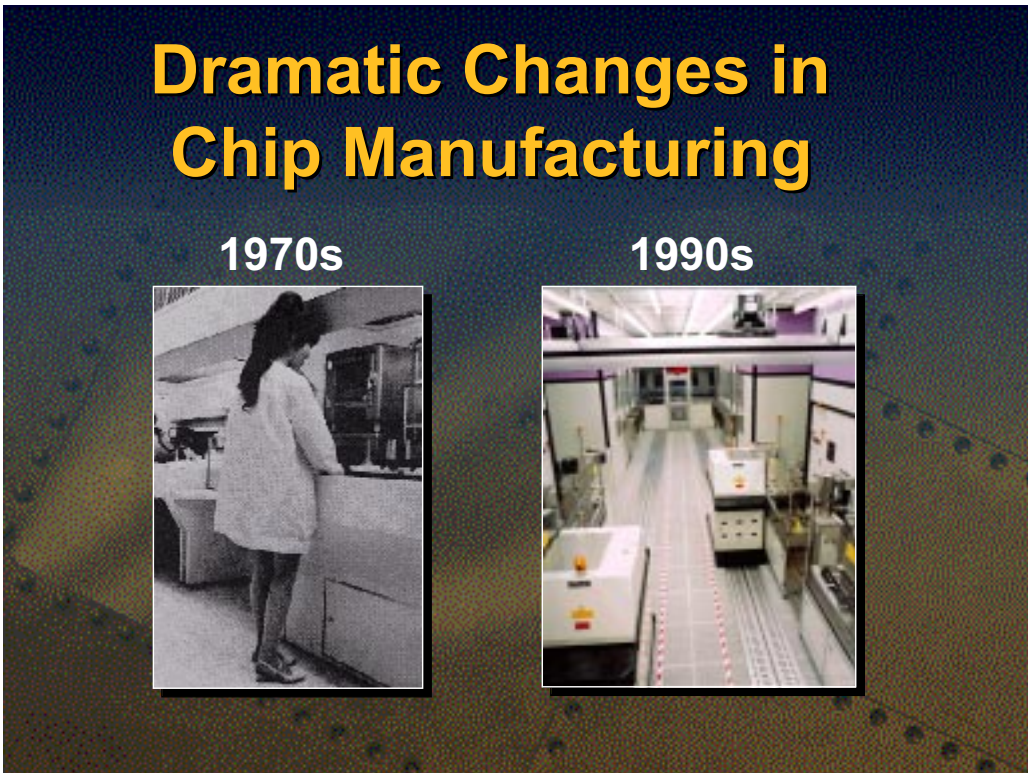

Horizontal PC Industry Structure
Apps
OS
Platform
Processor

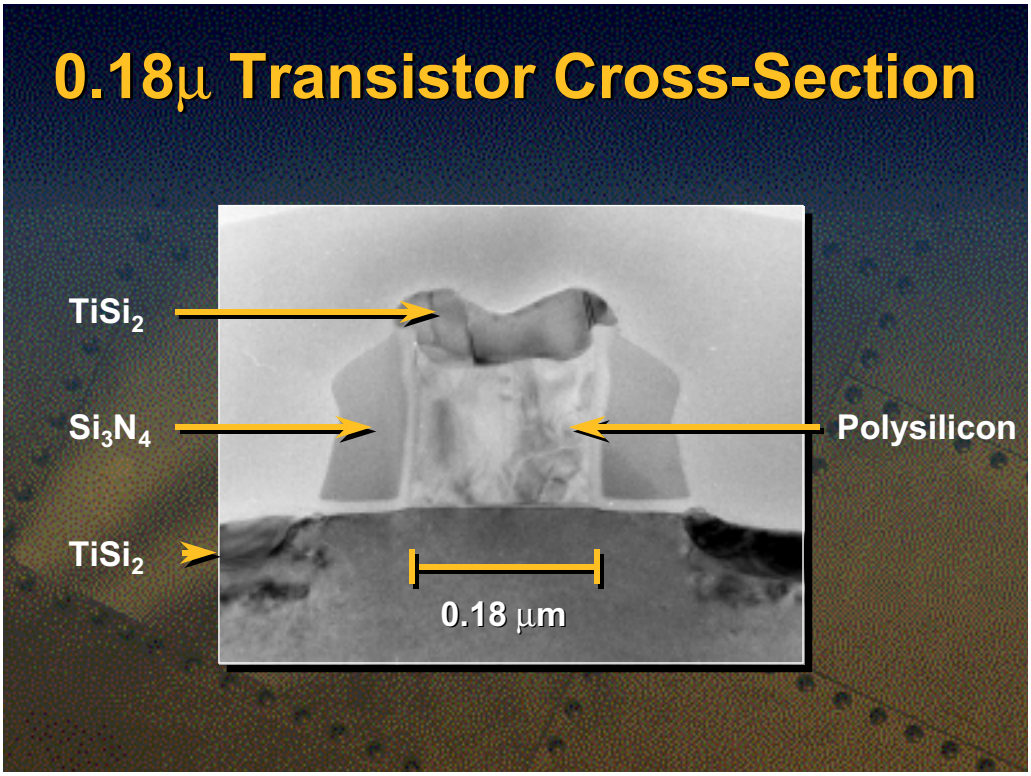

The "Virtuous Cycle"
innovation
competition
standards
value
volume


Recent Industry Changes
Established IA OEM's
COMPAQ
IBM
NCR
UNISYS
DELL
ICL
ALR
AST COMPUTER
Data General
NETPOWER
digital
SEQUENT
New IA Entrants
HEWLETT PACKARD
NEC
TANDEM
Worldwide Information Systems
SiliconGraphics
HITACHI
Bull
RISC OEM's moving to IA with IA-64™

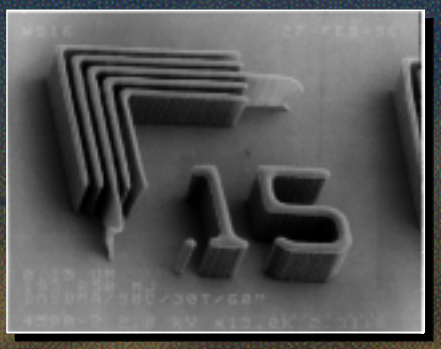

And Then There Was One...
Sun microsystems
Apps
OS
Platform
Processor
VS.
COMPAQ
IBM
Unisys
DELL
ICL
ALR
AST
SEQUENT
Apps
OS
Platform
Processor


Transistor Count Continues to Follow Moore's Law
K Transistors
1,000,000
100,000
10,000
1,000
100
10
1
1 Billion Transistors
Pentium® II Processor
Pentium® Pro Processor
Pentium® Processor
i486™ Processor
i386™ Processor
80286
8086
'75
'80
'85
'90
'95
'00
'05
'10
'15
Source: Intel


Dramatic Changes in
Chip Manufacturing
1970s
1990s


0.18μ Transistor Cross-Section
TiSi2
Si3N4
TiSi2
Polysilicon
0.18 μm

193nm Challenges



Imaging of Intel's ISI 193nm MicroStepper with TSI Resist Process

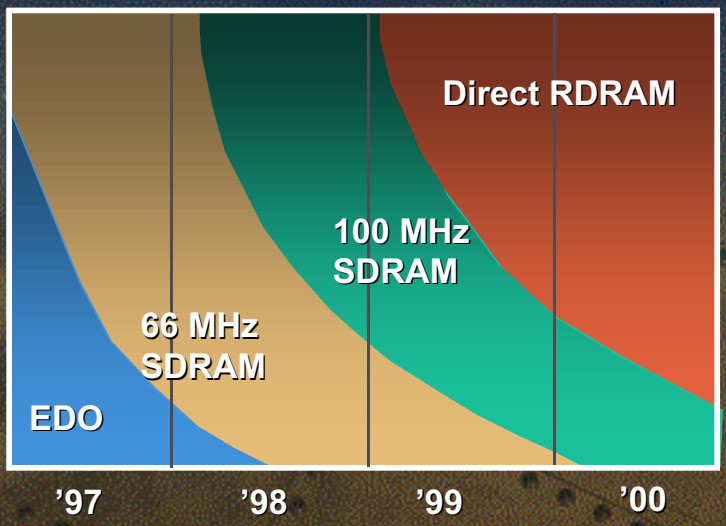
- Optical materials and lens design
- Resist materials
- Exposure tools
- Timing and funding

Memory Technology Roadmap

Server & Wkstn

Professional

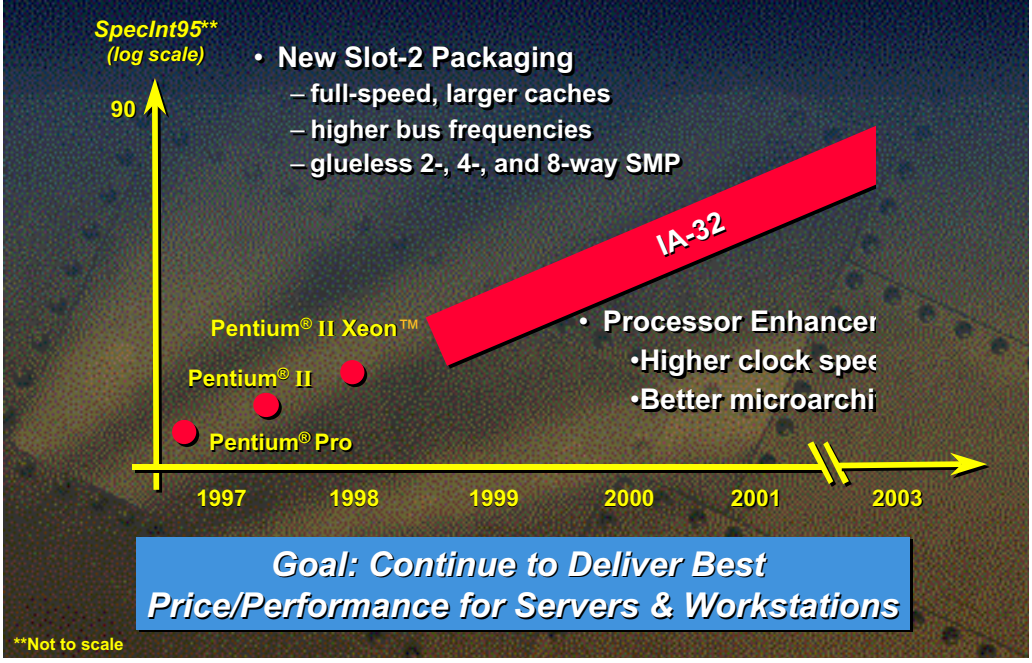
Volume

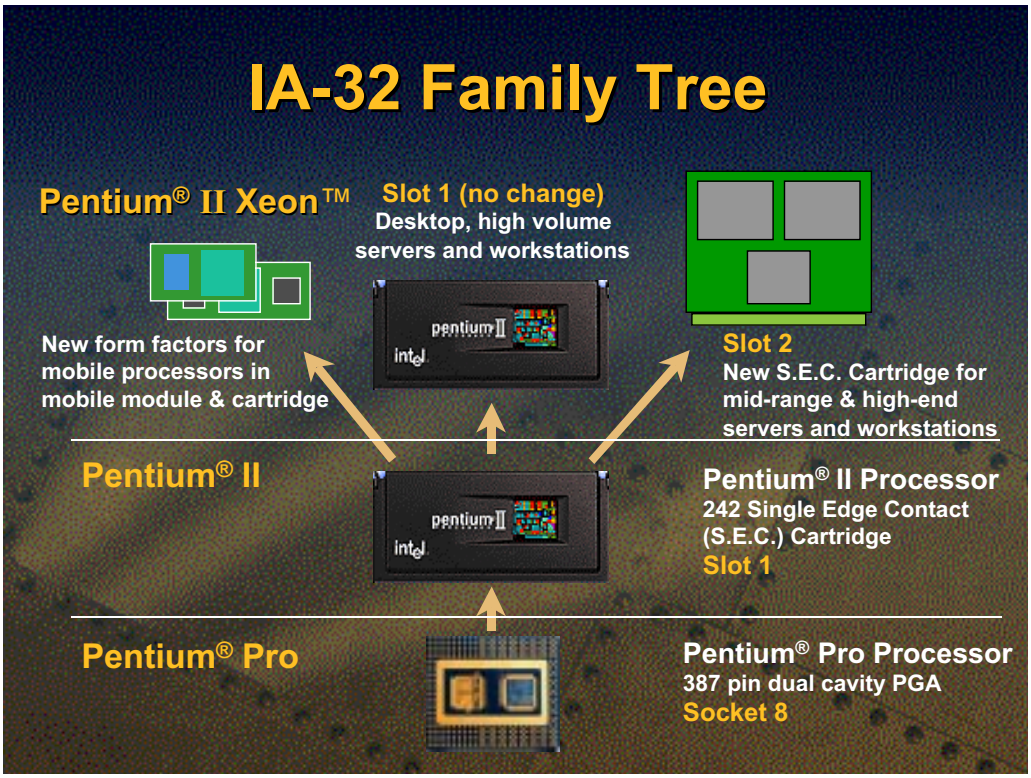


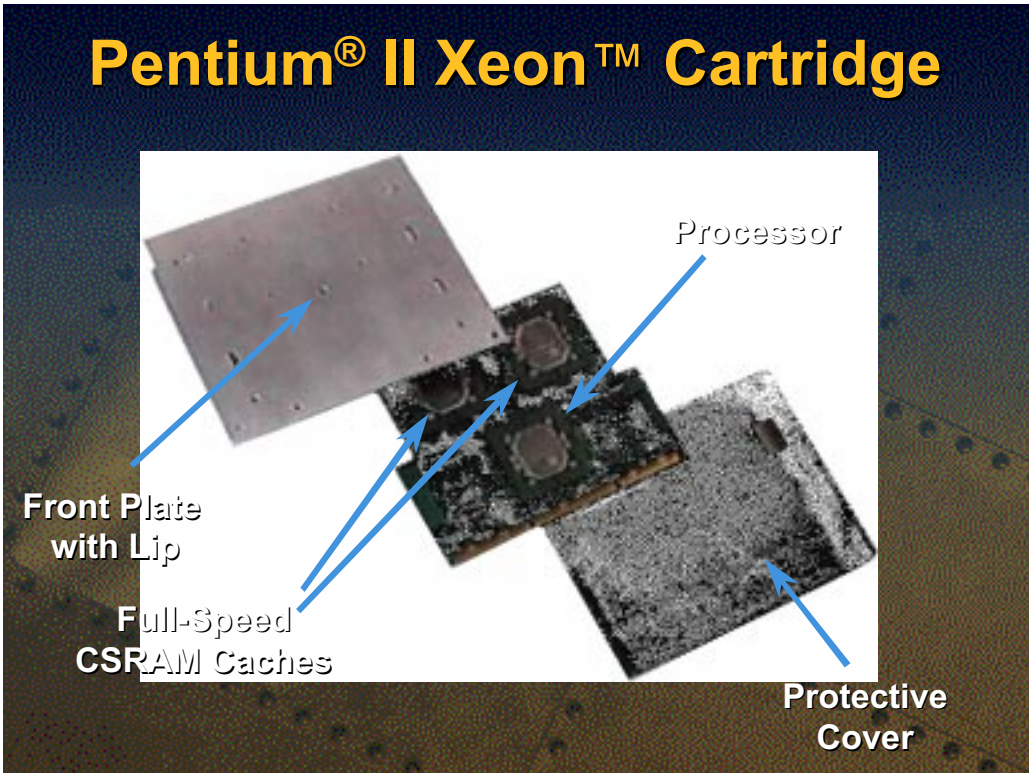
What this means for microprocessors ...

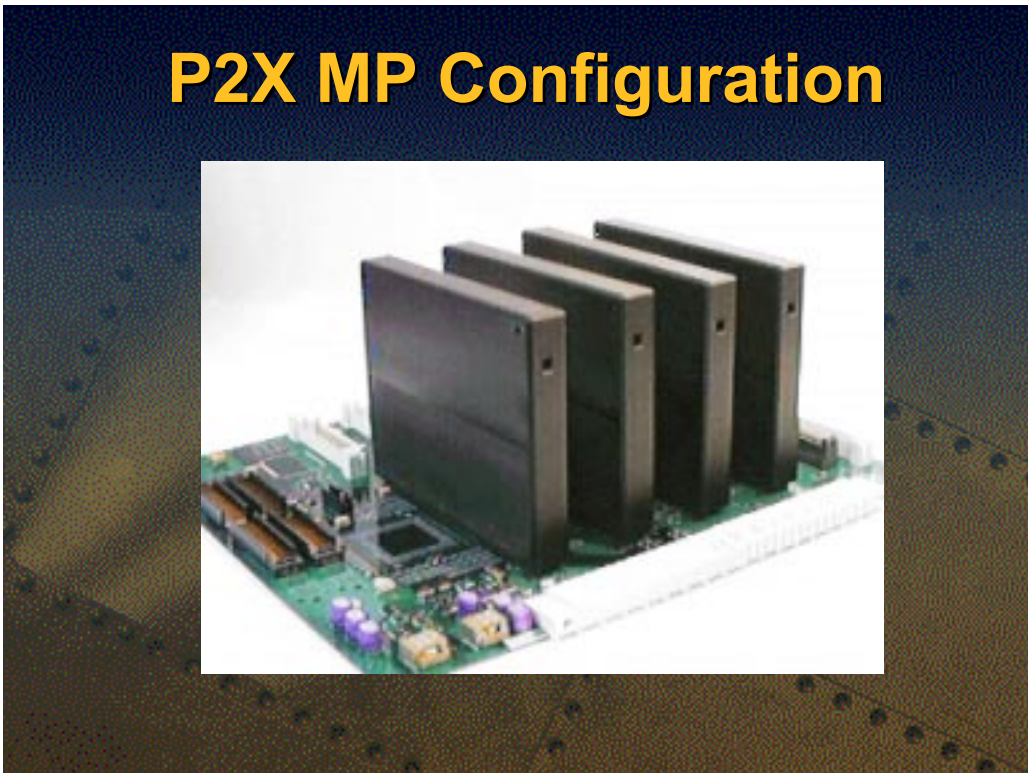
- No let up in chip technology progress
- MIPS will keep doubling every 18 mo's
- More function driven into silicon
- Product life cycles on Internet time

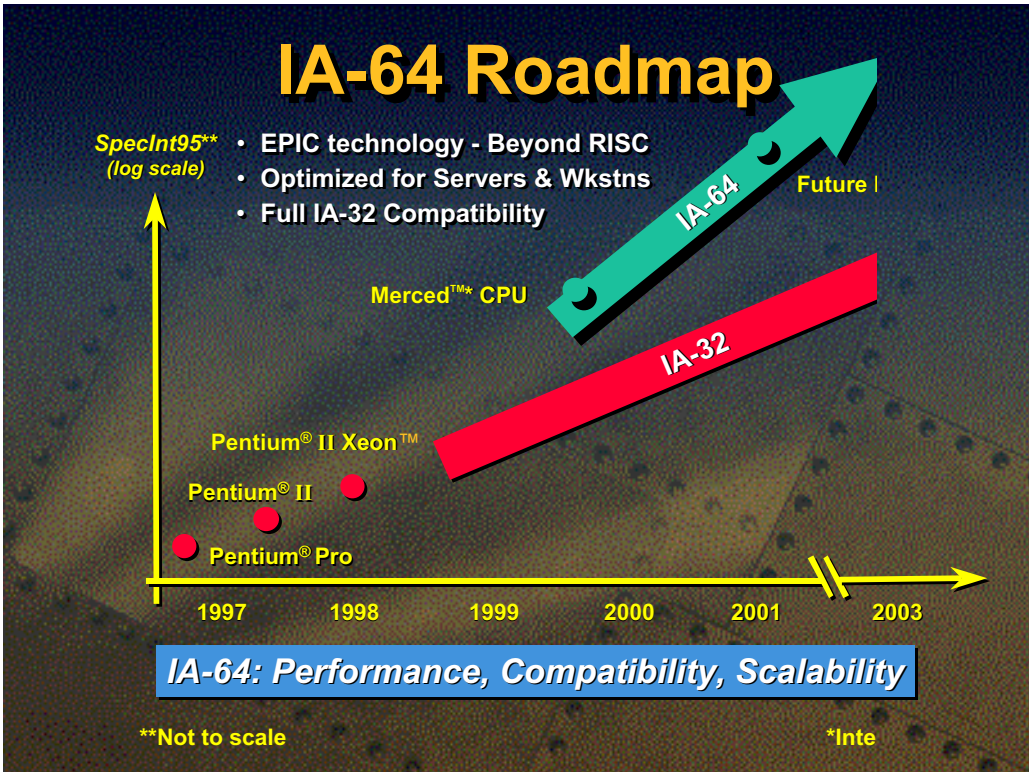
IA-32 Server & Workstation




IA-32 Family Tree
Pentium® II Xeon™
Slot 1 (no change)
Desktop, high volume servers and workstations
New form factors for mobile processors in mobile module & cartridge
Slot 2
New S.E.C. Cartridge for mid-range & high-end servers and workstations
Pentium® II
Pentium® II Processor
242 Single Edge Contact (S.E.C.) Cartridge
Slot 1
Pentium® Pro
Pentium® Pro Processor
387 pin dual cavity PGA
Socket 8


Pentium® II Xeon™ Cartridge
Processor
Front Plate with Lip
Full-Speed CSRAM Caches
Protective Cover


P2X MP Configuration


IA-64 Roadmap
SpecInt95**
(log scale)
• EPIC technology - Beyond RISC
• Optimized for Servers & Wkstns
• Full IA-32 Compatibility
IA-64
Future I
Merced™* CPU
IA-32
Pentium® II Xeon™
Pentium® II
Pentium® Pro
1997
1998
1999
2000
2001
2003
IA-64: Performance, Compatibility, Scalability
**Not to scale
*Inte

iA64 Objectives

- Enable industry leading system performance
- Enable compatibility with today's IA-32 software & PA-RISC software
- Allow scalability over a wide range of implementations
- Full 64-bit computing

