21st Century Engines of Discovery

David H. Bailey

Chief Technologist, Computational Research Dept

Lawrence Berkeley National Laboratory

http://crd.lbl.gov/~dhbailey

Laplace Anticipates Modern High-End Computers

"An intelligence knowing all the forces acting in nature at a given instant, as well as the momentary positions of all things in the universe, would be able to comprehend in one single formula the motions of the largest bodies as well as of the lightest atoms in the world, provided that its intellect were sufficiently powerful to subject all data to analysis; to it nothing would be uncertain, the future as well as the past would be present to its eyes."

-- Pierre Simon Laplace, 1773

Who Needs High-End Computers?

Expert predictions:

- (c. 1945) Thomas J. Watson (CEO of IBM):
  "World market for maybe five computers."
- (c. 1975) Seymour Cray:
  "Only about 100 potential customers for Cray-1."
- *(c. 1977) Ken Olson (CEO of DEC):*
  "No reason for anyone to have a computer at home."
- (c. 1980) IBM study:
  "Only about 50 Cray-1 class computers will be sold per year."

Present reality:

- Many homes now have 5 Cray-1 class computers.
- Latest PCs outperform 1990-era supercomputers.

Early Days of Parallel Computing (1985-1995)

- Practical parallel systems became commercially available for the first time.
- Some scientists obtained remarkable results.
- Many were enthused at the potential for this technology for scientific computing.

BUT

- Numerous faults and shortcomings of these systems (hardware and software) were largely ignored.
- Many questionable performance claims were made, both by vendors and by scientific users – “scientific malpractice.”

Where Are We Today?

- Several commercial vendors are producing robust, high-performance, well-supported systems.
- New systems achieve up to 50 Tflop/s (5×10^{13} flops/sec) peak performance, and 5-20 Tflop/s sustained performance on real scientific computations.
- We are on track to achieve 1 Pflop/s by 2009 or 2010.
- Many scientists and engineers have converted their codes to run on these systems.
- Numerous large industrial firms are using highly parallel systems for real-world engineering work.

But numerous challenges remain.

LBNL's Seaborg System

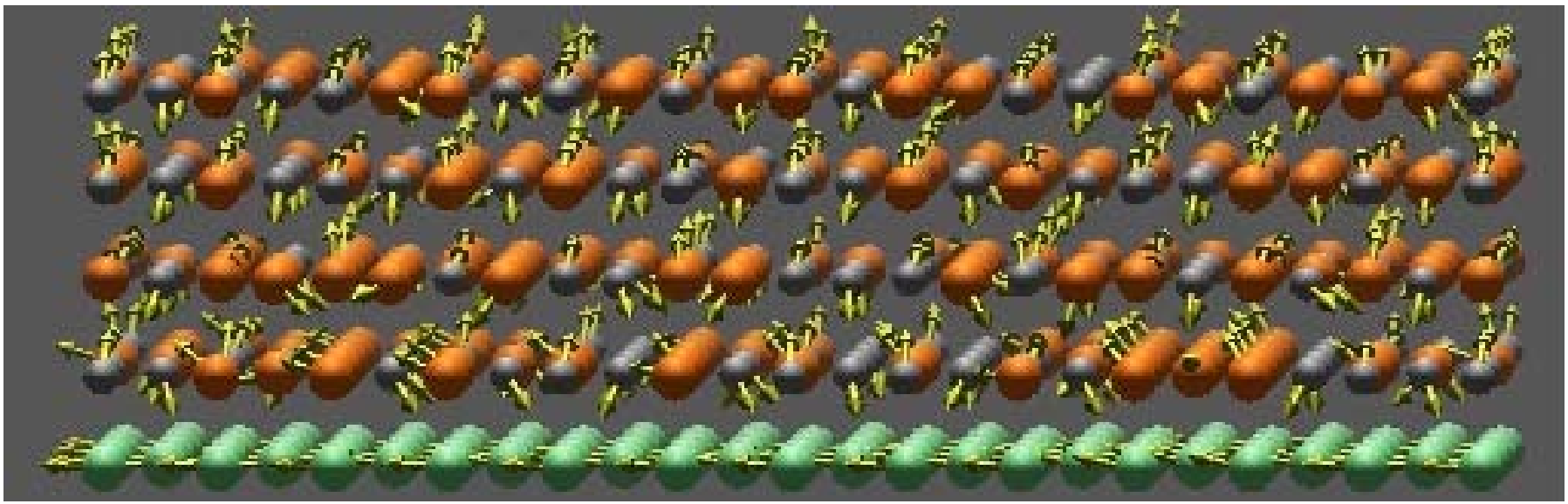
- 6000-CPU IBM SP: 10 Tflop/s (10 trillion flops/sec).
- Currently #21 on Top500 list of most powerful systems.

Applications for Future Petaflops-Scale Computers

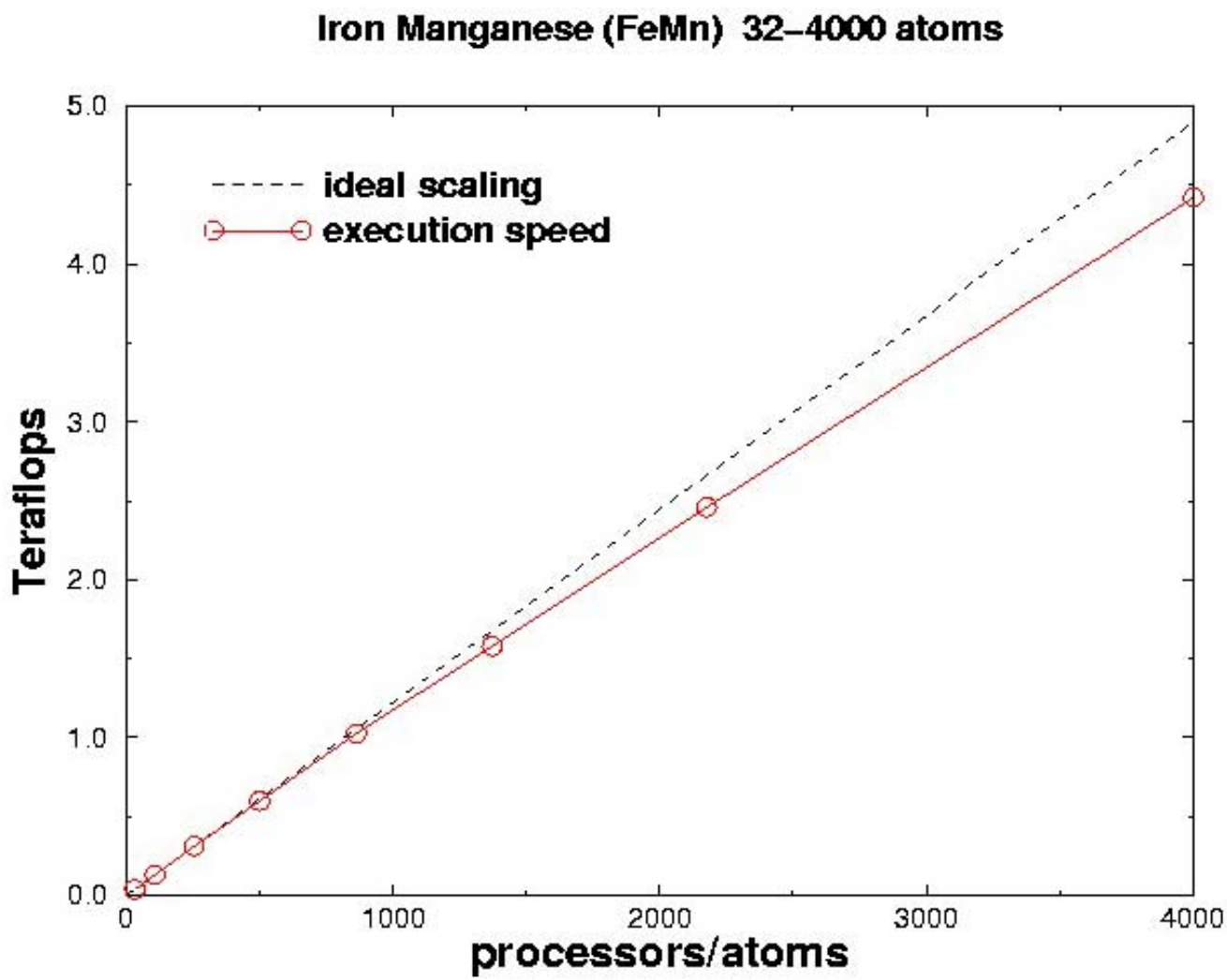
- Protein folding.
- Weather forecasting.
- Business data mining.
- DNA sequence analysis.
- Experimental mathematics.
- Inter-species DNA analyses.
- Medical imaging and analysis.
- Nuclear weapons stewardship.
- Multiuser immersive virtual reality.
- National-scale economic modeling.
- Climate and environmental modeling.
- Molecular nanotechnology design tools.
- Cryptography and digital signal processing.

