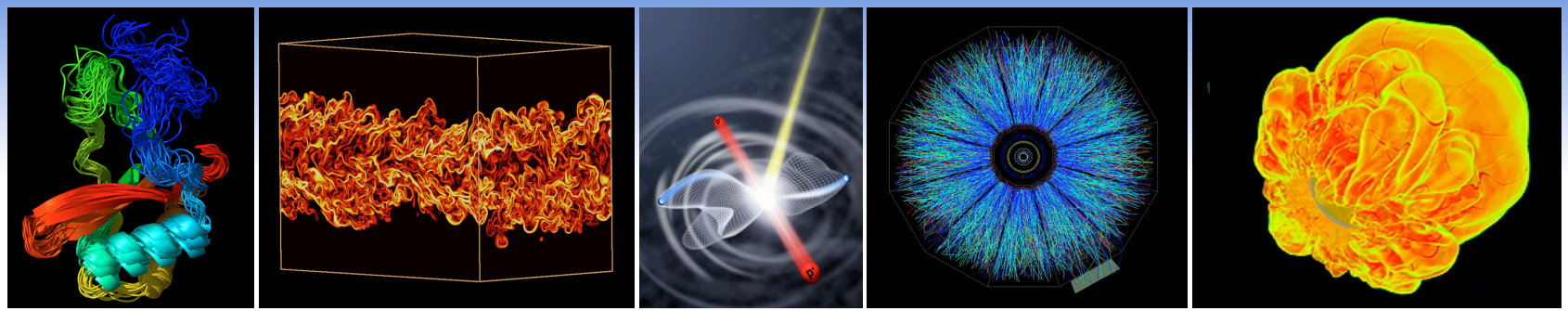


NERSC Workload Analysis

Harvey Wasserman

NERSC Science Driven System Architecture Group

http://www.nersc.gov/projects/SDSA/

NATIONAL ENERGY RESEARCH SCIENTIFIC COMPUTING CENTER

Science Driven Evaluation

- **Translate scientific requirements into computational needs**
  - **and then to hardware and software attributes required to support them.**
- **How do we represent these needs so that we can communicate them to others?**
  - **Answer: a set of carefully chosen benchmark programs.**

Office of Science
U.S. DEPARTMENT OF ENERGY

BERKELEY LAB

NERSC Benchmarks Serve 3 Critical Roles

- Carefully chosen to represent characteristics of the expected NERSC-6 workload.
- Give vendors opportunity to provide NERSC with concrete performance and scalability data;
  - Measured or projected.
- Part of the acceptance test and a measure of performance throughout the operational lifetime of NERSC-6.