The Merced™ Processor

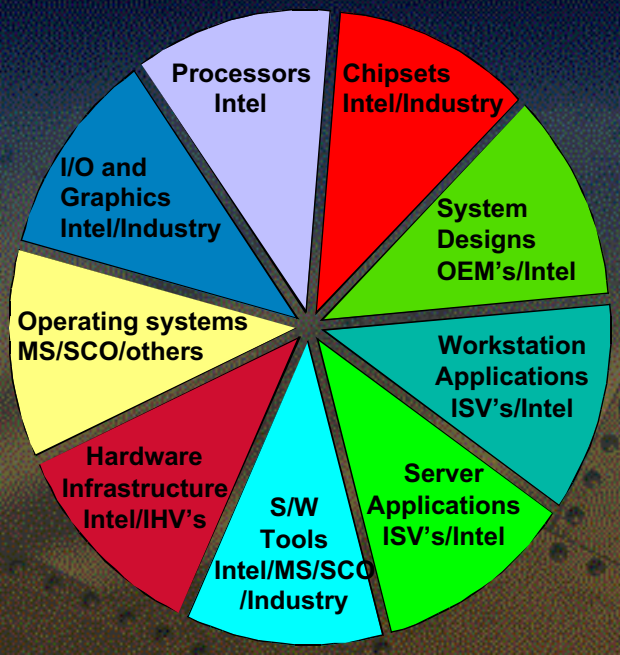
- Code name for Intel's first IA-64 CPU
- Industry leading performance and features for Servers and Workstations
- Full IA-32 binary compatibility in H/W
- Intel's 0.18 micron process technology
- Multiple product configurations

Merced™ Program Update

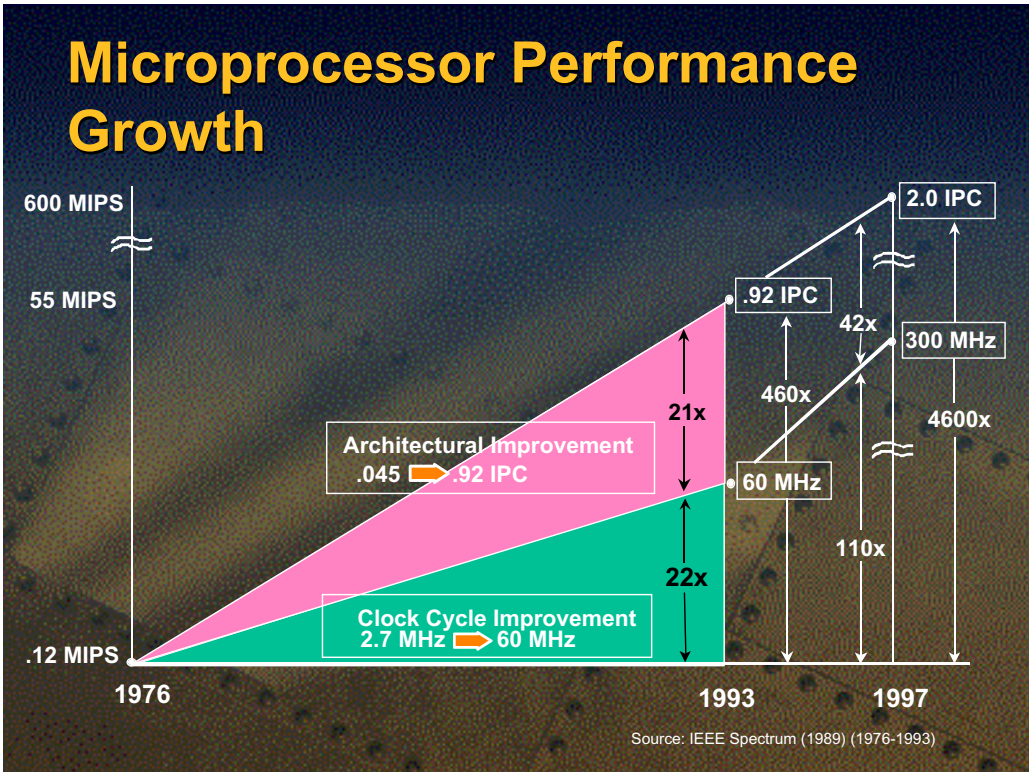
- Complete solution stack at launch
- Processor and chipset design teams on track for ’99 production
- OEM system design teams are making significant progress
- Operating systems and key applications on track for ’99 launch

Merced™ Processor:
Full speed ahead for 1999 production

IA-64: An Entire Platform



IA-64 Industry Commitment
COMPAQ
NEC
ORACLE
IBM
Microsoft
HEWLETT PACKARD
SAS
SAP
NCR
SCO
SIEMENS NIXDORF
UNISYS
ICL
NETPOWER
SEQUENT
HITACHI
Stratus
Worldwide Information Systems
Bull
DELL
AST COMPUTER
ALR
Avid
Adobe
Phoenix
Data General
PARAMETRIC TECHNOLOGY CORPORATION
SOFTIMAGE
INFORMIX
Unprecedented Industry Momentum
All third party marks and brands are the property of their respective owners


Microprocessor Performance Growth
600 MIPS
55 MIPS
.12 MIPS
1976
1993
1997
2.0 IPC
.92 IPC
300 MHz
60 MHz
42x
460x
4600x
21x
22x
110x
Architectural Improvement
.045 .92 IPC
Clock Cycle Improvement
2.7 MHz 60 MHz
Source: IEEE Spectrum (1989) (1976-1993)

Why a New Architecture?

- Insatiable demand for more performance
- Faster and faster processors, slow memory requiring heroic measures to compensate
  - Higher bandwidth
  - More registers, register renaming
  - Techniques to hide memory latency
- Need for greater number of instructions per clock cycle (more efficient use of silicon)
- Diminishing gains from growing processor design complexity

Architecture Challenges

- Mainstream: Superscalar RISC/CISC
  - Improved parallelism, but:
    - Significant hardware complexity
    - Not transparent to compiler
    - Base technology is 15-20 years old
- Peripheral: VLIW, Vector Architectures
  - Even higher parallelism cabability using:
    - Explicit parallelism
    - Parallelizing compilers
  - But, limited applicability, scalability, and no compatibility

A clear need for a new technology

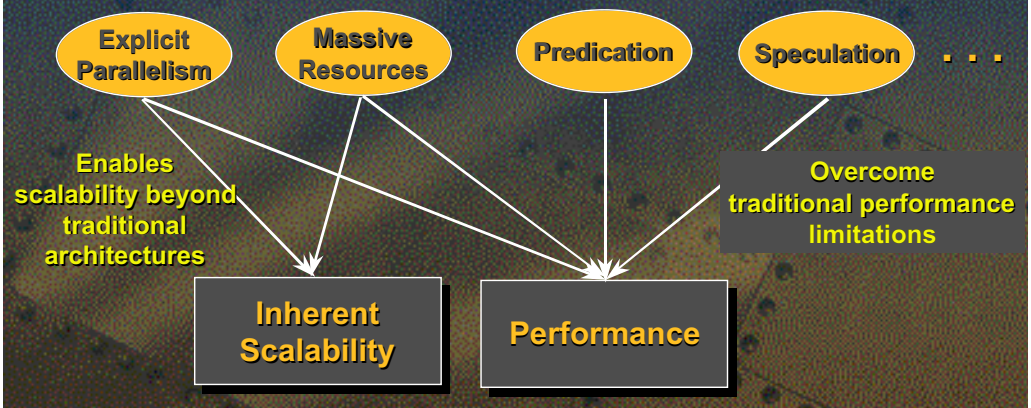
Greater Performance and Headroom Requires a New Approach . . .

… and more effective use of the hardware (silicon)

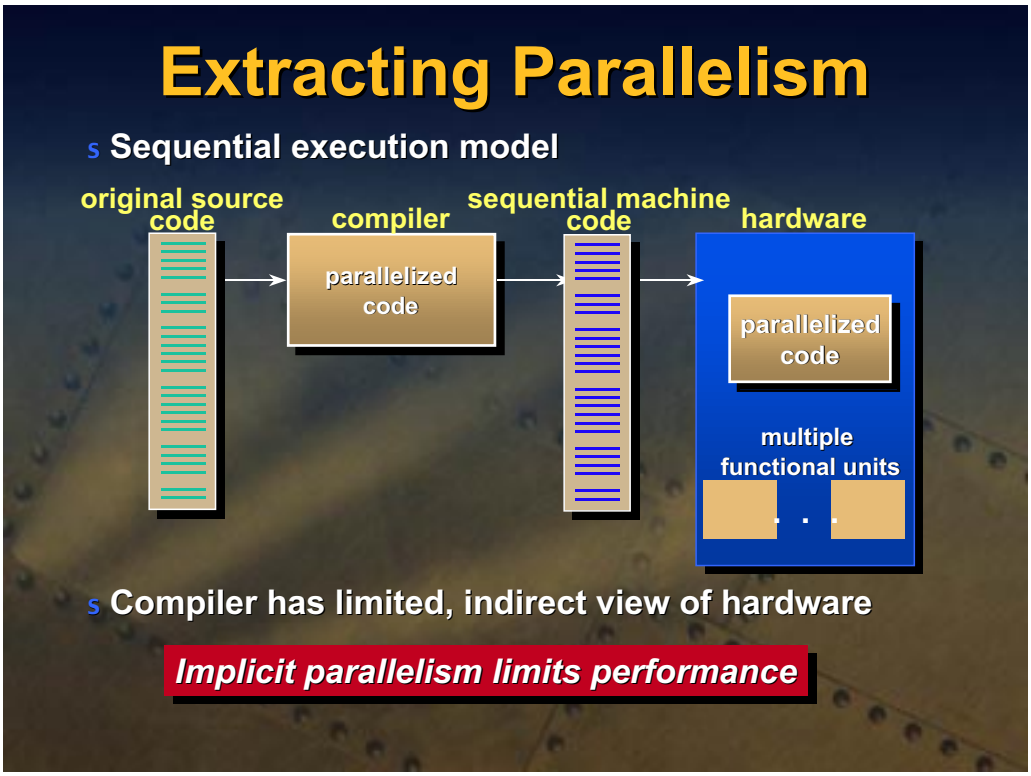
- Must explicitly schedule instructions
  - Advanced compiler technology
- Must enable greater parallelism
  - Next generation architecture technology
- Must be scalable across implementations and applications
  - Full compatibility

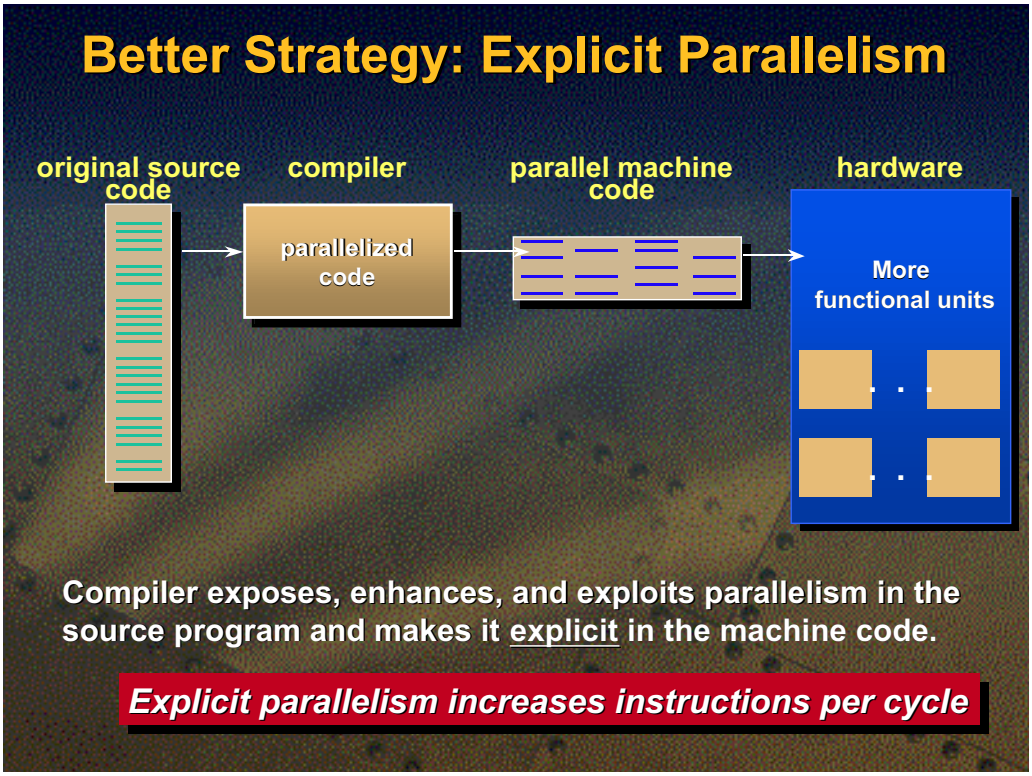
EPIC: Next Generation Architecture Technology

***E**xplicitly **P**arallel **I**nstruction **C**omputing*



IA-64: EPIC Technology Delivered


Extracting Parallelism
s Sequential execution model
original source code
compiler
sequential machine code
hardware
parallelized code
parallelized code
multiple functional units
s Compiler has limited, indirect view of hardware
Implicit parallelism limits performance


Better Strategy: Explicit Parallelism
original source code
compiler
parallel machine code
hardware
parallelized code
More functional units
Compiler exposes, enhances, and exploits parallelism in the source program and makes it explicit in the machine code.
Explicit parallelism increases instructions per cycle

Branches Limit Performance

if

Load a[i].ptr
p1, p2 = cmp a[i].ptr != 0
branch if p2

then

Load a[i].l
store b[i]
branch

else

Load a[i].r
store b[i]

i = i + 1

```
If a[i].ptr != 0
        b[i] = a[i].l;
else
        b[i] = a[i].r;
i = i + 1
```

Traditional Architectures: 4 basic blocks

Control flow introduces branches

Predication

if

Load a[i].ptr
p1, p2 = cmp a[i].ptr != 0
~~branch if p2~~

then

<p1> Load a[i].l
<p1> store b[i]
~~branch~~

else

<p2> Load a[i].r
<p2> store b[i]

i = i + 1

```
If a[i].ptr != 0
        b[i] = a[i].l;
else
        b[i] = a[i].r;
i = i + 1
```

Predication can remove branches

Predication Enhances Parallelism

Traditional Architectures: 4 basic blocks

if

```
Load a[i].ptr
p1, p2 = cmp a[i] != 0
jump if p2
```

then

```
Load a[i].l
store b[i]
jump
```

else

```
Load a[i].r
store b[i]
```

```
i = i + 1
```

EPIC Architectures: 1 basic block

```
Load a[i].ptr
p1, p2 = cmp a[i] != 0

<p1> Load a[i].l  <p2> Load a[i].r
<p1> store b[i]   <p2> store b[i]

i = i + 1
```

Predication enables more effective use of parallel hardware

Memory Latency Causes Delays

- **Loads significantly affect performance**
  - Often first instruction in dependency chain of instructions
  - Can incur high latencies

```
t1 = t1 + 1
If  t1 > t2
    j = a[t1 - t2]
    b[j] ++
```

Traditional Architectures

```
Add  t1 + 1
comp t1 > t2
branch
```

Barrier

```
Load a[t1-t2]
Load b[j]
add b[j] + 1
```

Load can cause exceptions

Speculation with EPIC Architecture

```
Add  t1 + 1
load.s a[t1-t2]
comp t1 > t2
jump
```

;Exception Detection

Propagate Exception

```
Check.s
Load b[j]
add b[j] + 1
```

;Exception Delivery

- **Separate load behavior from exception behavior**
  - Speculative load instruction (ld.s) initiates a load operation and detects exceptions
  - Propagate an exception "token" (stored with destination register) from ld.s to chk.s
  - Speculative check instruction (chk.s) delivers any exceptions detected by ld.s

Speculation Minimizes the Effect of Memory Latency

Traditional Architectures

```
Add  t1 + 1
comp t1 > t2
jump
```

Barrier

```
Load a[t1-t2]
Load b[j]
add b[j] + 1
```

EPIC Architecture

```
Add  t1 + 1
load.s a[t1-t2]
comp t1 > t2
jump
```

;Exception Detection

Propagate Exception

```
Check.s
Load b[j]
add b[j] + 1
```

;Exception Delivery

- **Give scheduling freedom to the compiler**
  - Allows ld.s to be scheduled above branches
  - chk.s remains in home block, branches to fixup code if an exception is propagated

Predication & Speculation

```
If a[i].ptr != 0
        b[i] = a[i].l;
else
        b[i] = a[i].r;
i = i + 1
```

With Predication

```
Load a[i].ptr
p1, p2 = cmp a[i].ptr != 0

<p1> Load a[i].l  <p2> Load a[i].r
<p1> store b[i]   <p2> store b[i]

i = i + 1
```

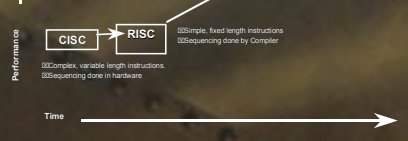
With Predication and With Speculation

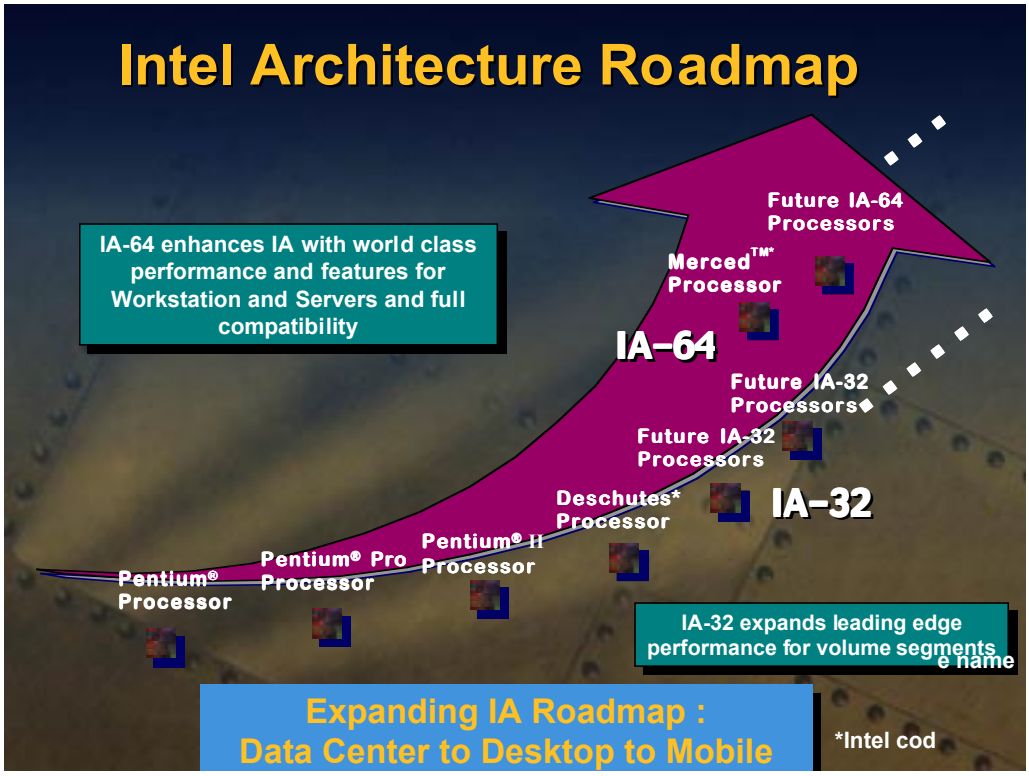
```
Load a[i]
load.s a[l].l load.s a[l].r
p1, p2 = cmp a[i] != 0

<p1> check.s     <p2> check.s
<p1> store b[i]  <p2> store b[i]

i = i + 1
```

EPIC: Next Generation Design Philosophy

Explicitly Parallel Instruction Computing

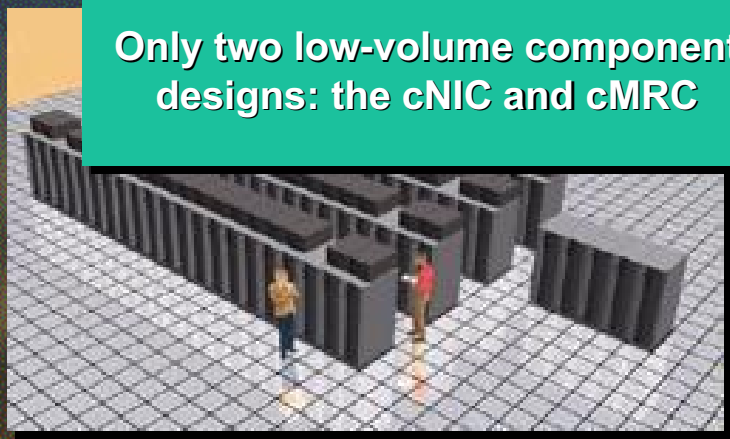
- **Explicit parallelism**
  - ILP is explicit in machine code
  - Compiler schedules across a wide scope
  - Binary compatibility across all family members
- **Features that enhance ILP**
  - Predication
  - Speculation
  - Others...
- **Resources for parallel execution**
  - Many registers
  - Many functional units
  - Inherently scalable




Intel Architecture Roadmap
IA-64 enhances IA with world class performance and features for Workstation and Servers and full compatibility
Future IA-64 Processors
Merced™* Processor
IA-64
Future IA-32 Processors
Future IA-32 Processors
Deschutes* Processor
IA-32
Pentium® II Processor
Pentium® Pro Processor
Pentium® Processor
IA-32 expands leading edge performance for volume segments
e name
Expanding IA Roadmap :
Data Center to Desktop to Mobile
*Intel cod

intel®

ASCI Option Red Machine



Only two low-volume component designs: the cNIC and cMRC

Virtual Interface Architecture

- Fully protected, direct user-level access to the network interface
  - exported by the VM sub-system
- *Illusion* of a dedicated network I/F
  - per thread, process, or application
- Applications transfer data without OS assistance
  - Zero copy

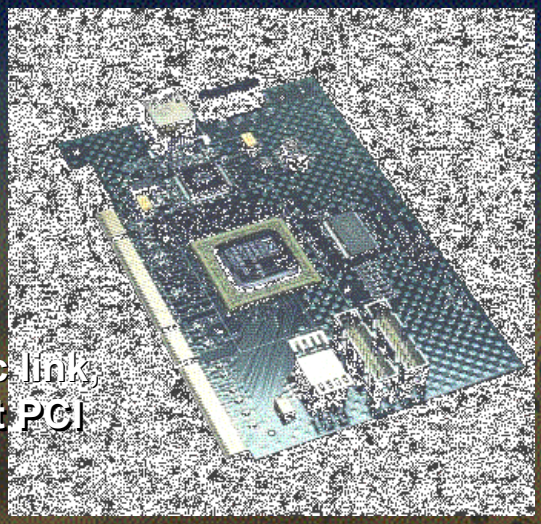
Rev. 1.0 Specification - End of 1997

VI Architecture Specification Update

- Version 1.0 of Specification Done
  - Released 12/16, publicly available
  - Find at http://www.viarch.org
- Development Tools Coming
  - Details: http://www.viarch-esg.org

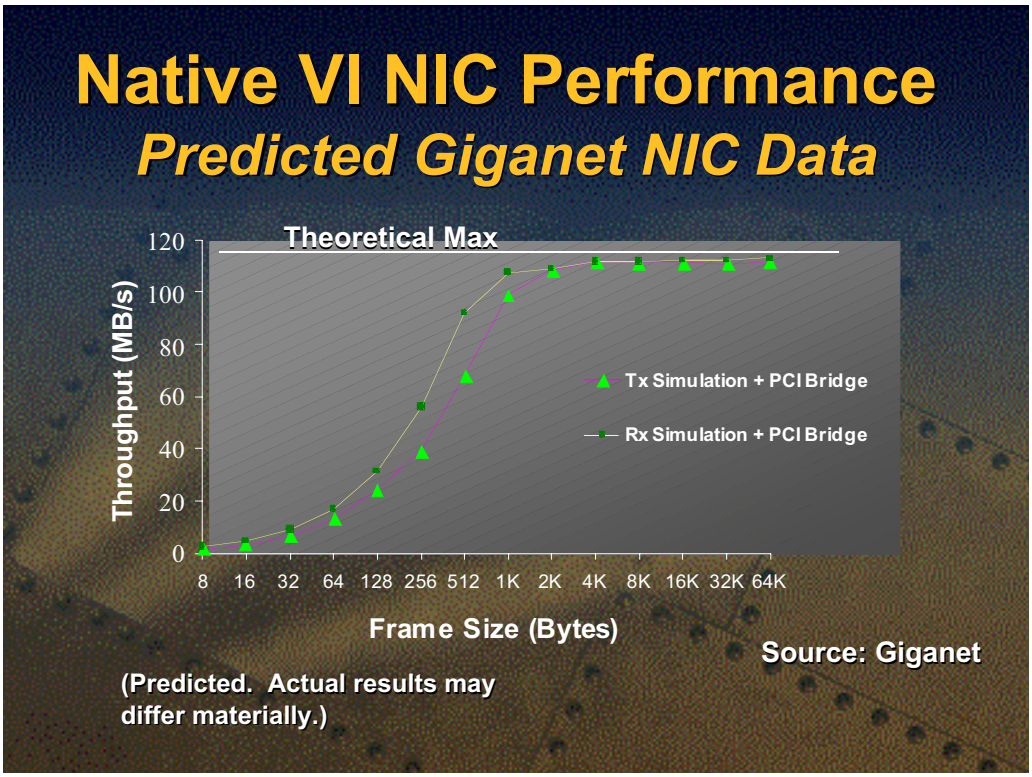
Get the Specs, Build the Products!