Nanoscience

Simulations of physical phenomena at the nanometer scale lead to future nanotech-produced materials and devices.



Iron-Manganese Runs on Seaborg



Nanoscience Code Characteristics

- Many-body methods:
  - Quantum Monte-Carlo.
  - Eigenfunction-type calculations (diagonalization of large matrices).
- Single-particle methods:
  - Wave functions are expanded in plane waves.
  - Large 3-D FFTs, dense linear algebra.
- Classical molecular dynamics codes:
  - Model inter-molecular forces using classical methods.
  - Used to study synthesis of nanostructures and large structures beyond the scope of quantum calculations.

Nanoscience Requirements

Electronic structures and magnetic materials:

- Current: ~500 atom; 1.0 Tflop/s, 100 Gbyte memory.
- Future (hard drive simluation): 5000 atom;
  30 Tflop/s, 4 Tbyte memory.

Molecular dynamics:

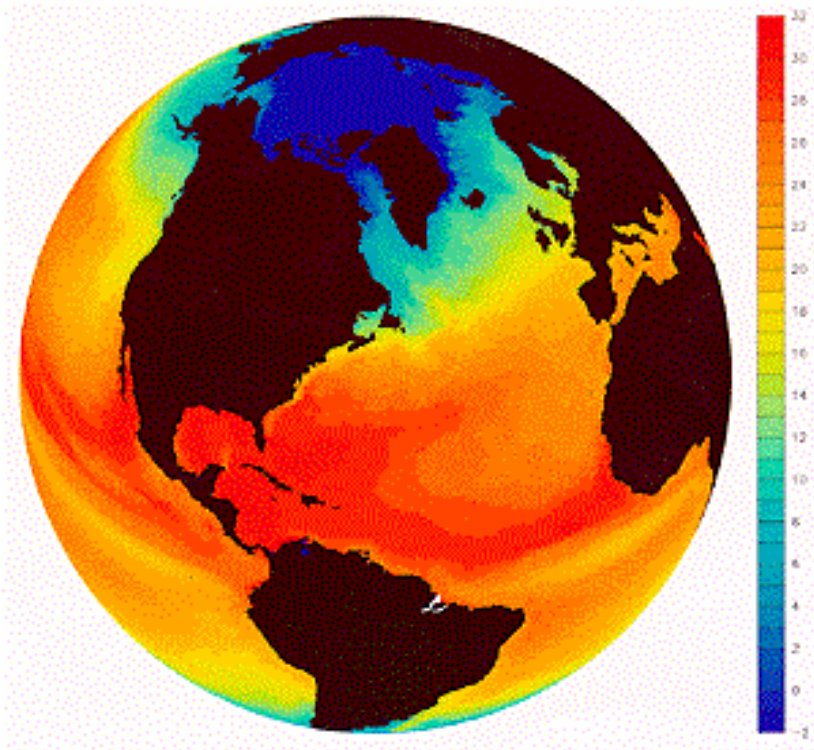
- Current: 10^9 atoms, ns time scale; 1 Tflop/s, 50 Gbyte mem.
- Future: alloys, us time scale; 20 Tflop/s, 5 Tbyte memory.

Continuum solutions:

- Current: single-scale simulation; 30 million finite elements.
- Future: multiscale simulations; 10 x current requirements.

Climate Modeling

Large community-developed codes, involving complex coupling of atmosphere, ocean, sea ice, land systems, are used to study long-term climate trends.



Climate Modeling Code Characteristics

- Solve equations of hydrodynamics, radiation transfer, thermodynamics, chemical reaction rates.
- Large finite difference methods, on regular spatial grids (require high main memory bandwidth).
- Short- to medium-length FFTs are used, although these may be replaced in future.
- Sea ice and land codes are difficult to vectorize.
- Scalability is often poor, due to limited concurrency.

Scientists would love to use finer grids, which would exhibit greater scalability, but then a century-long simulation would not be feasible.

Climate Modeling Requirements

Current state-of-the-art:

- Atmosphere: 1 horizontal deg spacing, with 29 vertical layers.
- Ocean: 0.25 x 0.25 degree spacing, 60 vertical layers.
- Currently requires one minute run time per simulated day, on 256 CPUs.

Future requirements (to resolve ocean mesoscale eddies):

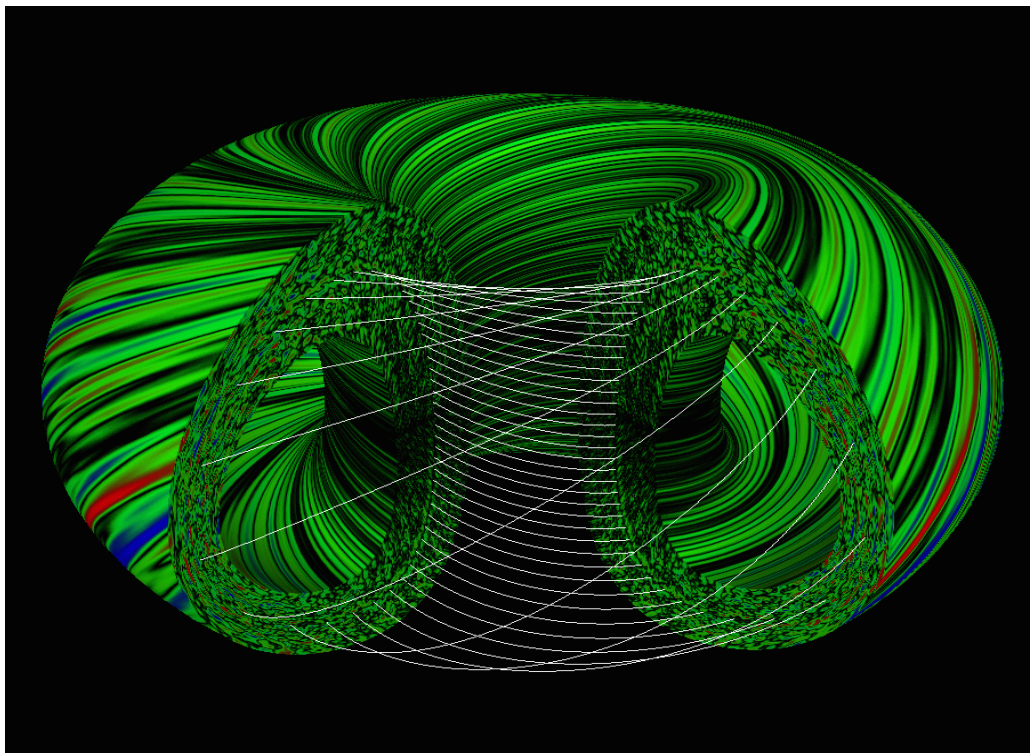
- Atmosphere: 0.5 x 0.5 deg spacing.
- Ocean: 0.125 x 0.125 deg spacing.
- Computational requirement: 17 Tflop/s.

Future goal: resolve tropical cumulus clouds:

- 2 to 3 orders of magnitude more than above.

Fusion Energy Research

Computational simulations help scientists understand turbulent plasmas in nuclear fusion reactor designs.



Fusion Code Characteristics

- Simulating the onset and evolution of instabilities is particularly difficult, because large time and length scales.
- Balance between heat generation of burning plasma and heat loss from electromagnetic turbulence needs to be better understood.
- Multi-physics, multi-scale computations.
- Numerous algorithms and data structures.
- Regular and irregular access computations.
- Adaptive mesh refinement.
- Advanced nonlinear solvers for stiff PDEs.

