HPC Challenge Benchmark Suite

Jack Dongarra (UTK/ORNL)

Motivation for Additional Benchmarks

Linpack Benchmark

- Good
  - One number
  - Simple to define & easy to rank
  - Allows problem size to change with machine and over time
  - Stresses the system with a long running job
- Bad
  - Emphasizes only "peak" CPU speed and number of CPUs
  - Does not stress local bandwidth
  - Does not stress the network
  - Does not test gather/scatter
  - Ignores Amdahl's Law (Only does weak scaling)
- Ugly
  - Benchmarketeering hype

- Perhaps there was a time when this was adequate.
- From Linpack Benchmark and Top500: "no single number can reflect overall performance"
- Clearly need something more than Linpack
- HPC Challenge Benchmark
  - Test suite stresses not only the processors, but the memory system and the interconnect.
  - The real utility of the HPCC benchmarks are that architectures can be described with a wider range of metrics than just Flop/s from Linpack.

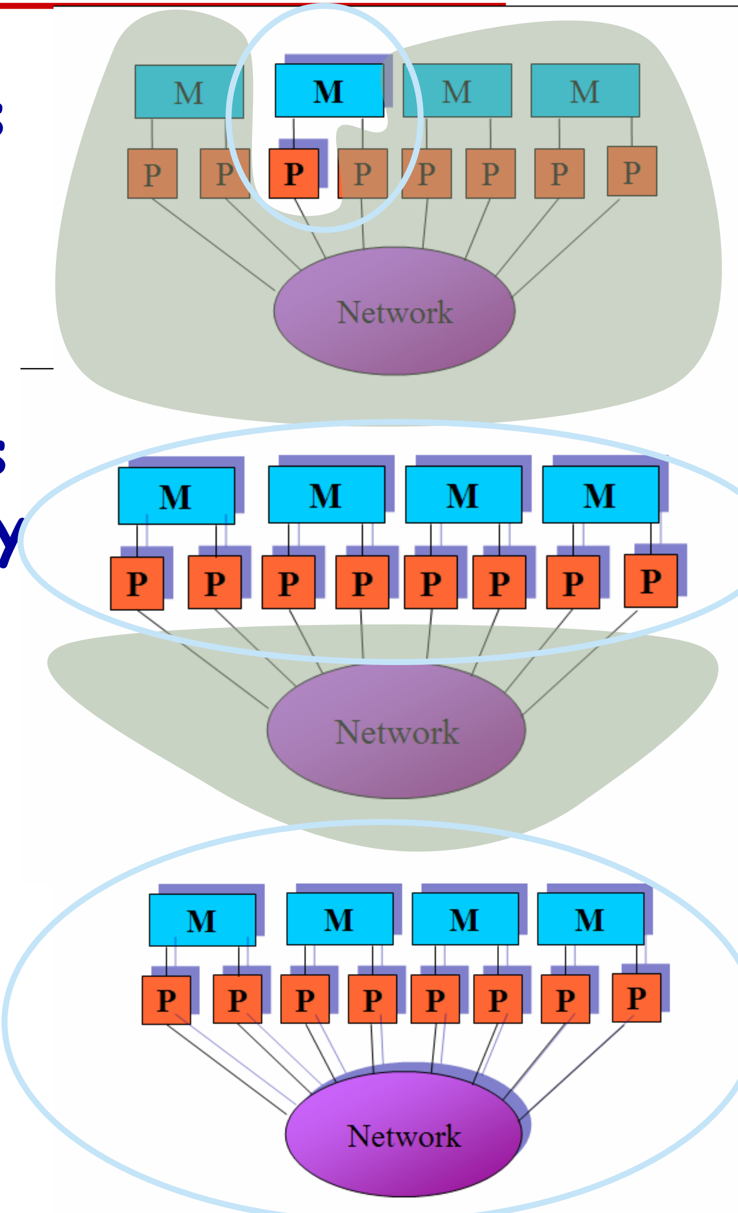
Goals HPC Challenge Benchmark

- Stress CPU, memory system, interconnect
- To *complement* the Top500 list
- To provide benchmarks that *bound* the performance of many real applications as a function of memory access characteristics
  - e.g., spatial and temporal locality

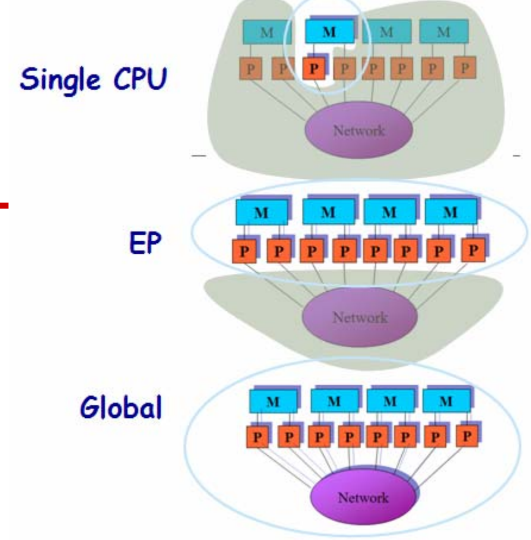
- Allow for optimizations
  - Record effort needed for tuning
  - Base run requires MPI and BLAS
- Provide verification of results
- Archive results