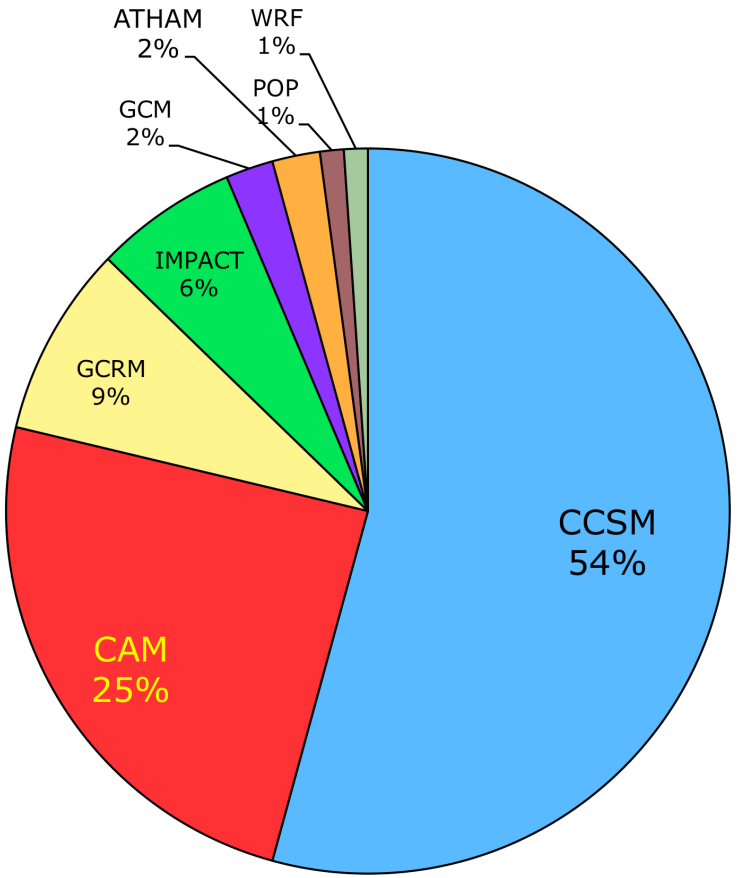
Source Information

- Documents
  - 2005 DOE Greenbook
  - 2006-2010 NERSC Plan
  - LCF Studies and Reports
  - Workshop Reports
  - 2008 NERSC assessment
- ERCAP analysis
- User discussion

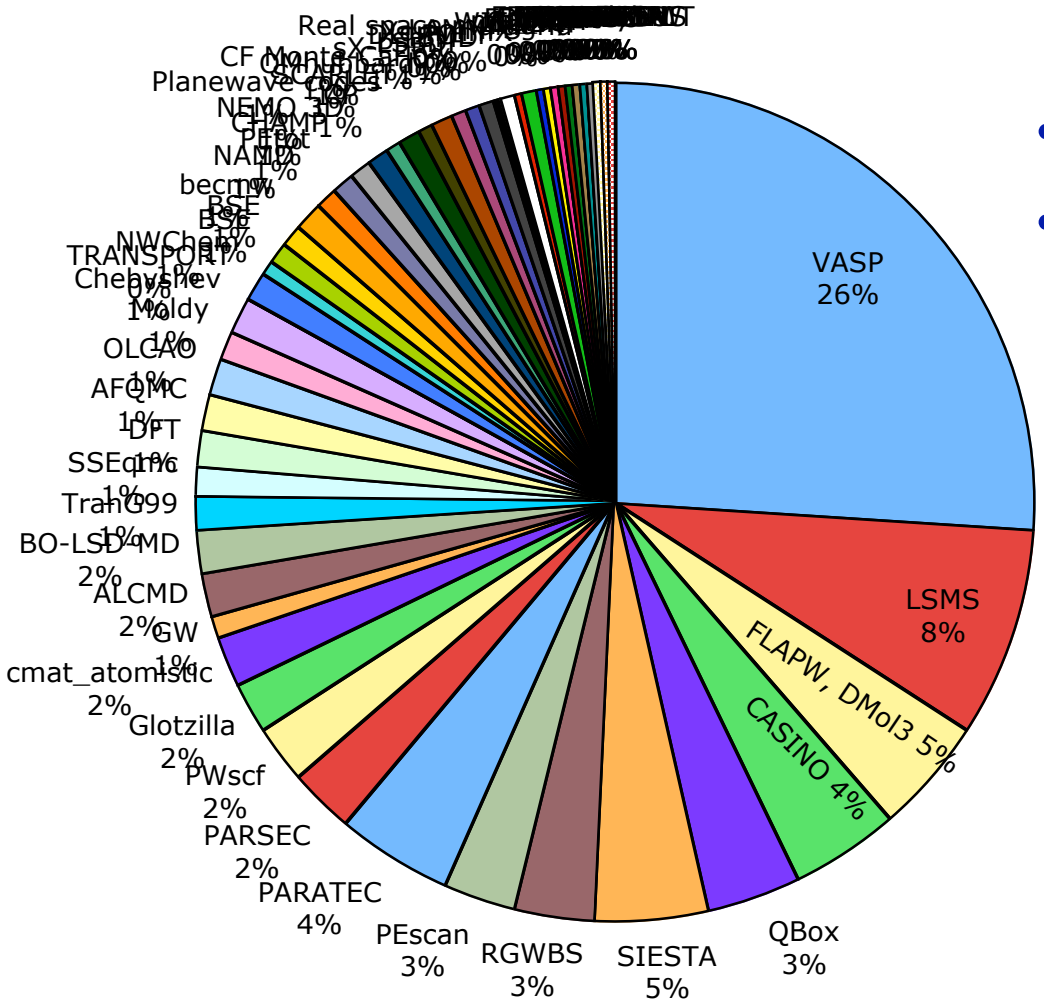
Example: Climate Modeling

- CAM dominates CCSM computational requirements.
- FV-CAM increasingly replacing Spectral-CAM in future CCSM calculations.
- Drivers:
  - Critical support of U.S. submission to the Intergovernmental Panel on Climate Change (IPCC).
    - Schedule coincident with arrival of NERSC-6 system.
  - V & V for CCSM-4
- Focus on ensemble runs - 10 simulations per ensemble, 5-25 ensembles per scenario, relatively small concurrencies.