Fusion Requirements

Tokamak simulation -- ion temperature gradient turbulence in ignition experiment:

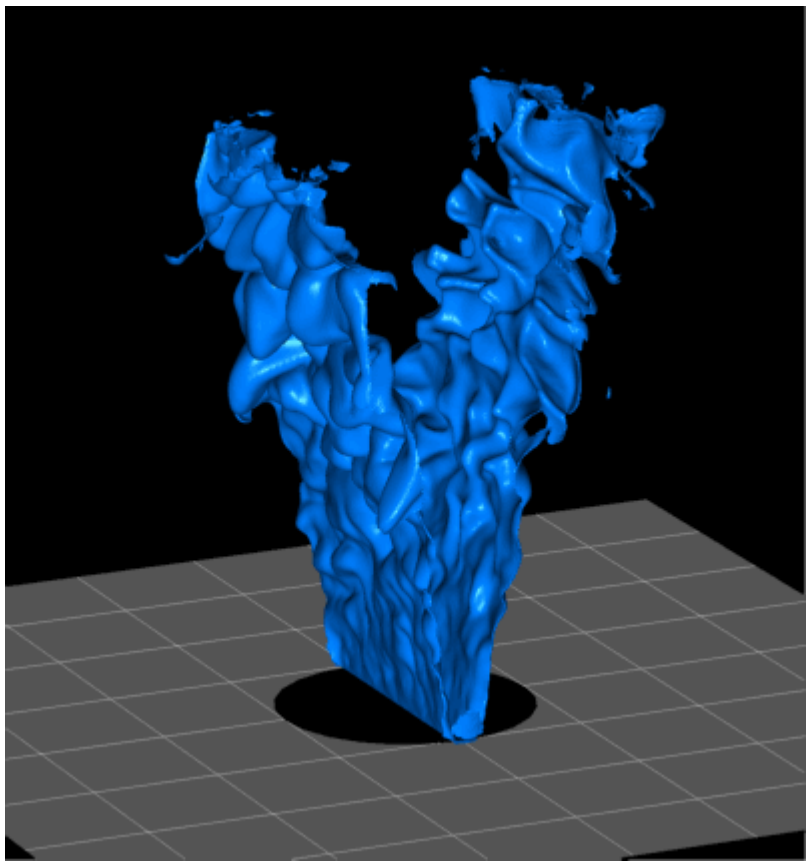
- Grid size: 3000 x 1000 x 64, or about 2 x 10^8 gridpoints.
- Each grid cell contains 8 particles, for total of 1.6 x 10^9.
- 50,000 time steps required.
- Total cost: 3.2 x 10^{17} flop (8 hours on 10 Tflop/s system), 1.6 Tbyte main memory.
- Improved plasma models will increase run time by 10X.

All-Orders Spectral Algorithm (AORSA) – to address effects of RF electromagnetic waves in plasmas.

- 120,000 x120,000 complex linear system.
- 230 Gbyte memory.
- 1.3 hours on 1 Tflop/s.
- 300,000 x 300,000 linear system requires 8 hours.
- Future: 6,000,000 x 6,000,000 system (576 Tbyte memory), 160 hours on 1 Pflop/s system.

Combustion

Simulations of internal combustion engines can be used to design systems with greater efficiency and lower pollution.



Combustion Code Characteristics

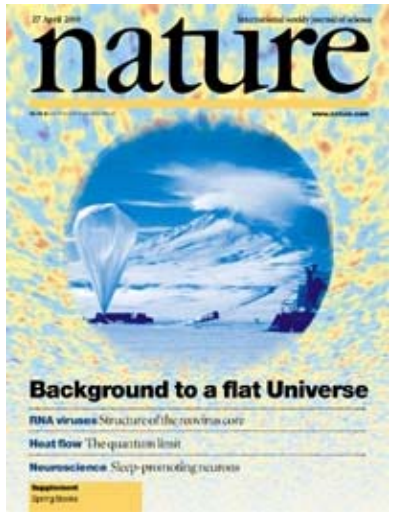
- Span huge range in time and length scales (10^9), requiring large and adaptively-managed meshes.
- Explicit finite difference, finite volume and finite element schemes for systems of nonlinear PDEs.
- Implicit finite difference, finite volume and finite element schemes for elliptic and parabolic PDEs (iterative sparse linear solvers).
- Zero-dimensional physics, including evaluation of thermodynamic and transport data.
- Adaptive mesh refinement.
- Lagrangian particle methods.

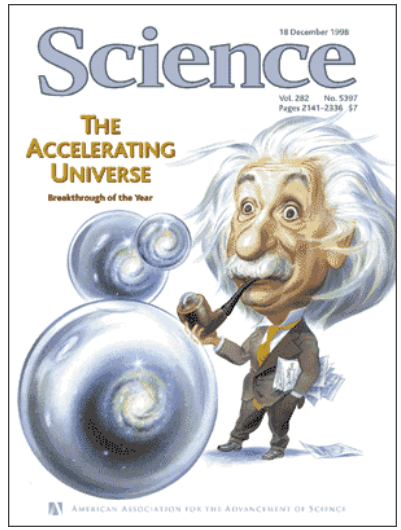
Combustion Requirements

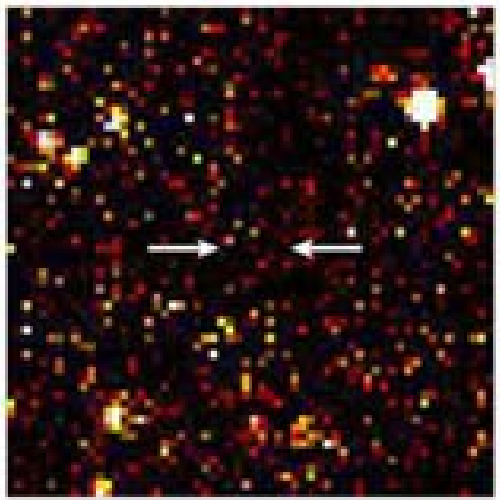
- Because of huge dynamic range of time and length scales, it is easy to construct problems several orders of magnitude beyond current capabilities.
- Current challenge is to find algorithms and techniques that produce reasonable approximations for certain specially constructed problems, in reasonable runtime.

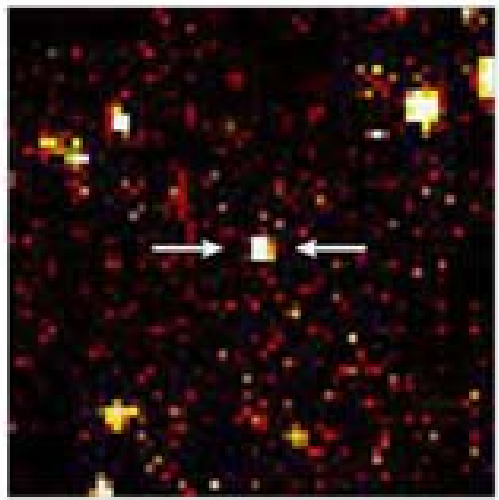
Astrophysics and Cosmology

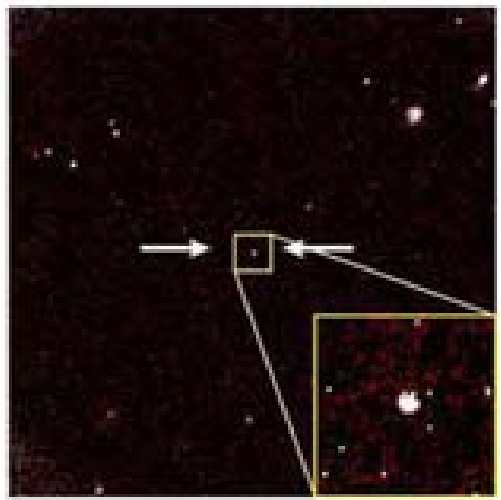
- The oldest, most distant Type 1A supernova was confirmed by computer analysis at NERSC.
- Supernova results point to an accelerating universe.
- Analysis at NERSC of cosmic microwave background data shapes concludes that geometry of the universe is flat.











Astrophysics Code Characteristics

- Undergoing a transformation from a data-starved discipline to a data-swamped discipline.
- Supernova hydrodynamics, energy transport, black hole simulations and universe field theory.
- Multi-physics and multi-scale phenomena.
- Large dynamic range in time and length.
- Requires adaptive mesh refinement.
- Dense linear algebra.
- FFTs.
- Spherical harmonic transforms.
- Operator splitting methods.

