

# Performance Evaluation and Analysis Consortium (PEAC) End Station

Presented by

**Patrick H. Worley**

Computational Earth Sciences Group  
Computer Science and Mathematics Division



# Overview

The PEAC End Station provides the performance evaluation and performance tool developer communities access to the Leadership Computing Facility (LCF) systems.

## Consortium goals

### System evaluation

- Evaluate the performance of LCF systems using standard and custom micro-, kernel, and application benchmarks

### Performance tools

- Port performance tools to LCF systems and make them available to National Center for Computational Sciences (NCCS) users
- Further develop the tools to take into account the scale and unique features of LCF systems

### Performance modeling

- Validate the effectiveness of performance modeling methodologies
- Modify methodologies as necessary to improve their utility for predicting resource requirements for production runs on LCF systems

# Overview (continued)

## Consortium goals (continued)

### Application analysis and optimization

- Analyze performance
- Help optimize current and candidate LCF application codes

### Performance and application community support

- Provide access to other performance researchers who are interested in contributing to the performance evaluation of the LCF systems or in porting complementary performance tools of use to the NCCS user community
- Provide access to application developers who wish to evaluate the performance of their codes on LCF systems

## All of this must be accomplished while adhering to the “Golden Rules” of the performance community:

- Low visibility (no production runs!)
- Open and fair evaluations
- Timely reporting of results

# Status as of 8/28/07

32 active users,  
39 active projects:

Consuming:

Contributing to:

- 13 application performance analysis and optimization
- 8 system evaluation
- 8 tool development
- 6 infrastructure development
- 4 application modeling

- XT4: 1,168,000 processor hours (exceeding 1,000,000 processor-hour allocation)

- 1 refereed journal paper
- 1 invited journal paper
- 6 refereed proceedings papers
- 10 proceedings papers
- 2 book chapters
- Numerous oral presentations

# System evaluation

LBNL	Memory, interprocess communication, and I/O benchmarks
	APEX-MAP system characterization benchmark
	Lattice-Boltzman kernels and mini applications
	Application benchmarks from Astrophysics (Cactus), Fluid Dynamics (ELBM3D), High Energy Physics (BeamBeam3D, MILC), Fusion (GTC), Materials Science (PARATEC), AMR Gas Dynamics (HyperCLaw)
ORNL	Computation, memory, interprocess comm., and I/O benchmarks
	Application benchmarks from Astrophysics (Chimera), Climate (CAM, CLM, FMS, POP), Combustion (S3D), Fusion (AORSA, GTC, GYRO, XGC), Molecular Dynamics (NAMD)
SDSC	Subsystem probes for system characterization needed for convolution-based performance modeling
Purdue Univ.	Computation, memory, and interprocess comm. benchmarks
	Application benchmarks from Chemistry (GAMESS), High Energy Physics (MILC), Seismic Processing (SEISMIC), Weather (WRF)

# Performance tools

<b>HPCToolkit</b>	Tool suite for profile-based performance analysis
<b>Modeling assertions</b>	Performance model specification and verification framework
<b>mpiP</b>	MPI profiling infrastructure
<b>PAPI</b>	Performance data collection infrastructure
<b>Scalasca</b>	Scalable trace collection and analysis tool
<b>SvPablo</b>	Performance analysis system
<b>TAU</b>	Performance analysis system
<b>MRNet</b>	Scalable performance tool infrastructure

# Application performance analysis and optimization

Chombo	AMR gas dynamics model
DeCart	Nuclear code
FACETS	Framework application for core-edge transport simulation
GADGET	Computational cosmology
GTC_s	Shape plasma version of GTC gyrokinetic turbulence code
NEWTRNX	Neutron transport code
PDNS3D/SBLI	Ab initio aeroacoustic simulations of jet and airfoil flows
PFLOTRAN	Subsurface flow model
PNEWT	Combustion code

# Application code scaling, optimization, and/or performance evaluation

POLCOMS	Coastal ocean model
S3D	Combustion model
TDCC-9d	Nuclear code
-	Lattice-Boltzman applications