Tests on Single Processor and System

- Local - only a single processor is performing computations.
- Embarrassingly Parallel - each processor in the entire system is performing computations but they do no communicate with each other explicitly.
- Global - all processors in the system are performing computations and they explicitly communicate with each other.



HPC Challenge Benchmark



Consists of basically 7 benchmarks;

- Think of it as a framework or harness for adding benchmarks of interest.

1. HPL (LINPACK) — MPI Global (Ax = b)

2. STREAM — Local; single CPU
   *STREAM — Embarrassingly parallel

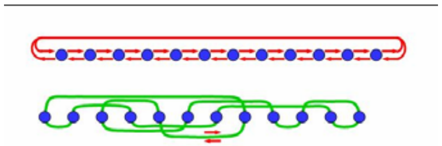
name	kernel	bytes/iter	FLOPS/iter
COPY:	a(i) = b(i)	16	0
SCALE:	a(i) = q*b(i)	16	1
SUM:	a(i) = b(i) + c(i)	24	1
TRIAD:	a(i) = b(i) + q*c(i)	24	2

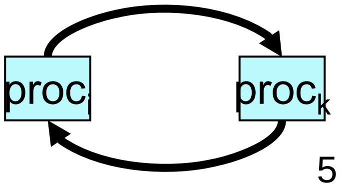
3. PTRANS ($A \leftarrow A + B^T$) — MPI Global