Astrophysics Requirements

Supernova simulation:

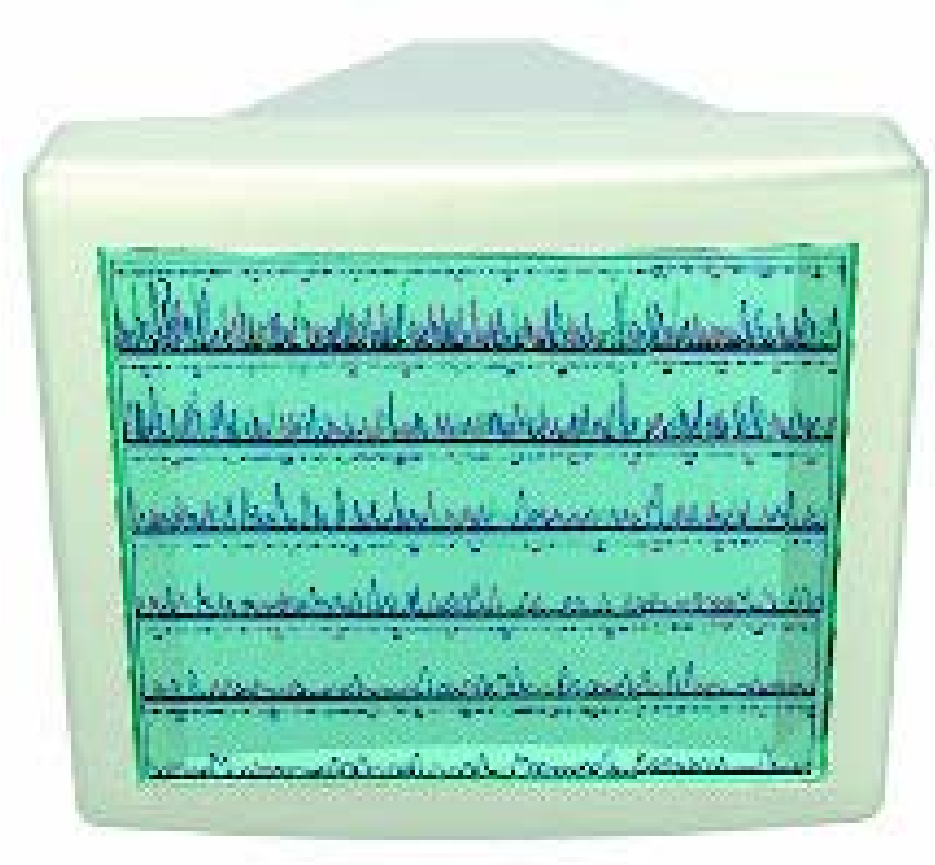
- Current models are only 2-D – require 1,000 CPU-hours.
- Initial 3-D model calculations will require 1,000,000 hours per run, on a system 10X as powerful as Seaborg.
- Will require 50 Tflop/s sustained system to perform meaningful 3-D calculations.

Analysis of cosmic microwave background data:

♦ MAXIMA data	5.3×10^{16} flops	100 Gbyte mem
♦ BOOMERANG data	1.0×10^{19} flops	3.2 Tbyte mem
♦ Future MAP data	1.0×10^{20} flops	16 Tbyte mem
♦ Future PLANCK data	1.0×10^{23} flops	1.6 Pbyte mem

Computational Biology

Computational simulations are yielding new insights into biology, and are basis for new high-tech pharmaceuticals.



Biology Applications

- Protein folding
  - Determines the equilibrium 3-D shape of a protein molecule.
  - Tremendous potential for drug design, pathogen analysis and studies of human biology.
  - This is a very challenging problem, even assuming massive computer power, although numerous advances have been made.
  - One approach – create an extensive repository of protein fold structures.
- Analysis of photosynthesis
  - Better understand carbon sequestration and energy transfer systems.

Biology Characteristics and Requirements

Key computational algorithms:

- Quantum Monte-Carlo.
- Classical molecular dynamics.
- Genome analysis algorithms (BLAST, etc).

Requirements:

- Hundreds of runs are currently planned, using up to 1000 CPUs.
- Up to 10X the current allocation will be required in the future.

Experimental Mathematics: Discovering New Theorems by Computer

High-precision (multi-hundred-digit) calculations of series and integrals, combined with integer-relation detection algorithms, have yielded some remarkable new results in mathematics. Examples (2005):

$$\frac{24}{7\sqrt{7}}\int_{\pi/3}^{\pi/2}\log\left|\frac{\tan t+\sqrt{7}}{\tan t-\sqrt{7}}\right|dt \stackrel{?}{=} \sum_{n=0}^{\infty}\left[\frac{1}{(7n+1)^2}+\frac{1}{(7n+2)^2}-\frac{1}{(7n+3)^2}+\frac{1}{(7n+4)^2}-\frac{1}{(7n+5)^2}-\frac{1}{(7n+6)^2}\right]$$

$$\zeta(10) \stackrel{?}{=} \frac{36\cdot 512}{11143}\left[\sum_{k=1}^{\infty}\frac{1}{k^{11}\binom{2k}{k}}+\frac{9}{4}\sum_{k=1}^{\infty}\frac{1}{k^{6}\binom{2k}{k}}\sum_{j=1}^{k-1}\frac{1}{j^4}+\frac{3}{2}\sum_{k=1}^{\infty}\frac{1}{k^{2}\binom{2k}{k}}\sum_{j=1}^{k-1}\frac{1}{j^8}+\frac{9}{4}\sum_{k=1}^{\infty}\frac{1}{k^{4}\binom{2k}{k}}\sum_{j=1}^{k-1}\frac{1}{j^6}+\frac{27}{8}\sum_{k=1}^{\infty}\frac{1}{k^{2}\binom{2k}{k}}\sum_{j=1}^{k-1}\frac{1}{j^4}\sum_{i=1}^{j-1}\frac{1}{i^4}\right]$$

Each of these identities has been verified to multi-thousand-digit accuracy, but proofs are not yet known.

Some Supercomputer-Class Experimental Math Computations

- Identification of B_4, the fourth bifurcation point of the logistic iteration:
  - Integer relation of size 121; 10,000-digit arithmetic.
- Identification of Euler-zeta sums:
  - Hundreds of integer relation problems, each of size 145 and requiring 5,000-digit arithmetic.
- Finding relation involving root of Lehmer's polynomial:
  - Integer relation of size 125; 50,000-digit arithmetic.
  - Requires 16 hours on a 64-CPU IBM SP parallel computer.
- Verification of the first identity on previous viewgraph.
  - Evaluation of the integral of a function with a nasty singularity, to 20,000-digit accuracy.
  - Requires one hour on a 1024-CPU parallel computer.

Research Questions for Future High-End Computing

- Can reliable, highly integrated systems be designed with 100,000 to 1,000,000 CPUs?
- Will exotic new device technology require different hardware architectures?
- When will 128-bit floating-point hardware be required?
- Can existing system software be used on these large systems?
- Will new programming models and/or languages be required?
- How well with real-world scientific applications scale on such systems?