Climate without INCITE



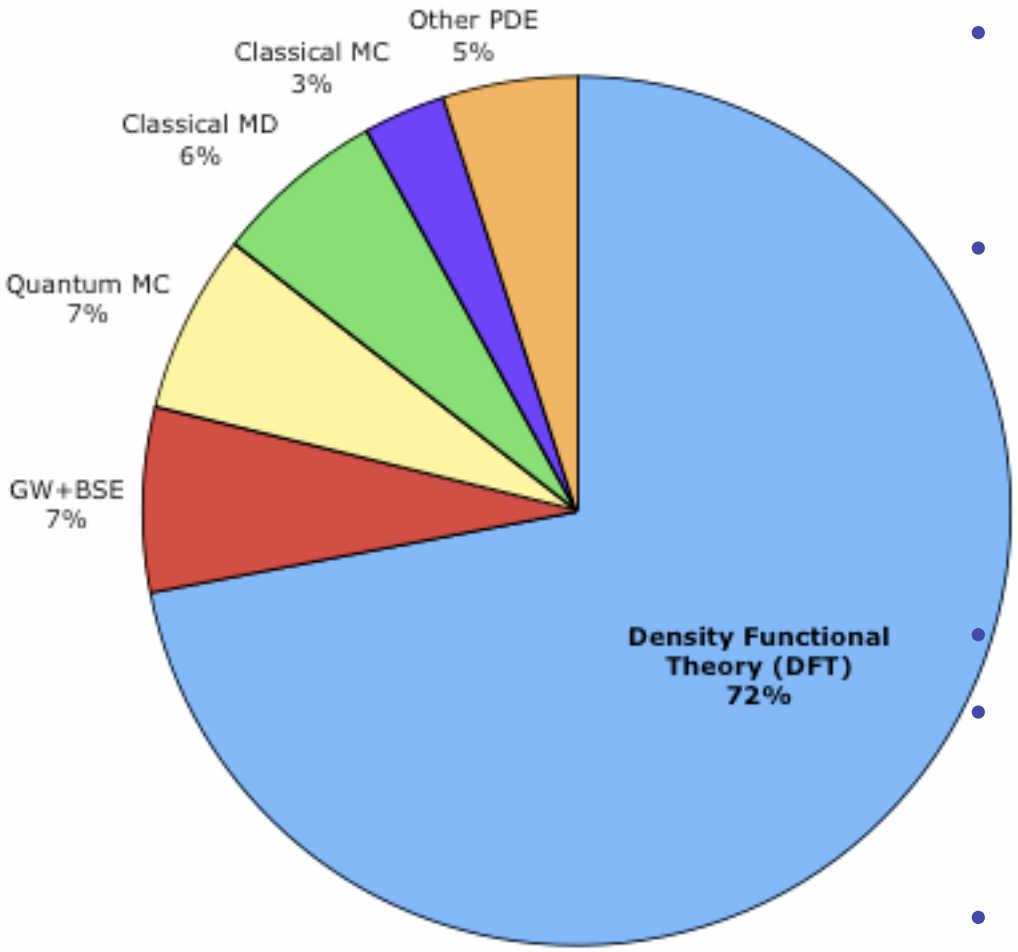
Material Science by Code



- **7,385,000 MPP hours awarded**
- **62 codes, 65 users**
- **Typical code used in 2.15 allocation requests**

	Code	MPP Hours	Percent	Cumulative%
1	VASP	1,992,110	26%	26%
2	LSMS	600,000	8%	34%
3	FLAPW, DMol3	350,000	5%	39%
4	CASINO	312,500	4%	43%
5	QBox	262,500	3%	46%
6	SIESTA	346,500	5%	51%
7	RGWBS	232,500	3%	54%
8	PEscan	220,000	3%	57%
9	PARATEC	337,500	4%	61%
10	PARSEC	182,500	2%	64%
	Other	167,300	34%	66%

