



Analysis of OpenMP and MPI Codes on Sun Fire Systems

Nawal Copty

Larry Meadows et.al

**EWOMP'03
Aachen
Sept.,2003**



Outline

- Presenter:
 - Nawal Copty
- Authors:
 - L. Meadows, M. Lee, D. Paulraj, S. Goil, CPE
 - B. Whitney, SAE
- Hardware Used
- Benchmarks Used
- Performance Results
- Conclusions

Hardware Used

- System 1: Single SF15K
 - 1, 72-cpu SF 15K, 900MHz US3cu, 144GB, backplane interconnect
- System 2: 8-node SF V880 Cluster
 - 8, 8-cpu SF V880, 1050MHz US3cu, 16GB, Myrinet interconnect
- System 3: 3-node SF 6800 Cluster
 - 3, 24-cpu SF 6800, 900MHz US3cu, 96GB, Sun Fire Link interconnect

Bandwidth and Latency

System Interconnect	SF 15K Backplane	SF V880 Cluster Myrinet	SF 6800 Cluster Sun Fire Link
MPI Latency (us)	4.1	14	5.5
MPI Bandwidth (MB/Sec)	600	130	520
Memory Latency (ns)			
Local	250	205	230
Remote	470	205	270
Memory Bandwidth (MB/Sec)	1600-2400*		

* Max memory bandwidth for 1 cpu = 2400 MB/Sec; achieved depends On compiler options, O/S version, page size, etc.

MPI: Point to Point; Memory latency: lat_mem_rd()

Benchmarks: SPEC HPC2002

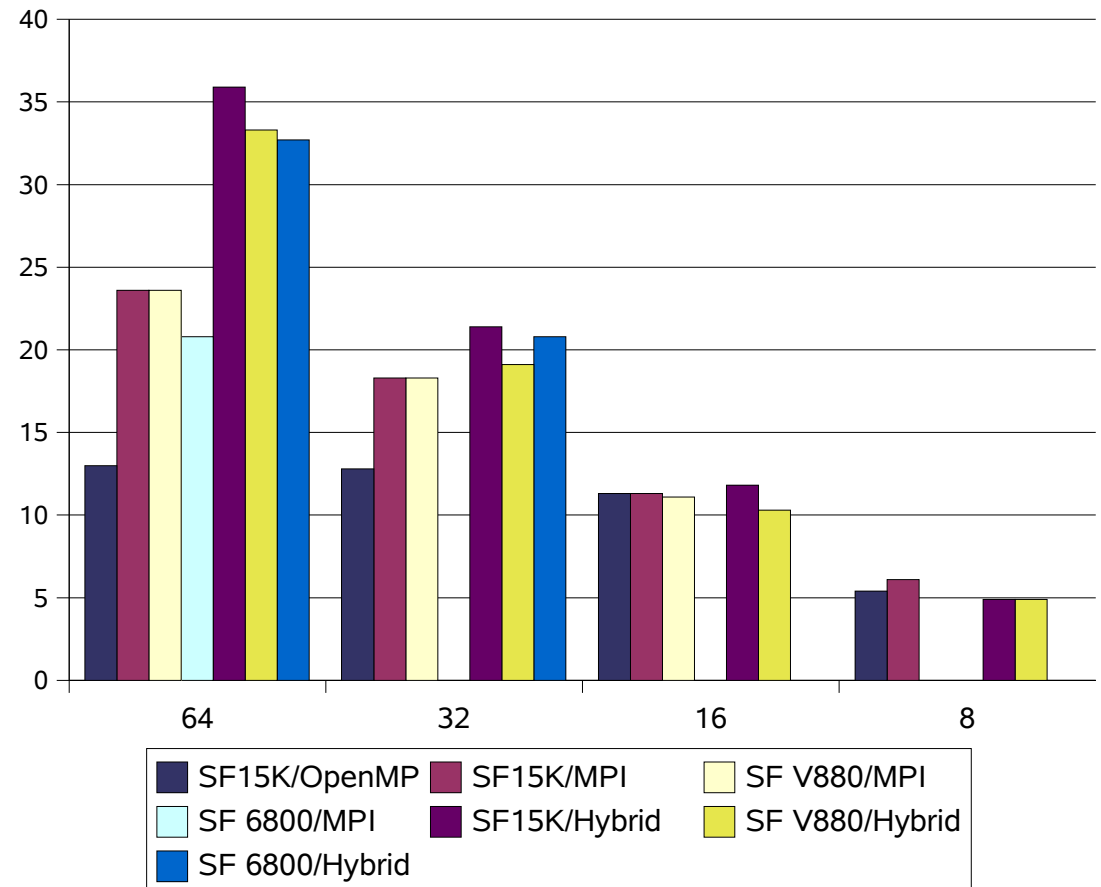
- SPEC CHEM2002 (garnet) MPI, OpenMP, Hybrid
 - MPI: Master/Slave + Global reduction
 - OpenMP: loop-level, dynamic scheduling
- SPEC ENV2002 (wrf) MPI, OpenMP, Hybrid
 - MPI: 2-d nearest neighbor
 - OpenMP: loop-level, static scheduling
- SPEC SEIS2002 (seis) MPI, OpenMP
 - MPI: point-to-point with an all-to-all pattern
 - OpenMP: SPMD at same level as MPI