The Japanese Earth Simulator System

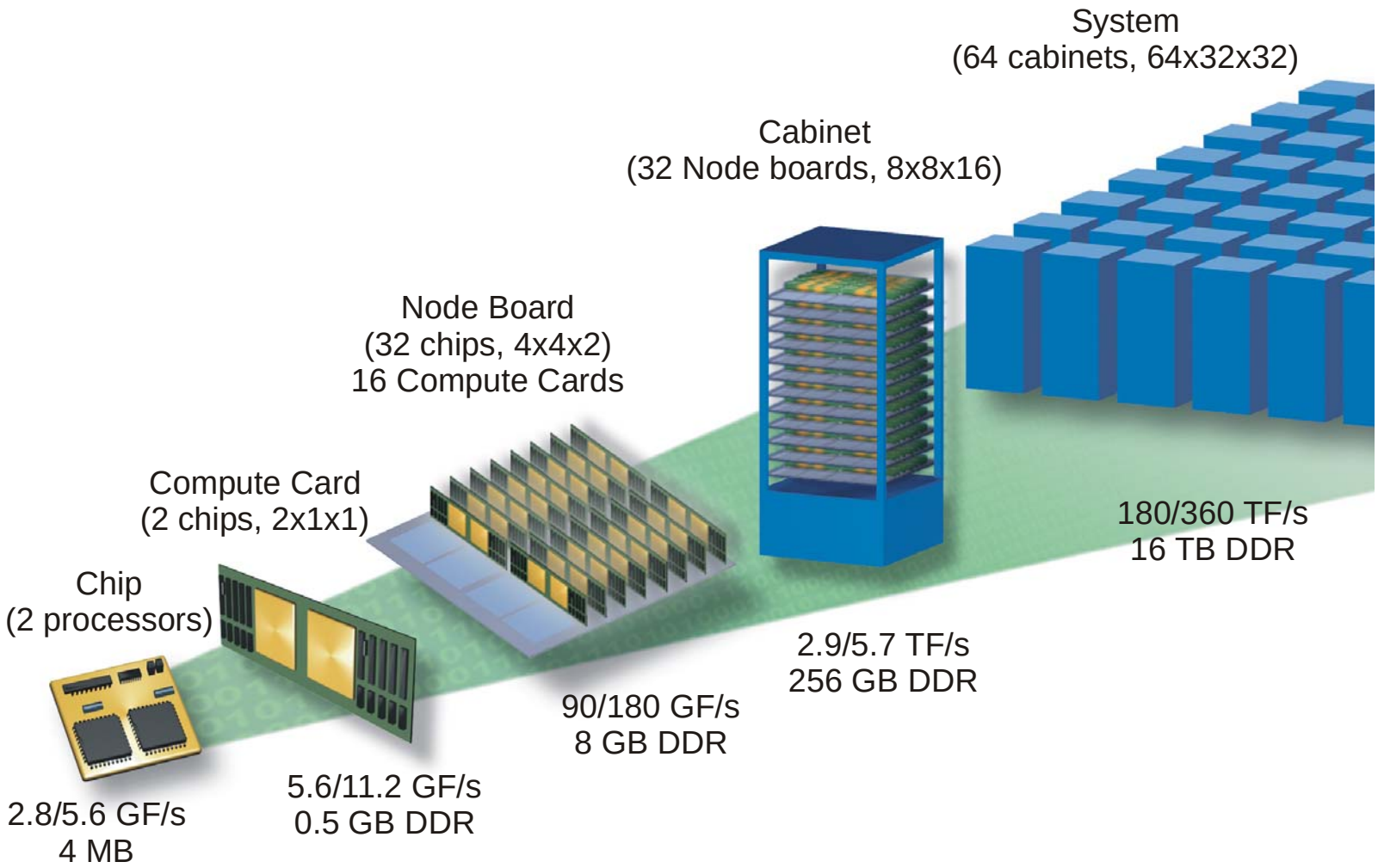
System design:

- Architecture: Crossbar-connected multi-proc vector system.
- Performance: 640 nodes x 8 proc per node x 8 Gflop/s per proc = 40.96 Tflop/s peak
- Memory: 640 nodes x 16 Gbyte per node = 10.24 Tbyte.

Sustained performance:

- Global atmospheric simulation: 26.6 Tflop/s.
- Fusion simulation (all HPF code): 12.5 Tflop/s.
- Turbulence simulation (global FFTs): 12.4 Tflop/s.

IBM's Blue Gene/L System



Other Future High-End Designs

- Processor in memory
  - Currently being pursued by a team headed by Prof. Thomas Sterling of Cal Tech.
  - Seeks to design a high-end scientific system based on special processors with embedded memory.
  - Advantage: significantly greater processor-memory bandwidth.
- Streaming supercomputer
  - Currently being pursued by a team headed by Prof. William Dally of Stanford.
  - Seeks to adapt streaming processing technology, now used in game market, to scientific computing.
  - Projects 200 Tflop/s, 200 Tbyte system will cost $10M in 2007.

40 Years of Moore's Law

Gordon Moore, 1965:

> *"The complexity for minimum component costs has increased at a rate of roughly a factor of two per year… Certainly over the short term this rate can be expected to continue, if not to increase. Over the longer term, the rate of increase is a bit more uncertain, although there is no reason to believe it will not remain nearly constant for at least 10 years."*
> [Electronics, Apr. 19, 1965, pg. 114-117.]

April 19, 2005: 40th anniversary of Moore's original prediction. No end in sight – progress is assured for at least another ten years.

Beyond Silicon: Sooner Than You Think

Nanotubes:

- Can function as conductors, memory and logic devices.
- Nantero, a venture-funded firm, has devices in development.

Molecular self-assembly:

- Researchers at HP have created a memory device with crisscrossing wires 2 nm wide, 9 nm apart.
- 2004 goal: 1000 bits/μ^2 (compared to 10 bits/μ^2 for DRAM).
- Zettacore, a venture-funded firm, is pursuing related ideas.

Molecular electronics:

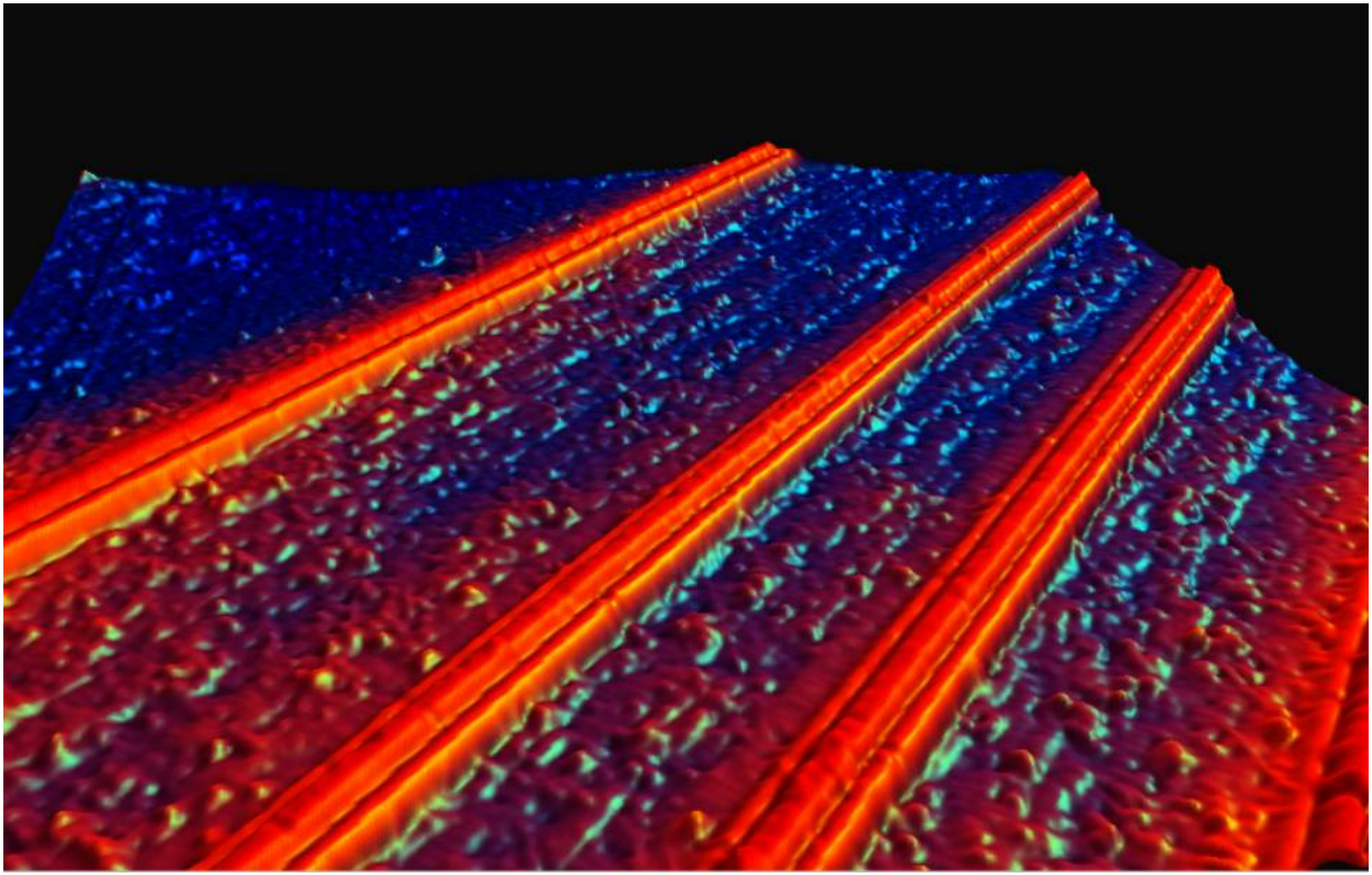
- Researchers Mitre and UCLA have demonstrated organic molecules that act as electronic logic devices.