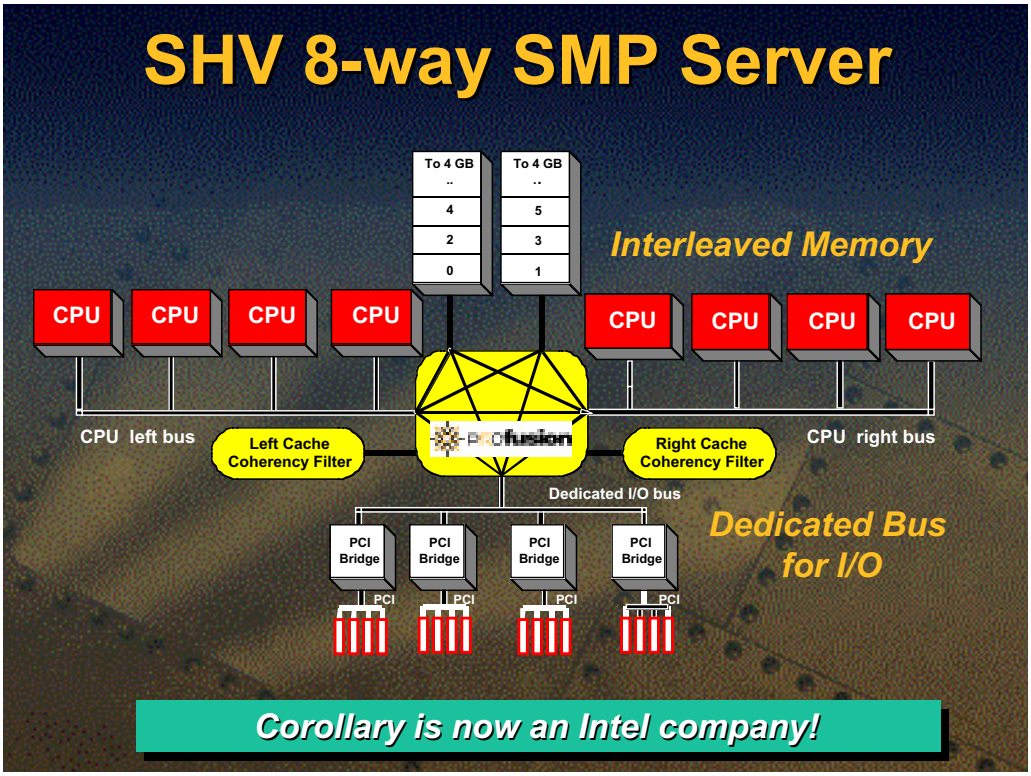
Giganet VI NIC



1Gbit/sec link,
32/64-bit PCI

Source: Giganet


Native VI NIC Performance
Predicted Giganet NIC Data
Theoretical Max
Throughput (MB/s)
120
100
80
60
40
20
0
8
16
32
64
128
256
512
1K
2K
4K
8K
16K
32K
64K
Frame Size (Bytes)
Tx Simulation + PCI Bridge
Rx Simulation + PCI Bridge
Source: Giganet
(Predicted. Actual results may differ materially.)


SHV 8-way SMP Server
To 4 GB
4
2
0
To 4 GB
5
3
1
Interleaved Memory
CPU
CPU
CPU
CPU
CPU
CPU
CPU
CPU
CPU left bus
Left Cache Coherency Filter
profusion
Right Cache Coherency Filter
CPU right bus
Dedicated I/O bus
PCI Bridge
PCI Bridge
PCI Bridge
PCI Bridge
PCI
PCI
PCI
PCI
Dedicated Bus for I/O
Corollary is now an Intel company!

High Performance Computing and the NCAR Procurement – Before, During, and After

Bill Buzbee	The Decision to Compete
Jim Hack	The Requirement for High Speed Processors in Climate Models
Steve Hammond	Highly Parallel, Microprocessor Systems: Performance and Other Considerations

Salishan Conference April'98 #1

The Decision to Compete

- January '92 visit with Seymour
  - Test load for Cray-3 OS
  - Showcase Cray-3 at NCAR
    - Installed Spring '93
    - Reliable and Popular by Fall '93
- Cray-4
  - Evolution of Cray-3
  - One GigaHertz Clock
  - Up to 32 Processors
  - Sticker Price[1]
    - $6M for Cray 4/8-512 (16 Gflop peak)

[1] Press Release, Cray Computer Corporation, Nov. 10, 1994

Salishan Conference April'98 #2

NCAR CCM2 Performance (64 bit) on Highly Parallel Systems (Gflops)

	TMC CM-5 *512 PE*	*Intel Paragon* *1024 PE*	*IBM SP-2* *128 PE*
T42	0.87	2.16	2.27
T170	2.30	3.18	2.40

Hack, J. J., J. M. Rosinski, D. L. Williamson, B. A. Boville, and J. E. Truesdale, 1995: Computational Design of the NCAR Community Climate Model. **Parallel Computing**, *21*, 1545-1569.

Drake, J., I. Foster, J. Michalakes, B. Toonen, and P. Worley, 1995: Design and performance of a scalable parallel community climate model. **Parallel Computing**, *21*, 1571-1591.

Hammond, S. W., R. D. Loft, J. M. Dennis, and R. K. Sato, 1995: Implementation and performance issues of a massively parallel atmospheric model. **Parallel Computing**, *21*, 1593-1619.

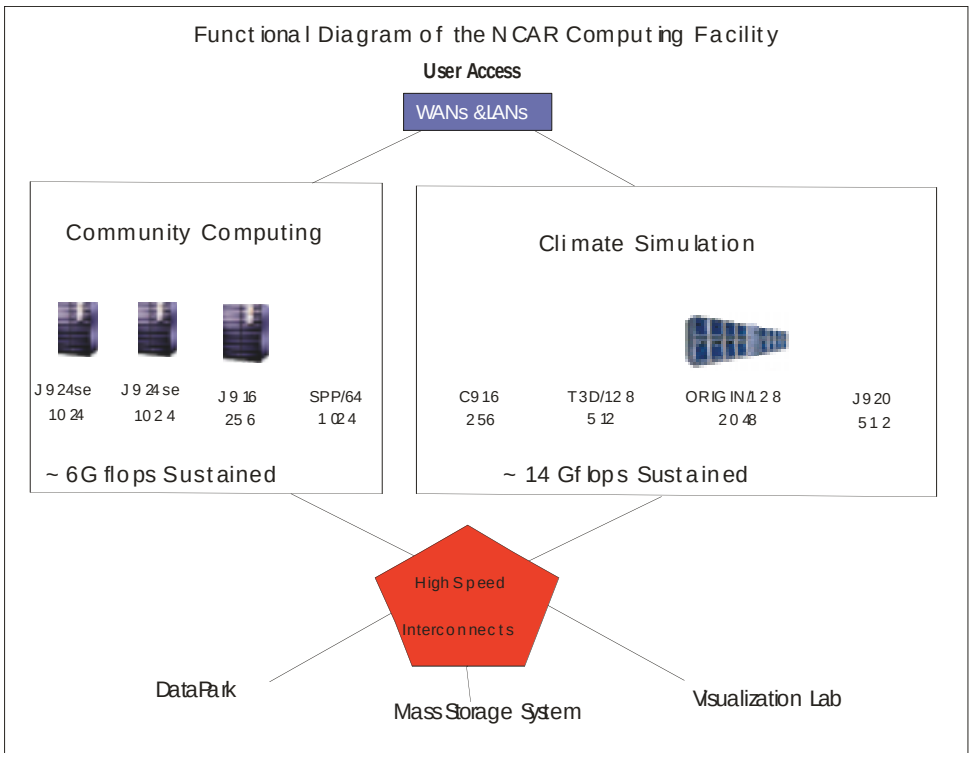
Rationale for SX-4 Selection

- Demonstrated dramatically superior performance (15-25 Gflops ***sustained*** in 1996!!)
- Thus, from the point of "What's best for science?", we had to go for it.
- Would have opened a new era in U.S. atmospheric modeling.
- Yes, we knew we were taking on the beltway, but there was always an outside chance of success.

Our Strategy

When the Department of Commerce (DOC) began its antidumping investigation in mid-96, NCAR adopted the following strategy:

1. Prepare to switch to highly parallel, Distributed Shared Memory (DSM) microprocessor systems, e.g. SGI Origin, HP Exemplar, etc.,
2. Continue acquiring U.S. manufactured Parallel Vector Processor (PVP) systems, and
3. Seek partnerships with national and international organizations.

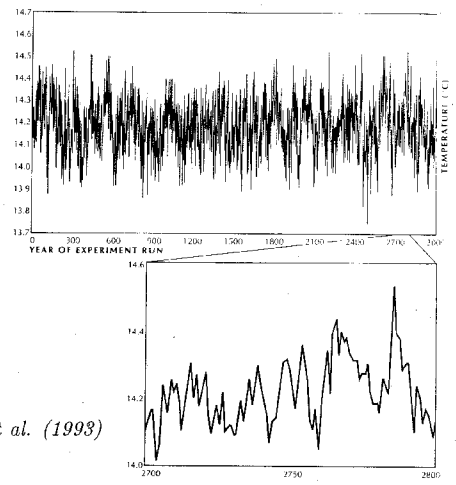


A check of realities shows:

- Leadership in high performance computing has moved offshore.
- Our international colleagues can, and are, doing simulations that cannot be done in the U.S.
- Highly parallel, microprocessor systems are progressing, but significant work remains to be done before they are ready for primetime computing.

Atmospheric General Circulation Models and Climate Simulation

- Reduced models of Climate System
  - apply "averaging operator" to governing equations
- Atmospheric General Circulation Models
  - simulate detailed "weather" fluctuations in the fluid system
  - day-to-day solution details are non-deterministic (Lorenz, 1962)
  - apply "averaging operator" to detailed solution sequence
  - *utility lies in prediction of future climate state statistical properties*
    chronological order of intermediate atmospheric states irrelevant

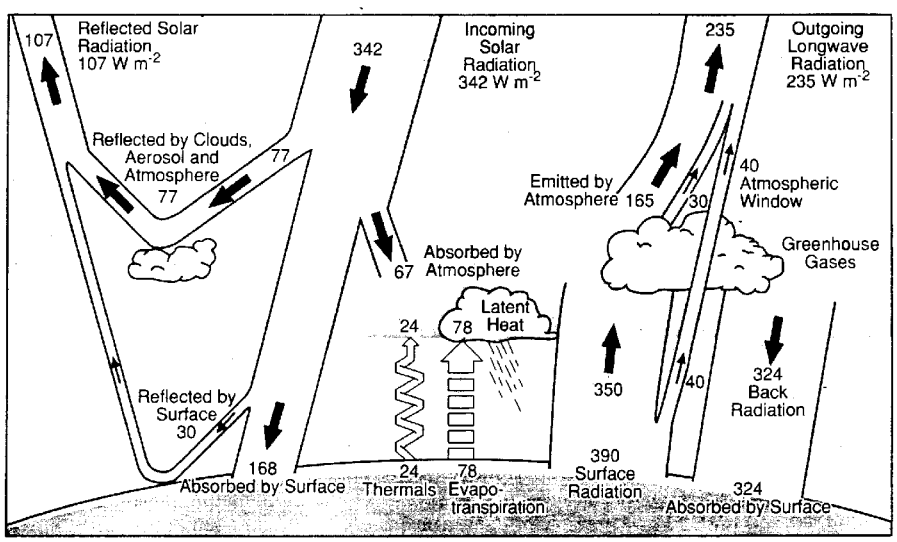


Hansen et al. (1993)

J. J. Hack

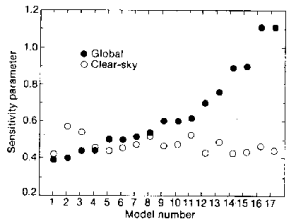
Fundamental element of global climate modeling

Accurately capture the details of the global energy budget!



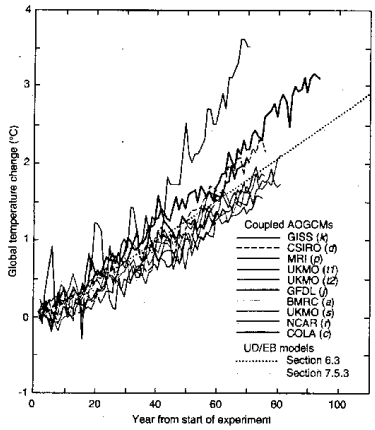
Climate Simulation Uncertainties

- Cloud feedback sensitivity



Cess et al. (1990)

- IPCC projections

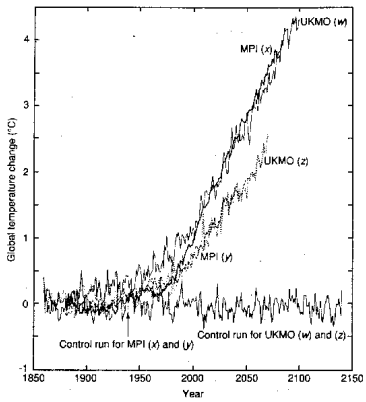


J. J. Hack

The role of sulphate aerosols on climate

- Natural sources
  volcanoes, deserts
- Anthropogenic sources (sulphur dioxide emissions)
  power plants, biomass burning
- Direct effects
  – conservative scattering of solar radiation
- Indirect effects
  – aerosol-induced changes to cloud albedo

Simulation estimates of the direct effects



J. J. Hack

Simulation requirements to examine the role of sulphate aerosol

- Physical parameterization enhancements
  - prognostic cloud water
  - imbedded sulphate chemistry component
  - $\Longrightarrow$ 60% increase in computational cost
- Simulation strategy to address global energy budget contributions
  - direct and indirect effects
    anthropogenic versus natural forcing
  - hierarchical evaluation of parametric complexity
  - ensemble approach to facilitate attribution
  - $\Longrightarrow$ 10's of numerical integrations required

***These types of studies will likely be completed by overseas colleagues*†**

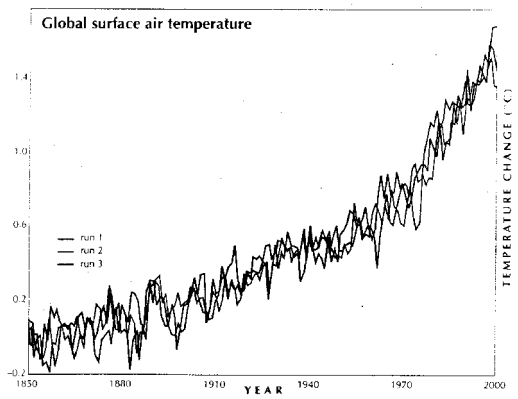
† *MP1 (80 GF); UKMO; (45 GF); ...*

J. J. Hack

Why ensembles?

- Large natural variability
  - required to address problems of detection and attribution

Example from Hansen et al. (1993)



J. J. Hack

Quantifying what we'd like to do (what our colleagues are doing)

- fully integrated chemistry
  - simple model for CO_2, CH_4, N_2O, CFC-11 and CFC-12, DMS, SO_2, SO_4
  - comprehensive model of stratospheric O_3

    $\Longrightarrow \sim$ 5X - 10X increase in operation count

- incorporation of a real middle atmosphere (standard configuration)
  - problem of vertical extent, resolution, and parameterized physics

    $\Longrightarrow \sim$ 4X increase in operation count

- increased resolution for standard configuration
  - T63L30; $\sim$ 200 km transform grid at equator

    $\Longrightarrow \sim$ 5X increase in operation count
  - nominal (2/3)°ocean

    $\Longrightarrow \sim$ 8X increase in operation count

- standard simulation strategy which inclndes ensembles
  - allow identification of sensitivities
  - allow attribution to physical/dynamical processes

J. J. Hack

Climate Modeling Considerations

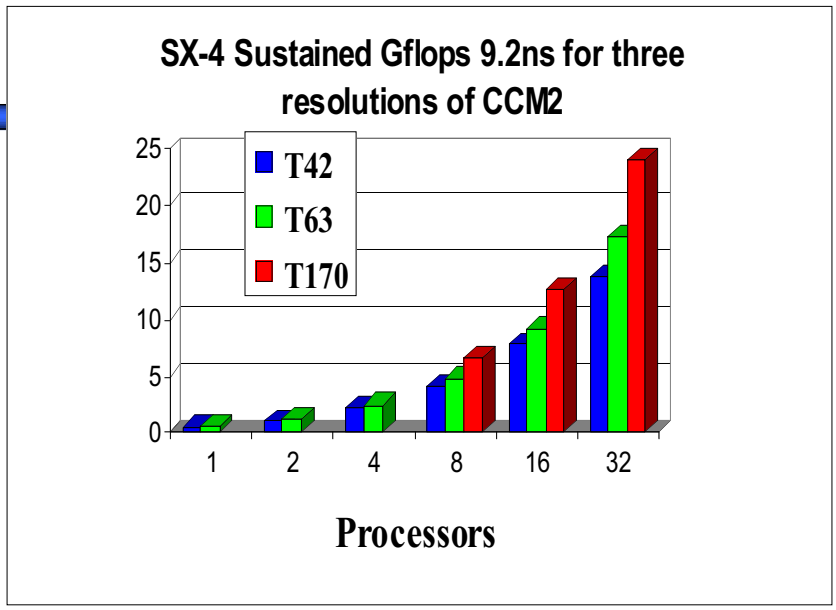
- 1 model year per day equals 365x reality
  - 3 months to simulate one century.
- 3 model years per day equals 1000x reality
  - 1 month to simulate one century.
  - currently stuck here.
  - C9/16 ~ 2000x reality for NCAR's CSM.
- 30 model years per day equals 10000x reality
  - 3 days to simulate one century.
  - Minimum capability: "sweet spot" for scientific productivity.

1

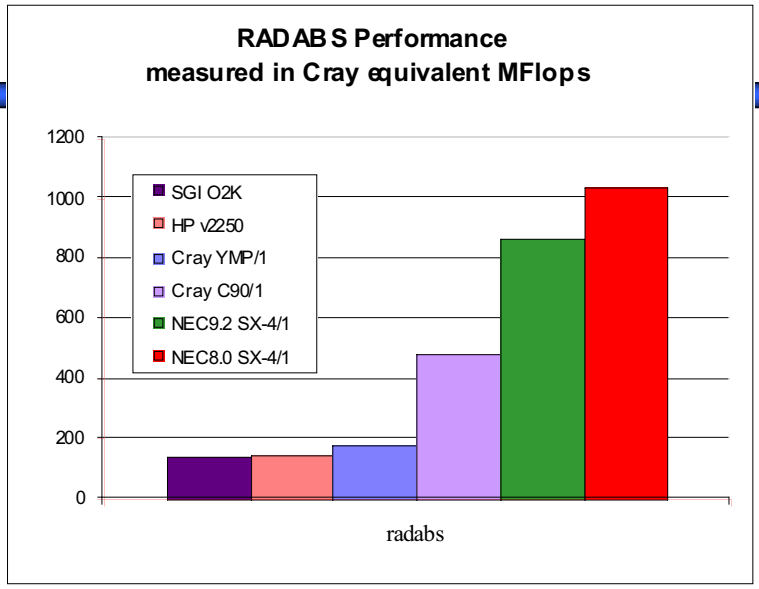
Climate Modeling Considerations – increased resolution

- What about increasing the resolution to increase available parallelism?
- Operation counts increase by $O(n^3)$ or more if increased vertical levels.
- Operation counts thus increase faster than exploitable parallelism.
- One must balance model resolution and complexity versus compute capability to remain in sweet spot.