Measurement Methodology

- Run Serial version on 900MHz V880 (1 cpu)
- Run Parallel versions as described; normalize to 900 Mhz clockrate.
- Divide by the Serial time to get a speedup factor
- In the following three graphs:
 - Y axis is speedup factor as described above
 - X axis is number of CPUs
 - Not all runs were done on all systems
 - Hybrid times are best times for N total cpus
 - See paper for details

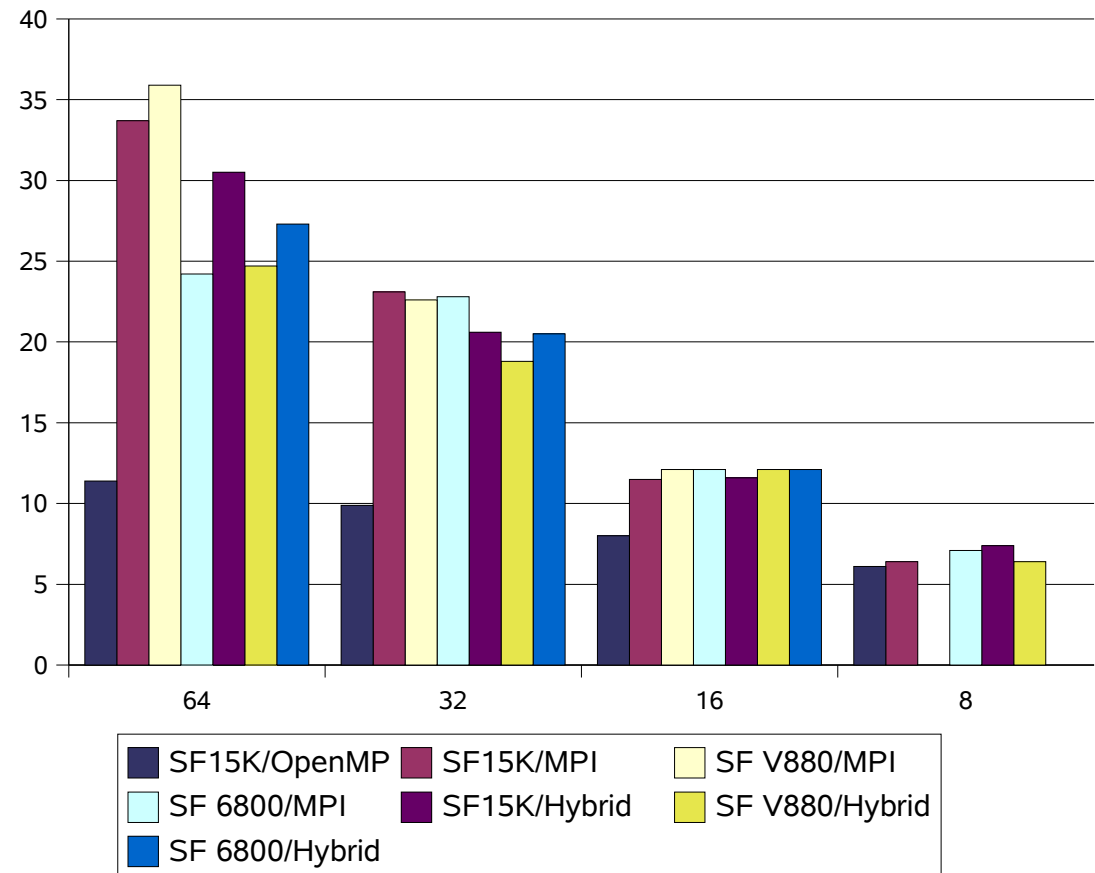
Games Performance

- Best times are with hybrid
 - Nx4 best
 - 2-level parallelism improves load-balancing
- OpenMP doesn't scale
 - Not enough work
- MPI Latency/BW not a major factor
- 15k fastest (Memory BW?)



