

Directions in High-End Computing

Rusty Lusk

Acting Director, Mathematics and Computer Science Division

Argonne National Laboratory



Outline

- The NE code base
- Progress in algorithms as well as computer speed
- Parallelism now “standard”
 - A standard programming model
 - Commodity parallel machines
- Supercomputing not as hard as it used to be
- The very high end
 - Office of Science computing centers
 - Availability
- Some examples from other fields
- Conclusion – imagine more!

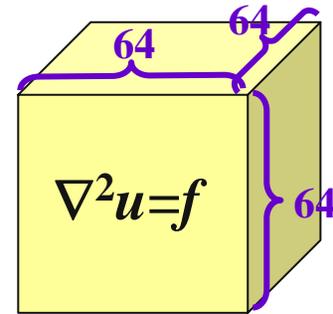
NE Codes

- Old (some ~20 years)
 - Reliable, well understood
 - Certified
- In use, and delivering results, but
 - Not taking advantage of enormous recent progress in
 - *Computer power*
 - *Algorithms for the mathematics behind the physics*
 - *Enabling software technology*

The power of optimal algorithms

- Advances in algorithmic efficiency can rival advances in hardware architecture
- Consider Poisson's equation on a cube of size $N=n^3$

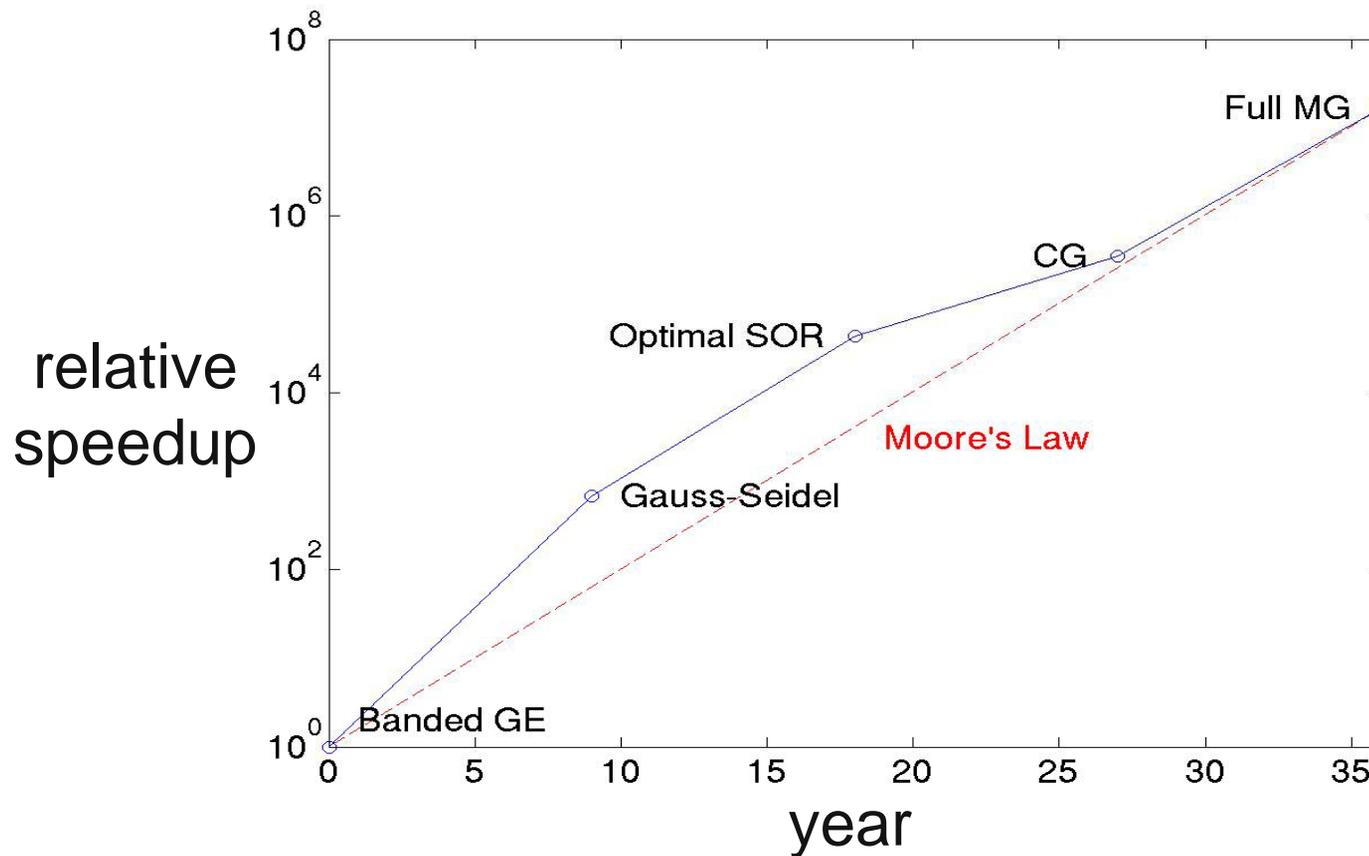
Year	Method	Reference	Storage	Flops
1947	GE (banded)	Von Neumann & Goldstine	n^5	n^7
1950	Optimal SOR	Young	n^3	$n^4 \log n$
1971	CG	Reid	n^3	$n^{3.5} \log n$
1984	Full MG	Brandt	n^3	n^3



- If $n=64$, this implies an overall reduction in flops of ~16 million (6 months reduced to one second)