2


SX-4 Sustained Gflops 9.2ns for three resolutions of CCM2
25
20
15
10
5
0
T42
T63
T170
1
2
4
8
16
32
Processors

3


RADABS Performance measured in Cray equivalent MFlops
1200
1000
800
600
400
200
0
SGI O2K
HP v2250
Cray YMP/1
Cray C90/1
NEC9.2 SX-4/1
NEC8.0 SX-4/1
radabs

4

Sustained Mflops/PE on SGI O2K, 195MHz

				8	16	32	64	128
CCM3.2	NCAR	Atm. GCM		78.5	71.8	61.4	56.0	
MM5	NCAR/PSU	Mesoscale			111.1	98.4	73.5	48.4
POP	LANL	Ocean GCM		51.6	54.6	53.0	42.4	
SC-MICOM	U. Miami	Ocean				85.3	74.6	59.8
MP-MICOM	U. Miami	Ocean		92.3	94.6	93.7	93.6	
CFL3DHP	NASA Ames	CFD		35.8	35.4			
TLNS3DMPI	NASA Ames	CFD			50.6	39.2		
Overflow	NASA Ames	CFD					51.5	
ARC3D	NASA Ames	CFD					98.4	
avg				64.6	69.7	71.8	70.0	54.1

Scientific Computing Division
National Center for Atmospheric Research

5

US HPC Observations

- Currently, must get US “supercomputing” from workstation vendors.
- Ratio of vector processor to microprocessor sustained performance is *O*(10:1).
- To achieve NEC-class capability (15-25 sustained Gflops) requires a system with at least *O*(300) microprocessors.
- Add full atmospheric chemistry and increased model sophistication and need *O*(1000) processors to stay in sweet spot.

Scientific Computing Division
National Center for Atmospheric Research

6

Scalable Systems with 100s to 1000s of PEs

- Need **scalable systems** to be able to answer what-if questions and support policy decisions.
- Not where the US "supercomputing" vendors are headed:
  - Loosely clustered 64 to 128-way DSMs.
- Significant penalty in bandwidth and latency when running a single application across cluster relative to within a single DSM.
- Some codes that scale well on scalable computers will not scale on clusters.
  - Lots of short messages.

Scientific Computing Division
National Center for Atmospheric Research

7

Scalable Systems, cont.

- Climate Modeling
  - Need to run very fast at relatively low resolutions.
  - Low resolution and large numbers of PEs means small messages.
- Implications: clustering network bandwidth and latencies must scale with the rest of the system.
- Need focused effort to improve clustering interconnect.

Scientific Computing Division
National Center for Atmospheric Research

8

Other Challenges

- Tremendous turnover in HPC industry.
- HPC Industry amnesia ….
  - Earlier lessons learned lost as companies go out of business.
  - CM-5 had integrated/scalable tools, rich scientific subroutine libraries, scalable I/O, etc.
- Currently, odds of things working within a DSM are good.
- Odds of these same facilities working across DSMs are poor.
- Print statements are still the best debugging tools.

9

Some Reasons for Optimism

- Today's DSMs are vastly superior to MPPs of a few years ago
  - comparable MPI/shared memory performance.
  - no per processor memory limitations.
  - don't have to parallelize code as very first step to get started.
- Industry (ASCI) focus on improving interconnect.
- Tools for large systems are improving.
- **Still a lot of work to do.**

10

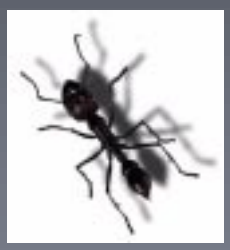
Economies of Scale
HPC into the Next Millennium
Gary Smaby
CEO/Principal Analyst
SMABY GROUP

Liberal Artistic License
SMABY GROUP

Game Plan

Supercomputing

- ✓ ‘What It Was Ain’t What It Is’
- ✓ Fifty Years of Evolution
- ✓ Mainstreaming HPC
- ✓ 2001 and Beyond
- ✓ Q & A



Conference on High Speed Computing - April 23, 1998
Salishan Lodge

SMABY GROUP

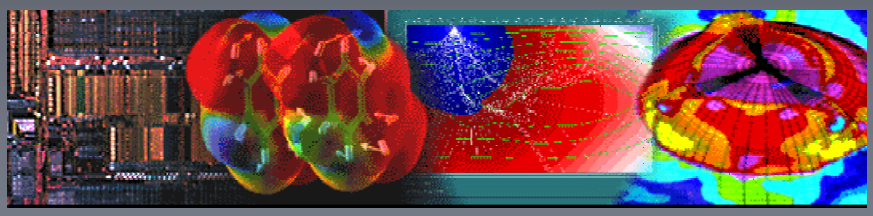
What *Was* a Supercomputer?

The world’s fastest computer
. . .at any given point in time.

SMABY GROUP

Today's Supercomputer

A platform-independent computational visualization tool for the imagination . . . . a time machine for simulating natural phenomena.

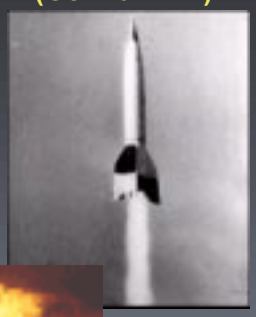


SMABY GROUP

Original Apps

✓Calculate trajectories

(German V-2)



✓Crack codes

(Colossus)

✓Design weapons

(Manhattan Project)

SMABY GROUP

The Evolution of Supercomputing The Forties

von Neumann defines first stored-program computer . . .

Grace Hopper discovers the first computer bug - a moth stuck between relays

1945

ENIAC - first digital computer fills 30x50 room, weighs 30 tons, 18,000 vacuum tubes, 1000 memory bits, 7500 ops

1946

Semiconductor revolution begins - transistor soon supplant vacuum tube

1947

SMABY GROUP

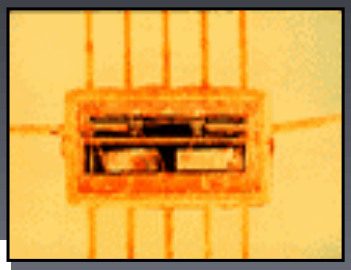
"Computers in the future may weigh no more than one-and-a-half tons."

Popular Mechanics (1949)

SMABY GROUP

The Evolution of Supercomputing The Fifties

Two start-ups form:
Digital Equipment
and
Control Data



First commercial FORTRAN program

1957

Texas Instrument builds first integrated circuit

1958

L.R. Johnson coins the term “architecture” to describe the IBM 7030 “Stretch”.

1959

SMABY GROUP

The Evolution of Supercomputing The Sixties

Digital’s PDP-1 is cheap ($120,000) and small (250 lbs.).

1960

Seymour bulds world’s fastest “super” computer - CDC 6600 - 3X faster than IBM’s 1 MIPS Stretch

1964

UNIX is developed by Bell Labs on a spare VAX

1969

SMABY GROUP

"But what the hell is it good for?"

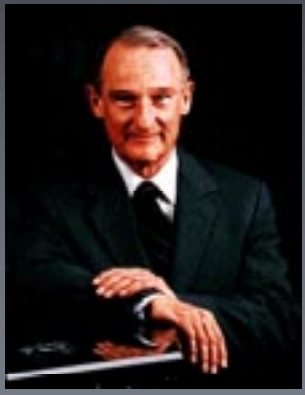
Anonymous IBM Engineer
Commenting on the computer chip (1968)

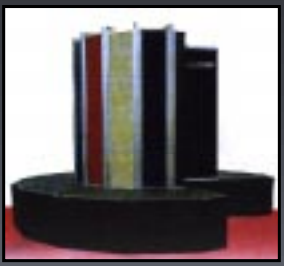
SMABY GROUP

The Evolution of Supercomputing — The Seventies

Seymour Cray leaves Control Data to start Cray Research. . .

1972





Cray-1 debuts. . . all R&D costs recovered in first sale to Los Alamos

1976

SMABY GROUP

The Evolution of Supercomputing: The Eighties

IBM introduces the PC

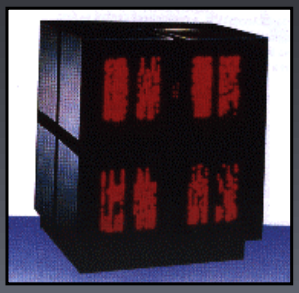
1981

Cray-XMP deploys parallelism to double speed

1982

CRAY-2 packs six miles of wiring in 4' tall chassis. . . Japanese enter the fray. . . minisupers debut

1985



Thinking Machine's 16,000 CPU MPP executes one *gflops*

1986

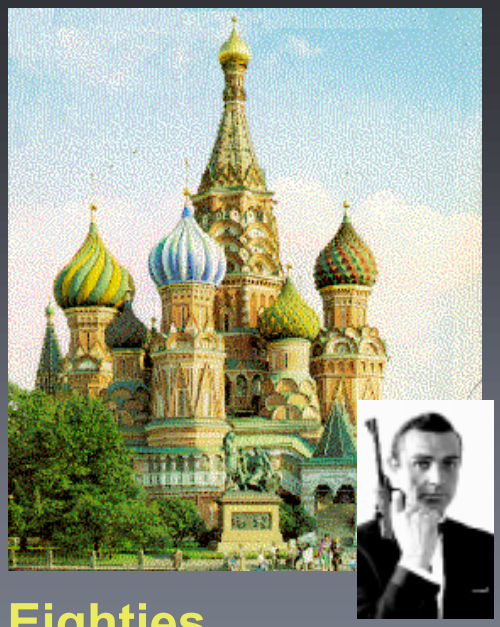
SMABY GROUP

The Evolution of Supercomputing: The Eighties

Other People's Money

SMABY GROUP

Collapse of the Evil Empire

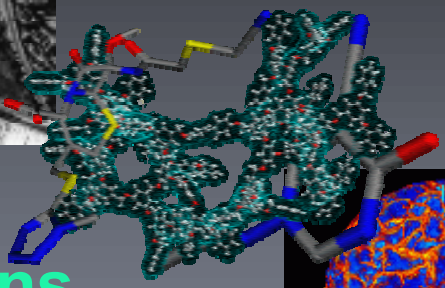


The Evolution of Supercomputing **The Eighties**

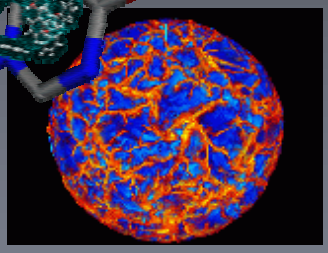
SMABY GROUP

The Evolution of Applications

From Nukes & Spooks

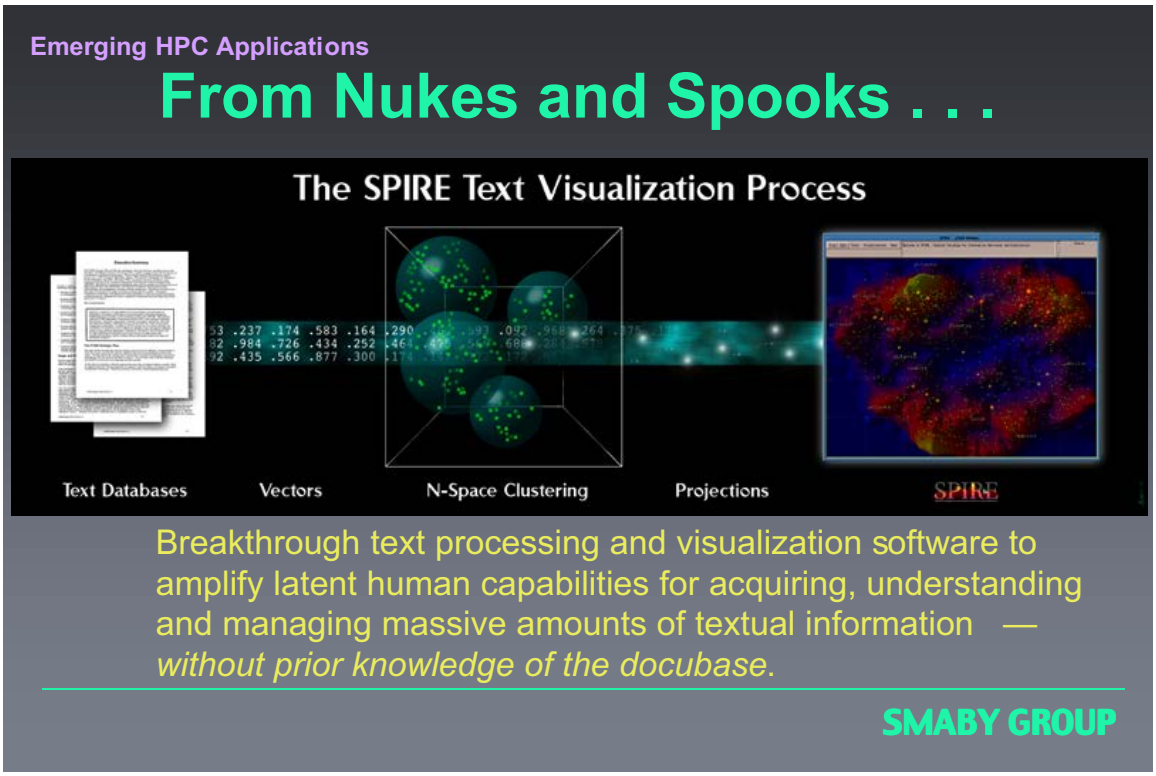


To Genes & Greens

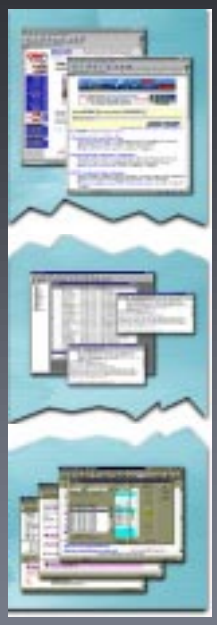


SMABY GROUP

From Nukes and Spooks . . .
SPIRE
Spatial Paradigm for Information Retrieval and Exploration
SMABY GROUP


Emerging HPC Applications
From Nukes and Spooks . . .
The SPIRE Text Visualization Process
Text Databases
Vectors
N-Space Clustering
Projections
SPIRE
Breakthrough text processing and visualization software to amplify latent human capabilities for acquiring, understanding and managing massive amounts of textual information — without prior knowledge of the docubase.
SMABY GROUP

Emerging HPC Applications



Competitive Intelligence

A systematic business methodology for gathering and analyzing information about your competitors, customers, suppliers, and markets to further your own company goals.

SMABY GROUP

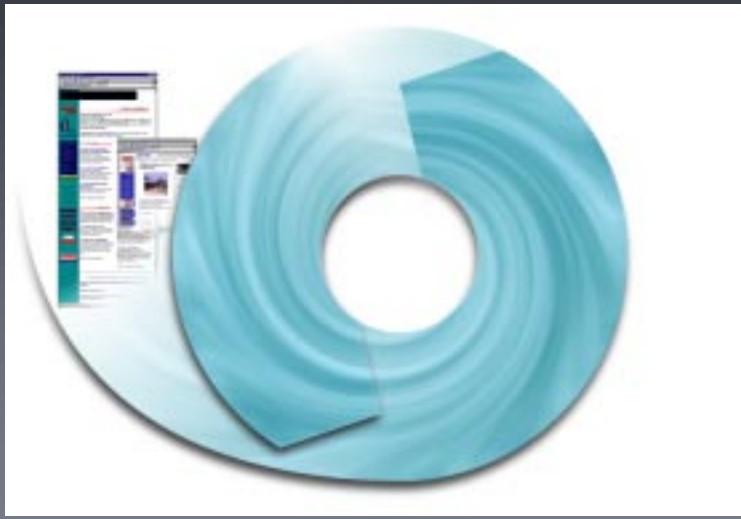
ForeKnowledge™

Self-organizing
and Interactive
Web Content Maps

SMABY GROUP

Common Knowledge™

A self-organizing knowledge reservoir that dynamically receives and maps intellectual corporate assets sent to it via email.

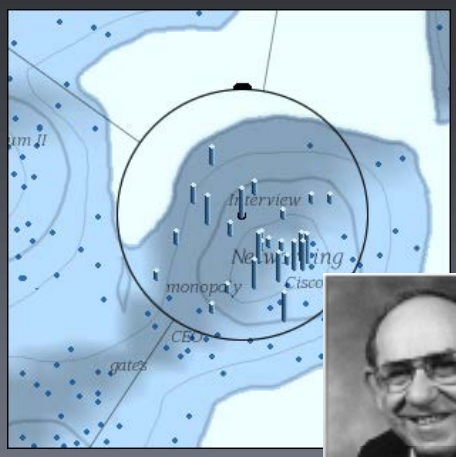


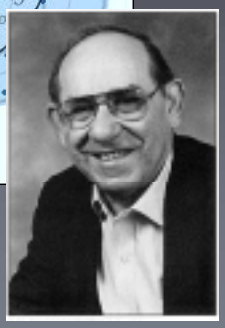
SMABY GROUP

Emerging HPC Applications

Competitive Intelligence

"You can see alot by watching."



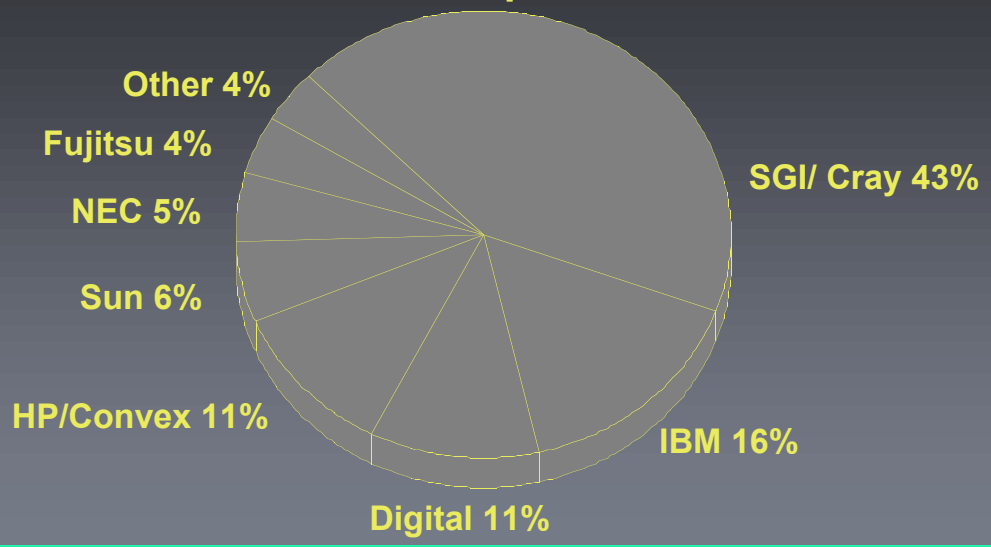


Yogi Berra

SMABY GROUP

Market Scale

1997 - $2.18B



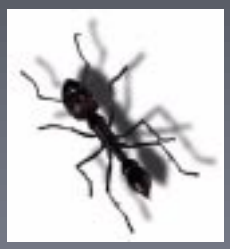
Scientific/Engineering/Technical HPC Market Size - Worldwide

SMABY GROUP

The Evolution of Supercomputing

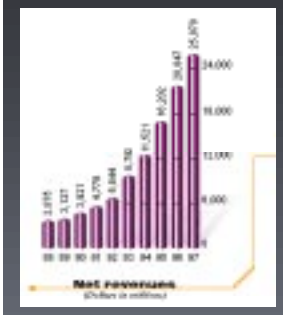
Predictions: 2001 and Beyond

1) The Ants Prevail

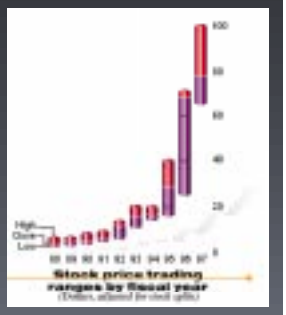


SMABY GROUP

Economies of Scale



intel



Merced Chip (P7)

Debuts in 1999 at 600 Mhz
SPECfp95 > 100
8 Instructions @ Cycle
Scales 1Ghz to 4Ghz
Design/Fab Cost >$2.5B
Cost to Produce ~$100
Intel Market Cap ~$125B

SMABY GROUP

The Evolution of Supercomputing

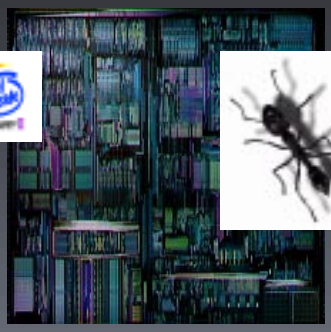
Predictions: 2001 and Beyond

2) Cray Makes Like Harley

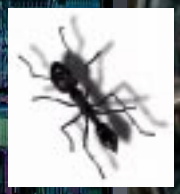
SMABY GROUP

The Evolution of Supercomputing: Today

Cray's Origin2000 scalable system stacks processors like LEGO blocks to achieve increased performance



Emergence of Intel's off-the-shelf Pentium and IA-64 processors

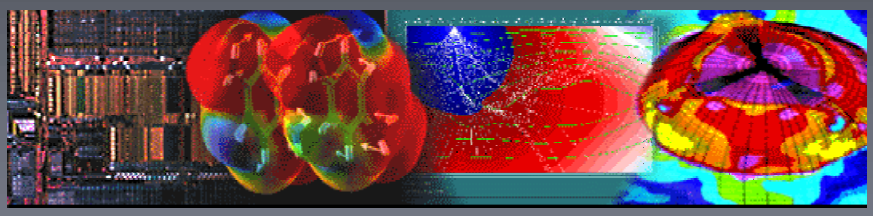


Sandia links 9072 Intel Pentium Pro chips to top one *teraflops*

SMABY GROUP

The Evolution of Supercomputing
Predictions: 2001 and Beyond

3) Apps Rule



SMABY GROUP

Hot HPC Apps

- ✓Model new drugs
- ✓Design exotic securities
- ✓Map the human genome
- ✓Create Hollywood special effects
- ✓Web-based knowledge management
- ✓Designing stuff (cars, airplanes,golf clubs)

SMABY GROUP

The Evolution of Supercomputing

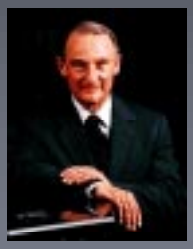
Predictions: 2001 and Beyond

4) Money Still Talks

"The real limitations come from money, not physics."

SMABY GROUP

PETAFLOP COMPUTING?



"We don't have too far to go (with circuit technology) until we get to the size of biological molecules . . . I think we'll be coming face-to-face with the life force."

Seymour Cray (5/30/96)

SMABY GROUP

The Evolution of Supercomputing

2001 and Beyond

"The future isn't what it used to be."

Arthur C. Clarke

SMABY GROUP

Q & A
or write me at
gary@smaby.com
SMABY GROUP

Conference Presentation by William Trimmer
The Other Side of Computing by William Trimmer Belle Mead Research, Inc.

Phone 908 359 0012, Fax 908 359 2094 58 Riverview, Belle Mead, NJ 08502 USA
Web trimmer.net, Email W.Trimmer@IEEE.org

Abstract

Computation takes us to a world of its own. An intangible world where incantations create, modify, destroy, and build castles in the sky. In this world huge power can be generated, information neatly trimmed for a purpose, and couriers sent to other castles at nearly the speed of light.