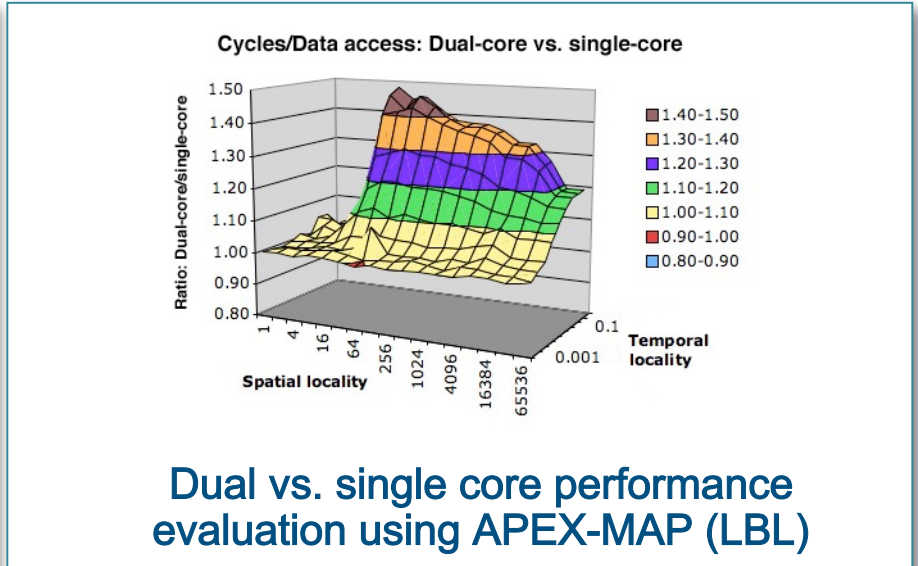
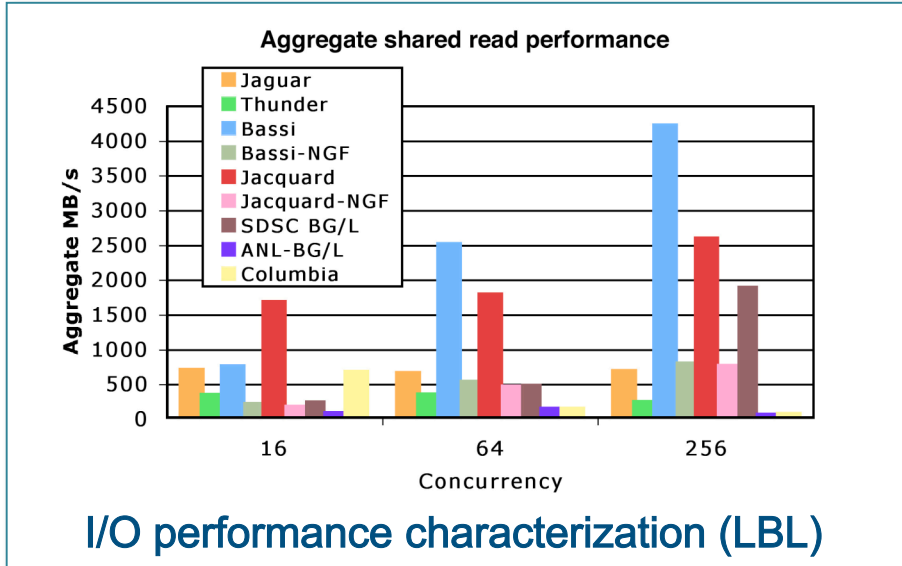
# System infrastructure

cafc	Co-array Fortran compiler for distributed-memory systems
GASNet	Runtime networking layer for UPC and Titanium compilers
PETSc	Toolset for numerical solution of PDEs
PVFS/Portals	PVFS file system implementation on native Portals interface
UPC	Extension of C designed for high-performance computing on large-scale parallel systems
-	Reduction-based communication library

# Performance modeling

PMAC	Genetic algorithm-based modeling of memory-bound computations
ORNL	NAS parallel benchmarks; HYCOM ocean code
Texas A&M Univ.	GTC fusion code
Univ. of Wisconsin	Reusable analytic model for wavefront algorithms, applied to NPB-LU, SWEEP3D, and Chimaera
	LogGP model for MPI communication on the XT4