4. RandomAccess — Local; single CPU
   *RandomAccess — Embarrassingly parallel
   RandomAccess — MPI Global

Random integer read; update; & write

5. BW and Latency – MPI

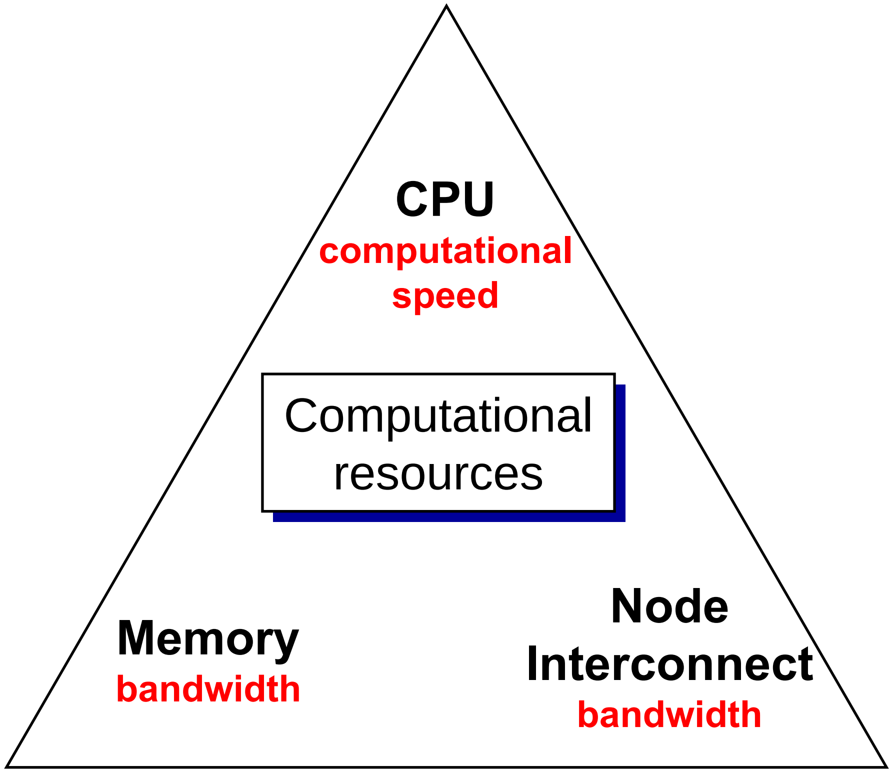




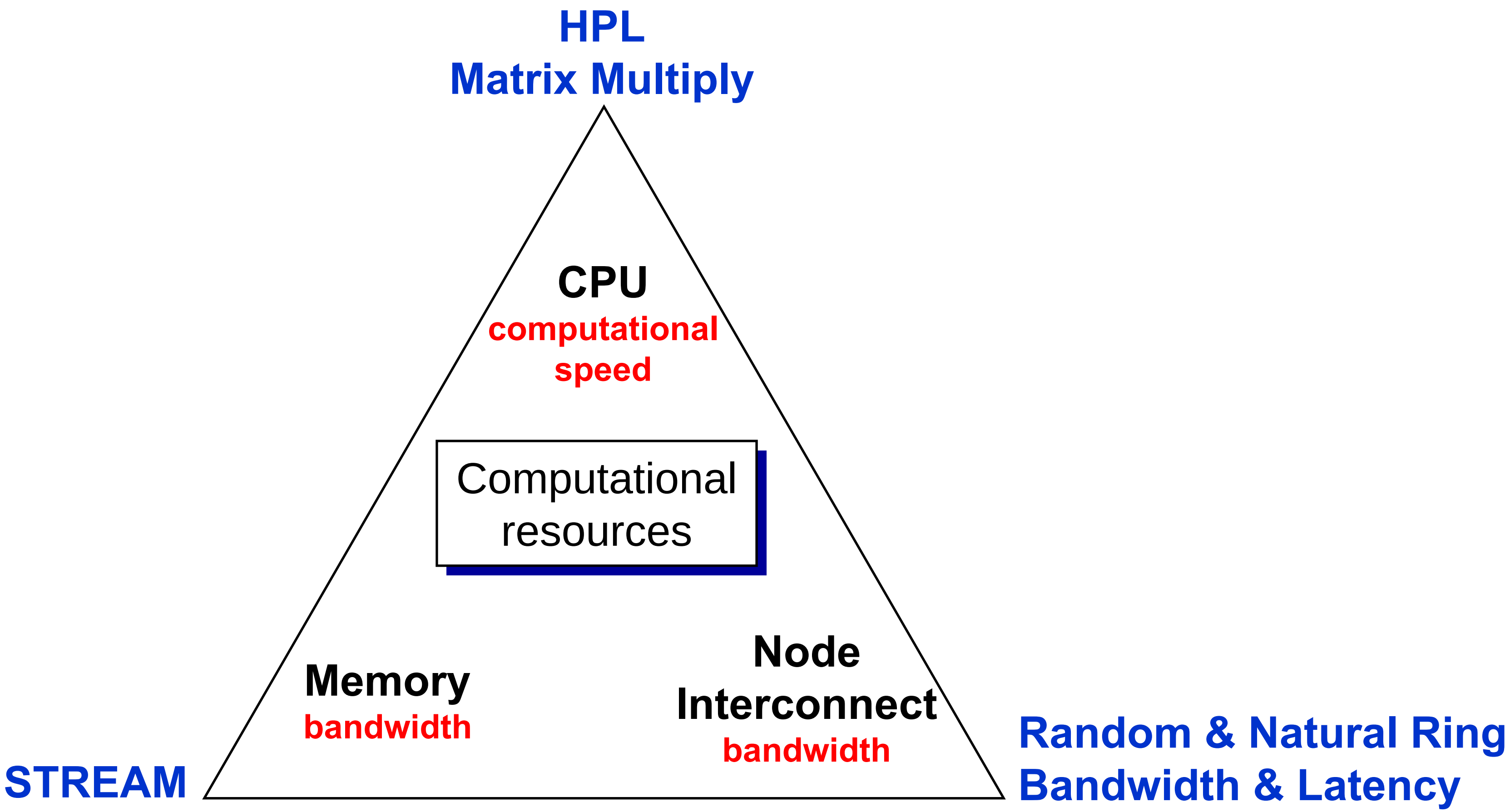
6. FFT - Global, single CPU, and EP

7. Matrix Multiply – single CPU and EP

Computational Resources and HPC Challenge Benchmarks



Computational Resources and HPC Challenge Benchmarks



How Does The Benchmarking Work?

- Single program to download and run
  - Simple input file similar to HPL input
- Base Run and Optimization Run
  - Base run must be made
    - User supplies MPI and the BLAS
  - Optimized run allowed to replace certain routines
    - User specifies what was done
- Results upload via website
- html table and Excel spreadsheet generated with performance results
  - Intentionally we are not providing a single figure of merit (no over all ranking)
- Goal: no more than 2 X the time to execute HPL.

http://icl.cs.utk.edu/hpcc/ web

HPC CHALLENGE

HPCS

- Home
- Rules
- News
- Download
- FAQ
- Links
- Collaborators
- Sponsors
- Upload
- Kiviat Diagram
- Results

HPC Challenge Benchmark

The HPC Challenge benchmark consists of basically 7 benchmarks:

1. HPL - the Linpack TPP benchmark which measures the floating point rate of execution for solving a linear system of equations.
2. DGEMM - measures the floating point rate of execution of double precision real matrix-matrix multiplication.
3. STREAM - a simple synthetic benchmark program that measures sustainable memory bandwidth (in GB/s) and the corresponding computation rate for simple vector kernel.
4. PTRANS (parallel matrix transpose) - exercises the communications where pairs of processors communicate with each other simultaneously. It is a useful test of the total communications capacity of the network.
5. RandomAccess - measures the rate of integer random updates of memory (GUPS).
6. FFTE - measures the floating point rate of execution of double precision complex one-dimensional Discrete Fourier Transform (DFT).
7. Communication bandwidth and latency - a set of tests to measure latency and bandwidth of a number of simultaneous communication patterns; based on b_eff (effective bandwidth benchmark).