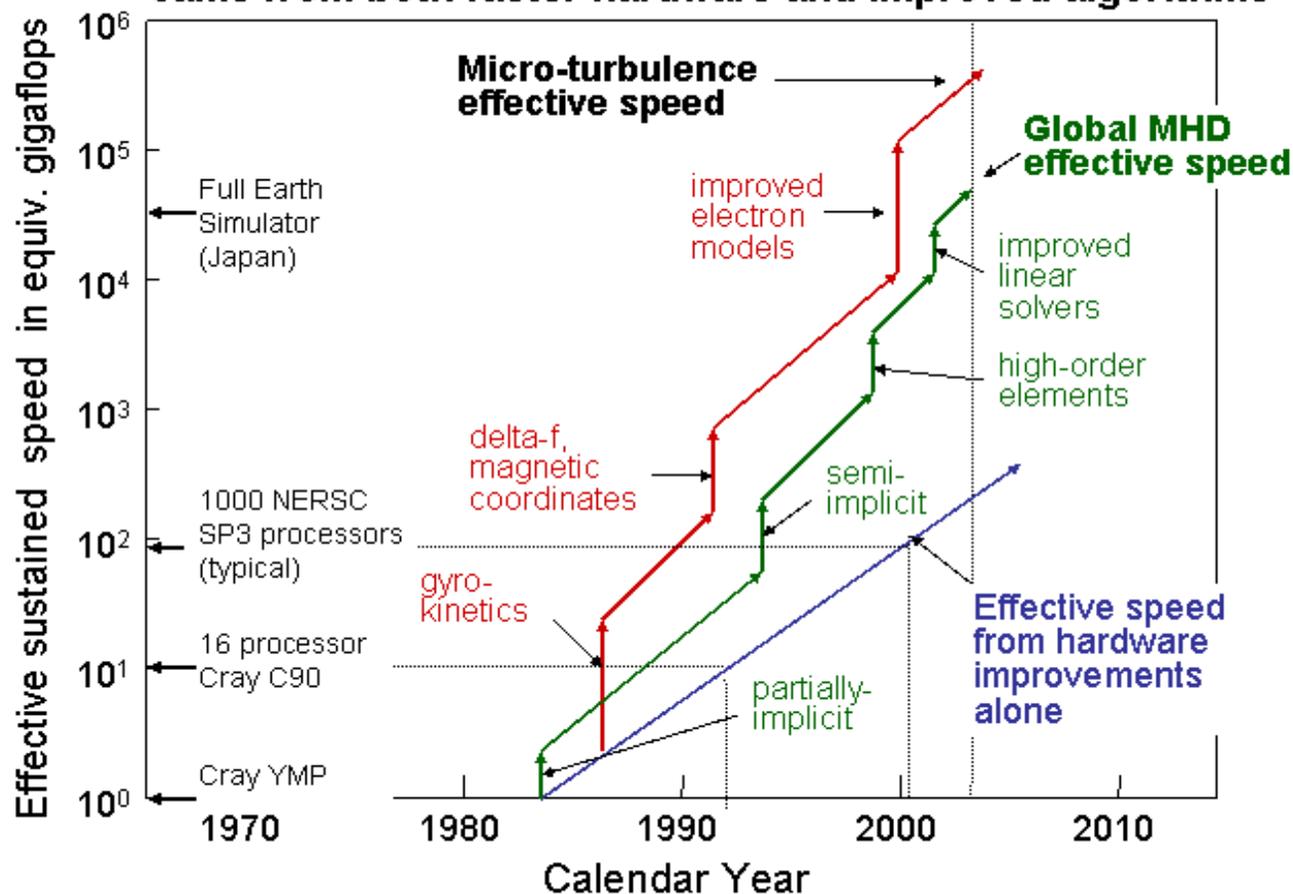
Algorithms and Moore's Law

- This advance took place over a span of about 36 years, or 24 doubling times for Moore's Law
- $2^{24} \approx 16$ million \Rightarrow the same as the factor from algorithms alone!



“Moore’s Law” for MHD simulations

Magnetic Fusion Energy: “Effective speed” increases came from both faster hardware and improved algorithms



“Semi-implicit”:
All waves treated implicitly, but still stability-limited by transport

“Partially implicit”:
Fastest waves filtered, but still stability-limited by slower waves

Figure from SCaLeS report, Volume 2

“Moore’s Law” for combustion simulations

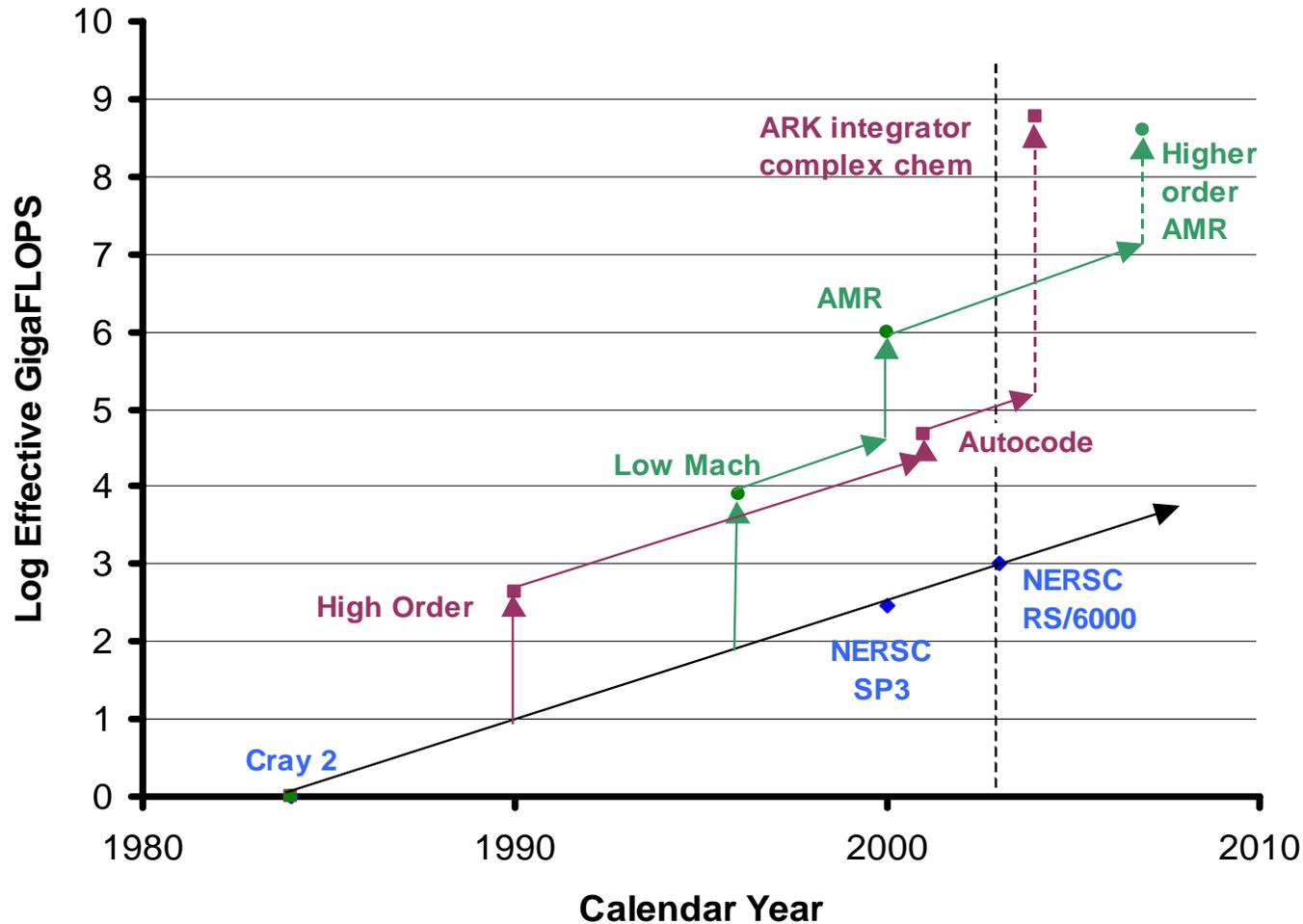
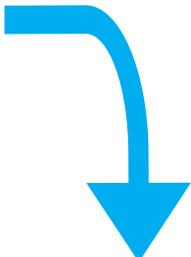