What computation and software can not do is interact with the real world. Interaction with the tangible takes forces and photons and electromagnetic fields and sensors.

Information and computational ability is intriguing and of great utility, but it can not tie your shoe, or fry an egg.

The original ingenious and intelligent computing systems were mechanical, things such as clocks that chimed and displayed dancing figures on the hour. Moving electrons in micro computers are now doing a superb job of providing the intelligence. Complex calculations and decisions are inexpensive. It is now the mechanical devices needed to interface electronics to the world that are expensive.

But things are about to change. This talk discusses the proliferation of micro mechanical devices. A complete gear syste
m, or a motor, or a chemical sensor can be made that is 1/1000 the size and cost of a CPU or memory chip. A revolution in our ability to interface computing power to the real world is about to begin.

Computation is to micro mechanical devices as Yin is to Yang. One is soft and intangible, the other active and tangible. The real promise is in their combination.

Prelude

There are more things in heaven and earth, Horatio, than are dreamt of in your philosophy. [Ref 1]

Alessandro Volta, an Italian Physicist (1745 to 1827), experimented with dissimilar metals in aqueous solutions, and found a weak source of current. Andre Marie Ampere, a French Physicist (1775 to 1836), passed currents through wires and discovered the relationship between currents and magnetic fields. Imagine now, if you can, these gentlemen walking through your house. What wonders we take for granted. Did their genius prepare them to comprehend the computer on your desk?

The world of science and engineering is still unfolding wonders. One recent development is the ability to make micro mechanical devices. Motors the diameter of a human hair and sensors the size of a grain of salt are enabling inexpensive systems that can interact with our world on a new scale.

Surprisingly, the time scale from conception to utilization has been collapsing. Nikola Tesla and Thomas Alva Edison developed practical electric motors in the 1880's, about a hundred years after Volta and Ampere's work. The micro comb drive motor was described in 1989, and is currently being used in automobiles as an airbag sensor. This micro development took less than a dozen years from conception to full scale implementation.

There are several reasons for the rapid advances of micro devices (also called MicroElectroMechanical Systems, MEMS, micromechanics, Micro System Technologies, mst, and Micro Machines). One, manufacturers of these micro devices are using well established technologies. The electronics industry provides sophisticated process chambers for making micro comb drive motors and a host of other micromachined devices. The machine tool industry has developed micro Electro Discharge Machining, EDM, techniques and single point diamond machining tools that can fabricate minuscule mechanical devices. The plating and molding communities have enabled LIGA and other replication processes that can replicate extremely fine structures. Two, there are a number of people skilled in these base technologies who are looking for new challenges. And three, established companies and the investment community recognize the potential of new technologies. The micro technologies are exploding forth using a well established infrastructure.

Volta and Ampere worked quietly in their labs. Edison worked with a small group of researchers. Currently I estimate there are 10,000 people working on micro-mechanical projects. Much of this current effort is developing products.

While there are more things in the universe that one can imagine, we can push the boundaries. Your help exploring the applications and engineering and science of micro devices is welcome.

Introduction

Micromechanics is an extremely broad field, a field that will touch most aspects of our grandchildren's lives. This field encompasses all of the current technologies -- only it is concerned with a smaller dimensional scale. And micromechanics promises applications in all disciplines. Richard Feynman well conveyed the excitement of our new discipline:

"I imagine experimental physicists must often look with envy at men like Kamerlingh Onnes, who discovered a field like low temperature, which seems to be bottomless, and in which one can go down and down. Such a man is then a leader and has some temporary monopoly in a scientific adventure. Percy Bridgman, in designing a way to obtain high pressures, opened up another new field and was able to move into it and lead us all along. The development of ever higher vacuum was a continuing development of the same kind.

"I would like to describe a field, in which little has been done, but in which an enormous amount can be done in principle. This field is not quite the same as the others in that it will not tell us much of fundamental physics (in the sense of, "what are the strange particles?" but it is more like solid-state physics in the sense that it might tell us much of great interest about the strange phenomena that occur in complex situations. Furthermore, a point that is most important is that it would have an enormous number of technical applications.

"What I want to talk about is the problem of manipulating and controlling things on a small scale." [Ref2]

Below is discussed the earlier perceptions that hindered the development of small mechanisms, our current field and needs, and a look at our future.

The Genesis

Perhaps things normally start small, and grow. Man's habitats have grown from houses, to buildings, to skyscrapers. Our ability to travel has increased from a few miles on foot, to horses, to trains, and now we can encircle the world in a few days. Individually we work to make large accomplishments in hopes of enormous success. We are enthralled with the big and sign)ficant and substantial.

The insignificant, insubstantial, and minuscule is usually beneath our concern.

And yet.

A dozen years ago, I was trying to persuade a machinist to build a very small structure. He listened patiently for awhile and then said, “Why do you want something small, a toy? I can make you something that is big and good.” In his mind, most people’s minds, small things were cheap and no more than a toy. When H. A. Rowland (1848 to 1901, professor of physics at the Johns Hopkins University, Baltimore) went to make very small and accurate grooves for diffraction gratings, he used large machines and buried them in even larger vaults for thermal stability. Ten years ago an eminent colleague at Bell Laboratories looked me in the eye, and said, “Your micro things will never amount to anything. Large objects will always do a better job at a lower cost. “ This was very strongly the feeling at this time.

Even Feynman responded with good natured jesting to critics of small machines. In his famous talk There’s Plenty of Room at the Bottom, given at the American Physical Society meeting in 1959 he says “What would be the utility of such machines? Who knows? Of course, a small automobile would only be useful for the mites to drive around in, and I suppose our Christian interests don’t go that far.” [Ref 3] And in his 1983 talk, Infinitesimal Machinery, at the Jet Propulsion Laboratory he says, “I also talked in the 1960 lecture about small machinery and was able to suggest no particular use for the small machines. You will see there has been no progress in that respect. “ [Ref 4]

Originally, the ingenious and intelligent systems were mechanical, things such as clocks that chimed and displayed dancing figures on the hour. Electronics is now doing a superb job of providing this intelligence. Complex calculations and decisions have now become inexpensive. Presently, it is the mechanical devices required to interface electronics to the world that are expensive. Fortunately the new micromechanical devices integrate well with electronics: one providing the intelligence and one providing the hands.

Electronics has led much of the recent development of micromechanical devices by providing many of the tools and techniques, making the rapid advances possible. This partnership is to great advantage.

Surprisingly, mechanical systems can now be smaller and less expensive than electrical systems.

There is an increasing breadth of microfabrication techniques enriching our capabilities. Examples include LIGA, EDM, precision machining, plating, and molding. To ignore the wide range of fabrication techniques available is to limit oneself.

Yet how did things insignificant in size gain a purpose?

Perhaps Johann Gutenberg gave an indication of the usefulness of small mechanical devices. Gutenberg means good mountain, and indeed, in 1456 he set in motion a mountain of small mechanical devices (individual movable type) for the good of mankind. One interesting aspect of his work is the interchangeability - a few standardized units are made that can be combined to meet most needs. This concept may be useful for our micro devices.

Ideas, excellent ideas, often seem to gain a life of their own. Like grass growing through the pavement, they seem to search for fertile minds and the correct opportunities. Often key ideas are invented in several different places. For example, Pi Sheng of China made movable type of Chinese characters from clay in 1040. And Korea molded metal type in sand in 1361. Gutenberg was just one of several who expressed the idea of moveable type. If you have a good idea, I encourage you to develop it now; there is a high probability others share your idea.

Until recently, minute mechanical systems have developed at a stately pace. For years the watch makers' art has represented the limits of our micro excursion. And the practitioners ofthe watch industry have succeeded admirably. For example, the motor in a wrist watch has high efficiency, runs for years (even after being dropped), and costs less than a cup of coffee. Yet, when I was talking with a gentleman who had designed many of the watches we wear, he said, "I have spent my life trying to make smaller mechanisms, and when you show me something really smaller, I do not know what to do with it." This was a common response to motors the diameter of a human hair.

The Present

The rapid race to more clever micro machines has just begun.

The earlier disdain for the small and insignificant is gone. Now there is a growing excitement about the micro.

Gone are my fears that the micro field would grow on 'isn't that neat' and then die when no purpose was found. Enough people now recognize the importance of micro science and engineering and product development to ensure the field.

Things insignificant in size do have a grand purpose.

Yet it is difficult to realize the breadth of this field of micro mechanical devices.

Most advances represent a specific technology. The Scanning Tunneling Microscope for example, gives us the ability to detect and perhaps manipulate atoms. High temperature superconductors hold the promise of efficient power transmission and novel electronic

circuits. The diesel engine gives us a source of mechanical power. Each of these is an important advance of a single thing.

The field we are contemplating here today is vast beyond our normal concerns. It is the science and engineering and development and commercialization of a whole new realm of human enterprise.

I defy you to think of a large scale, macro discipline in science or engineering that does not have a small scale, micro equivalent. Your challenge, should you decide to accept it, is helping to image the macro into the micro.

The Future

What will the field of micro mechanical devices be like in a thousand years?

The first thing to do after such a question is to recover from the shock of being asked.

Yet, it is useful to ponder the potential.

To stir the debate, several predictions are given below.

One, micromechanical devices will become omnipresent. They will fill the niches of our lives. Can you find fundamental limitations that will keep microdevices from becoming inexpensive and readily available? If not, why not their proliferation?

Clayton Teague of the National Institute of Standards and Technology in the U.S. gave an interesting presentation [Ref 5] on Feynman's tiny hands. Feynman proposed small hands manufacturing smaller hands, which in turn manufacture even smaller hands. This ever smaller procession of tiny hands can then be used to manufacture large numbers of useful micro devices. In this talk, Clayton Teague also discussed John von Neumann's conjecture on the self replication of complex systems. [Ref 6] At what point can our micro devices start to self replicate themselves? A self replicating micro system needs careful consideration.

Two, unless there is a need for something to be large, it will be small. There are many reasons for this. Material costs are less for small systems. The systems use less space, and small systems can perform functions more rapidly. Because of the small size and low cost, multiple systems can be used for one function, increasing the robustness and reliability. Because of their size, these microsystems can be dispersed; instead of the large devices used now, many micro devices can be used to give a finer "rained sensing and manipulation of our world. Very few things need to be large. But hopefully in the year 3000, dinner will still be large.

Three, the worlds of the micro (millimeter to micron) and the nano (micron to Angstrom), electronics, and genetic engineering will evolve into closely interrelated fields. Already micro devices are helping to image and handle molecules, and nano technology is making small tubes and balls that hold the promise of mechanical structures.

Conclusion

As micro electronics has made possible much of the micro mechanical revolution, so now micro mechanical devices will extend the reach of electronics, and carry electronics into new places.

Micro computers & computation are to micro mechanical sensors & actuators as Yin is to Yang. One is soft and intangible, the other active and tangible. The real promise is in their combination.

Please join us in this great adventure.

William Trimmer

References

1. William Shakespeare, Hamlet, Prince of Denmark Act I, Scene V.

2. “There’s Plenty of Room at the Bottom,” by R. Feynman, Caltech’s Engineering & Science magazine, February, 1960. (Reprinted in Micromechanics and MEMS: Classic and Seminal Papers to 1990, Edited by W. Trimmer, the IEEE Press PC4390-QCL, ISBN 0-7803-1085-3, January 1997, page 3.)

3. “There’s Plenty of Room at the Bottom,” by R. Feynman, Caltech’s Engineering & Science magazine, February, 1960. (Reprinted in Micromechanics and MEMS: Classic and Seminal Papers to 1990, Edited by W. Trimmer, the IEEE Press PC4390-QCL, ISBN 0-7803-1085-3, January 1997, page 7.)

4. “Infinitesimal Machinery, “ by Richard Feynman, Journal of Microelectromechanical Systems, Volume 2, Number 1, March 1993. (Reprinted in Micromechanics and MEMS: Classic and Seminal Papers to 1990, Edited by W. Trimmer, the IEEE Press PC4390-QCL, ISBN 0-7803-1085-3, January 1997, page 11.)

5. “1/N Feynman Machines as a Path to Ultraminiaturization?” by E. Clayton Teague, SPIE Conference on Microelectronic Manufacturing, Austin, Texas, October 23 to 24, 1995.

6. “Theory of Self-Reproducing Automata,” Edited by John von Neumann and completed by Arthur W. Burks, University of Illinois Press, Urbana, IL, 1966.

Portions of this talk are adaptations from the following sources:

a. “Grand in Purpose, Insignificant in Size,” by William Trimmer, given at the Tenth Annual International Workshop on Micro Electro Mechanical Systems, Nagoya, Japan, January 26 to 30, 1997, please see page 9 of the proceedings.

b. “Large Prospects for Small Systems Using Microelectromechanics,” by William Trimmer, given at GOMAC ’98, March 16 to 19, 1998, Arlington, Virginia, USA.

c. Editorial by William Trimmer in the Journal of Microelectromechanical Systems, Volume 6, Number 4, December 1997, pages 290 to 293.

Reference Sources

Micromechanics and MEMS, Classic and Seminal Papers to 1990, Edited by William S. Trimmer, IEEE, Piscataway, New Jersey, ISBN 0-7803-1085-3, IEEE Number PC4390, 1997. For a well written introduction to this field see the papers in Chapter 1.

The Journal of Microelectromechanical Systems, an IEEE / ASME Publication on Microstructures,

Microactuators, Microsensors, and Microsystems.

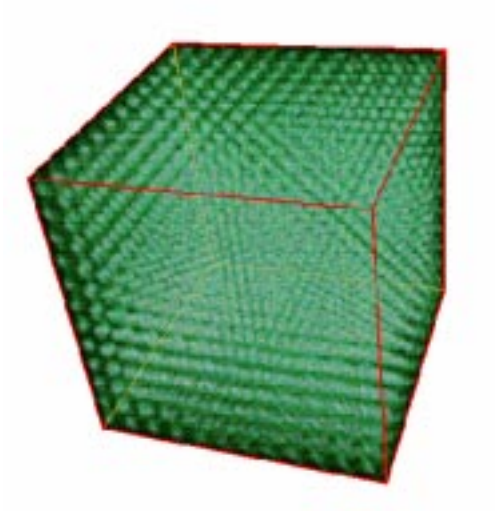
The Journal of Micromechanics and Microengineering, an Institute of Physics Publication on Structures, Devices and Systems.

Sensors and Actuators A, an Elsevier Sequoia Publication on solid-state devices for transducing physical signals.

For list of web sites, conferences, and other sources of information on Micromechanics and MEMS please see the web site: http://www. trimmer. net

“The Other Side of Computing” - Paper for the Conference on High Speed Computing Salishan Lodge, Gleneden Beach, Oregon, April 20 to 23, 1998, Presentation by William Trimmer

Crystalline Computation



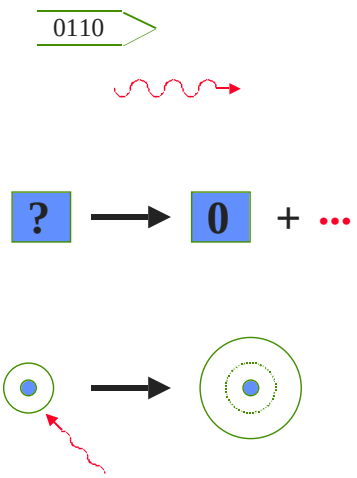
Norman Margolus
BU CCS, MIT AI Lab

Plan of the talk

- Fundamental Constraints $\Longrightarrow$ computers will imitate physics
- Uniform Crystal: general purpose, practical, scalable
- Crystalline Algorithms: what changes?
- Near Term: 3D bit-mapped computations
- Conclusions

Fundamental physical constraints

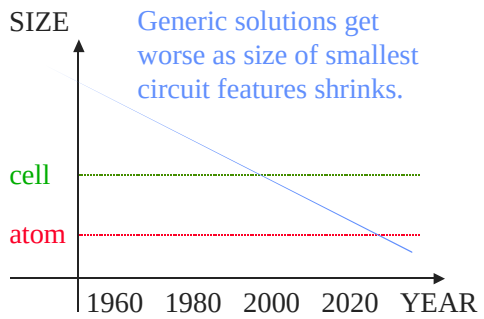
- Locality: no information can travel faster than light
- Invertibility: no information is ever erased from the world
- QM: the laws of physics act differently at small scales



Fundamental physical constraints

Handled in hardware:

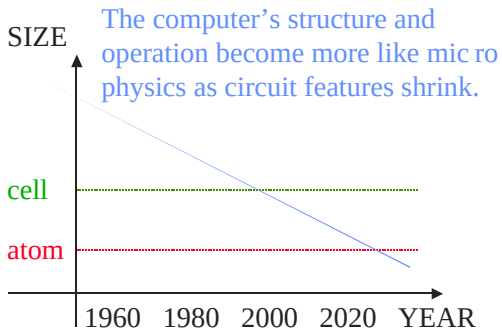
- Locality: use fast/short wires (gets harder)
- Invertibility: throw away info as heat (~ density)
- QM: space and time averaging (stats get bad)



Fundamental physical constraints

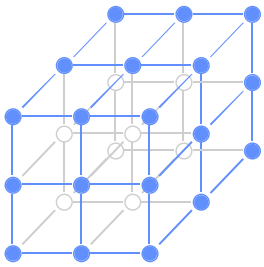
Incorporate constraints:

- Locality: mesh architecture (--> no wires)
- Invertibility: reversible logic (--> no heat)
- QM: exploit digital character (--> no stats)



Why a simple uniform crystal?

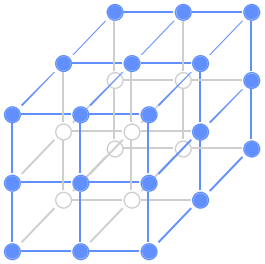
- Practical: easier to design, build, test and control
- Scalable: just make more -- as asymptotically good as anything
- Fast: high processing density and fast cycle time
- Fair: doesn't try to favor any class of (classical) computations



Why a simple uniform crystal?

Why not a ...

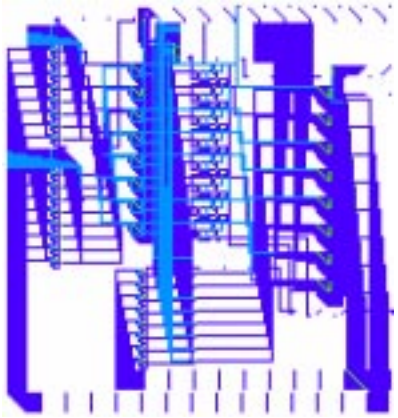
- Coherent quantum computer: too much overhead for general purpose computing
- Amorphous computer: we gain computing power from predictability



Crystalline algorithms

SPACE:

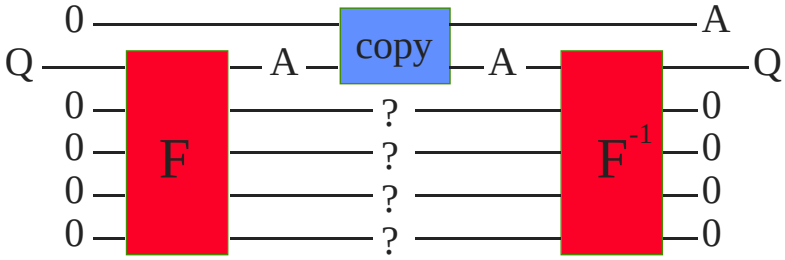
- The computer designer just provides a large-scale digital spatial medium: a 3D FPGA
- Assigning which positions will perform which operations is much like building hardware
- Traditional architectural ideas for mapping computation into space remain interesting



Crystalline algorithms

REVERSIBLE FUNCTIONS:

We can avoid the need to throw away information by "uncomputing" partial results.



Crystalline algorithms

MACROSCOPIC SCALING:

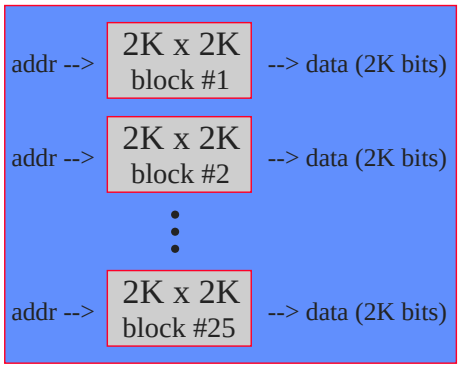
Large scale combinatorial computations become practical.

Crystalline computers today

- Mainstream make crystalline look hard

Possible today:

- Terabits/sec/chip for large-scale meshes
- Virtual processor SIMD using embedded DRAM

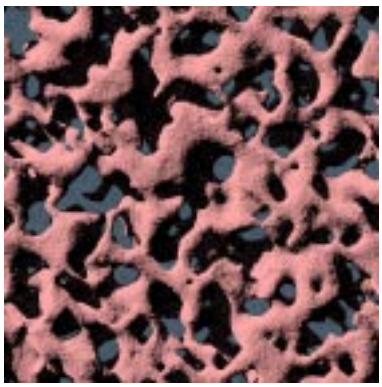


One chip, with 50ns row
of 50Kbits/row = 1 Tbit/

Crystalline computers today

Immediate applications:

- 3D bitmap-based approach to “virtual reality”
- 2D/3D image segmentation, manipulation and rendering
- physics simulations
- cryptography
- levelized logic simulation



Conclusions

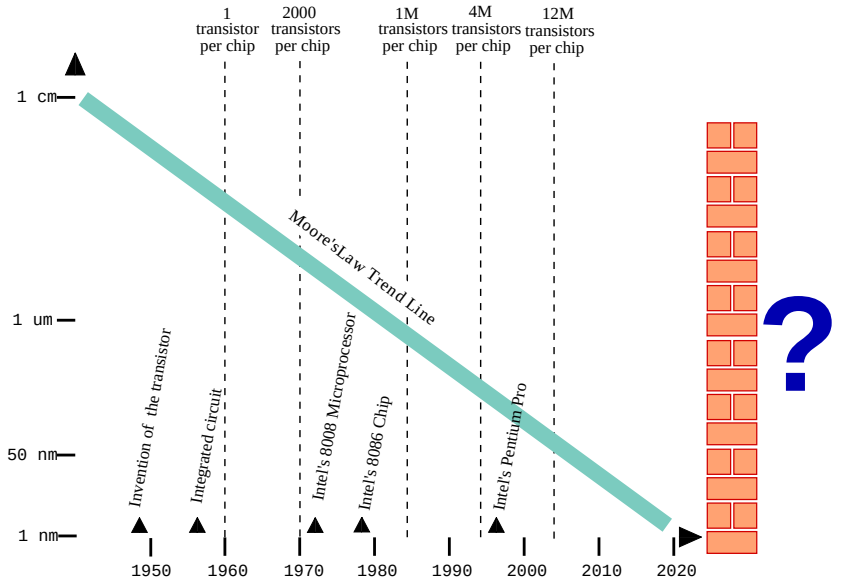
- Algorithms and computers will become more physics-like
- Crystalline computers will be the best possible general purpose computers
- Some of the power of crystalline computations can be harnessed today

Quantum Computing

E. Knill

- Quantum computing is inevitable.
- Applications of quantum computers.
- What are quantum computers?
- The power of quantum algorithms.
- Requirements for quantum computing.
- Quantum noise control.
- Proposed quantum computing devices.
- Prospects.