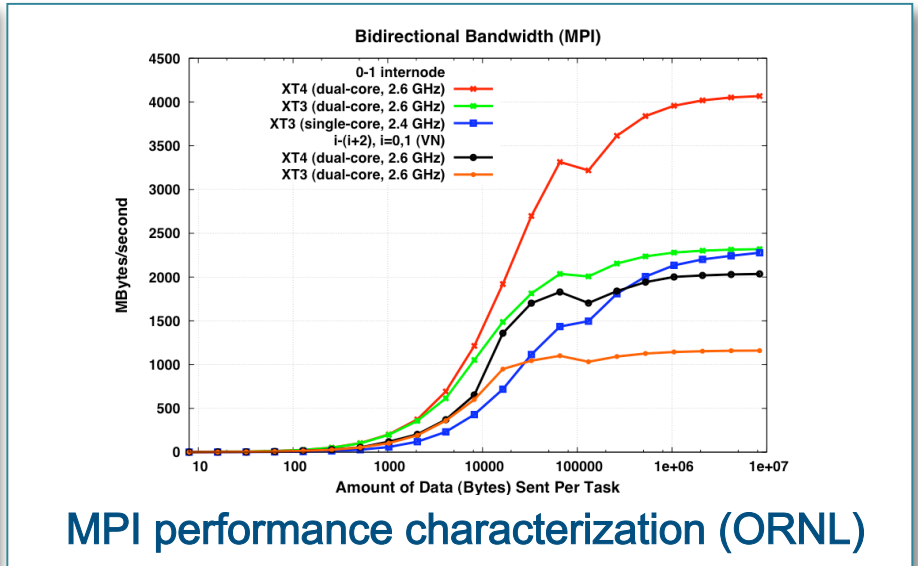
# Subsystem evaluations



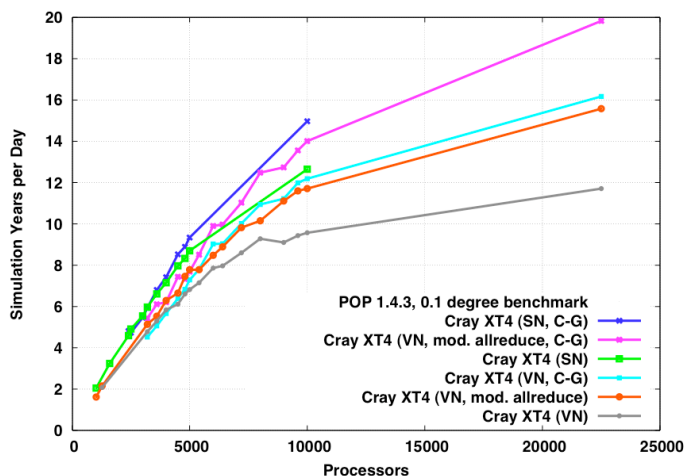
**Ratio of time for all processes sending in halo update to time for a single sender**

System	4 neighbors		8 Neighbors	
		Periodic		Periodic
BG/L	2.24		2.01	
BG/L, VN	1.46		1.81	
XT3	7.5	8.1	9.08	9.41
XT4	10.7	10.7	13.0	13.7
XT4 SN	5.47	5.56	6.73	7.06

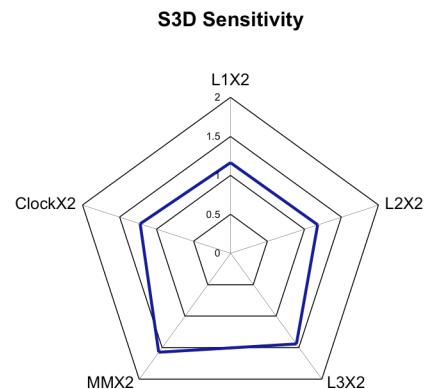
**Identifying performance anomalies (ANL)**



# Application analyses and benchmarks



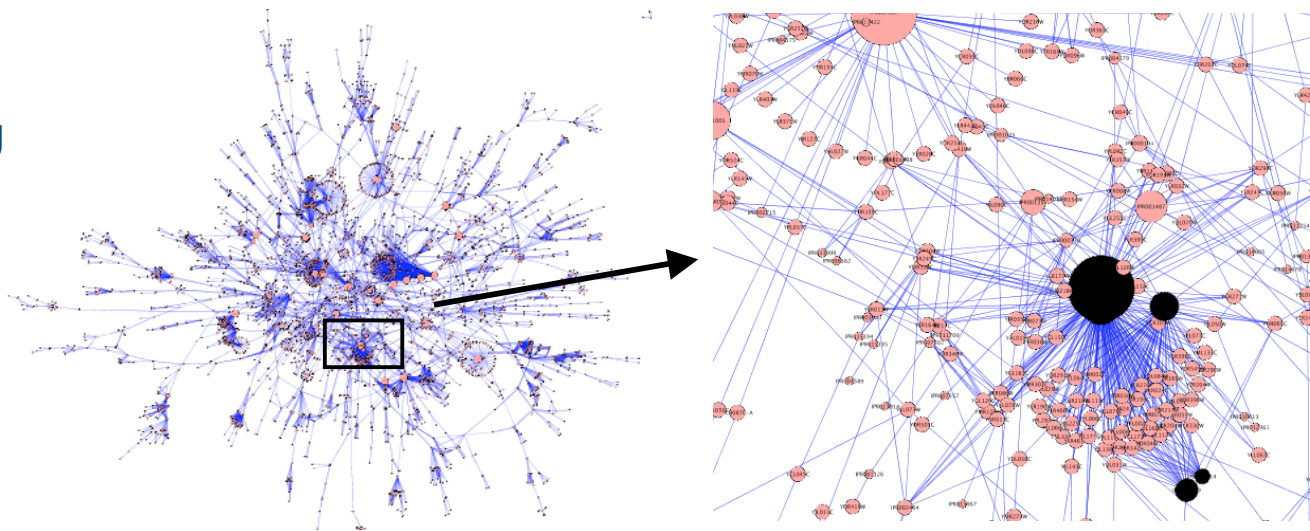
Scalability optimizations (ORNL)