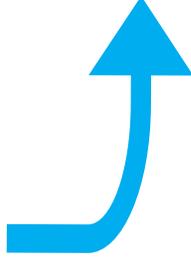
Figure from SCaLeS report, Volume 2

Gordon Bell Prize “price performance”

<i>Year</i>	<i>Application</i>	<i>System</i>	<i>\$ per Mflops</i>
1989	Reservoir modeling	CM-2	2,500
1990	Electronic structure	IPSC	1,250
1992	Polymer dynamics	cluster	1,000
1993	Image analysis	custom	154
1994	Quant molecular dyn	cluster	333
1995	Comp fluid dynamics	cluster	278
1996	Electronic structure	SGI	159
1997	Gravitation	cluster	56
1998	Quant chromodyn	custom	12.5
1999	Gravitation	custom	6.9
2000	Comp fluid dynamics	cluster	1.9
2001	Structural analysis	cluster	0.24



Four orders
of magnitude
in 12 years



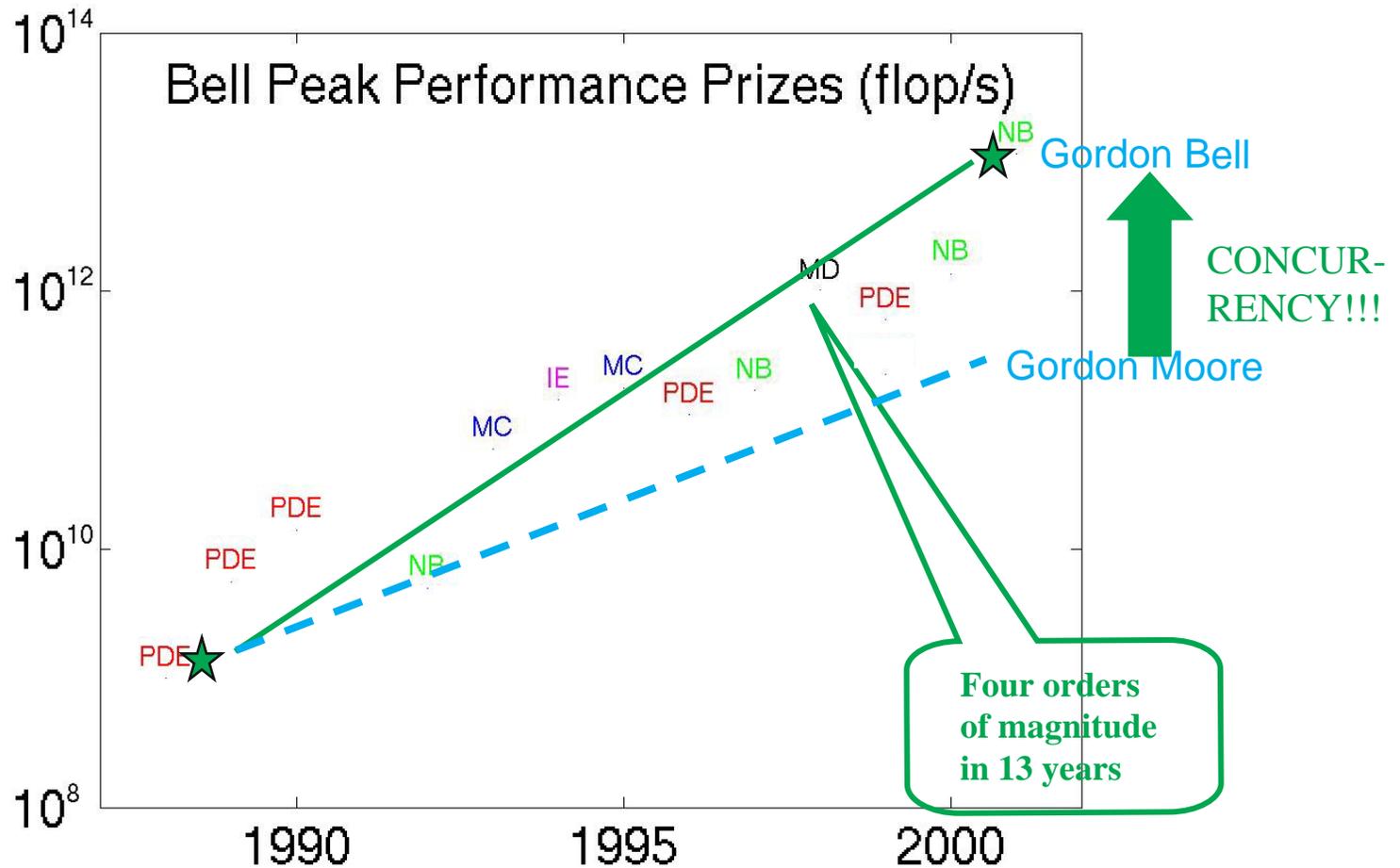
Price/performance has stagnated and no new such prize has been given since 2001.