Quantum dots:

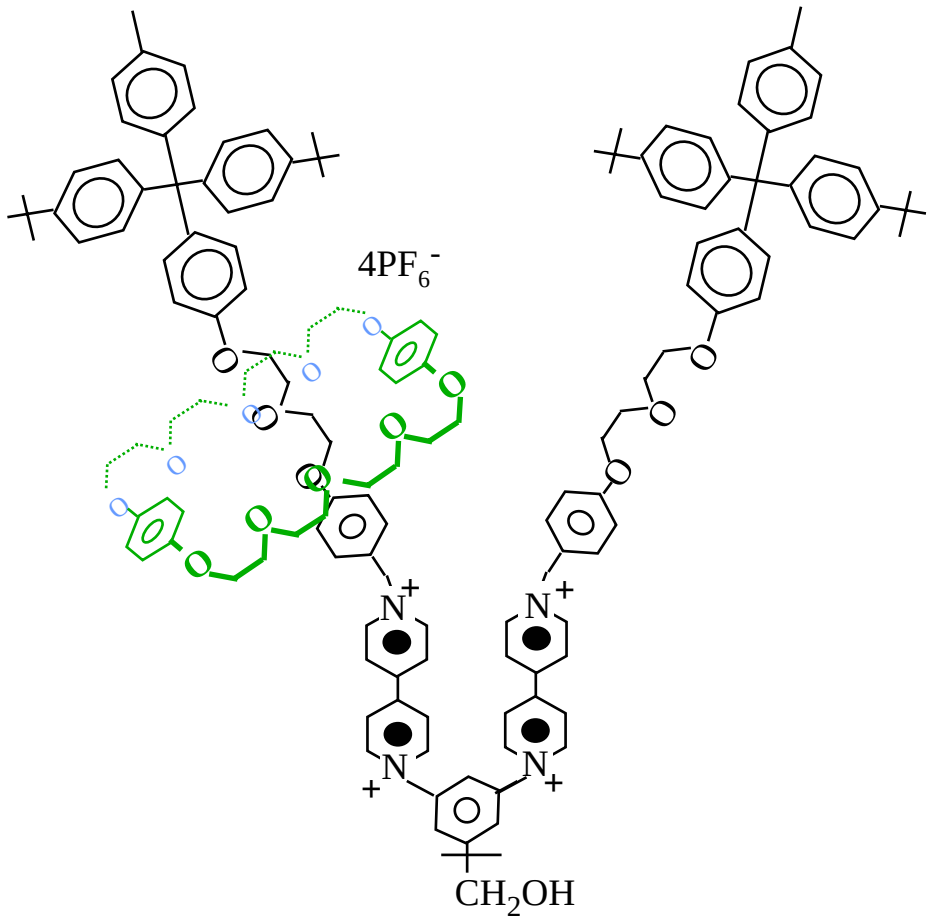
- Atomic-scale devices can do logic and store data.

Self-Assembled Wires 2nm Wide

[P. Kuekes, S. Williams, HP Labs]



Rotaxane Molecular Switch
[Prof. Fraser Stoddart, UCLA]



Fundamental Device Limits

Assume:

- Power dissipation of 1 watt at room temperature,
- A spatial volume of 1 cm^3.

Q: How many bit operation/second can be performed by a nonreversible computer executing Boolean logic?

A: $P/kT \log(2) = 3.5 \times 10^{20}$ bit ops/s

Q: How many bits/second can be transferred?

A: $\text{sqrt}(cP/kTd) = 10^{18}$ bit/s

“There’s plenty of room at the bottom”

-- Richard Feynman, 1959.

Amdahl's Law and Little's Law

Amdahl's Law: On a highly parallel system, sequential and modestly parallel code will dominate run time, limiting scalability.

Little's Law: Concurrency = latency x aggregate bandwidth. In practice, 10 times this much concurrency is needed.

Conclusion: We must identify and expose 10^9-way concurrency in every significant step of a large computation, or else we will not be able to productively utilize a 1 Pflop/s system with 10^6 CPUs.

The Performance Evaluation Research Center (PERC)

One of five Integrated Software Infrastructure Centers funded through the DoE SciDAC program.

Research thrusts:

- Tools for performance monitoring and code tuning.
- Performance modeling (predicting performance, based on characteristics of application and system).
- Generic (semi-automatic) code optimization.
- Applying tools to large-scale scientific codes.

Performance Modeling

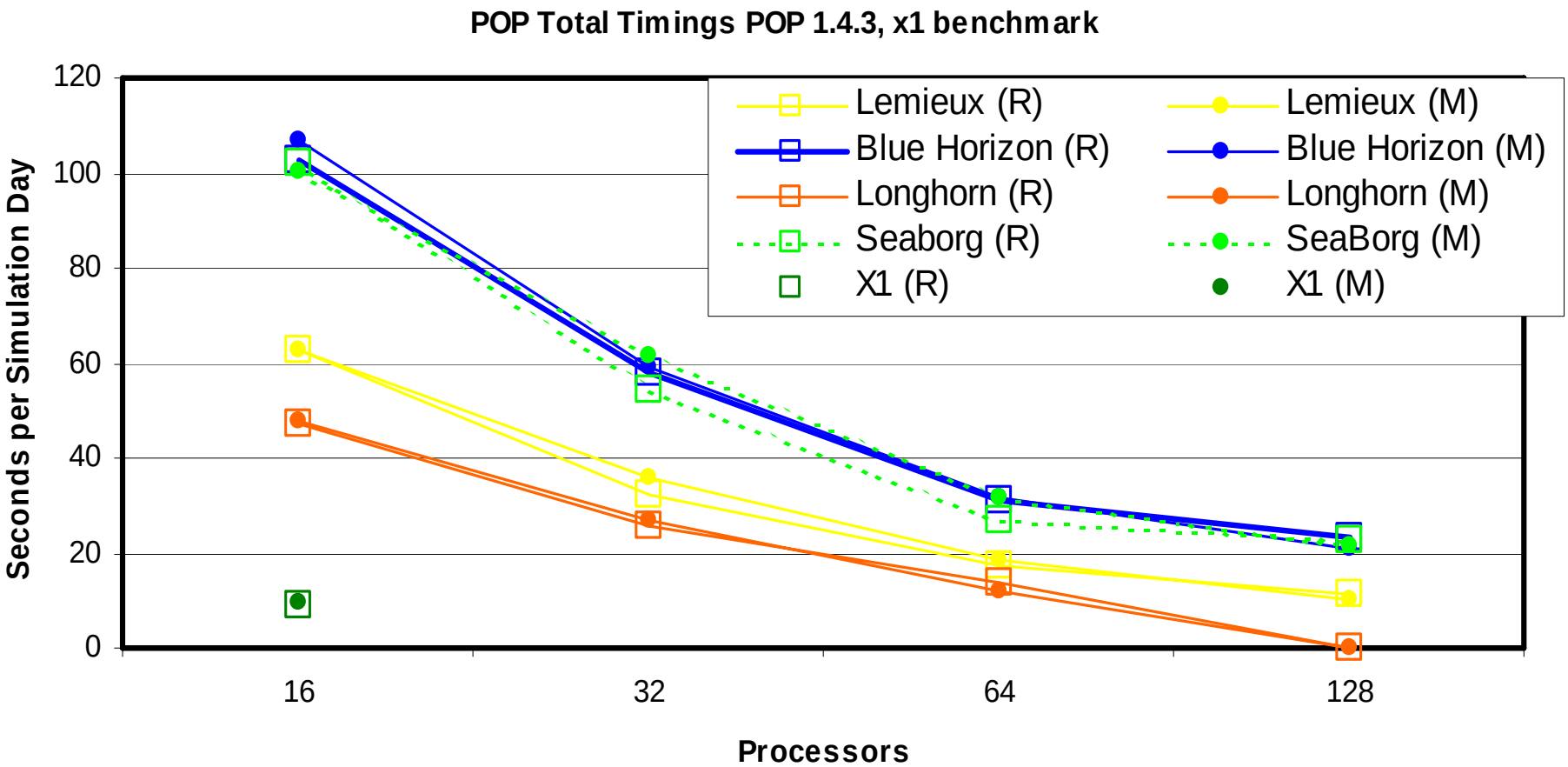
Methodology:

- Use semi-automatic tools to obtain “application signatures” of scientific applications.
- Use semi-automatic tools to obtain “machine signatures” of computer systems.
- “Convolve” these signatures to produce predictions of performance of a specific application on a specific system.