Materials Science by Algorithm



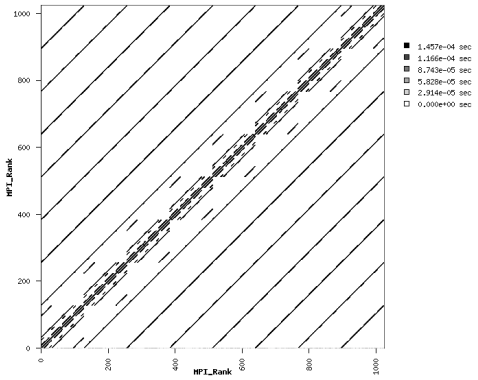
- Density Functional Theory codes
  - >70% of the MatSci. workload!
  - Majority are planewave DFT.
- Common requirements for DFT:
  - 3D global FFT
  - Dense Linear Algebra for orthogonalization of wave basis functions and calculating pseudopotential
- Dominant Code: VASP
- Science driver: nanoscience, ceramic crystals, novel materials, quantum dots, …
- Similar Codes (planewave DFT)
  - Qbox, PARATEC
  - PETOT/PESCAN

Analysis by Lin-Wang Wang

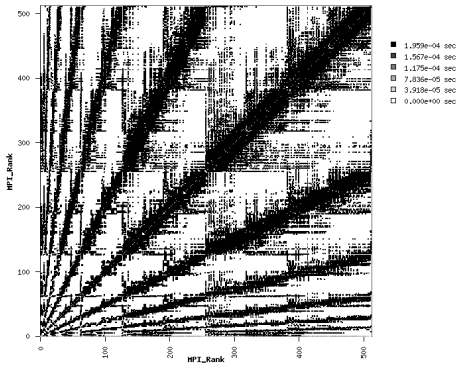
NERSC-6 Application Benchmarks

Benchmark	*Science Area*	*Algorithm Space*	*Base Case Concurrency*	*Problem Description*	*Lang*	*Libraries*
CAM	Climate (BER)	Navier Stokes CFD	56, 240 Strong scaling	D Grid, (~.5° resolution); 240 timesteps	F90	netCDF
GAMESS	Quantum Chem (BES)	Dense linear algebra	384, 1024 (Same as Ti-09)	DFT gradient, MP2 gradient	F77	DDI, BLAS
GTC	Fusion (FES)	PIC, finite difference	512, 2048 Weak scaling	100 particles per cell	F90	
IMPACT-T	Accelerator Physics (HEP)	PIC, FFT	256,1024 Strong scaling	50 particles per cell	F90	
MAESTRO	Astrophysics (HEP)	Low Mach Hydro; block structured -grid multiphysics	512, 2048 Weak scaling	16 32^3 boxes per proc; 10 timesteps	F90	Boxlib
MILC	Lattice Gauge Physics (NP)	Conjugate gradient, sparse matrix; FFT	256, 1024, 8192 Weak scaling	8x8x8x9 Local Grid, ~70,000 iters	C, assemb.	
PARATEC	Material Science (BES)	DFT; FFT, BLAS3	256, 1024 Strong scaling	686 Atoms, 1372 bands, 20 iters	F90	Scalapack, FFTW