Gordon Bell Prize “peak performance”

<i>Year</i>	<i>Type</i>	<i>Application</i>	<i>No. Procs</i>	<i>System</i>	<i>Gflop/s</i>
1988	PDE	Structures	8	Cray Y-MP	1.0
1989	PDE	Seismic	2,048	CM-2	5.6
1990	PDE	Seismic	2,048	CM-2	14
1992	NB	Gravitation	512	Delta	5.4
1993	MC	Boltzmann	1,024	CM-5	60
1994	IE	Structures	1,904	Paragon	143
1995	MC	QCD	128	NWT	179
1996	PDE	CFD	160	NWT	111
1997	NB	Gravitation	4,096	ASCI Red	170
1998	MD	Magnetism	1,536	T3E-1200	1,020
1999	PDE	CFD	5,832	ASCI BluePac	627
2000	NB	Gravitation	96	GRAPE-6	1,349
2001	NB	Gravitation	1,024	GRAPE-6	11,550
2002	PDE	Climate	5,120	Earth Sim	26,500
2003	PDE	Seismic	1,944	Earth Sim	5,000
2004	PDE	CFD	4,096	Earth Sim	15,200
2005	MD	Solidification	131,072	BGL	101,700

Four orders
of magnitude
in 13 years

Gordon Bell Prize outpaces Moore's Law



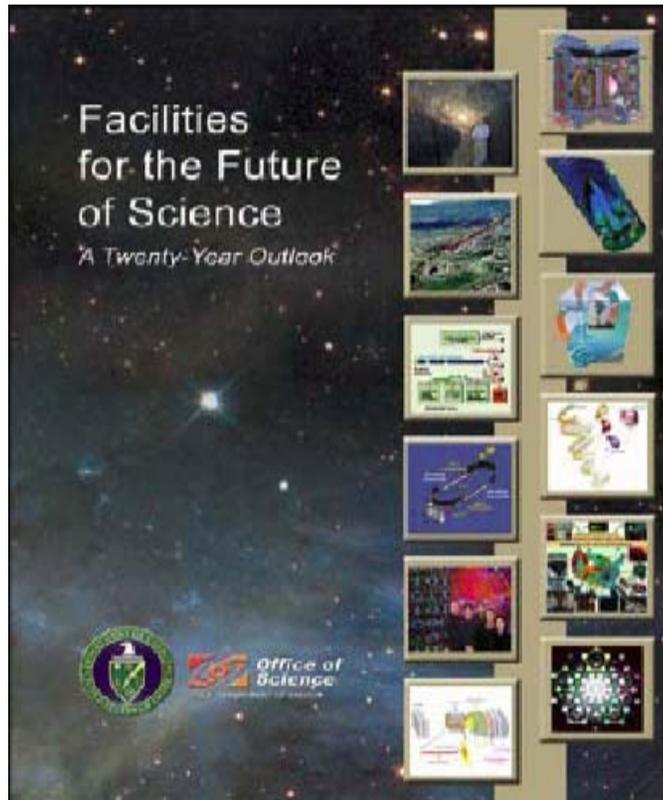
Parallelism Now Ubiquitous

- A standard parallel programming model
 - MPI library available on all parallel machines
 - Expressive, efficient, widely taught
 - Has spurred the growth of portable parallel libraries
- Commodity parallel computers
 - Clusters leverage the mass market
 - Linux clusters
 - Even Windows clusters
- Capacity vs. capability
 - Small and medium-sized clusters provide capacity for modest levels of parallelism, large numbers of serial and parallel jobs.
 - Capability machines for computations that cannot be done on “everyday parallel computers

Supercomputing is Easier Than It Used To Be

- Standards
 - A code written in MPI + Fortran(-90 or -77) will run on everything from your laptop to BlueGene
- Libraries
 - Standards have encouraged the development of libraries so much code need not be written at all
- I/O
 - Parallel I/O standards (MPI) have encouraged high-level parallel I/O libraries (HDF5, PnetCDF) so applications need not manage thousands of separate files in parallel applications.
 - Parallel file systems can deliver high bandwidth to disk.
- Graphics
 - Many choices in graphics libraries (e.g. VisIT)
- Frameworks
 - Can help combine, leverage existing software.

Leadership Computing Facilities (LCF)



- **November, 2003:** Then Secretary Abraham announces 20 Year Science Facility plan with #2 Near Term Priority – UltraScale Scientific Computing Capability.
- **February, 2004:** DOE Office of Science issued a call for proposals to SC laboratories to provide Leadership Class Computing Capability for Science.
- **May 12, 2004:** Following a peer review, then Secretary Abraham announces award to the partnership of Oak Ridge, Argonne and Pacific Northwest National Laboratories.