HPCCHALLENGE

Home | Rules | News | Download | FAQ | Links | Collaborators | Sponsors | Upload | Results

Condensed Results - Base and Optimized Runs - 76 Systems - Generated on Tue Sep 20 03:47:25 2005

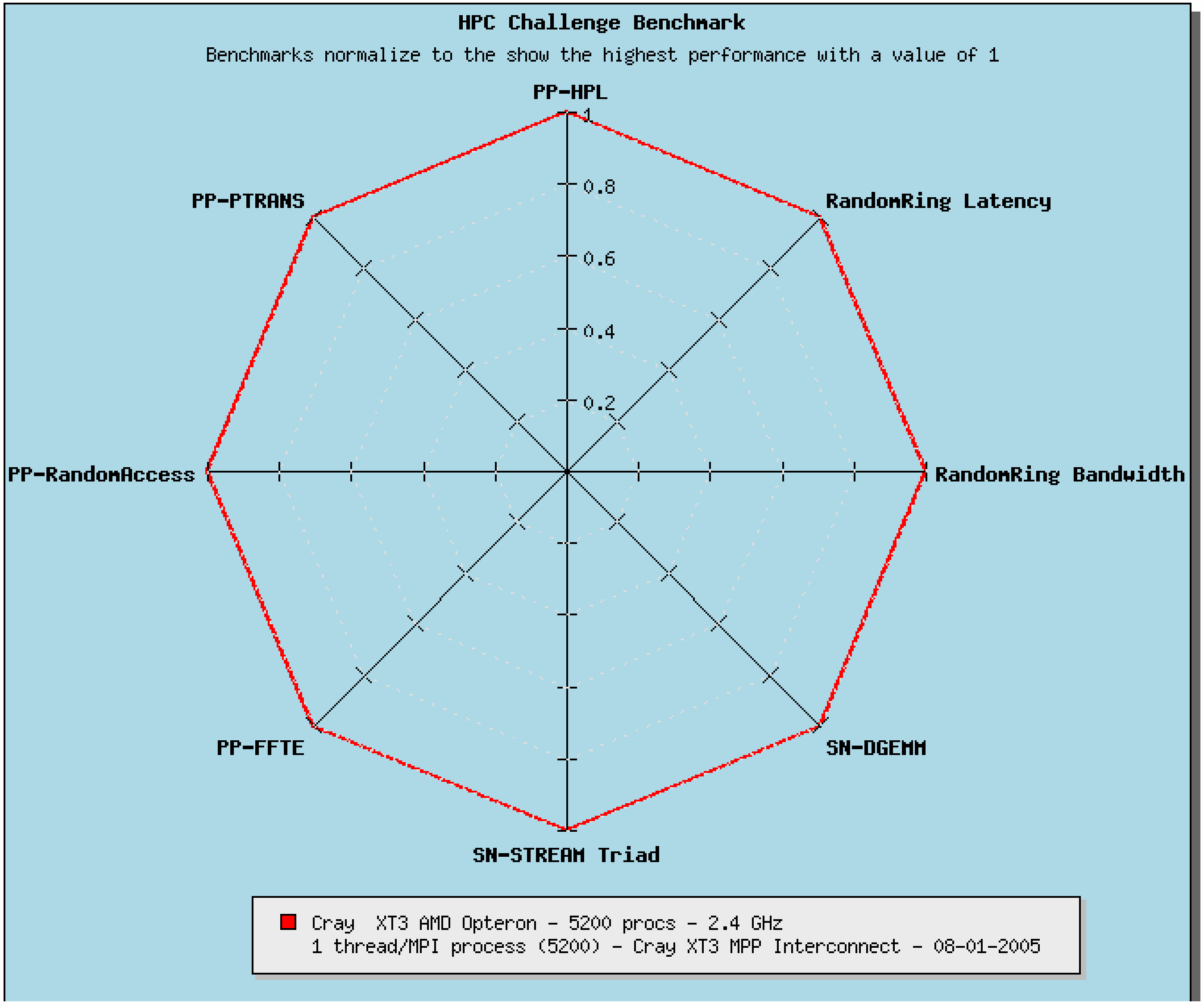
System Information System - Processor	Speed	Count	Threads	Processes	Run	G-HPL	G-PTRANS	G-Random Access	G-FFTE	G-STREAM Triad	EP-STREAM Triad	EP-DGEMM	RandomRing Bandwidth	RandomRing Latency
MA/PT/PS/PC/TH/PR/CM/CS/IC/IA/SD					Type	TFlop/s	GB/s	Gup/s	GFlop/s	GB/s	GB/s	GFlop/s	GB/s	usec
Atipa Conquest cluster AMD Opteron	1.4GHz	128	1	128	base	0.2526	3.247			208.5	1.629		0.03627	23.68
Clustervision BV Beastie AMD Opteron	2.4GHz	32	1	32	base	0.1038	0.816	0.000235	2.15	107.0	3.342	4.195	0.02648	53.23
Cray X1 MSP	0.8GHz	64	1	64	base	0.5216	3.229			959.3	14.990		0.94074	20.34
Cray X1 MSP	0.8GHz	60	1	60	base	0.5778	30.431			898.4	14.974		1.03291	20.83
Cray X1 MSP	0.8GHz	120	1	120	base	1.0610	2.460			1019.5	8.496		0.83014	20.12
Cray T3E Alpha 21164	0.6GHz	1024	1	1024	base	0.0482	10.277			529.2	0.517		0.03174	12.09
Cray X1 MSP	0.8GHz	252	1	252	base	2.3847	97.408			3758.4	14.914		0.42899	22.27
Cray X1 MSP	0.8GHz	252	1	252	opt	2.3678	96.137			5478.7	21.741		0.43828	22.64
Cray X1 MSP	0.8GHz	60	1	60	opt	0.5789	31.072			1306.1	21.768		1.00986	21.16
Cray X1 MSP	0.8GHz	124	1	124	base	1.2054	39.525			1856.7	14.973		0.70857	20.15
Cray X1 MSP	0.8GHz	124	1	124	opt	1.1820	39.383			2697.3	21.752		0.80388	20.85
Cray X1 MSP	0.8GHz	124	1	124	opt	1.1820	39.383			2697.3	21.752		0.80388	20.85
Cray X1 MSP	0.8GHz	60	1	60	base	0.5087	1.634	0.003075	3.14	894.1	14.902	10.915	1.16779	14.66
Cray T3E Alpha 21164	0.675GHz	512	1	512	base	0.2232	9.774	0.028946	15.48	272.2	0.532	0.661	0.03571	8.14
Cray XD1 AMD Opteron	2.2GHz	64	1	64	base	0.2239	10.592	0.022397	16.36	170.0	2.656	4.034	0.22697	1.63
