SPEC MPI2007
Benchmarks for HPC Systems

Ron Lieberman
Chair, SPEC HPG
HP-MPI Performance
Hewlett-Packard Company

Dr. Tom Elken
Manager, Performance Engineering
QLogic Corporation

Dr William Brantley
Manager HPC Performance
AMD

Dr. Matthias S. Mueller
Vice Chair, SPEC HPG
Deputy Director, CTO
Center for Information Services and High Performance Computing (ZIH)
Dresden University of Technology

Dr Matthijs van Waveren
Secretary, SPEC HPG
Fujitsu Systems Europe Ltd

CAUTIONS

- The information contained within this presentation is a forward looking statement.
- Additionally, any slides with performance data are to be considered ‘ESTIMATES’ and are labeled as such.

SPEC MPI2007

- An application benchmark suite that measures CPU, memory, interconnect, compiler, MPI, and file system performance.
- Search program ended 3/31/06
- Candidate codes in the areas of Comp. Chemistry, Weather, HE Physics, Oceanography, CFD, etc.

CPU2006/MPI2007 Similarities

- Same tools used to run the benchmarks
- Similar run and reporting rules
- Uses geometric mean to calculate overall performance relative to a baseline system
- Similar output format

Comparison of benchmark characteristics

Characteristic	CPU2006	OMPM2001	MPI2007 (est)
Max. working set	0.9/1.8 GB, 32/64-bit	1.6 GB	to be decided
Memory needed	1 or 2 GB. 32- or 64-bit	2 GB	1 GB per benchmark
Benchmark runtime	20 min @ 2 GHz	5 hrs @ 300 MHz	to be decided
Language	C, C++, F95	C, F90, OpenMP	C, C++, F95, MPI
Focus	Single CPU	< 16 CPU system	> 16 CPU system
System type	Desktop	MP workstation	Engineering cluster
Runtime	50-60 hours	34 hours	to be decided
Runtime 1 CPU	50-60 hours	140 hours	to be decided
Run modes	Single and rate	Parallel	Parallel
Number benchmarks	29	11	to be decided
Iterations	Median of 3	Worst of 2, median of 3 or more	Worst of 2, median of 3 or more
Source mods	Not allowed	Allowed	Not allowed
Baseline flags	Any, same for all	Any, same for all	Any, same for all
Reference system	1 CPU @ 300 MHz	4 CPU @ 350 MHz	16 cores @ 2.2 GHz

SPEC MPI2007 Development

- Participating Members
  - AMD, Fujitsu, HP, IBM, INTEL,
  - QLogic (PathScale), SGI, SUN,
  - University of Dresden
- Current release targeted for March-June 2007
  - ISC'07 in Dresden June 2007 most likely release.
- We are always looking for new members to help develop benchmarks

SPEC MPI2007 Benchmark Goals

- Runs on Clusters or SMP's
- Validates for correctness and measures performance
- Supports 32-bit or 64-bit OS/ABI.
- Consists of applications drawn from National Labs and University research centers
- Supports a broad range of MPI implementations and Operating systems including Windows, Linux , Proprietary Unix

SPEC MPI2007 Benchmark Goals

- Scales up and scales out
- Has a runtime of ~1 hour per benchmark test at 16 ranks using GigE with 1 GB memory footprint per rank
- Is extensible to future large and extreme data sets planned to cover larger number of ranks.