Modes of Impact for Leadership Computing

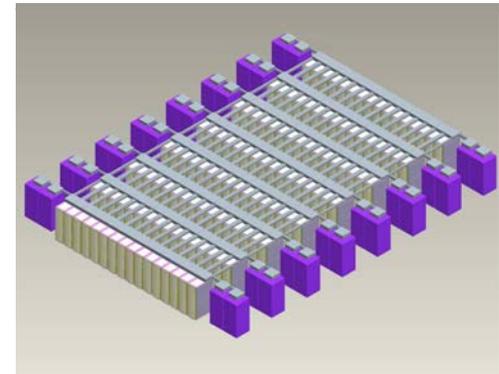
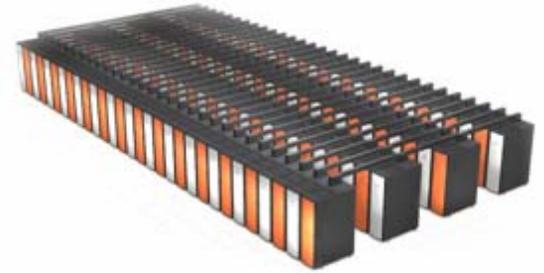
1. Generation of significant datasets via simulation to be used by a large and important scientific community
 - Example: Providing a high-resolution first principles turbulence simulation dataset to the CFD and computational physics community
2. Demonstration of new methods or capabilities that establish feasibility of new computational approaches that are likely to have significant impact on the field
 - Example: Demonstration of the design and optimization of a new catalyst using first principles molecular dynamics and electronic structure codes
3. Analysis of large-scale datasets not possible using other methods
 - Example: Computationally screen all known microbial drug targets against the known chemical compound libraries
4. Solving a science or engineering problem at the heart of a critical DOE mission or facilities design or construction project
 - Example: Designing a passively safe reactor core for the Advanced Burner Reactor Test Facility

DOE Leadership Computing Facility Strategy

- DOE selected the ORNL, ANL and PNNL team (May 12, 2004) based on a competitive peer review of four proposals to develop the DOE SC Leadership Computing Facilities
 - ORNL will deploy a series of systems based on Cray's X1 and XT architectures
 - ANL will deploy a series of systems based on IBM's BlueGene architecture
 - PNNL will contribute software technology for programming models (Global Arrays) and parallel file systems
- DOE SC will make these systems available as capability platforms to the broad national community via competitive awards (INCITE)
 - Each facility will target ~20 large-scale production applications teams
 - Each facility will also support order 100 development users
- DOE's LCC facilities will complement the resources at NERSC
 - Large number of projects (200 – 300)
 - Medium- to very-large-scale projects that occasionally need a very high capability

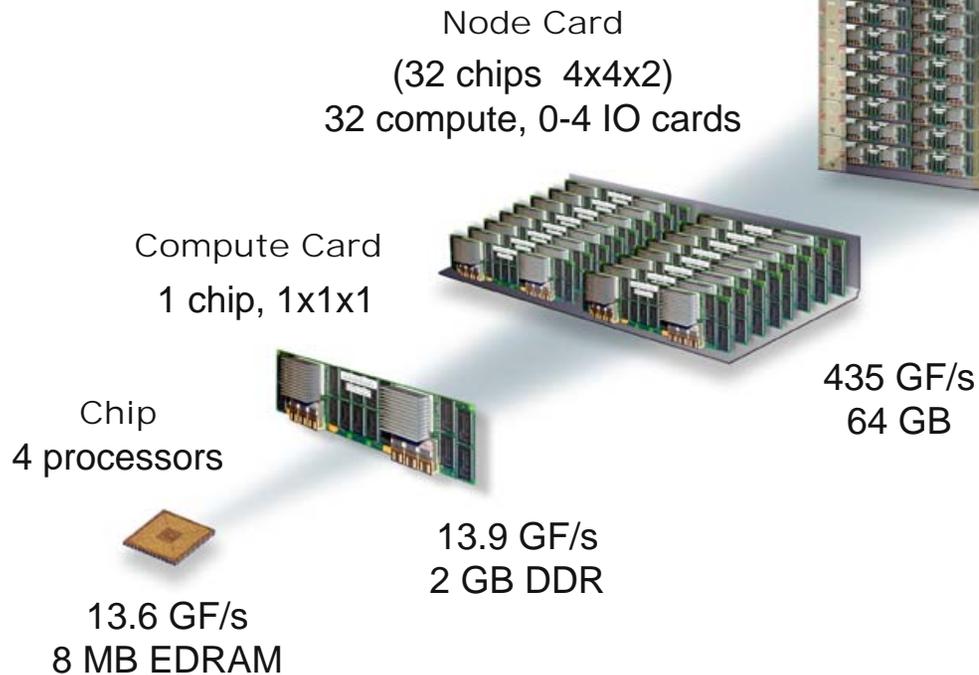
Office of Science Facilities Plan

- NERSC – delivery of NERSC-5 in FY 2007
- LCF at Oak Ridge
 - Cray XT3 upgrade
 - *Now: 25 teraflop/s to 50 teraflop/s*
 - *By end of 2007: 250 teraflop/s*
 - Cray Baker acquisition by end of 2008: 1 PF
- LCF at Argonne
 - IBM BluGene/P in FY 2007: 100 TF
 - Upgrade to 250-500 TF in 2008
 - Planning Petaflop BG/Q by end of decade



BlueGene/P

- Processors + memory + network interfaces are all on the same chip.
- Faster multi-core processors with larger memory
- 5 flavors of network, with faster signaling, lower latency



System
72 Racks



1 PF/s
144 TB

- High packaging density
- High reliability
- Low system power requirements
- XL compilers, ESSL, GPFS, LoadLeveler, HPC Toolkit
- MPI, MPI2, OpenMP, Global Arrays