Cray X1 MSP	0.8GHz	32	1	32	base	0.2767	32.661	0.001662	2.96	475.8	14.870	8.258	1.41269	14.94
System Information System - Processor - Speed - Count - Threads - Processes					Run	G-HPL	G-PTRANS	G-Random Access	G-FFTE	G-STREAM Triad	EP-STREAM Triad	EP-DGEMM	RandomRing Bandwidth	RandomRing Latency
MA/PT/PS/PC/TH/PR/CM/CS/IC/IA/SD					Type	TFlop/s	GB/s	Gup/s	GFlop/s	GB/s	GB/s	GFlop/s	GB/s	usec
Cray XT3 AMD Opteron	2.6GHz	1100	1	1100	base	4.7823	217.923	0.137002	266.66	5274.7	4.795	4.811	0.28638	25.94
Cray mfeg8 X1E	1.13GHz	248	1	248	opt	3.3889	66.010	1.854750	-1.00	3280.9	13.229	13.564	0.29886	14.58
Cray XD1 AMD Opteron	2.4GHz	128	1	128	base	0.5021	13.515	0.066672	35.52	500.1	3.907	4.334	0.25919	2.06
Cray X1E X1E MSP	1.13GHz	252	1	252	base	3.1941	85.204	0.014868	15.54	2440.0	9.682	14.185	0.36024	14.93
Cray XT3 AMD Opteron	2.4GHz	3744	1	3744	base	14.7040	608.506	0.220296	417.17	18146.4	4.847	4.413	0.16164	25.32
Cray XT3 AMD Opteron	2.4GHz	5200	1	5200	base	20.5270	874.899	0.268583	644.73	26020.8	5.004	4.395	0.14682	25.80
Cray xt3 AMD Opteron	2.4GHz	32	1	32	base	0.1388	7.376	0.060602	9.37	156.4	4.888	4.776	0.57281	8.74
Cray X1E	1.13GHz	32	4	32	base	0.3376	18.920	0.008969	5.20	307.6	9.611	11.606	1.40487	12.21
Cray XT3 AMD Opteron	2.6GHz	4096	1	4096	base	16.9752	302.979	0.533072	905.57	20656.5	5.043	4.782	0.16896	9.44
Dalco Opteron/QsNet Linux Cluster AMD Opteron	2.2GHz	64	1	64	base	0.2180	6.320	0.004700	13.55	153.4	2.397	3.879	0.17003	11.46
Dell PowerEdge 1850 cluster Intel Xeon EM64T	3.4GHz	64	1	64	base	0.3489	1.892	0.004256	10.39	81.0	1.266	6.081	0.14386	9.81
Dell PowerEdge 2650 Cluster Intel Xeon	2.4GHz	32	1	32	base	0.0966	0.910	0.000276	1.94	18.5	0.579	3.818	0.03788	42.23
Dell PowerEdge 2650 Cluster Intel Xeon	2.4GHz	32	1	32	base	0.1002	1.145	0.002238	2.20	18.7	0.583	3.762	0.04771	8.91