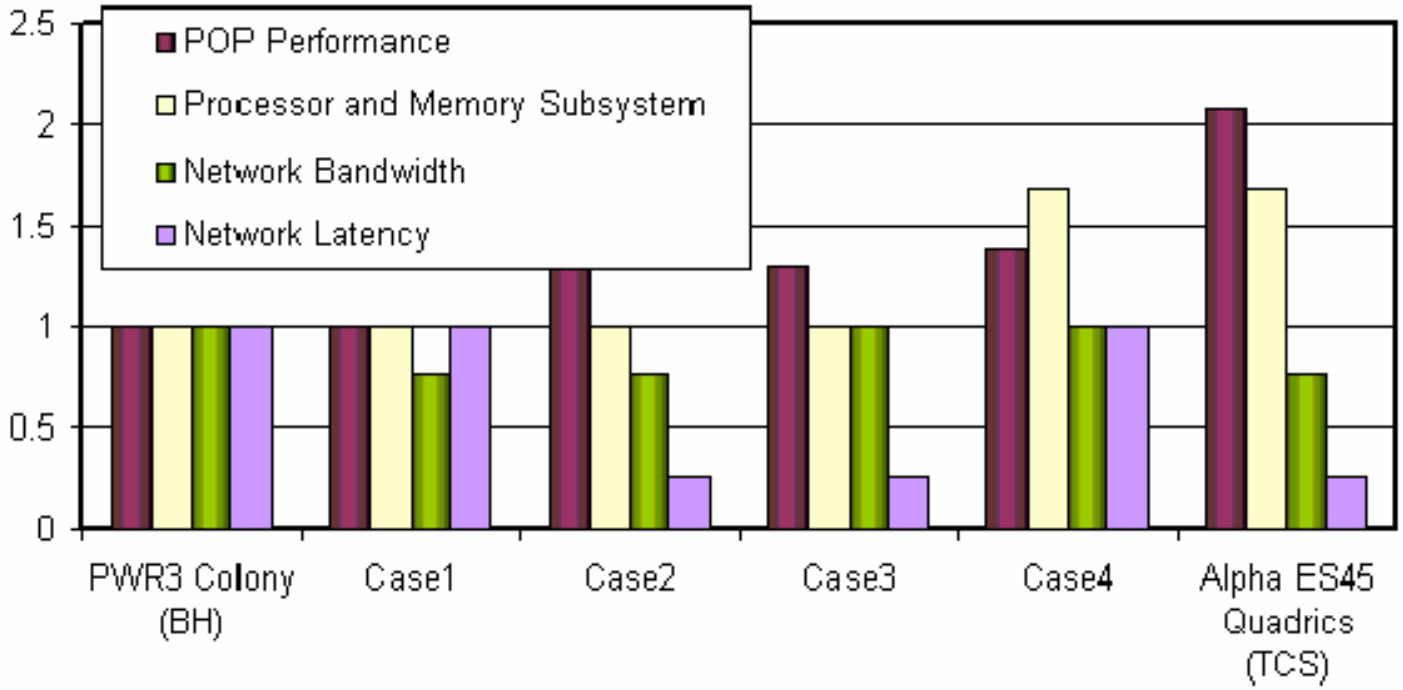
Uses:

- Help scientists understand performance behavior.
- Help computing centers acquire the best system for their workloads, and project future requirements.
- Help system architects design improved future systems.
- Perform “what-if” analyses, varying various parameters.

Performance Modeling: Prediction Versus Measurement



“What-If” Analysis for IBM Power 3 System with Colony Switch



Case 1: Colony switch latency, but single-rail Quadrics switch bandwidth.

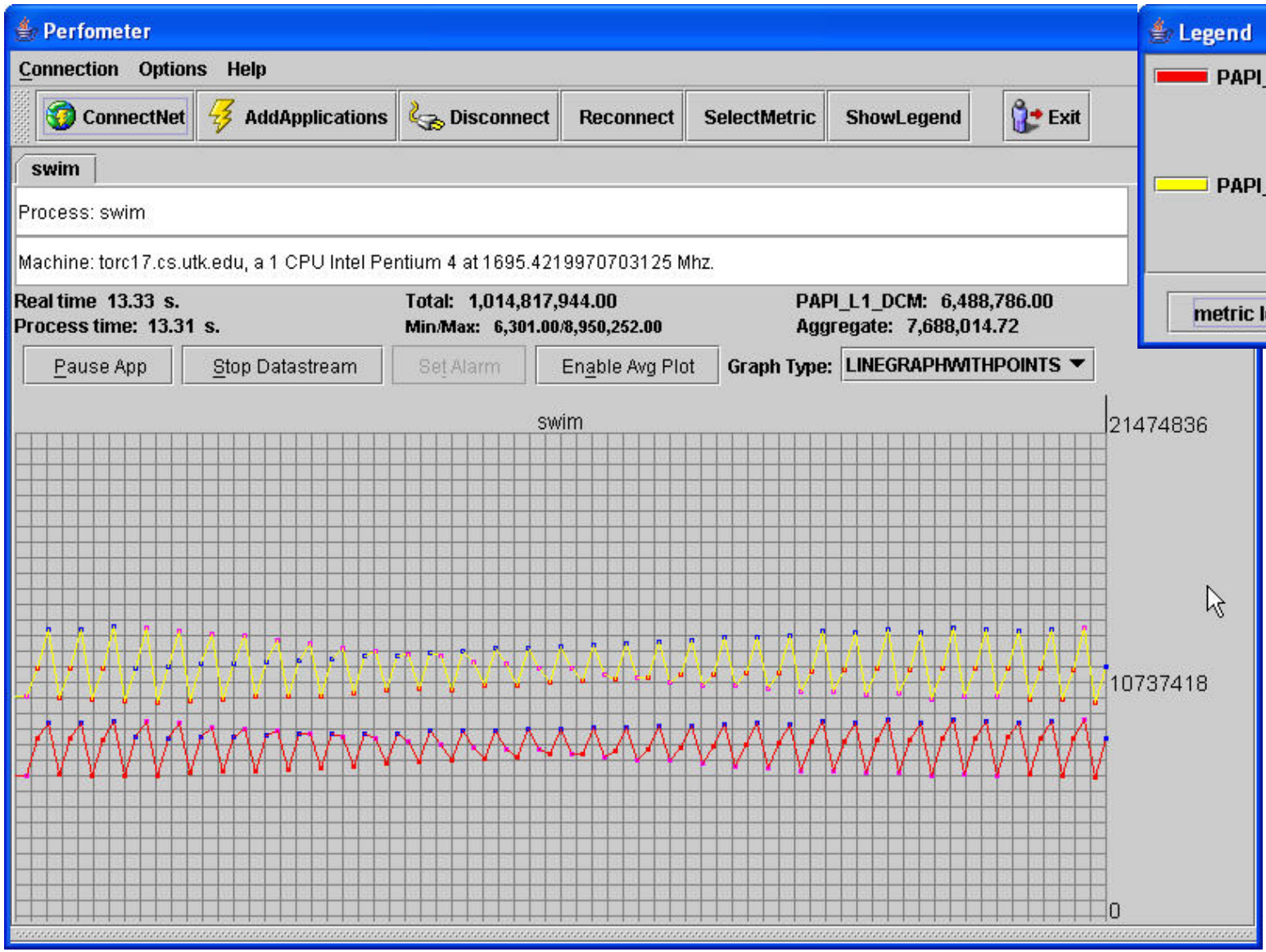
Case 2: Quadrics switch latency and bandwidth.