BlueGene community knowledge base is preserved

Example Leadership Science Applications

- **Qbox** — FPMD solving Kohn-Sham equations, strong scaling on problem of 1000 molybdenum atoms with 12,000 electrons (86% parallel efficiency on 32K cpus @ SC05), achieved 207 TFs recently on BG/L
- **ddcMD** — many-body quantum interaction potentials (MGPT), 1/2 billion atom simulation, 128K cpus, achieved > 107 TFs on BG/L via fused dgemm and ddot
- **BlueMatter** — scalable biomolecular MD with Lennard-Jones 12-6, P3ME and Ewald, replica-exchange 256 replicas on 8K cpus, strong scaling to 8 atoms/node
- **GAMESS** — *ab initio* electronic structure code, wide range of methods, used for energetics, spectra, reaction paths and some dynamics, scales $O(N^5-N^7)$ in number of electrons, uses DDI for communication and pseudo-shared memory, runs to 32,000 cpus
- **FLASH3** — produced largest weakly- compressible, homogeneous isotropic turbulence simulation to date on BG/L, excellent weak scaling, 72 million files 156 TB of data

INCITE – Innovative and Novel computational Impact on Theory and Experiment

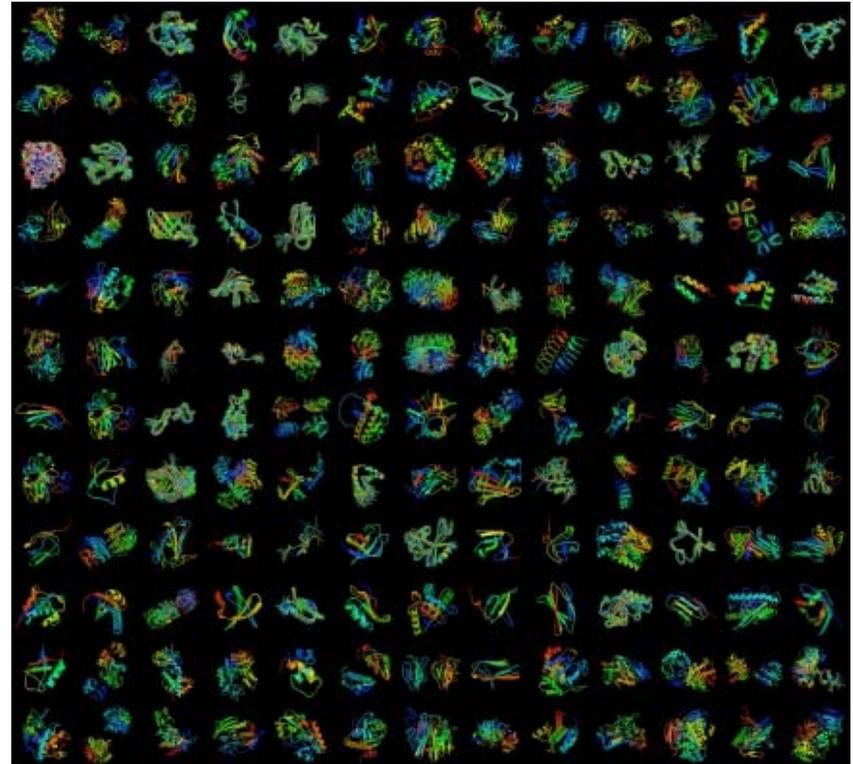
- Initiated in 2004
- Provides Office of Science computing resources to a small number of computationally intensive research projects of large scale that can make high-impact scientific advances through the use of a large allocation of computer time and data storage
- Open to national international researchers, including industry
- No requirement of DOE Office of Science funding
- Peer-reviewed
- 2004 awards: 4.9 million processor hours at NERSC awarded to three projects
- 2005 awards: 6. million processor hours at NERSC awarded to three projects

INCITE (continued)

- 2006 – expanded to include SC high end computing resources at PNNL, ORNL, and ANL as well as LBNL
 - Multiple-year requests
 - 15 awards for 18.2 million processor hours
- 2007 – expanded to include 80% of leadership class facilities at ORNL and ANL plus 10% of NERSC and 5% of PNNL
 - See <http://hpc.science.doe.gov>

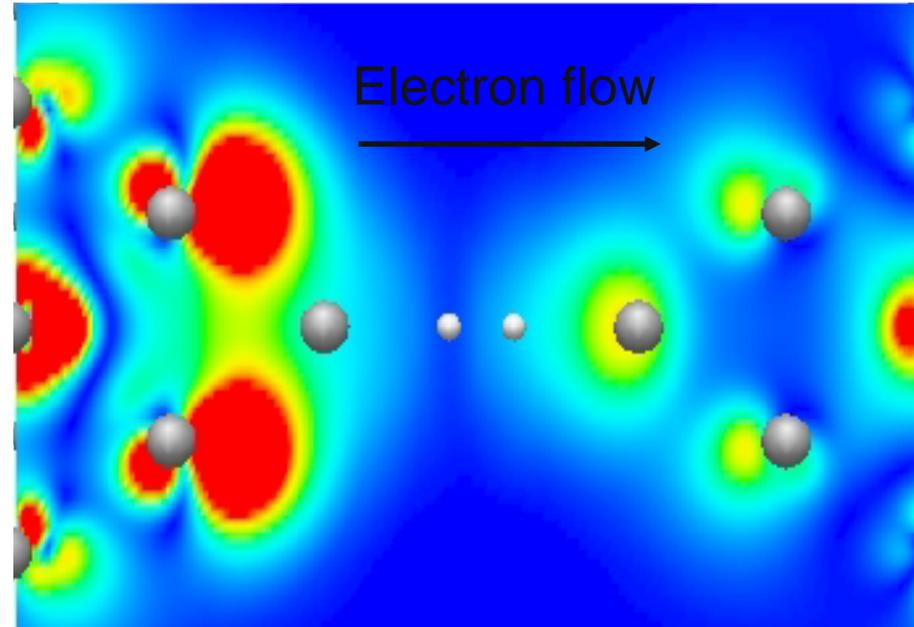
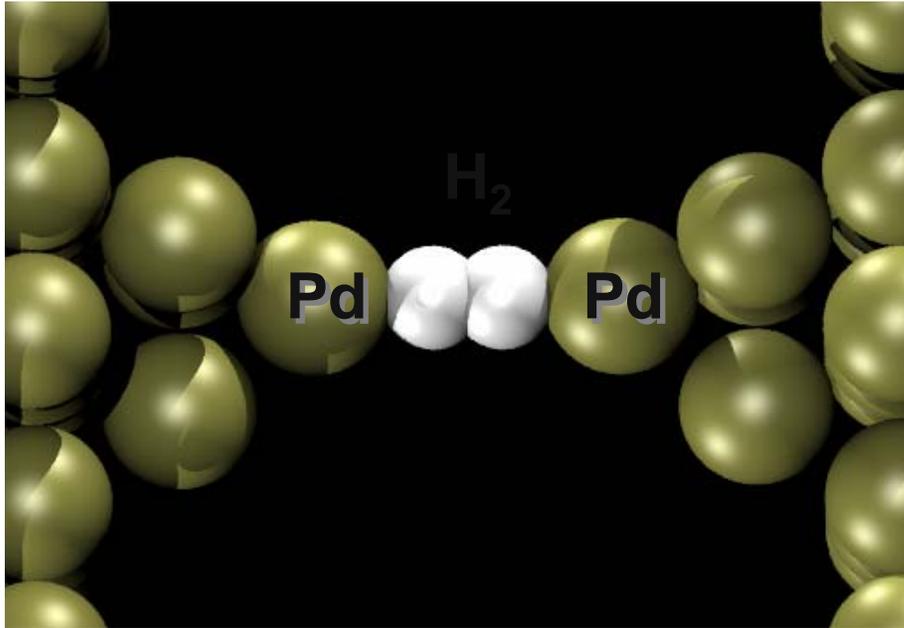
INCITE Project: Molecular Dynamics