SPEC MPI2007 - Current Status

- Runs on range of architectures
  - Opteron, Xeon, Itanium2, PA-Risc, Power5, Sparc,
- Ported to variety of operating systems
  - Linux (RH/XC, SuSE, FC), Windows CCS, HPUX, Solaris, AIX
- Broad range of MPI's evaluated
  - HP-MPI, MPICH, MPICH2, Open MPI, IBM-MPI, Intel MPI, MPICH-GM, MVAPICH, Fujitsu MPI, InfiniPath MPI, SGI MPT

SPEC MPI2007 - Current Status

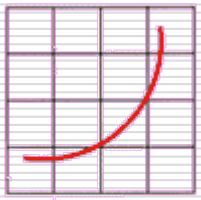
- Scalable from 16 to 128 ranks (processes) for medium data set. 16 of 18 benchmarks run at 512 ranks.
- Runtime of 1 hour per benchmark test at 16 ranks using GigE on an unspecified reference cluster.
- Memory footprint should be < 1GB per rank at 16 ranks.
- Exhaustively tested for each rank count
  - 12
  - 15 -> 130
  - 140, 160, 180, 200, 225, 256, 512

MPI2007 Performance Dimensions

Scale out/up	Clusters, SMPs, Fatnode clusters
Launch strategies	affinity, process placement
MPI Distributions	open source, industrial, collective algorithms
Operating systems	distributions, kernel revisions, tunables
Interconnects	hardware, protocol, drivers, multi-rail
Hardware	CPU, memory, motherboards
Compilers	optimization, correctness
File Systems	Disks, Software, Network

SPEC MPI2007 - Likely Uses

- Customer RFP's
- Marketing messages as it relates to publication on SPEC HPG Web site.
- Academic Research
- Product Qualification and Performance evaluation
  - Evaluate new releases, interconnects, OS's…



SPEC MPI2007 Benchmark Characteristics

Category	Language	#LOC
Physics	C	18K
CFD	FORTRAN	11K
CEM	FORTRAN	22K
CFD	FORTRAN & C	45K
Bioinformatics	C	11K
Bioinformatics	C++	1,421K
Oceanography	FORTRAN	71K
Ray Tracing	C	16K
Molecular Dynamics	C++	58K
Weather Forecasting	FORTRAN & C	218K
FEM (HT)	FORTRAN & C	31K
Hydrodynamics	FORTRAN	7K
Chemistry	FORTRAN & C	93K
Hydrodynamics	FORTRAN	45K
Abinitio	C	260K
Ocean & Atm.	FORTRAN & C	41K
Gravitation	C	24K
CFD	FORTRAN	6K

SPEC MPI2007 (32 ranks) Characteristics -- ESTIMATES

Elapsed Time	%User Time	%MPI Time	Gbytes-Xfer
2142.44	82%	18%	142
3997.10	72%	28%	214
1682.58	67%	33%	134
1926.18	91%	9%	0
1142.03	92%	8%	1
2269.12	50%	50%	0
2016.27	64%	36%	497
2034.54	99%	1%	1
1841.00	94%	6%	133
3085.30	74%	26%	440
653.17	86%	14%	38
1116.59	85%	15%	142
1203.73	96%	4%	140
1400.41	83%	17%	91
580.05	86%	14%	6
2180.32	62%	38%	876
920.04	80%	20%	22
733.14	94%	6%	67