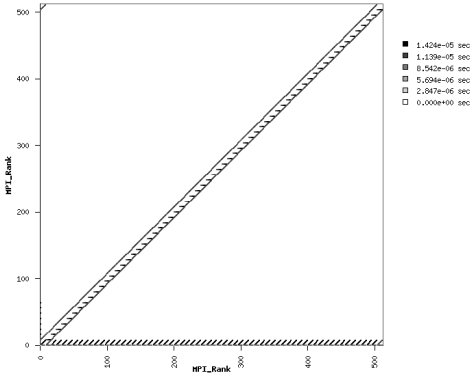
Communication Topology



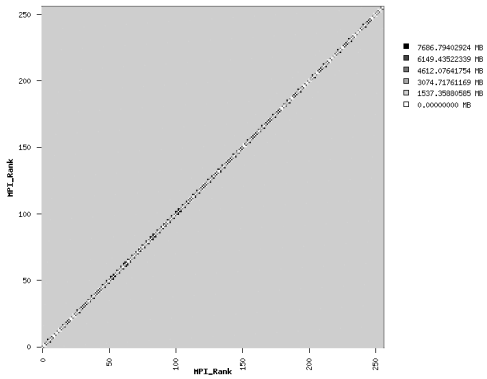
MILC



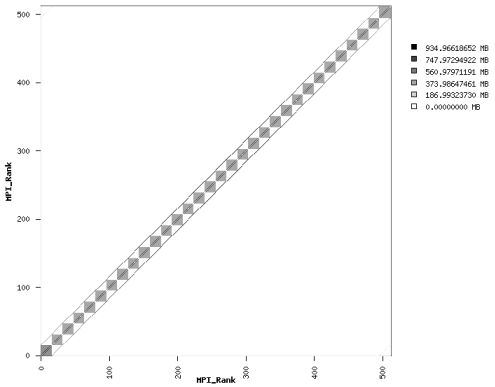
MAESTRO



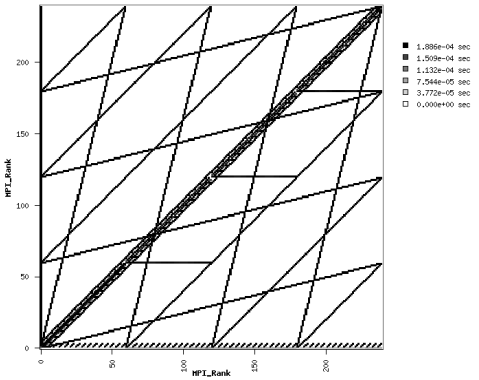
GTC



PARATEC



IMPACT-T



CAM

Summary: CI & %MPI

	CAM	GAMESS	GTC	IMPACT-T	MAESTRO	MILC	PARATEC
CI*	0.67	0.61	1.15	0.77	0.24	1.39	1.50
Cray XT4 %Peak per Core (largest case)	13%	12%	24%	14%	5%	14%	44%
Cray XT4 %MPI Medium	29%		4%	9%	20%	20%	27%
Cray XT4 %MPI Large	35%		6%	40%	20%	23%	64%
Cray XT4 %MPI ExtraL	n/a	n/a	n/a	n/a	n/a	30%	n/a
Cray XT4 Avg Msg Size Med	113K	n/a	1 MB	35KB	2K	16KB	34KB

*CI is the computational intensity, the ratio of # of Floating Point Operations to # of memory operations.

Benchmark Evolution

- Two new applications
- Concurrency increased ~4X
- Strong scaling in 4 codes
- MAESTRO emphasis on implicit time-stepping, AMR/Multiscale Physics

Recent Workload Publication

Antypas, K., Shalf, J., and Wasserman, H. (2008).
NERSC-6 Workload Analysis and Benchmark Selection Process.
LBNL-1014E.

Available from

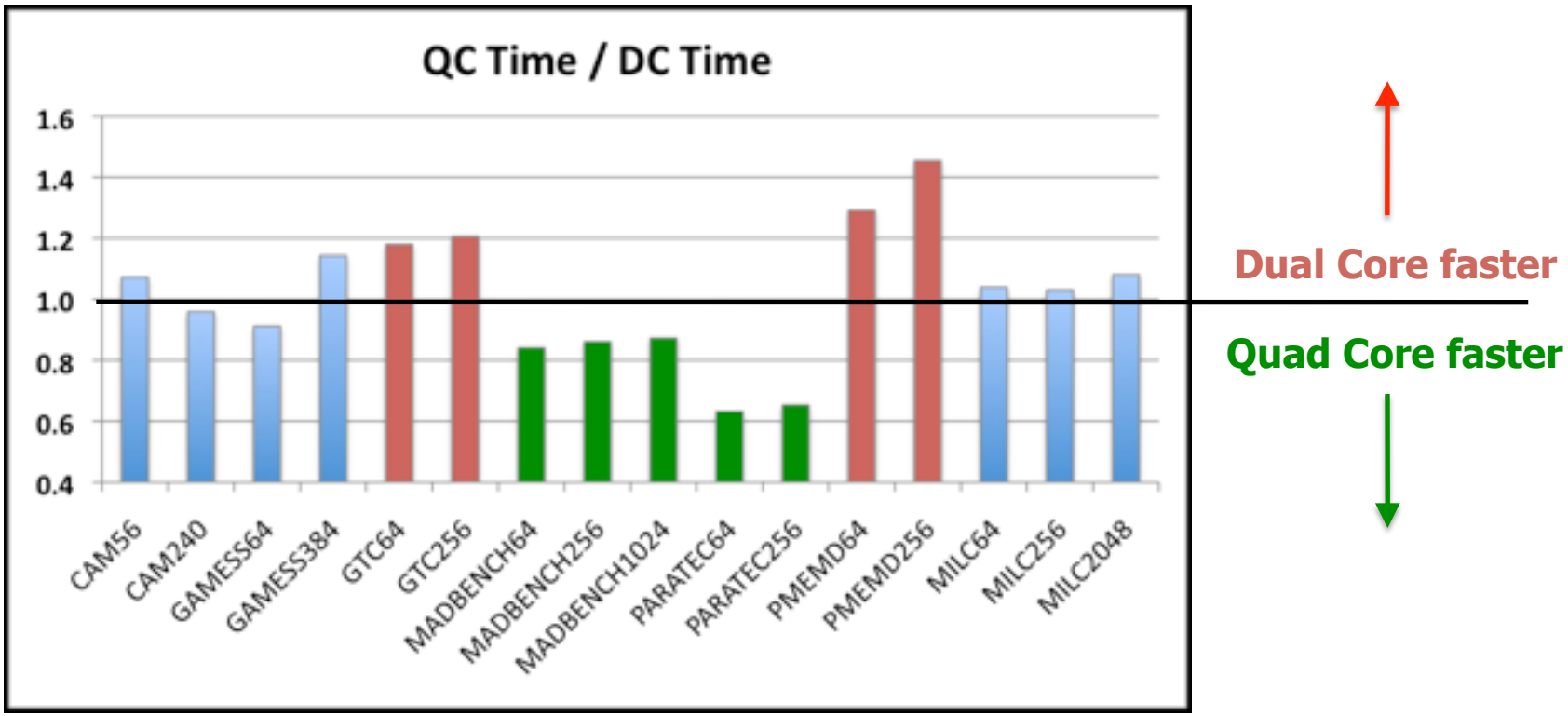
http://www.nersc.gov/projects/SDSA/reports/