- V. Daggett, U Washington, 2M NERSC proc. hours
- Understand protein folding pathways by 'unfolding' proteins at high temp.
- Computed unfolding of 151 most common fold structures at different temperatures
- Multiple runs of MD calculation for each fold/temp. pair



The first 156 protein targets

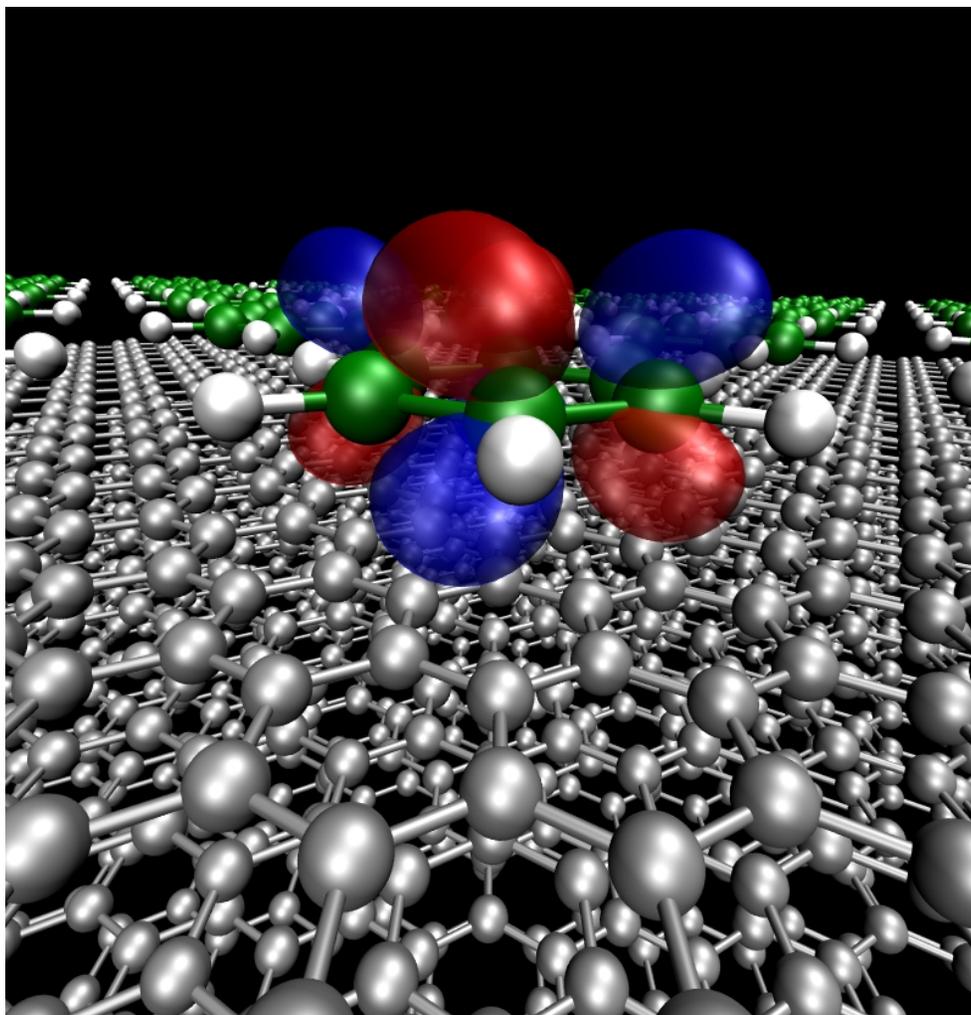
Exploring the Limits of Nanoelectronics with Theory: Single Molecule Electrical Junctions



Shown at left is a single hydrogen molecule (in white) bridging palladium point contacts. At right, a density plot of the dominant transmitting electronic state reveals a significant reflection of charge at the left Pd contact, leading to a high resistance, consistent with recent experiments. (Red is high electronic density in the plot, blue is low.)