Moore's Law

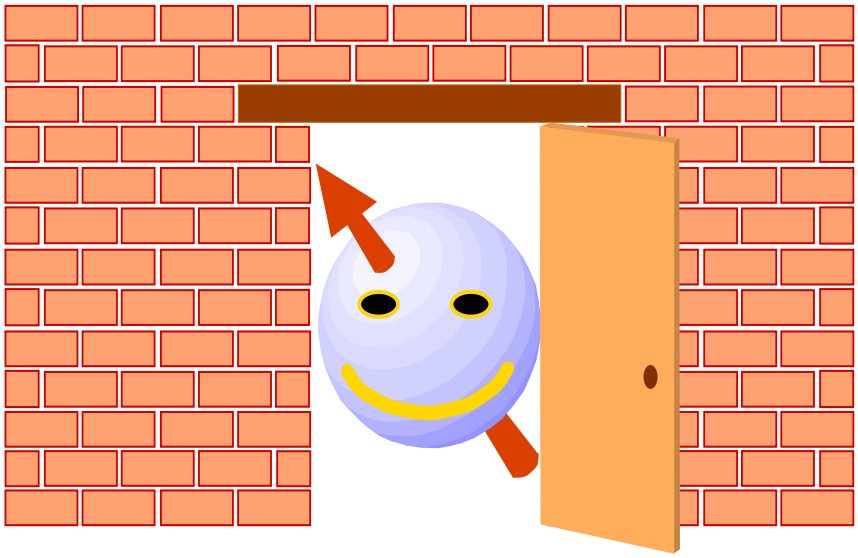


Moore[27], Montemerlo&*al*.[26]

LOS ALAMOS National Laboratory

1

Behind the Wall?



Lloyd 1996

Quantum Computer Applications

- Long distance quantum cryptography.
- Number theory algorithms: Factoring, discrete logarithm. Shor 1994[31, 34]
- Accelerated combinatorial searching. Grover 1995[12, 13]
- Code breaking: Public key cryptography, DES.
- Physics simulations.

  Feynman 1984[10], Lloyd 1996[22], Wiesner 1996[37], Zalka 1996[39]

Specifying a Model of Computation

- State space.
- Manipulating states.
- Initial state.
- Read out.

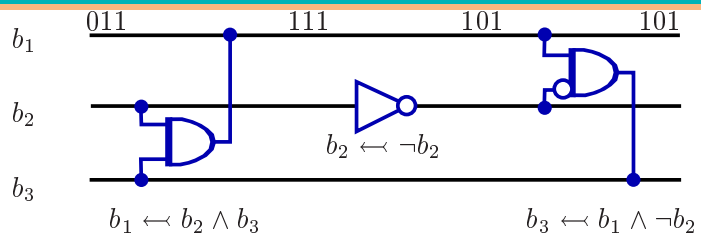
- Quantum computing is an *extension* of classical computing.

State Space

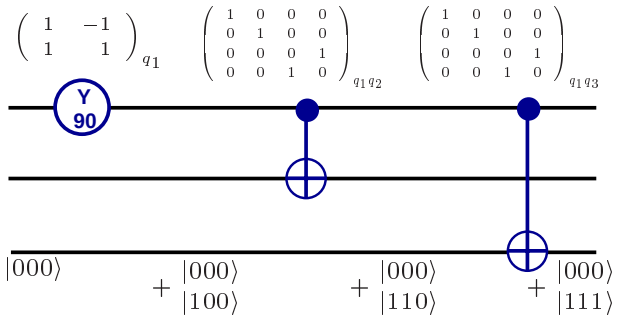
	1 bit	n bits
Classical	$0, 1$	$000\ldots, 100\ldots,$ $010\ldots, 110\ldots,$ $001\ldots, \ldots$
Quantum	$\alpha\|0\rangle + \beta\|1\rangle$	$\alpha_{000\ldots}\|000\ldots\rangle +$ $\alpha_{100\ldots}\|100\ldots\rangle +$ $\alpha_{010\ldots}\|010\ldots\rangle +$ $\alpha_{110\ldots}\|110\ldots\rangle +$ $\alpha_{001\ldots}\|001\ldots\rangle + \ldots$

Processing Information

Classical



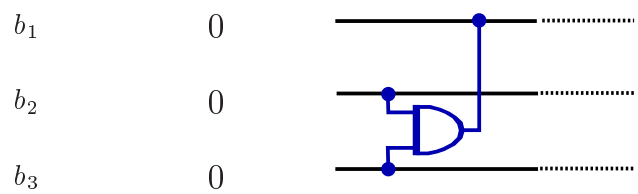
Quantum



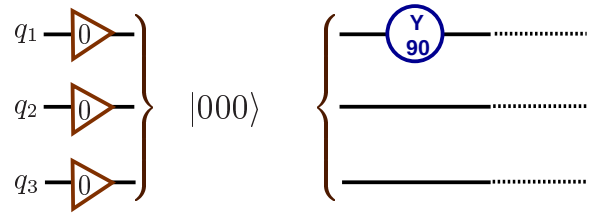
Initial State

prepared state

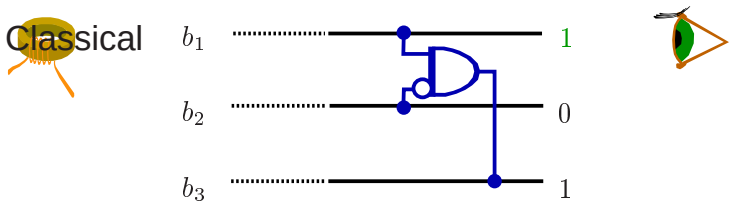
Classical

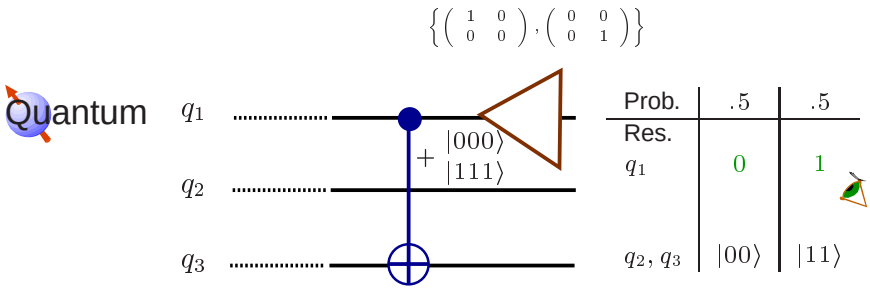


Quantum



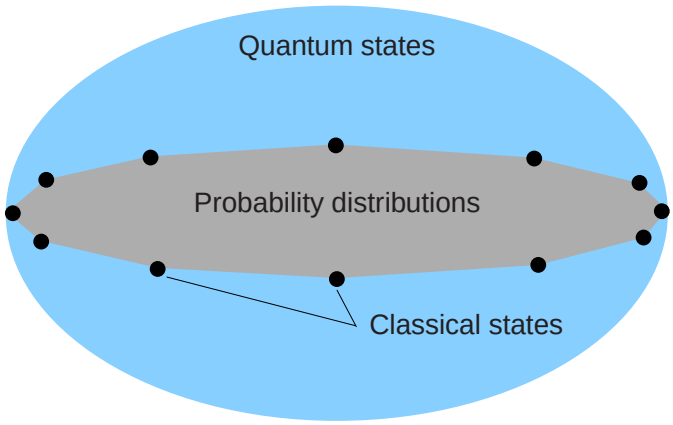
Readout





Prob.	.5	.5
Res.		
q_1	0	1
q_2, q_3	$\|00\rangle$	$\|11\rangle$

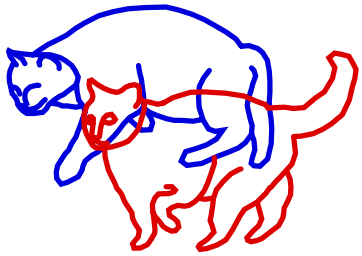
Power of QC†: State Space?



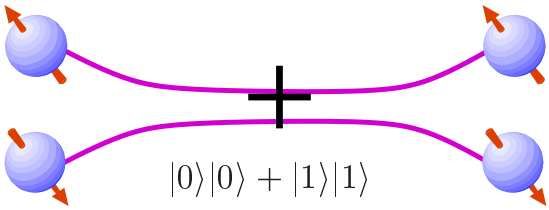
- Exponential dimension in terms of physical resources.
- No feasible classical simulation for > 40 qubits?

† Quantum Computing.

Power of QC†: Entanglement?

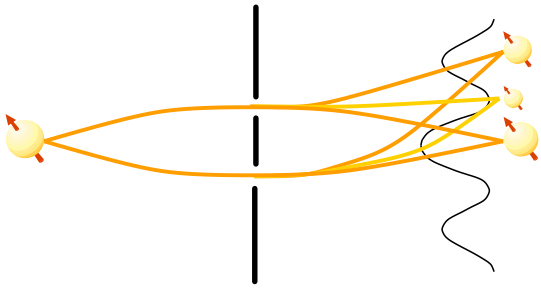


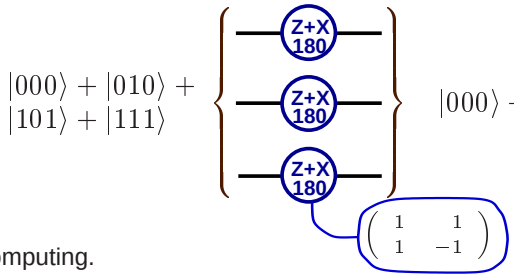
- For quantum communication.



† Quantum Computing.

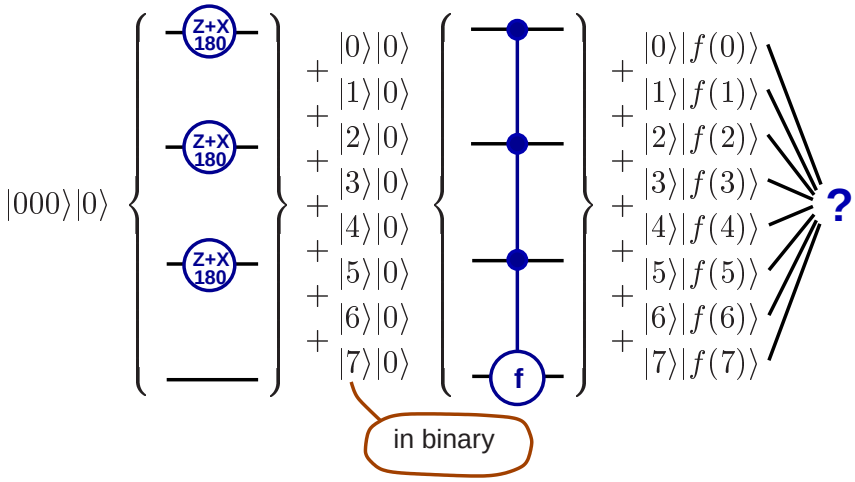
Power of QC†: Interference effects?





† Quantum Computing.

Power of QC†: Quantum Parallelism?



† Quantum Computing.

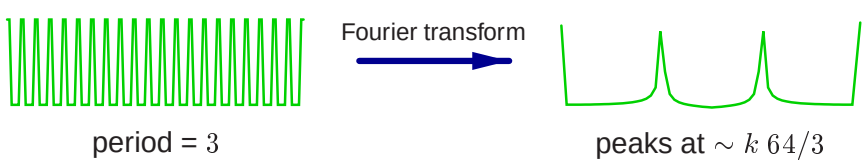
Period Determination

- f is periodic with period p if for all k, $f(k+p) = f(k)$.

Example:

$$f(3k) = 1, \quad f(3k+1) = 0, \quad f(3k+2) = 0$$
$$k \in \{0, \ldots, 63\}$$

- Problem: Determine the unknown period of a periodic function.



Quantum Fourier Transform

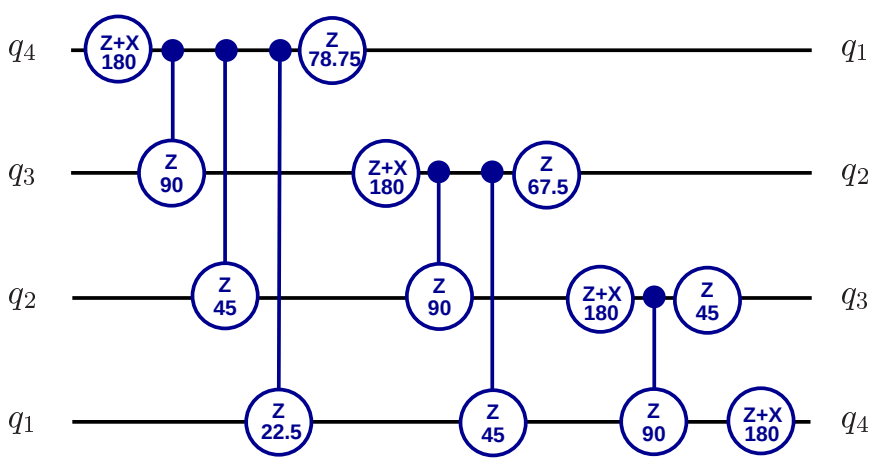
$$\mathcal{F} : \sum_{n=0}^{N} \alpha_n |n\rangle \rightarrow \sum_{k=0}^{N} \left(\sum_{n} e^{2\pi n k i/N} \alpha_n \right) |k\rangle$$

amplitudes

- Example:

$$\begin{array}{rrl} & 1 & |0\rangle \\ + & 0 & |1\rangle \\ + & 1 & |2\rangle \\ + & 0 & |3\rangle \\ + & 1 & |4\rangle \\ + & 0 & |5\rangle \\ + & 1 & |6\rangle \\ + & 0 & |7\rangle \end{array} \longrightarrow \begin{array}{rrl} & 1 & |0\rangle \\ + & 0 & |1\rangle \\ + & 0 & |2\rangle \\ + & 0 & |3\rangle \\ + & 1 & |4\rangle \\ + & 0 & |5\rangle \\ + & 0 & |6\rangle \\ + & 0 & |7\rangle \end{array}$$

Network for the QFT†



q_4: most significant bit. q_1: least significant bit.

† Quantum Fourier Transform.

Pseudo Code for the QFT†

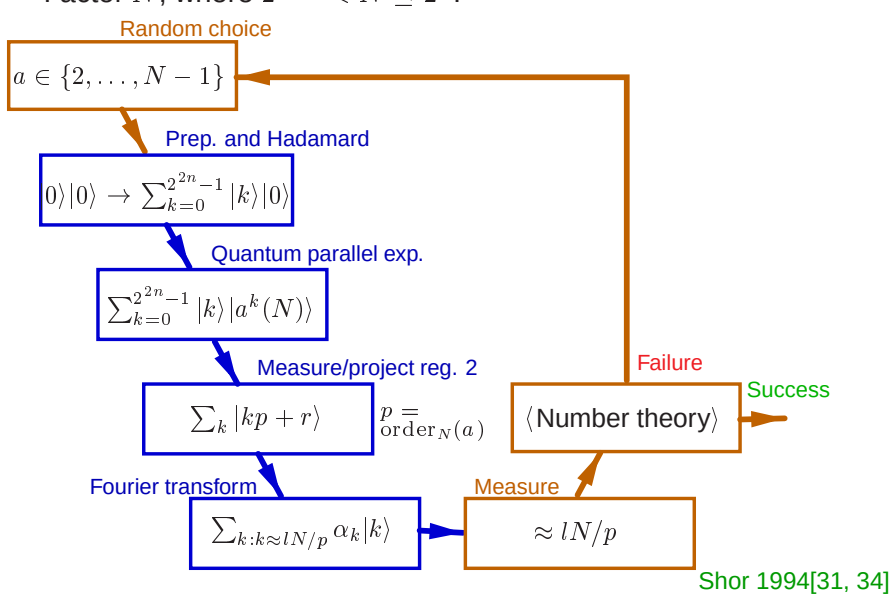
FOURIER($\ulcorner a \urcorner$)
Input: A quantum register $\ulcorner a \urcorner$ with d qubits. The most significant qubit has index $d-1$.
Output: The amplitudes of $\ulcorner a \urcorner$ are Fourier transformed over $\mathbb{Z}_{2^d}$. The most significant bit in the output has index 0, i.e. the ordering is reversed.

$d \leftarrow$ **length** $(\ulcorner a \urcorner)$
$\omega \leftarrow e^{i2\pi/2^d}$
for $i = d-1$ **to** $i = 0$
  for $j = d-1$ **to** $j = i+1$
    if $\ulcorner a_j \urcorner$ **then** $\mathcal{R}_{\omega^{2^{d-i-1+j}}}(\ulcorner a_i \urcorner)$
  end
  $\mathcal{H}(\ulcorner a_i \urcorner)$
end

† Quantum Fourier Transform.

Quantum Factoring

- Factor N, where $2^{n-1} < N \leq 2^n$:



Requirements for Quantum Computing

- State space:
  - Controlled system space.
  - Scalability.
- State preparation:
  - Fiducial pure initial state.
  - Initial states on demand.
- State control:
  - Single system unitary rotations.
  - Conditional dynamics.
  - Complete set of gates.
- Readout:
  - Reliable projective measurement *or* reliable expectations of system observables.
- Noise:
  - No memory errors *or* high parallelism.
  - Error per operation $<$ threshold.

DiVincenzo&Loss 1997[9], Knill&Laflamme 1996[19], Preskill 1997[28]

LOS ALAMOS 18

Noise in Quantum Computing

Memory: Most superpositions are fragile.

- Control vs. environment: Strong interactions for control, but weak interactions with environment.

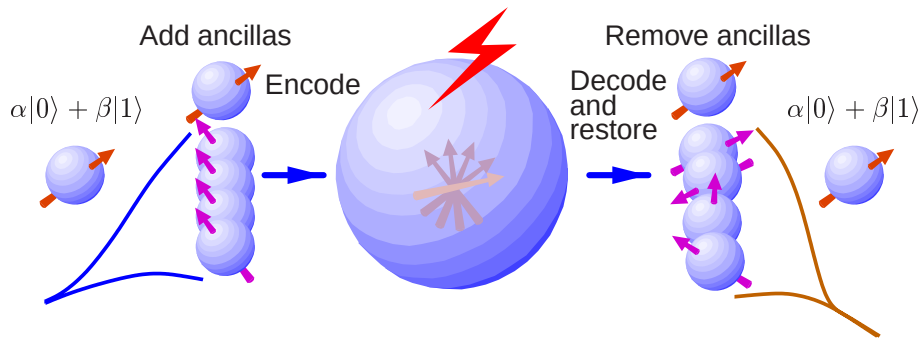
Operations: Gates are inherently “analogue” and hence inaccurate.

- Error space: Continuity of possible errors.

LOS ALAMOS 19

Quantum Error Correction

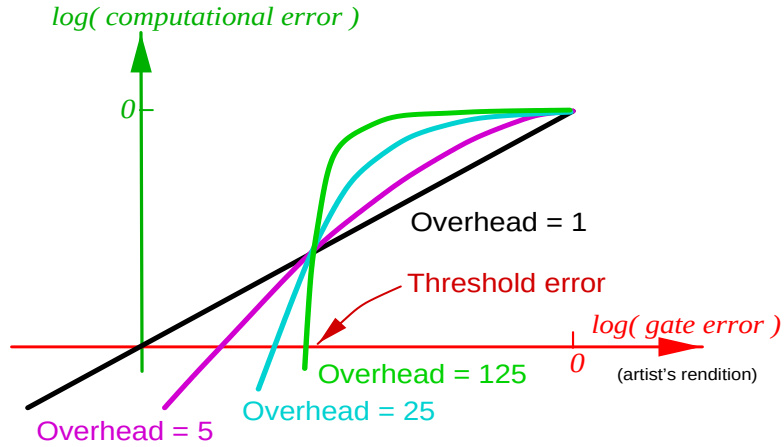
- Discretize errors.
- Errors are local and independent.



Shor 1995[32], Steane 1995[35]

The Accuracy Threshold Theorem

Theorem 1. *If the error amplitude per gate (including "no-op") is less than a threshold, then it is possible to quantum compute and communicate arbitrarily accurately.*



Shor 1996[33]

Knill&*al.* 1996[20], Aharonov&Ben-Or 1996[1], Kitaev 1996[17], Preskill 1997[28]

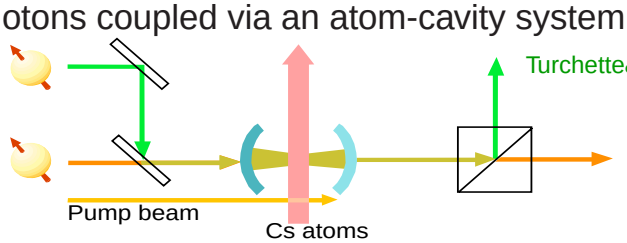
Proposed Quantum Computers

- Photonic QC[1]: Photons and nonlinear optics.
  Milburn 1988[24], Yamamoto&*al.* 1988[38]
- Ion trap: Ions in a linear trap, coupling via center of mass mode, control with lasers. Cirac&Zoller 1995[5]
- Cavity QED[2]: Photons in cavities and flying or trapped atoms.
  Turchette&*al.* 1995[36]
- Superconducting Josephson junctions: Superconducting domains, flux or cooper pairs. Bocko&*al.* 1997[2], Shnirman&*al.* 1997[30]
- Quantum dots: Trapped electrons in semiconductor nano-dots.
  Loss&DiVincenzo 1997[23]
- NMR[3]: Nuclear spins in molecules.
  Gershenfeld&Chuang 1996[11], Cory&*al.* 1996[6]
- Solid state NMR[3]: Implanted nuclear spins in silicon at low temperature, coupling via electrons.
  Kane 1997[15], Privman&*al.* 1997[29]
- Others: Optical lattices, crystal NMR[3], . . .

[1] Quantum Computing. [2] Quantum Electro-Dynamics. [3] Nuclear Magnetic Resonance.