The values plotted for HPL, PTRANS, RandomAccess, and FFTE are per processor. The values plotted for SN-DGEMM and SN-STREAM are per thread. The value plotted for RandomRing Latency is normalized using it's reciprocal. Only those systems that have values for all the tests plotted are available for the diagram. Use the left-hand column to select up to 6 systems to plot in the Kiviat diagram.

Systems for Kiviat Chart - Base Runs Only - 47 Systems - Generated on Mon Sep 26 08:12:42 2005

Plot	System Information System - Processor - Speed - Count - Threads - Processes					PP-HPL	PP-PTRANS	PP-Random Access	PT-SN-STREAM Triad	PP-FFTE	PT-SN-DGEMM	RandomRing Bandwidth	RandomRing Latency
	MA/PT/PS/PC/TH/PR/CM/CS/IC/IA/SD					TFlop/s	GB/s	Gup/s	GB/s	GFlop/s	GFlop/s	GB/s	usec
☐	Clustervision BV Beastie AMD Opteron	2.4GHz	32	1	32	0.003243	0.025498	(0.00000734)	3.339	(0.06709)	4.200	0.0265	53.23
☐	Cray X1 MSP	0.8GHz	60	1	60	0.008479	0.027237	(0.00005125)	16.211	(0.05241)	10.904	1.1678	14.66
☐	Cray T3E Alpha 21164	0.675GHz	512	1	512	0.000436	0.019090	(0.00005654)	0.542	(0.03023)	0.680	0.0357	8.14
☐	Cray XD1 AMD Opteron	2.2GHz	64	1	64	0.003498	0.165506	(0.00034995)	2.766	(0.25564)	3.980	0.2270	1.63
☐	Cray X1 MSP	0.8GHz	32	1	32	0.008647	1.020644	(0.00005194)	16.221	(0.09265)	8.459	1.4127	14.94
☐	Cray XT3 AMD Opteron	2.6GHz	1100	1	1100	0.004348	0.198112	(0.00012455)	4.989	(0.24242)	4.811	0.2864	25.94
☐	Cray XD1 AMD Opteron	2.4GHz	128	1	128	0.003922	0.105590	(0.00052088)	4.358	(0.27748)	4.334	0.2592	2.06
☐	Cray X1E X1E MSP	1.13GHz	252	1	252	0.012675	0.338111	(0.00005900)	23.129	(0.06165)	15.156	0.3602	14.93
☐	Cray XT3 AMD Opteron	2.4GHz	3744	1	3744	0.003927	0.162528	(0.00005884)	4.621	(0.11142)	4.414	0.1616	25.32
☐	Cray XT3 AMD Opteron	2.4GHz	5200	1	5200	0.003947	0.168250	(0.00005165)	4.720	(0.12399)	4.393	0.1468	25.80
☐	Cray xt3 AMD Opteron	2.4GHz	32	1	32	0.004337	0.230513	(0.00189380)	4.888	(0.29276)	4.773	0.5728	8.74
☐	Cray X1E	1.13GHz	32	4	32	0.010551	0.591247	(0.00028027)	5.710	(0.16258)	3.629	1.4049	12.21
☐	Cray XT3 AMD Opteron	2.6GHz	4096	1	4096	0.004144	0.073969	(0.00013014)	5.042	(0.22109)	4.775	0.1690	9.44
☐	Dalco Opteron/QsNet Linux Cluster AMD Opteron	2.2GHz	64	1	64	0.003407	0.098742	(0.00007344)	2.432	(0.21169)	3.893	0.1700	11.46
☐	Dell PowerEdge 1850 cluster Intel Xeon EM64T	3.4GHz	64	1	64	0.005451	0.029568	(0.00006650)	2.844	(0.16232)	6.152	0.1439	9.81
☐	Dell PowerEdge 2650 Cluster Intel Xeon	2.4GHz	32	1	32	0.003020	0.028422	(0.00000862)	1.167	(0.06052)	3.999	0.0379	42.23
Plot	System Information System - Processor - Speed - Count - Threads - Processes					PP-HPL	PP-PTRANS	PP-Random Access	PT-SN-STREAM Triad	PP-FFTE	PT-SN-DGEMM	RandomRing Bandwidth	RandomRing Latency
	MA/PT/PS/PC/TH/PR/CM/CS/IC/IA/SD					TFlop/s	GB/s	Gup/s	GB/s	GFlop/s	GFlop/s	GB/s	usec
☐	Dell PowerEdge 2650 Cluster Intel Xeon	2.4GHz	32	1	32	0.003130	0.035796	(0.00006994)	1.197	(0.06863)	4.025	0.0477	8.91
☐	Dell PowerEdge 2650 Cluster Intel Xeon	2.4GHz	32	1	32	0.003031	0.043274	(0.00006583)	1.182	(0.07353)	4.059	0.0659	19.00
☐	Dell PowerEdge 2650 Cluster Intel Xeon	2.4GHz	32	1	32	0.003169	0.063993	(0.00005507)	1.147	(0.10128)	3.999	0.1784	9.88
☐	HP Compaq SC45 Alpha 21264C	1GHz	484	1	484	0.001199	0.013162	(0.00001671)	1.642	(0.01035)	1.748	0.0226	39.63
☐	HP SC-40 Alpha 21264B	0.833GHz	484	1	484	0.000896	0.010390	(0.00001298)	1.219	(0.00932)	1.449	0.0173	50.10
☐	IBM eServer pSeries 655 Power 4+	1.7GHz	256	4	64	0.004197	0.092661	(0.00002149)	2.912	(0.04087)	4.706	0.7240	8.34
☐	IBM eServer pSeries 655 Power 4+	1.7GHz	128	4	32	0.004154	0.060726	(0.00002211)	3.009	(0.03424)	4.733	0.7472	7.94
☐	IBM eServer pSeries 655 Power 4+	1.7GHz	64	4	16	0.004085	0.062680	(0.00002282)	2.847	(0.03058)	4.725	0.7483	7.57