Steven Louie,
Marvin Cohen,
UC Berkeley
Jeff Neaton,
Molecular Foundry

Excited electronic states at metal-organic interfaces

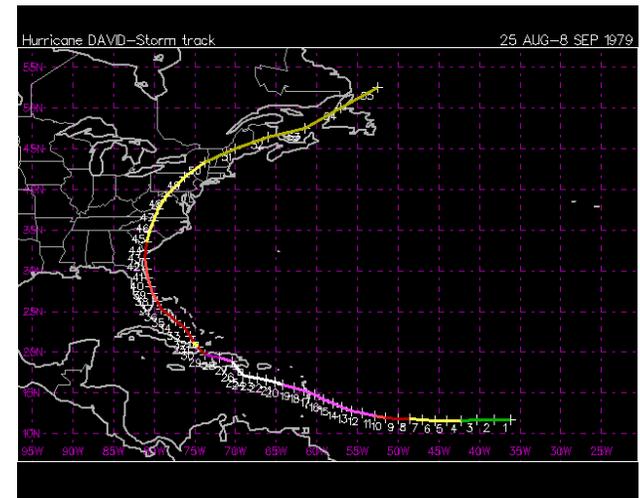
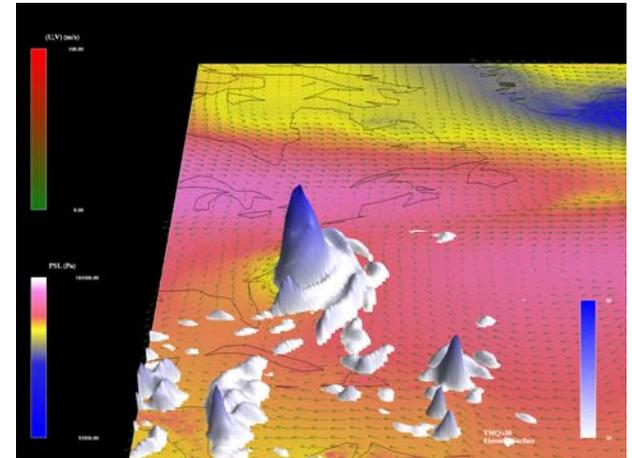


**Mark Hybertsen &
George Flynn
Columbia University
Jeff Neaton
Molecular Foundry**

Lowest unoccupied molecular orbital of a benzene molecule physisorbed on a graphite surface. Our calculations predict that, relative to the gas-phase, orbital energies are strongly modified by the surface.

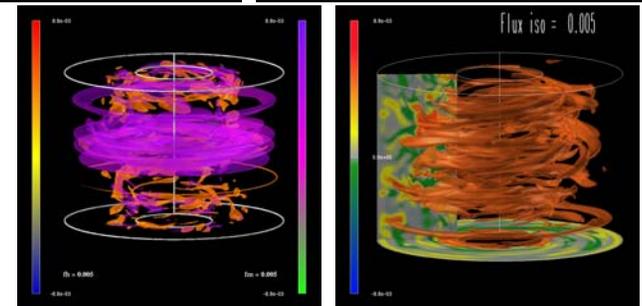
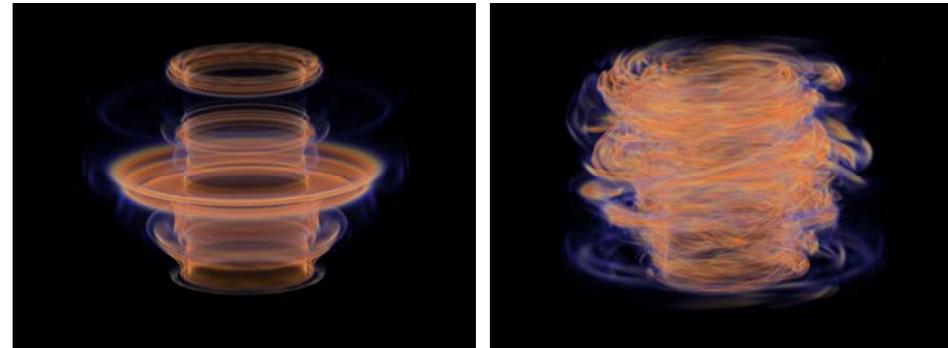
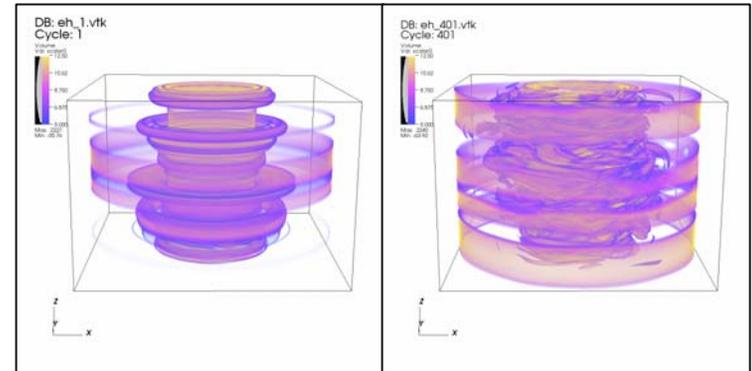
Science Driven Analytics: Comparing Real and Simulated Storm Data

- NERSC designed a prototype workflow enabling fast qualitative comparisons between simulated storm data and real observations
- By using the NERSC Global Filesystem (NGF) the most appropriate resource can be used at each stage
 - IBM P5 (Bassi) for large-scale parallel computing
 - Linux cluster (Jacquard) for data reduction
 - Visualization server



INCITE 4 – Magneto-Rotational Instability and Turbulent Angular Momentum Transport

- Visual Analytics support for collaboration with F. Catteano, Univ. of Chicago
- iterative, investigatory approach to explore alternative methods in order to determine which one provides the best visual and scientific results.
- Top row: hydro enstrophy from two different timesteps.
- Middle row: magnetic enstrophy from two different timesteps
- Bottom row: hydro and magnetic flux (left), total advective radial flux of axial angular momentum (right).
- [Movie](#) of time-evolving magnetic enstrophy.



Conclusion

- Introduction of GNEP coincides with availability of
 - New levels of computing power
 - New algorithms, libraries, and technologies to take advantage of it
- Computing advances have enabled significant progress in other fields
- The time is right for mathematics and computer science to play a significant role in the next generation of nuclear physics and engineering codes.