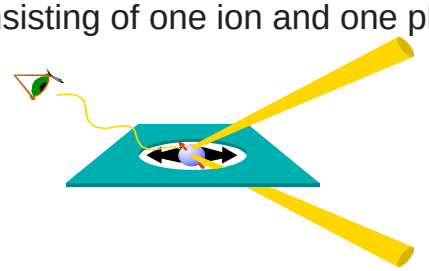
Implemented Quantum Gates I

- Cavity QED: Conditional phase shift between photons coupled via an atom-cavity system.



Turchette&*al.* 1995[36]

- Ion traps: Conditional dynamics on two bits consisting of one ion and one phonon in a Paul trap.



Monroe&*al.* 1995[25]

Requirements for QC†: Ion Traps I

- Initial state, control and readout: Confirmed by experiments using an ion and a phonon.
- Time per operation: $\gtrsim 1\mu s$, depending on phonon frequency.
- Robustness:
  - Achievable: $R \equiv$ operations/error $\lesssim 1000$?
  - Currently: $R \approx 10$.
  - Little parallelism.

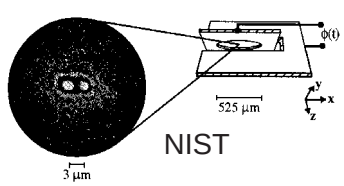
† Quantum Computing.

Requirements for QC: Ion Traps II

- State space and scalability: Single traps are limited to $10 \ldots 80$ ions in practice.

Elliptical trap:

QC ready, in ground state.

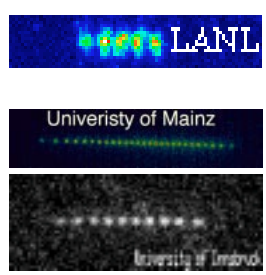


NIST

King&*al* 1998[16]

Linear traps:

Crystals, not in ground state



- In theory, parallel, coupled ion traps are possible.

Implemented Quantum Gates II

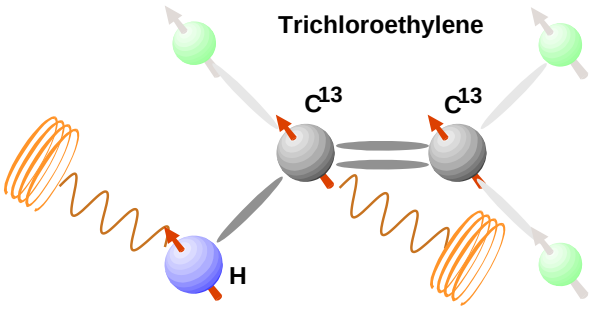
- Liquid state NMR:

- Two bit algorithms (Pseudo-pure state, EPR, Grover, Deutsch-Josza) in chloroform and cystosine.

  Knill&*al.* 1997[18], Chuang&*al.* 1997[3], 1998[4], Jones&Mosca 1998[14]

- Three bit algorithms (Toffoli gate, GHZ, quantum error correction) in trichloroethylene and alanine.

  Laflamme&*al.* 1997[21], Cory&*al.* 1997[8], 1998 [7]



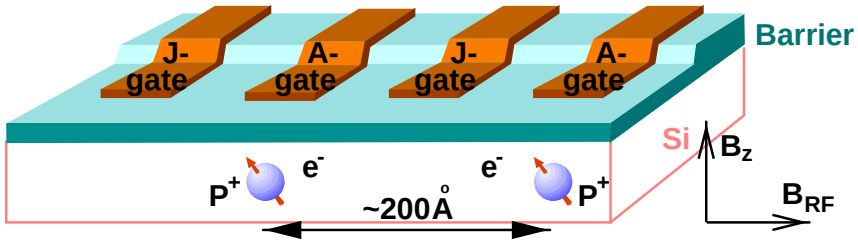
Requirements for QC: Liquid State NMR

- Initial state, control and readout:
  - Experimentally, up to three bits.
  - Exponential loss of signal for state preparation.
- Time per operation: $\gtrsim 10ms$.
- Robustness:
  - Achievable: $R \gtrsim 100$?
  - Currently: $R \approx 20$.
  - But threshold theorem does not apply.
- State space and scalability:
  - State space is scalable, but state preparation is not.
  - Limited to $\lesssim 10$ bits ($\lesssim 100$ for physics simulation?).

Requirements for QC: Solid State NMR

Silicon based quantum computer.

(Theoretical proposal, experiments in progress.)



- Initial state, control and readout: Possible in theory.
- Time per operation: $\gtrsim 2\mu s$.
- Robustness: $R \gg 1000$?
- State space and scalability: Few limitations.

Kane 1997[15]

Quantum Computing Scenarios

- Minimal utilization:
  - Device-level quantum control and error-correction.
  - Long distance quantum cryptography.

+ Anticipated utilization: “Quantum coprocessors” for
  - “Number theory” calculations.
  - Accelerated combinatorial searching.

++ Maximal utilization:
  - Most computing and communication is fully quantum.

Conclusion

- Quantum computing is inevitable.
  - Needed for atomic scale control and information processing.
- Quantum computing is useful.
  - Quantum communication.
  - Quantum control and error-correction.
  - Optimization and “number theory”.
  - Physics simulation.
- Quantum computing is robust.
  - The accuracy threshold theorem.
- Quantum computing “exists”.
  - Experiments up to three bits in CQED, ion trap and NMR.

LOS ALAMOS 30

Quantum Computing [E. Knill, http: / /www. c3 . lam . gov/ ~knill]

References

[1] D. Aharonov and M. Ben-Or. Fault-tolerant quantum computation with constant error. *In Proceedings of the 29’th Annual ACM Symposium on the Theory of Computing*, New York, New York, 1996. ACM Press. quant-ph/9611025.

[2] M. F. Bocko, A. M. Herr, and M. J. Feldman. Prospects for quantum coherent computation using superconducting electronics. *IEEE Trans. App. Supercond.*, 7:3638-3641, 1997.

[3] I. L. Chuang, N. Gershenfeld, M. Kubinec, and D. Leung. Bulk quantum computation with nuclear magnetic resonance: Theory and experiments. *Proceedings of the Royal Society of London A*, 1998.

[4] I. L. Chuang, L. M. K. Vandersypen, X. Zhou, D. W. Leung, and S. Lloyd. Experimental realization of a quantum algorithm. quant-ph/9801037, 1998.

[5] J. Cirac and P. Zoller. Quantum computations with cold trapped ions. *Physical Review Letters*, 74:4091-, 1995.

[6] D. G. Cory, A. F. Fahmy, and T. F. Havel. Ensemble quantum computing by nmr-spectroscopy. *Proceedings of the National Academy of Sciences of the United States of America*, 94: 1634-1639,1997.

17] D. G. Cory, W. Maas, M. Price, E. Knill, R Laflamme, W. H. Zurek, T. F. Havel, and S. S. Somaroo. Experimental quantum error correction. quant-ph/9802018, 1998.

[8] D. G. Cory, M. D. Price, and T. F. Havel. Nuclear magnetic resonance spectroscopy: An experimentally accessible paradigm for quantum computing. *Physica D*., To appear, 1997. quant-ph/9709001.

[9] D. P. DiVincenzo and D. Loss. Quantum information is physical *Superlattices and Microstructures*, to appear:?, 1997. cond-mat/9710259.

[10] R. P. Feynman. Quantum-mechanical computers. *Journal of the Optical Society of America B-Optical Physics*, 1:464, 1984.

[11] N. A. Gershenfeld and I. L. Chuang. Bulk spin resonance quantum computation. *Science*, 275:350-356, 1997.

[12] L. K. Grover. A fast quantum mechanical algorithm for database search. quant-ph/9605043.

[13] L. K. Grover. Quantum mechanics helps in searching for a needle in a haystack. *Physical Review Letters*, 79:325-328, 1997.

[14] J. A. Jones and M. Mosca. Implementation of a quantum algorithm to solve deutsch's problem on a nuclear magnetic resonance quantum computer. quant-ph/9801027, 1998.

[15] B. E. Kane. A silicon-based nuclear spin quantum computer. *Nature*, to appear:?, 1998.

[16] B. E. King, C. S. Wood, C. J. Myatt, Q. A. Turchette, D. Leibfied, W. M. Itano, C. Monroe, and D. J. Wineland. Initializing the collective motion of trapped ions for quantum logic. quant-ph/ 9803023, 1998.

[17] A. Yu. Kitaev. *Russian Math. Surv*., 6, 1997.

[18] E. Knill, I. Chuang, and R. Laflamme. Effective pure states for bulk quantium computation. *Phys. Rev A.*, 80, 1998. quant-ph/9706053.

[19] E. Knill and R. Laflamme. Assumptions for fault tolerant quantum computing. Technical Report LAUR-96-2718, Los Alamos National Laboratory, http: / /www.c3.lanl.gov/ ~knill, 1996.

[20] E. Knill, R. Laflamme, and W. Zurek. Resilient quantum computation. *Science*, 279:342-345, 1998.

[21] R. Laflamme, E. Knill, W. H. Zurek, P. Catasti, and S. V. S. Mariappan. Nmr-ghz. quant-ph/9709025,~1997.

[22] S. Lloyd. Universal quantum simulators. *Science*, 273:1073-1078, 1996.

[23] D. Loss and D. P. DiVincenzo. Quantum computation with quantum dots. cond-mat/9701055, 1997.

[24] G. J. Milburn. Quantum optical fredkin gate. *Physical Review Letters*, 62:2124-2127,1988.

[25] C. Monroe, D. M. Meekhof, B. E. King, W. M. Itano, and D. J. Wineland. Demonstration of a universal quantum logic gate. *Physical Review Letters*, 75:4714-4717, 1995.

[26] M. S. Montemerlo, J. C. Love, G. J. Opitek, D. Goldhaber-Gorden, and J. C. Ellenbogen. Technologies and designs for electronic nanocomputers. Mitre preprint MTR 96W0000044, http://www.mitre.org/research/nanotech, 1996.

[27] G. Moore. An update on moore's law. Speech given at the INTEL Developer's Forum. http://www.intel.se/pressroom/archive/speeches/gem93097.htm, 1997.

[28] J. Preskill. Reliable quantum computers. *Proceedings of the Royal Society of London A*, 454:385-410,1998.

[29] V. Privman, I. D. Vagner, and G. Kventsel. Quantum computation in quantum-hall systems. quant-ph/9707017, 1997.

[30] A. Shnirman, G. Schon, and Ziv Hermon. Quantum manipulations of small josephson junctions. *Physical Review Letters*, 79:2371-2374,1997.

[31] P. W. Shor. Algorithms for quantum computation: Discrete logarithms and factoring. *In Proceedings of the 35'th Annual Symposium on Foundations of Computer Science*, pages 124-134, Los Alamitos, California, 1994. IEEE Press.

[32] P. W. Shor. Scheme for reducing decoherence in quantum computer memory. Phys. Rev. A, 52:2493-, 1995.

[33] P. W. Shor. Fault-tolerant quantum computation. *In Proceedings of the Symposium on the Foundations of Computer Science*, pages 56-65,LosAlamitos,California, 1996. IEEE press. quant-ph/9605011.

[34] P. W. Shor. Polynomial-time algorithms for prime factorization and discrete logarithms on a quantum computer. *SIAM J. Comput*., 26:1484-1509,1997.

[35] A. Steane. Multiple particle interference and quantum error correction. *Proceedings of the Royal Society of London A*, 452:2551-2577,1996.

[36] Q. A. Turchette, C. J. Hood, W. Lange, H. Mabuchi, and H. J. Kimble. Measurement of conditional phase shifts for quantum logic. *Physical Review Letters*, page 4710, 1995.

[37] S. Wiesner. Simulations of many-body quantum systems by a quantum computer. quant-ph/9603028, 1996.

[38] Y. Yamamoto, M. Kitagawa, and K. Igeta. *In Proceedings of the Third Asia-Pacific Physics Conference*, Singapore, 1988. World Scientific.

[39] C. Zalka. Threshold estimate for fault tolerant quantum computation. quant-ph/9612028, 1996.

List of Attendees

Tilak Agerwala
IBM Corporation
522 South Road
MS P962
Poughkeepsie, NY 12601
Phone (914) 433-7800
Fax (914) 433-8598
Email tilak@us.ibm.com

Bob Aiken
Network Research Program Mgr
ER-31, US DOE
19901 Germantown Road
Germantown, MD 20784-1290
Phone (301) 903-9960
Fax (301) 903-7774
Email aiken@er.doe.gov

Steve Ashby
Lawrence Livermore National Laboratory
P.O. Box 808, L-561
Livermore, CA 94551
Phone (925) 423-2462
Fax (925) 422-6675
Email sfashby@llnl.gov

Greg Astfalk
Hewlett-Packard
3000 Waterview Parkway
Richardson, TX 75080-1400
Phone (972) 497-4787
Fax (972) 497-4441
Email astfalk@rsn.hp.com

Robert Ayars
Los Alamos National Laboratory
CIC-7, MS B294
Los Alamos, NM 87545
Phone (505) 667-9047
Fax (505) 665-6333
Email rsa@lanl.gov

Robert Babb
University of Denver
Dept. of Math & Comp. Science
2360 S. Gaylord Street
Denver, CO 80208
Phone (303) 871-2460
Fax (3035 871-3010
Email babb@cs.du.edu

David H. Bailey
Lawrence Berkeley National Laboratory
1 Cyclotron Drive
MS 50B-4230
Berkeley, CA 94720
Phone (510) 486-7377
Email dhbailey@lbl.gov

Kimberly Barnes
Oak Ridge National Laboratory
Bldg 6010, MS 6414
Oak Ridge, TN 37831-6414
Phone (423) 574-3146
Fax (423) 574-0680
Email barneskd@ornl.gov

Richard Barrett
Los Alamos National Laboratory
CIC-8 MS B272
Los Alamos, NM 87545
Phone (505) 667-6845
Fax (505) 665-6333
Email rbarrett@lanl.gov

Forest Baskett
Silicon Graphics Inc.
P.O. Box 7311
2011 N. Shoreline Blvd.
Mt. View, CA 94043
Phone (650) 962-3608
Fax (650) 969-6289
Email baskett@sgi.com

Doug Beason
Lawrence Livermore National Laboratory
L-170
Livermore, CA 94550
Phone (925) 422-3717
Fax (925) 422-0779
Email dbeason@llnl.gov

Don Becker
NASA Goddard Space Flight Center
USRA/CESDIS, Code 930.5
Greenbelt, MD 20771
Phone (301) 286-4403
Fax (301) 286-1777
Email becker@cesdis.usra.edu

Peter Beckman
Los Alamos National Laboratory
CIC-ACL, MS B287
Los Alamos, NM 87545
Phone (505) 665-0800
Fax (505) 665-4939
Email beckman@acl.lanl.gov

'Doc' Bedard
Department of Defense
Fort Meade, MD 20755-6000
Phone (301) 688-0329
Fax (301) 688-0330
Email Doc@afterlife.ncsc.mil

Jacqueline Bell
Defense Special Weapons Agency
6801 Telegraph Road
Alexandria, VA 22310
Phone (703) 325-6763
Fax (703) 325-2954
Email jlb@lanl.gov

Charlie Bender
Ohio Supercomputer Center
Building 460
1224 Kinnear Road
Columbus, OH 43212
Phone (614) 292-6082
Email bender@osc.edu

Greg Bishop
Hewlett Packard
1591 Eudora Street
Denver, CO 80220
Phone (303) 649-5456
Fax (303) 649-5310
Email greg_bishop@hp.com

John Blaylock
Los Alamos National Laboratory
CIC-11, MS B269
Los Alamos, NM 87545
Phone (505) 667-8970
Fax (505) 667-0168
Email jwb@lanl.gov

Tor Bloch
UNI-C
DTU, Bldg 304
2800 LYNGBY
DENMARK
Phone +45 35 87 88 89 60
Fax +45 35 87 89 90
Email Tor.Bloch@uni-c.dk

Bill Boas
ABBA Technologies
2403 San Mateo, NE
Suite P16
Albuquerque, NM 87059
Phone (505) 889-3337
Fax (505) 889-3338
Email bboas@abbatech.com

Bob Borchers
National Science Foundation
4201 Wilson Blvd.
Room 1122
Arlington, VA 22230
Phone (703) 306-1970
Fax (703) 306-0632
Email borchers@nsf.gov

Joe Brandt
Lawrence Livermore National Laboratory
P. O. Box 808, L-66
Livermore, CA 94551
Phone (925) 422-7043
Fax (925) 423-4820
Email jbrandt@llnl.gov

Al Brenner
Institute for Defense Analysis
1801 N. Beauregard Street
Alexandria, VA 22311
Phone (703) 845-6603
Fax (7035 845-6848
Email abrenner@ida.org

Forrest Brown
P.O. Box 9049
Niskayuna, NY 12309
Phone (518) 395-7370
Fax (518) 395-6171
Email brownf@rpi.edu

Susan Brown
Center for Human Motor Res.
Division of Kinesiology
Univ. of Michigan
Ann Arbor, MI 48109-2214
Phone (313) 763-6755
Fax (313) 936-1925
Email shcb@umich.edu

Donald Burton
Los Alamos National Laboratory
X-CI, MS F663
Los Alamos, NM 87545
Phone (505) 667-8746
Fax (505) 665-5553
Email burton@lanl.gov

Bill Buzbee
National Center for Atmospheric Research
P.O. Box 3000
Boulder, CO 80303-3000
Phone (303) 497-1206
Fax (303) 497-1298
Email buzbee@ucar.edu

Andy Byers
University of Colorado at Boulder
Dept. of Elect. & Comp. Engr
Room ECEElB55
Boulder, CO 80309
Phone (303) 492-8719
Fax (303) 492-5323
Email byersa@ucsub.colorado.edu

William J. Camp
Sandia National Laboratory
P.O. Box 5800 MS 0321
Albuquerque, NM 87185-0321
Phone (505) 845-8939
Fax (505) 844-2415
Email wicamp@sandia.gov

Curtis Canada
Los Alamos National Laboratory
CIC-ACL, MS B287
Los Alamos, NM 87545
Phone (505) 665-7453
Fax (505) 665-4939
Email cvd@ac.lanl.gov

Bill Carlson
Inst. for Defense Analysis
17100 Science Drive
Bowie, MD 20715-4300
Phone (301) 805-7486
Fax (301) 805-7502
Email wwc@super.org

Ralph Carlson
Lawrence Livermore National Laboratory
P.O. Box 808, L-711
Livermore, CA 94551
Phone (925) 422-1422
Fax (925) 423-8607
Email carlson10@llnl.gov

Michael Clover
Los Alamos National Laboratory
X-CI, MS F663
Los Alamos, NM 87545
Phone (505) 667-9797
Fax (505) 665-5553
Email mrc@lanl.gov

Sam Coleman
Lawrence Livermore National Laboratory
P. O. Box 808, L-60
Livermore, CA 94551
Phone (925) 422-4323
Fax (925) 423-8715
Email scoleman@llnl.gov

David Cooke
School of Physical Therapy
Elborn College
University of Western Ontario
London, Ontario, Canada N6G lH1
Phone (519) 661-3947
Fax (519) 661-3866
Email ahsjdc@julian.uwo.ca

David M. Cooper
Lawrence Livermore National Laboratory
P. O. Box 808; L-66
Livermore, CA 94550
Phone (925) 422-4010
Fax (925) 423-4820
Email dcooper@llnl.gov

George Cotter
National Security Agency
9800 Savage Road
Fort Meade, MD 20755-6217
Phone (301) 688-6434
Fax (301) 688-4980
Email grcotte@afterlife.ncsc.mil

Christie Davis
US Department of Energy
Forrestal Building, DP-51-lJ-077
1000 Independence Ave., SW
Washington, DCV 20585
Phone (202) 586-5950
Email cdm@lanl.gov

Jay Davis
Lawrence Livermore National Laboratory
P.O. Box 808, L-103
Livermore, CA 94550
Phone (925) 423-1414
Fax (925) 422-0096
Email davis23@11nl.gov

Hugo de Garis
Brain Builder Grp, Evolutionary Syst. Dept.
ATR Human Info. Proc. Labs
2-2 HIKARI DAI, Seika-cho Soraku-gun
Kansai Science City, Kyoto 619-02 Japan
Phone (+81) 774 95 1079
Fax (+81) 774 95 1008
Email degaris@hip.atr.co.jp

Tom DeBoni
Lawrence Berkeley National Laboratory
1 Cyclotron Road, 50C
Berkeley, CA 94720
Phone (510) 486-8617
Fax (510) 486-7891
Email tmdeboni@lbl.gov

Jim Decker
U.S. Department of Energy
1000 Independence Ave. SW
Suite 7B058
Washington, DC 20585
Phone (202) 586-5434
Fax (202) 586-4120
Email James.Decker@oer.doe.gov

David J. DiLeonardo
Bettis Atomic Power Laboratory
P.O. Box 79
West Mifflin, PA 15122-0079
Phone (412) 476-5463
Fax (412) 476-6924

Jack Dongarra
University of Tennessee
Computer Science Department
104 Ayres Hall
Knoxville, TN 37996-1301
Phone (423) 974-8295
Fax (423) 974-8296
Email dongarra@cs.utk.edu

Mark Duchaineau
Lawrence Livermore National Laboratory
P.O. Box 808, L-561
Livermore, CA 94550
Phone (925) 423-1320
Fax (925) 423-2993
Email duchaineaul@llnl.gov

Steve Elbert
U.S. Department of Energy, ER-31
Math., Information & Comp. Sci. Div.
19901 Germantown Road
Germantown, MD 20874-1290
Phone (301) 903-6904
Fax (301) 903-7774
Email elbert@er.doe.gov

Beth Erbert
Lawrence Livermore National Laboratory
P.O. Box 808, L-66
Livermore,CA 94551
Phone (925) 422-2614
Fax (925) 422-3149
Email berbert@llnl.gov

Rob Falgout
Lawrence Livermore National Laboratory
P.O. Box 808, L-561
Livermore, CA 94550
Phone (925) 422-4377
Fax (925) 423-2993
Email rfalgout@llnl.gov

John Feo
San Diego Supercomputer Center
P.O. Box 85608
San Diego, CA 92186-0906
Phone (619) 822-0952
Email john@tera.com

Larry Ferderber
Lawrence Livermore National Laboratory
P.O. Box 808, L020
Livermore, CA 94550
Phone (925) 423-7789
Fax (925) 424-3858
Email ferderber1@llnl.gov

Steve Finn
User Technology Associates
2242 Malcolm Ave.
Los Angeles, CA 90064
Phone (310) 470-2335
Fax (310) 470-2886
Email sf@sunmail.dswa.mil