MPI2007 Benchmark Message call counts

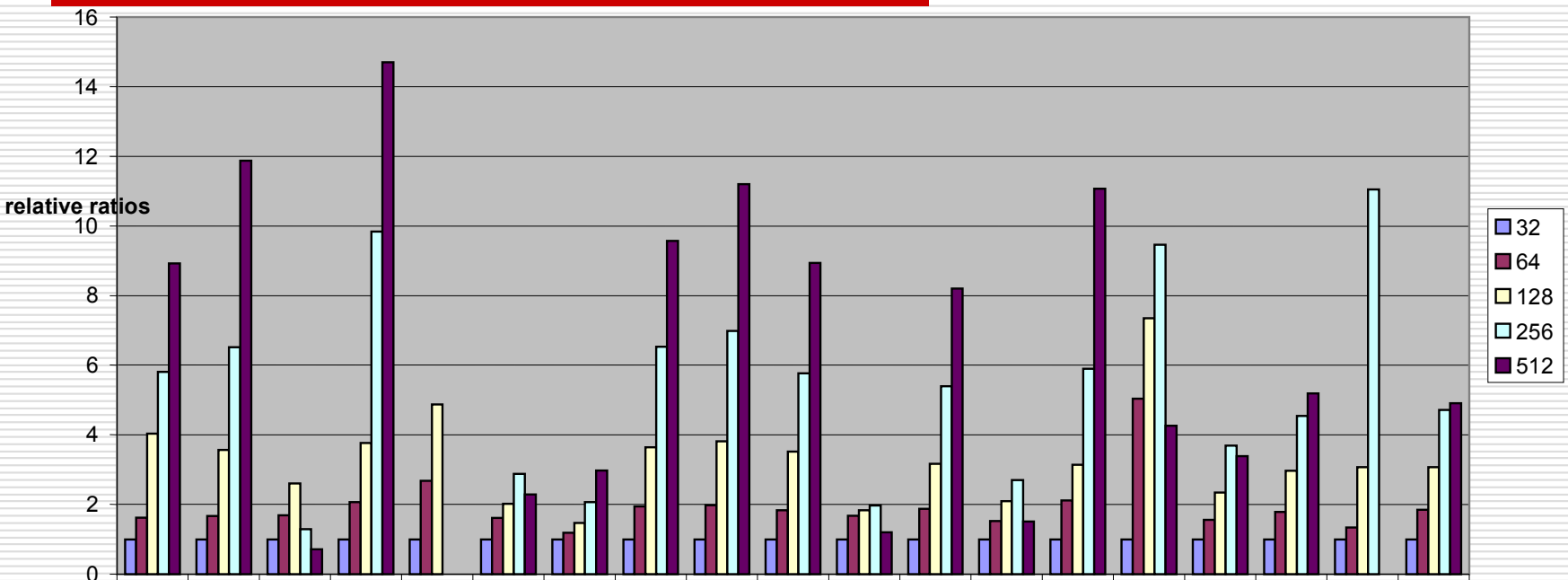
MPI_Allgather				303040	32			32	
MPI_Allgatherv									
MPI_Allreduce	17700	140832					23628416		1696
MPI_Barrier	62	1088	160	320			8640	32	64
MPI_Bcast	122		292000			256	9664		1888
MPI_Cart_create									32
MPI_Comm_create							96		
MPI_Comm_dup									
MPI_Comm_free									32
MPI_Comm_split	32		32						
MPI_Gather				8512					
MPI_Iprobe					163720822	90852			
MPI_Irecv	359340	3201600				18688	5.58E+08		196544
MPI_Irsend						18688			
MPI_Isend	359340					18688	5.58E+08		
MPI_Issend						18688			
MPI_Probe					202504	69100			
MPI_Recv			3270	35371	606434	106506			9152
MPI_Reduce		64	128						
MPI_Scan									32
MPI_Send		3201600	3270	35371	606434	69130			205696
MPI_Send_init								16158	
MPI_Sendrecv			1204000						
MPI_Ssend									
MPI_Start								16158	
MPI_Startall								1	
MPI_Test						724276			
MPI_Testany								522	
MPI_Wait	718680	3201600				18716			196544
MPI_Waitall				151264			3.19E+08	32	
MPI_Waitany									

MPI2007 Benchmark Message call counts

MPI_Allgather				512					
MPI_Allgatherv				7936					
MPI_Allreduce		2002016	60416	36992	12864	5376	190336		224
MPI_Barrier			15520	9760	96	28352	4224	2080	32
MPI_Bcast	67488	352	1184		1248	1152	340224		288
MPI_Cart_create					32				
MPI_Comm_create						26144			
MPI_Comm_dup		32		224					
MPI_Comm_free						848			
MPI_Comm_split					32				32
MPI_Gather									
MPI_Iprobe									
MPI_Irecv	6508380	6015144	1991616	5266164	845056			14774240	19000
MPI_Irsend									
MPI_Isend		6015144			845056		6231	7390144	
MPI_Issend									
MPI_Probe									
MPI_Recv	10106		360			1580	367784	280052	7600320
MPI_Reduce				1152	64				
MPI_Scan									
MPI_Send	6518486		1991976	5266164		1580	361553	7663614	7619320
MPI_Send_init							7243224		
MPI_Sendrecv									
MPI_Ssend								534	
MPI_Start									
MPI_Startall							14168576		
MPI_Test								1.13E+08	
MPI_Testany									
MPI_Wait	6508380		1991616					22163392	19000
MPI_Waitall		1394816			249888		14170586		
MPI_Waitany				5266164					