DC-QC Comparison

- Compare 2.6-GHz AMD64 with 2.3 GHz Barcelona-64
  - 12% difference in clock speed.
- Many core architectural differences.
- Initial comparison uses CLE 2.0 & 2.1
- SeaStar interconnect: Torus cleaved

Initial QC / DC Comparison

NERSC-5 Benchmarks



Compare time for *n* cores using DC sockets to time for *n* cores using QC sockets.

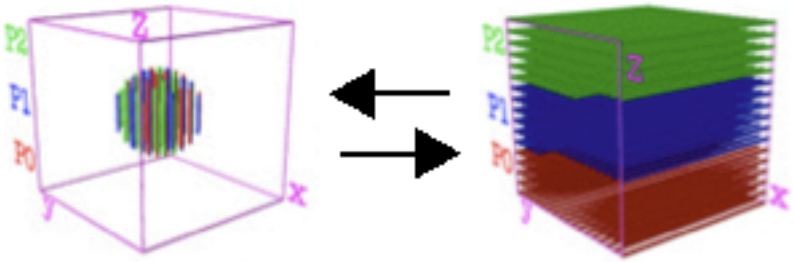
Data courtesy of Helen He, NERSC USG

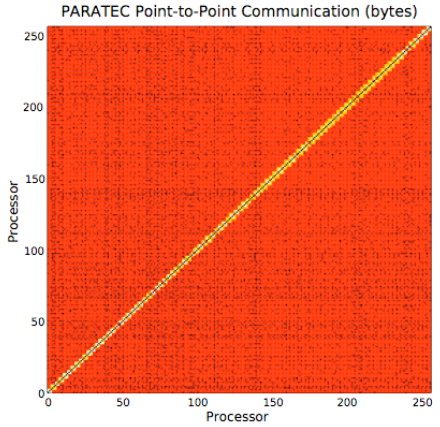
SSE and Quad Core

- Barcelona "SSE128" instruction generation handled by compiler.
- Compiler vectorizes loops; however, performance benefit generally ~small.
  - For user-level code.
  - Use `–fastsse`, `-scalarsse`
  - `pgf90 -fastsse -help`
- Big benefit for SciLib routines

PARATEC: Parallel Total Energy Code

- Captures the performance of ~70% of NERSC material science computation.
- Planewave DFT; calculation in both Fourier and real space; custom 3-D FFT to transform between.
- Uses MPI / SCALAPACK / FFTW / BLAS3
- All-to-all data transpositions dominate communications.





Communication Topology for PARATEC from IPM.