Performance sensitivities (SDSC)

Processing of genomes into domain maps: need improved load balancing that takes into account scale-free nature of the graphs.

Porting and optimizing new applications (RENCI/NCSA)



# Tool development

SvPablo source code-correlated performance analysis (RENCI)

**Project Description:** Lattice QCD - MILC

**Source Files:** control.c, update.c, ./generic\_ks/dslash\_fn2.c, ./generic\_ks/fermion\_force\_asqtad3.c

**Performance Contexts:** STD MILC, 2.4x2.4x2.4, 8 Procs; STD MILC, 2.4x2.4x2.4, 16 Procs; STD MILC, 2.4x2.4x2.4, 32 Procs; STK MILC, 2.4x2.4x2.4, 64 Procs; STD MILC, 2.4x2.4x2.4, 128 Procs; STD MILC, 2.4x2.4x2.4, 256 Procs

**Source File:** /autofs/spin/home/yingz/hewmilc/ks\_imp\_dyn/control.c

```

51  avspect_iters = av_s_iters = avboorr_iters = 0;
52  for( traj_done=0; traj_done < trajecs; traj_done++ ){
53
54      /* do the trajectories */
55      s_iters=update();
56
57      /* measure every "propinterval" trajectories */
58      if( (traj_done%propinterval)==(propinterval-1) ){
59
60          /* call gauge_variable fermion_variable measuring routines */
61          /* results are printed in output file */
62          urephase(OFF);
63          ug_measure( );
64          urephase(ON);
65
66          /* Measure bpb, etc */
67          #ifdef ONEMASS
68          if_meas_imp(F_OFFSET(phi), F_OFFSET(xxx), mass);
69          #else
70          if_meas_imp( F_OFFSET(phi1), F_OFFSET(xxx1), mass1 );
71          if_meas_imp( F_OFFSET(phi2), F_OFFSET(xxx2), mass2 );
72          #endif
73

```

callsites: 47

ID	Lev	File/Address	Line	Parent_Funct	MPI_Call
1	0	0x000000000206ed3		mod_xc_xcogetc_	Bcast
2	0	0x000000000206b49		mod_xc_xcaput4_	Waitall
3	0	0x00000000020d1c7		mod_xc_xcsum_	Bcast
4	0	0x000000000210e53		mod_xc_xctilr_	Startall
5	0	0x000000000210d4e		mod_xc_xctilr_	Request_free
6	0	0x000000000210e6c		mod_xc_xctilr_	...
7	0	0x000000000205e5a		mod_xc_xcaget4_	...
8	0	0x000000000210d21		mod_xc_xctilr_	...
9	0	0x000000000204c21		mod_xc_xcsumj_	...
10	0	0x000000000206b28		mod_xc_xcaput4_	...
11	0	0x000000000206a16		mod_xc_xcaput4_	...
12	0	0x000000000210d30		mod_xc_xctilr_	...
13	0	0x000000000210e3f		mod_xc_xctilr_	...
14	0	0x000000000210d3f		mod_xc_xctilr_	...
15	0	0x000000000210e06		mod_xc_xctilr_	...
16	0	0x00000000020d7e1		mod_xc_xcsumj_	Barrier
17	0	0x000000000211956		mod_xc_xctmrp_	Send
18	0	0x000000000206ed1		mod_xc_xctilr_	Barrier
19	0	0x000000000208f5f		mod_xc_xcmacr_1	Barrier

mpiP callsite profiling (LLNL/ORNL)