ESTIMATES

scaling 32-512 mpi2007



																			geom ean
32	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
64	1.62	1.67	1.69	2.07	2.68	1.61	1.18	1.94	1.98	1.83	1.68	1.87	1.52	2.11	5.04	1.55	1.78	1.33	1.85
128	4.04	3.57	2.60	3.76	4.87	2.02	1.46	3.64	3.81	3.52	1.83	3.17	2.10	3.14	7.35	2.34	2.96	3.07	3.07
256	5.81	6.51	1.29	9.83	0.00	2.87	2.07	6.53	6.98	5.77	1.97	5.39	2.69	5.90	9.46	3.69	4.53	11.05	4.71
512	8.92	11.88	0.71	14.70	0.00	2.28	2.97	9.57	11.21	8.94	1.20	8.21	1.51	11.07	4.26	3.39	5.19	0.00	4.91

SPEC MPI2007 Fair Use Policy

- **SPEC/HPG Fair Use Rule**
- Consistency and fairness are guiding principles for SPEC. To assure these principles are sustained, the following guidelines have been created with the intent that they serve as specific guidance for any organization (or individual) who chooses to make public comparisons using SPEC benchmark results.
- When any organization, including vendor or research oriented organizations, or any individual makes public claims using SPEC benchmark results, SPEC requires that the following guidelines be observed:
- [**1**] Reference is made to the SPEC trademark. Such reference may be included in a notes section with other trademark references (see **http://www.spec.org/spec/trademarks.html** for all SPEC trademarks and service marks).
- [**2**] The SPEC web site (**http://www.spec.org**) or a suitable sub page is noted as the source for more information.
- [**3**] If any public claims or competitive comparisons are made, the results stated or compared must be compliant with that benchmark's run and reporting rules and must cite the following: SPEC metric, CPU description (number of chips and cores), and number of OpenMP threads and/or MPI ranks.
- [**4**] If competitive comparisons are made the following rules apply:**a.** the basis for comparison must be stated, **b.** the source of the competitive data must be stated, **c.** the date competitive data was retrieved must be stated, **d.** all data used in comparisons must be publicly available (from SPEC or elsewhere) **e.** the benchmark must be currently accepting new submissions if previously unpublished results are used in the comparison.
- [**5**] Comparisons with or between non-compliant test results can only be made within academic or research documents or presentations where the deviations from the rules for any non-compliant results have been disclosed. A compliant test result is a test result that has followed the run rules, and has been submitted and approved by SPEC. SPEC HPG makes recommendations for the academic or research use of benchmark results in the document, "Guidelines for the Use of SPEC HPG Benchmarks in Research Publications.".

SPEC MPI2007 RunRules

- http://www.spec.org/mpi2007/docs/runrules.html
- This document specifies how the benchmarks in the MPI2007 suites are to be run for measuring and publicly reporting performance results, to ensure that results generated with the suites are meaningful, comparable to other generated results, and reproducible (with documentation covering factors pertinent to reproducing the results).
- Per the SPEC license agreement, all results publicly disclosed must adhere to the SPEC Run and Reporting Rules, or be clearly marked as estimates.

Acknowledgements

- Active members of SPEC HPG who make things happen with their dedication and passion
  - AMD, Fujitsu, HP, IBM, Intel, QLogic, SGI, SUN, University of Dresden
- SPEC OSG for allowing us to leverage CPU2006 benchmarks and tools
- Have I mentioned we are always looking for new members to help develop benchmarks?