PARATEC Library Usage

- **BLAS3 routines**
  - **ZSYMM (*C <— alphaAB+betaC*)**
  - **ZTRMM (*B <— alphaAB*)**
  - **ZGEMM (*C <— alphaAB+betaC*)**
  - **ZHER2K (*C <— alphaAB^T+alphaBA^T+betaC*)**
  - **ZHERK (*C <— alphaAA^T+betaC*)**

PARATEC: Performance

"Medium" Problem (64 cores)

	Dual Core	Quad Core	Ratio
FFTs[1]	425	537	1.3
Projectors[1]	4,600	7,800	1.7
Matrix-Matrix[1]	4,750	8,200	1.7
Overall[2]	2,900 (56%)	4,600 (50%)	1.6

- [1] Rates in MFLOPS/core from PARATEC output.
- [2] Rates in MFLOPS/core from NERSC-5 reference count.
- Projector/Matrix-Matrix rates dominated by BLAS3 routines.

=> SciLIB takes advantage of wider SSE in Barcelona-64.

PARATEC: Performance

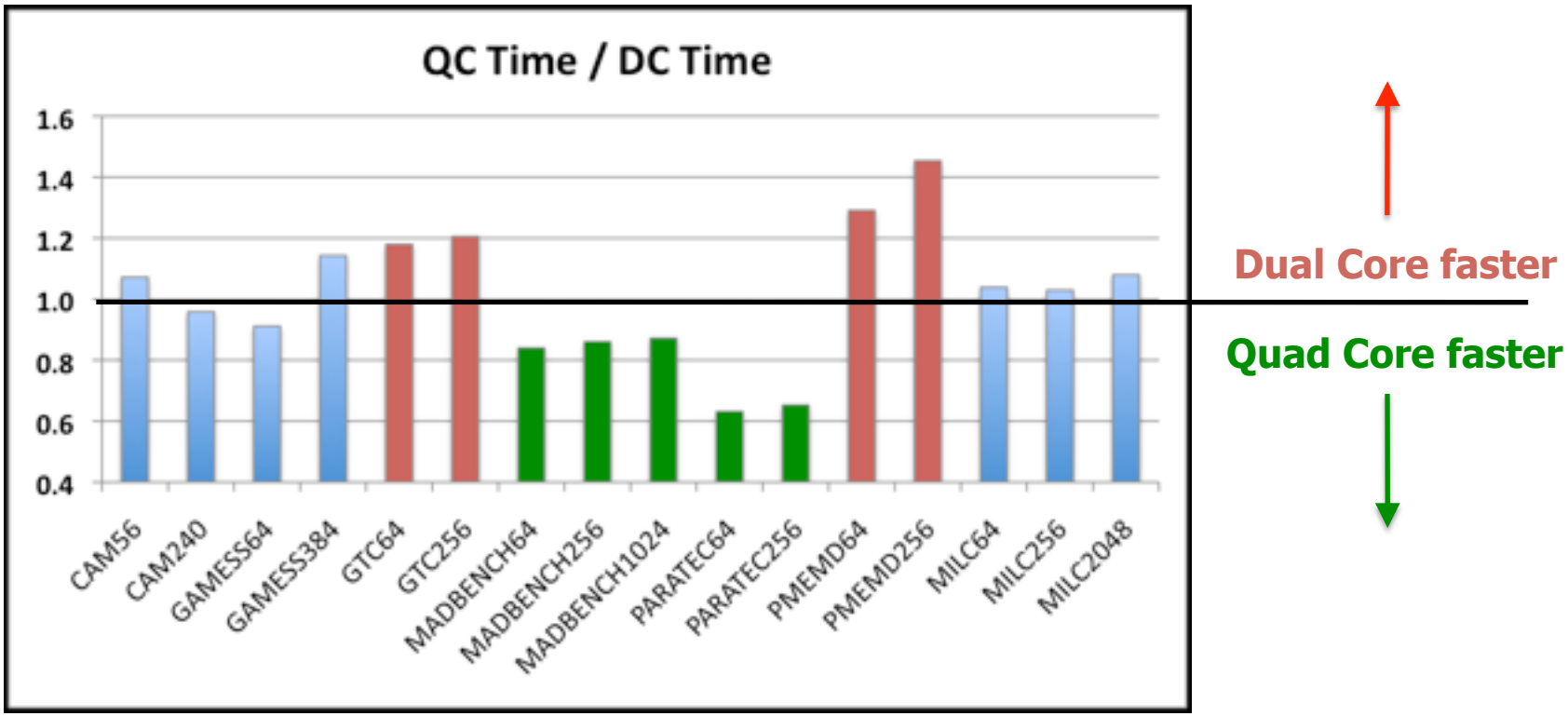
	FFT Rate	Projector Rate	Overall
Franklin Dual-Core	198	4,524	671 (50%)
Franklin Quad-Core	309	7,517	1,076 (46%)
Jaguar Quad-Core	270	6,397	966 (45%)
BG/P	207	567	532 (61%)
HLRB-II	194	993	760 (46%)
BASSI	126	1,377	647 (33%)