Janet Frame
Lawrence Livermore National Laboratory
P.O. Box 808, L-66
Livermore, CA 94550
Phone (925) 422-2732
Fax (925) 423-4820
Email frame2@llnl.gov

Richard Games
The MITRE Corporation
MS K312
202 Burlington Road
Bedford, MA 01730
Phone (781) 271-8081
Fax (781) 271-2780
Email rg@mitre.org

Don Geiser
Hewlett Packard
24 Inverness Place East
Englewood, CO 80112
Phone (303) 649-5614
Fax (303) 649-5318
Email donald-geiser@hp.com

Al Geist
Oak Ridge National Laboratory
P.O. Box 2008, Bldg. 6012
Oak Ridge, TN 37831-6367
Phone (423) 574-3153
Fax (423) 574-0680
Email gst@ornl.gov

Alan Geller
Silicon Graphics, Inc.
3000 Executive Parkway
Suite 410
San Ramon, CA 94583
Phone (925) 824-1791
Fax (925) 277-1982
Email geller@sanramon. sgi.com

Morven Gentleman
Institute for Information Technology
National Research Council
Montreal Road, Bldg M-50
Ottawa, ON, Canada K1A OR6
Phone (613) 993-9010
Fax (613) 952-0074
Email gentleman@iit.nrc.ca

Wendy Gentleman
HB 8000
Thayer School of Engineering
Dartmouth College
Hanover, NH 03755
Phone (603) 646-2685
Fax (603) 646-3856
Email wendy.gentleman@dartmouth.edu

Mike Gittings
Los Alamos National Laboratory
X-CI, MS F663
Los Alamos, NM 87545
Phone (505) 667-1453
Fax (5055 665-5553
Email gittings@lanl.gov

Joe Grisillo
2815 East Lake Ave., E.
Seattle, WA 98102
Phone (520) 297-3312
Email joeg@tera.com

John Gustafson
Ames Laboratory
Iowa State University
236 Wilhelm
Ames, IA 50011
Phone (515) 294-9294
Fax (515) 294-4491
Email gus@scl.ameslab.gov

Joyce Guzik
Los Alamos National Laboratory
X-TA, MS B220
Los Alamos, NM 87545
Phone (505) 667-8927
Fax (505) 665-2227
Email joy@lanl.gov

Jim Hack
National Center for Atmospheric Research
P.O. Box 3000
Boulder, CO 80307-3000
Phone (303) 497-1387
Fax (303) 497-1298
Email jhack@ucar.edu

Erik Hagersten
Sun Microsystems
MS UMPK12-302
901 San Antonio Ave.
Palo Alto, CA 94303
Phone (650) 786-6104
Fax (650) 568-9603
Email erik.hagersten@eng.sun.com

Steve Hammond
National Center for Atmospheric Research
Scientific Computing Division
P. O. Box 3000
Boulder, CO 80304
Phone (303) 497-1811
Fax (303) 497-1286
Email hammond@ncar.ucar.edu

Norman Hardy
143 Ramona Road
Portola Valley, CA
Phone (415) 851-2582
Email norm@netcom.com

Gary Haussmann
University of Colorado at Boulder
Dept. of Elect. & Comp. Engr
Room ECEElB55
Boulder, CO 80309
Phone (303) 492-8719
Fax (303) 492-2758
Email gary.haussmann@colorado.edu

Ann Hayes
Los Alamos National Laboratory
CIC-ACL, MS B287
Los Alamos, NM 87545
Phone (505) 665-4506
Fax (505) 665-4939
Email ahh@acl.lanl.gov

Andrew Heller
4501 Hennig Drive
Austin, TX 78738
Phone (512) 263-1666
Fax (512) 263-1665
Email heller@heller.com

Don Heller
Ames Laboratory
Iowa State University
327 Wilhelm Hall
Ames, Iowa 50011
Phone (515) 294-3784
Fax (5155 294-4491
Email heller@scl.ameslab.gov

Michael J. Henesey
IBM
P.O. Box 100
Somers, NY 10589
Phone (914) 766-3486
Fax (914) 766-3139

Kathy Hirons
Los Alamos National Laboratory
CIC-DO, MS B260
Los Alamos, NM 87545
Phone (505) 667-8167
Fax (5055 665-4361
Email kph@lanl.gov

Tad Hogg
Xerox Palo Alto Research Center
3333 Coyote Hill Road
Palo Alto, CA 94304
Phone (415) 812-4166
Fax (415) 812-4334
Email hogg@parc.xerox.com

Karin Hollerbach
Lawrence Livermore National Laboratory
P. O. Box 808, L-561
Livermore, CA 94550
Phone (925) 422-9111
Fax (925) 422-7819
Email hollerbach1@llnl.gov

Don Hossink
IBM
6001 Indian School Road NE
Albuquerque, NM 97110
Phone (505) 888-2326
Fax (505) 888-2781
Email dhossink@us.ibm.com

Jon Huppenthal
SRC Computers Inc.
4240 N. Nevada Ave.
Colorado Springs, CO 80907
Phone (719) 262-0213
Fax (7195 262-0223
Email hupp@srccomp.com

Ken Iobst
IDA/CCS
17100 Science Drive
Bowie, MD 20715-4300
Phone (301) 805-7375
Fax (301) 805-7602
Email ken@super.org

George Ivy
Logicon RDA
P. O. Box 92500
Los Angeles, CA 90009
Phone (310) 645-1122 x248
Fax (310) 645-2553
Email gmi@lanl.gov

Jay Johnson
Storage Tech
2270 South 88th Street
Louisville, CO 80028-4392
Phone (303) 661-5724
Fax (303) 661-6221
Email jay_johnson@storetech.com

Merritt Jones
The MITRE Corporation
MC-W-433
1820 Dolley Madison Blvd.
McLean, VA 22102
Phone (703) 883-5471
Fax (703) 883-3320
Email merritt@mitre.org

Chu Jong
University of New Mexico
Computer Science Department
Albuquerque, NM 87131
Phone (505) 277-3032
Fax (505) 277-6927
Email cjong@cs.unm.edu

Harry Jordan
Tera Computer Co.
2815 Eastlake Ave. E
Seattle,WA 98102-3027
Phone (206) 490-2105
Fax (206) 323-1318
Email Harry.Jordan@Colorado.edu

Ray Juzaitis
Los Alamos National Laboratory
X-DO, MS B218
Los Alamos, NM 87545
Phone (505) 667-5496
Fax (505) 665-4080
Email rjsq@lanl.gov

David Kahaner
ATIP
6-15-21 Roppongi, Harks Bldg 1F
Minato-ku, Tokyo, Japan
Phone +81-3-5411-6670
Fax +81-3-5411-6671
Email kahaner@xroads.cc.u-tokyo.ac.jp

Sid Karin
University of California at San Diego
9500 Gilman Drive
La Jolla, CA 92093-0505
Phone (619) 534-5075
Fax (619) 534-5056
Email skarin@ucsd.edu

James Kasdorf
Pittsburgh Supercomputing Center
Mellon Institute Building
4400 Fifth Avenue
Pittsburgh, PA 15213
Phone (412) 268-4960
Fax (412) 268-5832
Email kasdorf@psc.edu

Bruce Kasson
RCI Ltd.
2570 Bald Mountain Road
Vail, CO 81657
Phone (970) 476-6519
Fax (970) 476-2933
Email bkasson@vail.net

Dan Katz
NASA, JPL
4800 Oak Grove Drive
MS 168-522
Pasadena, CA 91109
Phone (818) 354-7359
Fax (818) 393-3134
Email daniel.s.katz@jpl.nasa.gov

Richard Kaufmann
Digital Equipment Corporation
8040 Girard Ave. #7
La Jolla, CA 92037
Phone (619) 917-3004
Email richard.kaufmann@digital.com

Gary Kent
U.S. Deptartment of Energy
10909 Outpost Drive
North Potomac, MD 20878
Phone (202) 586-8299
Fax (202) 586-1057
Email kent@dpmail.dp.doe.gov

Tom Kitchens
U.S. Department of Energy
Math., Information & Comp. Sci. Div.
19901 Germantown Road
Germantown, MD 20874-1290
Phone (301) 903-5152
Fax (301) 903-7774
Email kitchens@er.doe.gov

Ken Kliewer
Oak Ridge National Laboratory
P.O. Box 2008
Oak Ridge, TN 37831-6203
Phone (423) 241-2859
Fax (423) 241-2850
Email kliewer@ccs.ornl.gov

Robert Kluttz
Naval Security Group
9800 Savage Road
Suite 6685, OP Bldg #3
Fort Meade, MD 20755
Phone (301) 617-3500
Fax (301) 617-3114
Email ajbibk@fggm.dsis.gov

Manny Knill
Los Alamos National Laboratory
T-6, MS B265
Los Alamos, NM 87545
Phone (505) 665-8283
Email knill@lanl.gov

Kenneth Koch
Los Alamos National Laboratory
X-CI, MS F663
Los Alamos, NM 87545
Phone (505) 667-4288
Fax (505) 665-5553
Email krk@lanl.gov

Tom Kraay
15192 Harrison Hill Lane
Leesburg, VA 22075
Phone (703) 771-7056 (H)
Fax (703) 771-7056

Bill Kramer
Lawrence Berkeley National Laboratory
Building 50B - 4230
1 Cyclotron Drive
Berkeley, CA 94720
Phone (510) 486-7577
Fax (510) 486-4300
Email wtkramer@lbl.gov

Dale Land
Los Alamos National Laboratory
CIC-2, MS D445
Los Alamos, NM 87545
Phone (505) 667-9923
Email land@lanl.gov

Stephen Lee
Los Alamos National Laboratory
X-CI, MS F663
Los Alamos, NM 87545
Phone (505) 665-2989
Fax (505) 665-5553
Email srlee@lanl.gov

John Lekashman
NASA Ames Research Center
MS 258-5
Moffett Field, CA 94035
Phone (650) 604-4359
Fax (650) 604-4377
Email lekash@bayarea.net

Bill Lennon
Lawrence Livermore National Laboratory
P.O. Box 808, L-541
Livermore, CA 94551
Phone (925) 422- 1091
Fax (925) 424-2707
Email wjlennon@llnl.gov

Michael R. Leuze
P. O. Box 2008
Bldg 6012, MS 6367
Oak Ridge, TN 37831-6367
Phone (423) 574-3125
Fax (423) 574-0680
Email leuzemr@ornl.gov

Michael Levine
Pittsburgh Supercomputing Center
4400 Fifth Ave.
Pittsburgh, PA 15213
Phone (412) 268-4960
Fax (412) 268-5832
Email levine@a.psc.edu

Norm Margolus
545 Technology Square, Room 793
Cambridge, MA 02139
Phone (617) 253-2614
Fax (617) 258-7413
Email nhm@mit.edu

Ron Matlock
Fujitsu America
2500 City West Blvd.
Suite 300
Houston, TX 77042
Phone (713) 267-2350
Fax (713) 267-2305
Email ronm@fai.com

Mike McCoy
Lawrence Livermore National Laboratory
P.O. Box 808, L-68
Livermore, CA 94550
Phone (925) 422-4236
Fax (925) 422-3887
Email mccoy2@11nl.gov

Bill McCurdy
Lawrence Berkeley National Laboratory
MS 50B - 4230
Berkeley, CA 94720
Phone (510) 486-4283
Fax (510) 486-4300
Email cwmccurdy@lbl.gov

Charlie McMillan
Lawrence Livermore National Laboratory
P.O. Box 808, L-170
Livermore, CA 94551
Phone (925) 423-4994
Fax (925) 422-3389
Email mcmillan1@llnl.gov

Dan Meiron
Applied Mathematics, 217-50
Caltech
Pasadena, CA 91125
Phone (626) 395-4563
Fax (626) 584-5917
Email dm@ama.caltech.edu

William Mermagen
U. S. Army Research Laboratory
ATTN: AMSRL-CI
Aberdeen Proving Ground, MD 21005-5067
Phone (410) 278-6639
Fax (410) 278-5075
Email mermagen@arl.mil

Paul Messina
California Institute of Technology
1200 E. California Blvd. MC 158-79
Pasadena, CA 91125
Phone (626) 395-3907
Fax (626) 584-5917
Email messina@cacr.caltech.edu

George Michaels
Inst. for Comp. Sciences, MS SC3
George Mason University
Fairfax, VA 22030-4444
Phone (703) 993-1998
Fax (703) 993-1980
Email gmichael@gmu.edu

Doug Miller
Lawrence Livermore National Laboratory
P.O. Box 808, L-016
Livermore, CA 94550
Phone (925) 424-4238
Fax (925) 423-5112
Email dougmiller@llnl.gov

Ray Miller
Los Alamos National Laboratory
P.O. Box 1663
Los Alamos, NM 87545
Phone (505) 665-3222
Fax (505) 665-4361
Email rdm@lanl.gov

Ron Minnich
David Sarnoff Research Center
201 Washington Road
Princeton, NJ 08540
Phone (609) 734-3120
Fax (609) 734-2049
Email minnich@sarnoff.com

Ken Miura
Fujitsu Limited
Company Mail No./H185
1-1,Kamikodaraka 4-ChomeNakahara-ku,
211-8588, Japan
Phone (408) 456-7408
Fax (408) 456-7071
Email kenm@fai.fujitsu.com

Gloria Montoya-Rivera
Los Alamos National Laboratory
CIC-DO, MS B260
Los Alamos, NM 87545
Phone (505) 667-6165
Fax (5055 667-4361
Email gmon@lanl.gov

John Morrison
Los Alamos National Laboratory
DDCIC, MS B260
Los Alamos, NM 87545
Phone (505) 667-6164
Fax (505) 665-4361
Email jfm@lanl.gov

Ron Musick
Lawrence Livermore National Laboratory
P.O. Box 808, L-561
Livermore, CA 94550
Phone (925) 424-5015
Fax (9255 423-2993
Email rmusick@llnl.gov

David Nelson
Director
Office of Scientific Computing in Energy
U.S. Department of Energy, ER-30
Washington, DC 20585
Phone (301) 903-7486
Fax (301) 903-4846
Email nelson@er.doe.gov

Charles Nietubicz
US Army Research Laboratory
ATTN: AMSRL-CI-H
Aberdeen Proving Ground, MD 21005-5077
Phone (410) 278-3691
Fax (410) 278-5077
Email cjn@arl.mil

Steven Oberlin
SGI/Cray
1168 Industrial Blvd.
Chippewa Falls, WI 54729
Phone (715) 726-4335
Fax (715) 726-4537
Email smo@cray.com

Rod Oldehoeft
Los Alamos National Laboratory
ACL, MS B287
Los Alamos, NM 87545
Phone (505) 665-3663
Fax (505) 665-4939
Email rro@er.doe.gov

C. Edward Oliver
Oak Ridge National Laboratory
P. O. Box 2008, Bldg. 4500N, Room K-244
Oak Ridge, TN 37831-6259
Phone (423) 576-4666
Fax (423) 241-3249
Email oliverce@ornl.gov

Cherri M. Pancake
Oregon State University
Dept. of Computer Science
303 Dearborn Hall
Corvallis, OR 97331-3202
Phone (541) 737-2109
Fax (541) 737-3014
Email pancake@cs.orst.edu

Melinda Piket-May
University of Colorado
Dept. of Elect. & Comp.
Engr Room ECEElB55
Boulder, CO 80309-0425
Phone (303) 492-7448
Fax (3035 492-5323
Email mjp@hedgehog.colorado.edu

Milica Popovic
Electrical & Comp. Engr. Dept.
Northwestern University
2145 Sheridan Road
Evanston, IL 60208
Phone (847) 491-8887
Fax (8475 467-3217
Email poppy@nwu.edu

Terri Quinn
Lawrence Livermore National Laboratory
P.O. Box 808, L-61
Livermore, CA 94550
Phone (925) 423-2385
Fax (925) 423-8715
Email quinn1@llnl.gov

Justin Rattner
Intel Corporation
15201 N.W. Greenbrier Parkway
Beaverton, OR 97006
Phone (503) 677-7603
Email justin@ssd.intel.com

John Riganati
Sarnoff Corporation
CN 5300, 201 Washington Road
Princeton, NJ 08543-5300
Phone (609) 734-2170
Fax (609) 734-2049
Email riganati@sarnoff.com

Garry Rodrigue
Lawrence Livermore National Laboratory
P.O. Box 808, L-561
Livemmore, CA 94550
Phone (925) 422-4037
Fax (925) 422-7819
Email rodrigue@llnl.gov

Subhash Saini
NASA Ames Research Center
MS T27A-1
Moffett Field, CA 94035-1000
Phone (415) 604-4343
Fax (415) 604-3957
Email saini@nas.nasa.gov

Mark Seager
Lawrence Livermore National Laboratory
P. O. Box 808, L-060
Livermore, CA 94550
Phone (925) 423-3141
Fax (925) 423-8715
Email seager1@llul.gov

Ron Selva
Silicon Graphics, Inc.
600 Waterfront Drive
Pittsburgh, PA 15222
Phone (412) 320-2013
Fax (412) 320-2114
Email selva@pittsburgh.sgi.com

Bill Sembrat
SRC Computers, Inc.
395 Rider Ridge Road
Santa Cruz, CA 95065
Phone (408) 423-0697
Fax (408) 423-6822
Email sembrat@aol.com

Margaret Simmons
NAVO MSRC at Stennis
Building 1103, Room 248B
Stennis Space Center, MS 39592
Phone (601) 688-1583
Fax (601) 688-1064
Email mls@sdsc.edu

Horst Simon
Lawrence Berkeley National Laboratory
1 Cyclotron Drive
Building 50B - 4230
Berkeley, CA 94720
Phone (510) 486-7377
Fax (510) 486-4300
Email HDSimon@lbl.gov

Richard Sincovec
Division of Computer Science
Univ. of Texas at San Antonio
6900 North Loop 1604 West
San Antonio, TX 78249-0677
Phone (210) 458-4434
Fax (210) 458-4437
Email sincovec@cs.utsa.edu

Charles Slocomb
Los Alamos National Laboratory
CIC-DO, MS B260
Los Alamos, NM 87545
Phone (505) 667-6164
Fax (505) 665-4361
Email cas@lanl.gov

Gary Smaby
Smaby Group, Inc.
1315 Foshay Tower
Minneapolis, MN 55402
Phone (612) 333-0002
Fax (612) 333-0204
Email gary@smaby.com

Burton Smith
Tera Computer Company
2815 Eastlake Ave. East
Seattle, WA 98102-1318
Phone (206) 490-2000
Fax (206) 323-1318
Email burton@tera.com

Allan Snavely
San Diego Supercomputer Center
10100 Hopkins
La Jolla, CA 92186
Phone (619) 534-5158
Fax (619) 534-5152
Email allans@sdsc.edu

Patrick Soran
Los Alamos National Laboratory
X-CI, MS F663
Los Alamos, NM 87545
Phone (505) 667-8103
Fax (505) 665-5553
Email soran@lanl.gov

Will Stackhouse
1123 Stanford Ave.
Redondo Beach, CA 90278-4039
Phone (310) 374-6838

Gene Stokes
Defense Spec. Weapons Agency
6801 Telegraph Road
Alexandria, VA 22310-3398
Phone (703) 325-6414
Fax (703) 325-2954
Email gstokes@dswahome.dswa.mil

Jack Strauss
Spectral Systems, Inc.
5200 Springfield Pike, Suite 320
Dayton, OH 45431
Phone (937) 256-9729
Fax (937) 256-9317
Email jack_strauss@ spectral-sys.com

Steve Stringer
Silicon Graphics, Inc.
1742 Central Ave.
Los Alamos, NM 87544
Phone (505) 662-4000
Fax (505) 662-9457
Email sfs@sgi.com

Thomas M. Sutton
Knolls Atomic Power Lab
F3-7
P. O. Box 1072
Schenectady, NY 12301-1072
Phone (518) 395-7047
Fax (518) 395-4422
Email sutton@kapl.gov

Robert Tomlinson
Los Alamos National Laboratory
CIC-8 MS B272
Los Alamos, NM 87545
Phone (505) 665-6599
Fax (505) 665-6333
Email bob@lanl.gov

Bruce Trent
Los Alamos National Laboratory
X-NH, MS F664
Los Alamos, NM 87545
Phone (505) 667-8066
Fax (505) 665-3561
Email bct@lanl.gov

Bill Trimmer
Belle Mead Research, Inc.
58 Riverview Terrace
Belle Mead, NJ 08502-3218
Phone (908) 359-0012
Fax (908) 359-2094
Email W.Trimmer@EEE.org

Verna Van Aken
Los Alamos National Laboratory
CIC-13, MS P223
Los Alamos, NM 87545
Phone (505) 665-5779
Fax (505) 665-0218
Email vanakenv@lanl.gov

Manuel Vigil
Los Alamos National Laboratory
CIC-7, MS B294
Los Alamos, NM 87545
Phone (505) 667-5243
Fax (505) 665-6333
Email mbv@lanl.gov

Steve Wallach
CenterPoint Ventures
Two Galleria Tower
13455 Noel Road, Suite 1670
Dallas, TX 75240
Phone (972) 702-1132
Email wallach@centerpointvp.com

Bob Ward
University of Tennessee
107 Ayres Hall
Knoxville, TN 37996-5067
Phone (615) 9744400
Fax (615) 974-4404
Email ward@cs.utk.edu

Geert Wenes
IBM
23 Via Brisa
Sante Fe, NM 87501
Phone (505) 438-8286
Email gwenes@us.ibm.com

Paul Wheeler
Lawrence Livermore National Laboratory
P.O. Box 808, L-022
Livermore, CA 94550
Phone (925) 422-0535
Fax (925) 423-0925
Email wheeler2@llnl.gov

Elizabeth Williams
Center for Computer Sciences
17100 Science Drive
Bowie, MD 20715-4300
Phone (301) 805-7468
Fax (301) 805-7602
Email ew@super.org

Paul Woodward
Astronomy Department
Univ. of Minnesota
116 Church Street S.E.
Minneapolis, MN 55455
Phone (612) 624-4097
Fax (612) 626-2029
Email paul@lcse.umn.edu

Jack Worlton
2231 East 3980 South
Salt Lake City, UT 84124
Phone (801) 277-7442
Fax (801) 272-2234
Email worlton@aol.com

This report has been reproduced directly from the
best available copy.

It is available to DOE and DOE contractors from
the Office of Scientific and Technical Information,
P.O. Box 62,
Oak Ridge, TN 37831.
Prices are available from
(615) 576-8401.

It is available to the public from the
National Technical Information Service,
US Department of Commerce,
5285 Port Royal Rd.
Springfield, VA 22616.

Los Alamos
NATIONAL LABORATORY
Los Alamos, New Mexico 87545