Case 3: Quadrics latency but Colony bandwidth.

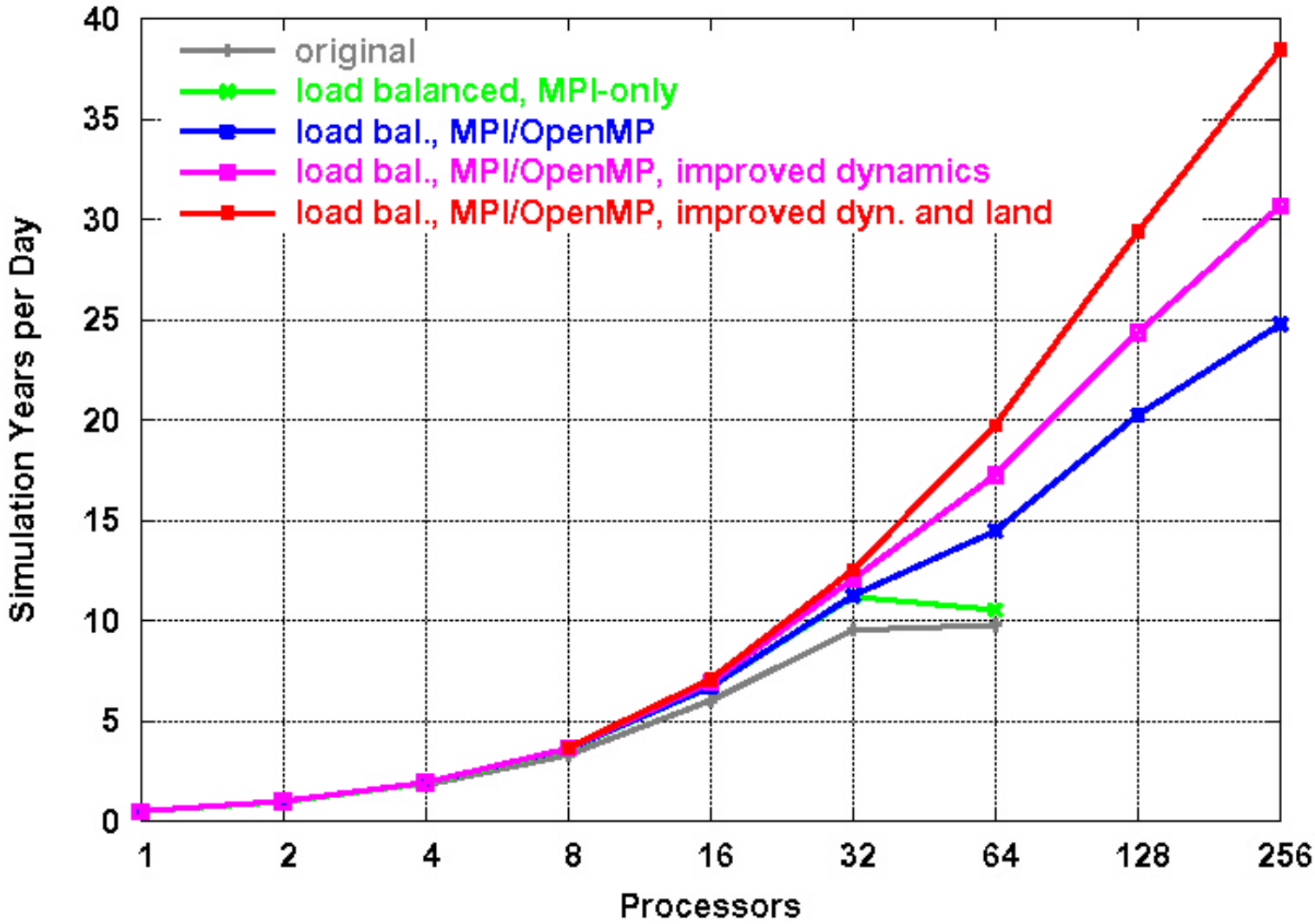
Case 4: Colony bandwidth and latency, but Alpha ES640 processors and memory.

Final Column: Alpha ES640 with Quadrics switch

Performance Tools: The PAPI Performeter



Improvement to Climate Modeling Code Using Performance Tools



Generic Code Optimization

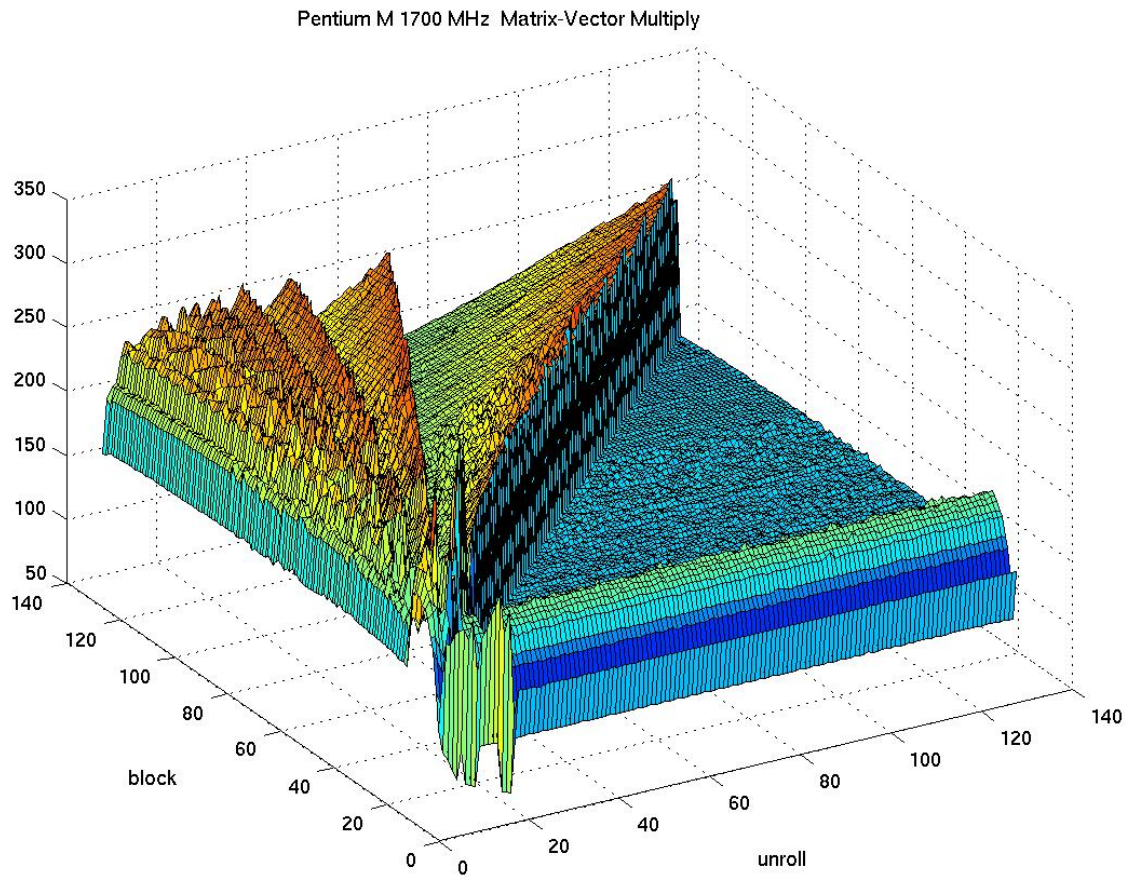
Experimentation with DGEMV (matrix-vector multiply):

128x128=16,384 cases.

Experiment took 30+ hours to run.

Best performance = 338 Mflop/s with
blocking=11
unrolling=11

Original performance = 232 Mflop/s



Conclusions

- Computers with 1 Pflop/s (10^{15} flop/sec) peak performance will be here in 4-5 years.
- These systems will employ hundreds of thousands of CPUs, joined with massive networks.
- Moore's Law has at least 10 more years to go.
- Exotic new technologies, such as nanotubes, molecular self-assembly and quantum dots, will likely be ready when conventional photolithography reaches its limit.
- There is no shortage of useful scientific applications for such systems.

Challenges:

- Will we be able to efficiently use systems with 10^5 to 10^6 CPUs?
- Will new programming models or system software be needed?
- How well will scientific applications scale in performance?