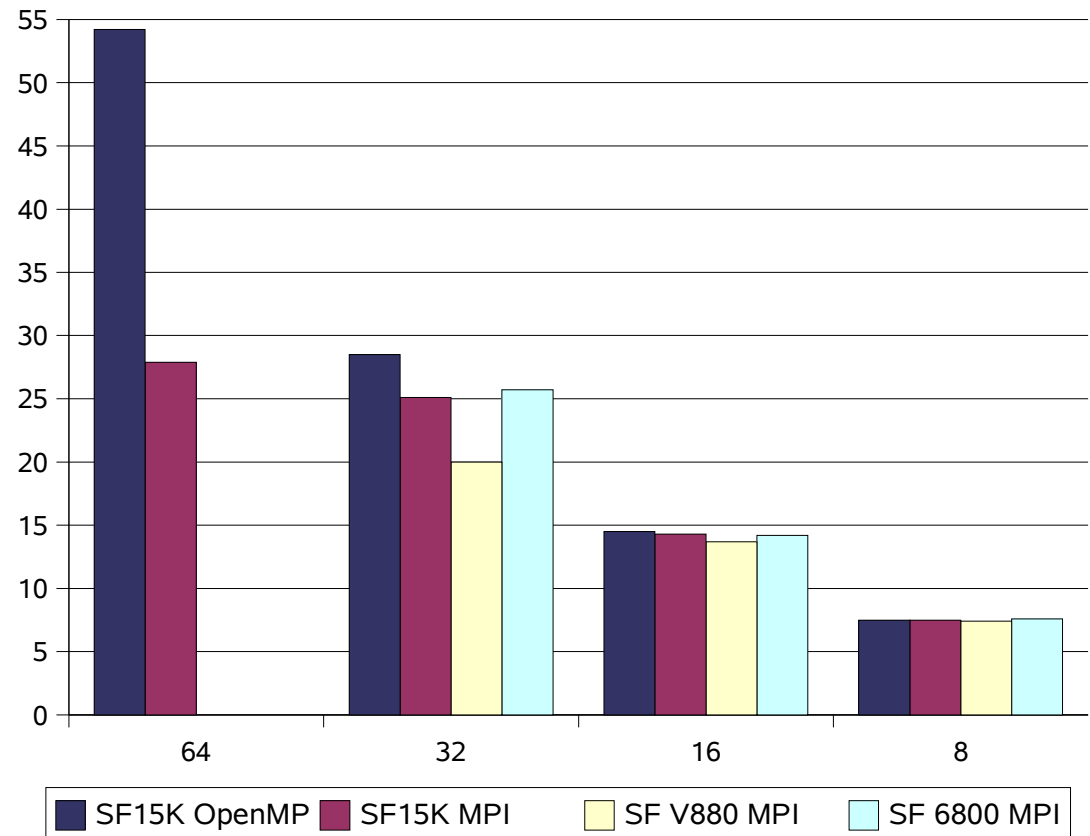
Wrf Performance

- MPI usually best
 - 6800 better in hybrid for 64; binding?
- MPI well balanced, so loop-level doesn't help
- OpenMP doesn't scale



Seis Performance

- OpenMP has the best performance
- MPI Scales poorly past 32
- MPI latency and BW important (see v880 @ 32)
- OpenMP can scale!



Conclusions (1)

- Codes requiring lots of bandwidth do best on SF15K (garness)
 - Nx4 is best – balances bandwidth and latency
 - Extra processor helpful for this code due to master-slave
 - Binding may be helping
- Hybrid codes may perform well, but
 - Assignment of threads to boards is needed
 - Codes with good load balancing don't need hybrid (wrf)

Conclusions (2)

- Codes requiring many messages depend on MPI latency & BW (Seis)
 - OpenMP SPMD style can perform well
 - Lots of small messages shows latency bottleneck
- Best machines

Benchmark/# cpus	64	32	16	8
Gamess	15K/Hybrid	15K/Hybrid	15K/Hybrid	V880/OpenMP
Wrf	V880/MPI	15K/MPI	V880/MPI or 6800/Hybrid	15K/Hybrid
Seis	15K/OpenMP	15K/OpenMP	15K/OpenMP	All



Nawal Copty & Larry Meadows

Nawal.Copty@sun.com

Larry.Meadows@sun.com