HPCC Kiviat Chart



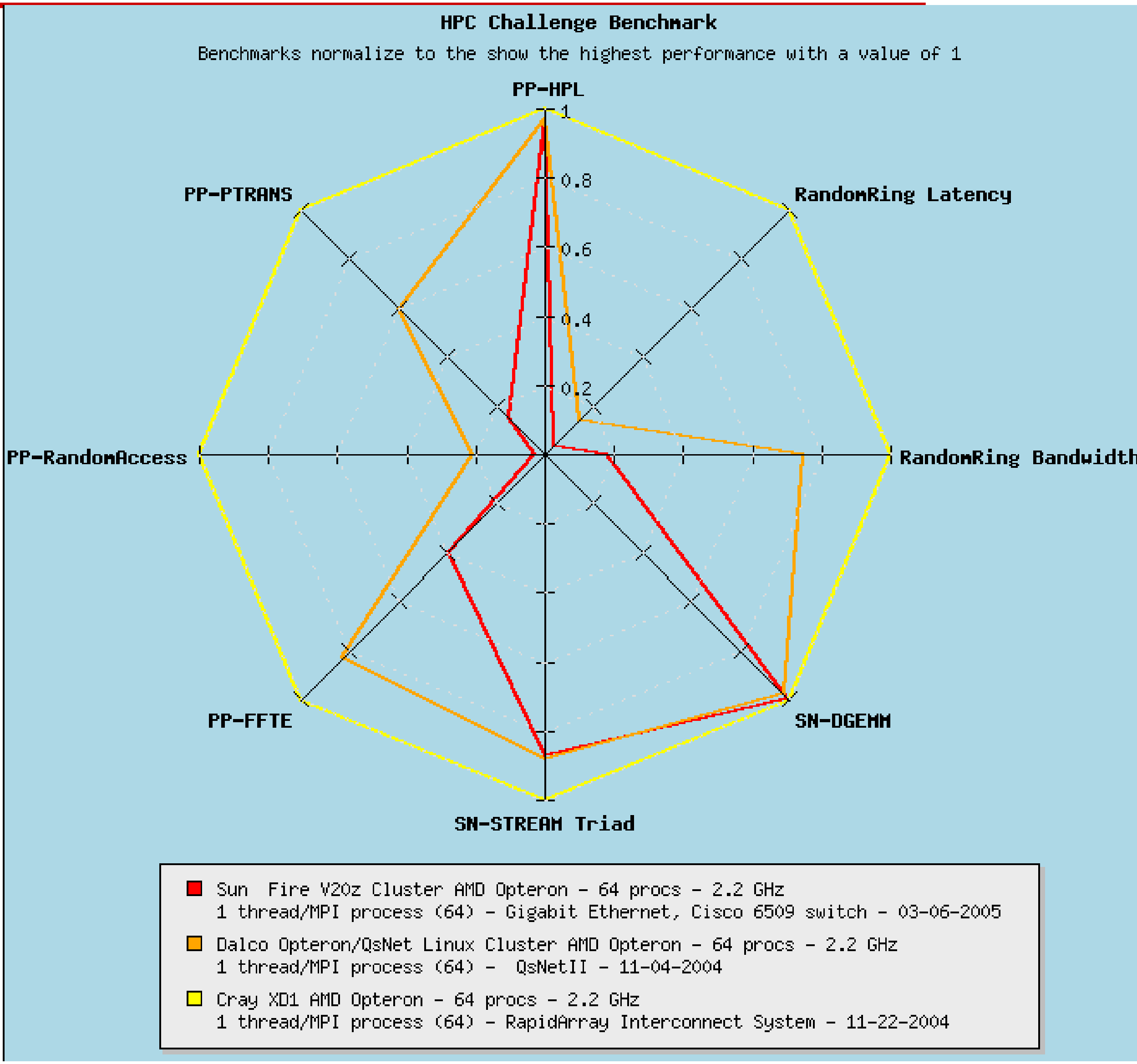
HPCC: A Comparison of 3 Systems

- Three systems using the same processor and number of processors.
  - AMD Opteron 64 processors 2.2 GHz

Cray XD1
Custom Interconnect

Dalco Linux Cluster
Quadrics Interconnect

Sun Fire Cluster
Gigabit ethernet Interconnect



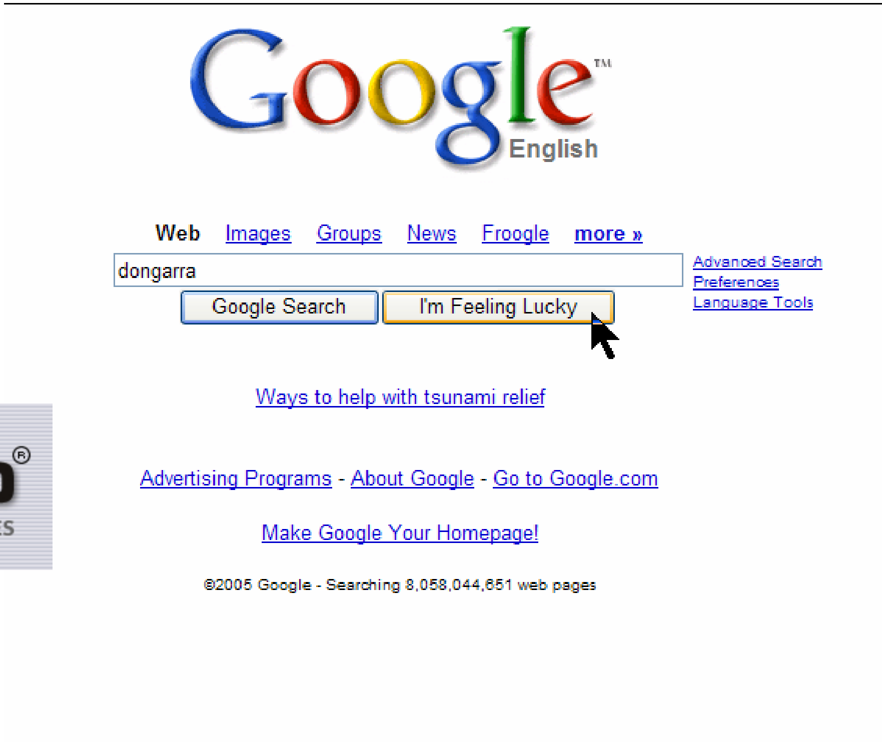
Future Directions

- Looking at reducing execution time
- Constructing a framework for benchmarks
- Developing machine signatures
- Plans are to expand the benchmark collection
  - Sparse matrix operations
  - I/O
  - Smith-Waterman (sequence alignment)

- Port to new systems
- Provide more implementations
  - Languages (Fortran, UPC, Co-Array)
  - Environments
  - Paradigms

Collaborators

- HPC Challenge
  - Piotr Łuszczek, U of Tennessee
  - David Bailey, NERSC/LBL
  - Jeremy Kepner, MIT Lincoln Lab
  - David Koester, MITRE
  - Bob Lucas, ISI/USC
  - Rusty Lusk, ANL
  - John McCalpin, IBM, Austin
  - Rolf Rabenseifner, HLRS Stuttgart
  - Daisuke Takahashi, Tsukuba, Japan



http://icl.cs.utk.edu/hpcc/