HLRB-II is an SGI Altix 4700 installed at LRZ, dual-core Itanium with NUMAlink4 Interconnect (2D Torus based on 256/512 core fat trees)

- NERSC-5 "Large" Problem (256 cores)
- FFT/Projector rates in MFLOPS per core from PARATEC output.
- Overall rate in GFLOPS from NERSC-5 official count
- Optimized version by Cray, un-optimized for most others
- Note difference between BASSI, BG/P, and Franklin QC

Initial QC / DC Comparison

NERSC-5 Benchmarks



Compare time for *n* cores using DC sockets to time for *n* cores using QC sockets.

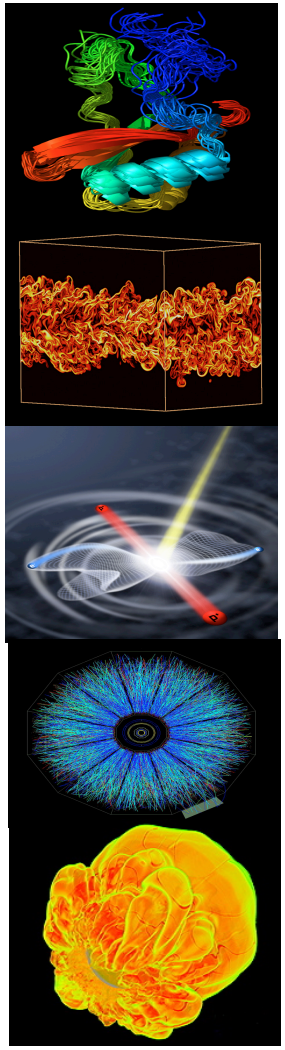
Data courtesy of Helen He, NERSC USG

MILC LG QCD Performance

- NERSC supplied Cray with generic x86 SSE assembler (~3,300 lines)
- Cray rewrote for AMD64 (2006)
- SC -> DC: 40% per-core perf. reduction
- Some updates for Barcelona-64 (2008)
- Prefetch code (~400 lines) is UGLY
- Moral: let the compiler do it

NATIONAL ENERGY RESEARCH SCIENTIFIC COMPUTING CENTER

About the Cover



Schematic representation of 2° secondary structure of native state simulation of the enzyme RuBisCO, the most abundant protein in leaves and possibly the most abundant protein on Earth. **http://www.nersc.gov/news/annual_reports/annrep05/research-news/11-proteins.html**

Direct Numerical Simulation of Turbulent Nonpremixed Combustion. Instantaneous isocontours of the total scalar dissipation rate field. (From E. R. Hawkes, R. Sankaran, J. C. Sutherland, and J. H. Chen, "Direct Numerical Simulation of Temporally-Evolving Plane Jet Flames with Detailed CO/H_2 Kinetics," submitted to the 31st International Symposium on Combustion, 2006.)

A hydrogen molecule hit by an energetic photon breaks apart. First-ever complete quantum mechanical solution of a system with four charged particles. W. Vanroose, F.Martín, T.N. Rescigno, and C. W. McCurdy, "Complete photo-induced breakup of the H_2 molecule as a probe of molecular electron correlation," Science 310, 1787 (2005)

Display of a single Au + Au ion collision at an energy of 200 A-GeV, shown as an end view of the STAR detector. K. H. Ackermann et al., "Elliptic flow in Au + Au collisions at = 130 GeV," Phys. Rev. Lett. **86, 402 (2001).**

Gravitationally confined detonation mechanism from a Type 1a Supernovae Simulation by D. Lamb et al, U. Chicago, done at NERSC and LLNL

Office of Science
U.S. DEPARTMENT OF ENERGY

BERKELEY LAB