Aggregate Time (top twenty, descending, milliseconds)

Call	Site	Time	App%	MPI%	COV
Waitall	18	1.96e+08	14.46	43.39	0.84
Allreduce	27	1.4e+08	10.31	30.93	0.21
Waitall	6	8.06e+07	5.95	17.84	1.28
Bcast	3	9.89e+06	0.73	2.19	0.92
Barrier	25	5.8e+06	0.43	1.28	0.04
Send	42	5.79e+06	0.43	1.28	0.04
Barrier	35	5.29e+06	0.39	1.16	0.04
Barrier	29	4.1e+06	0.30	0.91	0.04
Startall	28	1.42e+06	0.10	0.31	0.53
Startall	4	1.2e+06	0.09	0.27	0.28
Recv	41	8.64e+05	0.06	0.19	0.41
Recv	22	2.61e+05	0.02	0.06	1.45
Send	39	1.92e+05	0.01	0.04	0.44
Send	38	1.75e+05	0.01	0.04	0.46
Allreduce	19	1.74e+05	0.01	0.04	0.56
Request_free	24	6.43e+04	0.00	0.01	0.47
Recv	32	4.16e+04	0.00	0.01	0.00
Recv	43	3.74e+04	0.00	0.01	0.57
Waitall	2	2.9e+04	0.00	0.01	0.59

Metrics: Root percent, Selection percent, Peer percent

- 0.0 Time
- 34.8 Execution
- 0.0 MPI
- 0.0 Communication
- 0.0 Collective
- 0.0 Early Reduce
- 0.0 Early Scan
- 0.0 Late Broadcast
- 2.4 Wait at N x N
- 0.0 N x N Complete
- 14.5 Point-to-point
- 48.3 Late Sender
- 0.0 Late Receiver
- 0.0 Synchronization
- 0.0 MPI I/O
- 0.0 Init/Exit
- 0.0 Overhead
- 100.0 Visits
- 100.0 Communications
- 0.0 Synchronizations

Call Tree: 0.0 HYPRE\_StructSMGSolve, 0.0 hypre\_SMGSolve, 0.0 hypre\_StructMatrixDest, 0.0 hypre\_StructVectorDest, 0.0 hypre\_StructMatrixRef, 0.0 hypre\_StructVectorRef, 0.0 hypre\_StructInnerProd, 0.0 hypre\_SMGRelaxSetReg, 0.0 hypre\_SMGRelaxSetMas, 0.0 hypre\_SMGRelaxSetZer, 0.0 hypre\_SMGRelax, 0.0 hypre\_SMGRelaxSet, 0.0 hypre\_SMGResidual, 46.5 hypre\_SMGSolve, 0.0 hypre\_SMGSetStruct, 5.6 hypre\_SMGResidual, 1.3 hypre\_SemiRestrict, 30.2 hypre\_SemiInterp, 0.0 hypre\_StructAxy, 0.0 hypre\_SMGAxy

Peer percent: (31, 31, 21) to (31, 31, 2)

Summary: 91,140,454 (48.3%) 1.887e+05 42,345,535 (46.5%) 9.114e+04 <1.880 ± 19.2%> 3.239e+00

SCALASCA trace-based performance analysis (FZ-Jülich, UTenn)

# Co-principal investigators

Argonne  
National  
Laboratory  
William Gropp



Lawrence  
Berkeley  
National  
Laboratory  
David Bailey  
Leonid Olikier



Lawrence  
Livermore  
National  
Laboratory  
Bronis  
de Supinski



Oak Ridge  
National  
Laboratory  
Jeffrey Vetter  
Patrick Worley (PI)



Rice  
University  
John  
Mellor-  
Crummey



University of  
California–  
Berkeley  
Kathy Yelick



University of  
California–  
San Diego  
Allan Snavelly



University  
of Maryland  
Jeffrey  
Hollingsworth



University  
of North  
Carolina  
Daniel Reed



University  
of Oregon  
Allen Malony



UNIVERSITY  
OF OREGON

University of  
Tennessee  
Jack Dongarra



University of  
Wisconsin  
Barton Miller





# Contact

## **Patrick H. Worley**

Computational Earth Sciences Group  
Computer Science and Mathematics Division  
(865) 574-3128  
worleyph@ornl.gov

## **Barbara Helland**

DOE Program Manager  
Office of Advanced Scientific Computing Research  
DOE Office of Science

