NASA Ames Research Center

Performance Evaluation of Supercomputers using HPCC and IMB Benchmarks

Subhash Saini[1], Robert Ciotti[1], Brian T. N. Gunney[2], Thomas E. Spelce[2], Alice Koniges[2], Don Dossa[2], Panagiotis Adamidis[3], Rolf Rabenseifner[3], Sunil R. Tiyyagura[3], Matthias Mueller[4], and Rod Fatoohi[5]

[1]Advanced Supercomputing Division
NASA Ames Research Center
Moffett Field, California 94035, USA
{ssaini, ciotti}@nas.nasa.gov

[2]Lawrence Livermore National Laboratory
Livermore, California 94550, USA
{gunney, koniges, dossa1}@llnl.gov

[3]High-Performance Computing-Center (HLRS)
Allmandring 30, D-70550 Stuttgart, Germany
University of Stuttgart
{adamidis, rabenseifner, sunil}@hlrs.de

[4]ZIH, TU Dresden. Zellescher Weg 12,
D-01069 Dresden, Germany
matthias.mueller@tu-dresden.de

[5]San Jose State University
One Washington Square
San Jose, California 95192
rfatoohi@sjsu.edu

Abstract

The HPC Challenge (HPCC) benchmark suite and the Intel MPI Benchmark (IMB) are used to compare and evaluate the combined performance of processor, memory subsystem and interconnect fabric of five leading supercomputers - SGI Altix BX2, Cray X1, Cray Opteron Cluster, Dell Xeon cluster, and NEC SX-8. These five systems use five different networks (SGI NUMALINK4, Cray network, Myrinet, InfiniBand, and NEC IXS). The complete set of HPCC benchmarks are run on each of these systems. Additionally, we present Intel MPI Benchmarks (IMB) results to study the performance of 11 MPI communication functions on these systems.

1. Introduction:

Performance of processor, memory subsystem and interconnect is a critical factor in the overall performance of computing system and thus the applications running on it. The HPC Challenge (HPCC) benchmark suite is designed to give a picture of overall supercomputer performance including floating point compute power, memory subsystem performance and global network issues [1,2]. In this paper, we use the HPCC suite as a first comparison of systems. Additionally, the message-passing paradigm has become the de facto standard in programming high-end parallel computers. As a result, the performance of a majority of applications depends on the performance of the MPI functions as implemented on these systems. Simple bandwidth and latency tests are two traditional metrics for assessing the performance of the interconnect fabric of the system. However, simple measures of these two are not adequate to predict the performance for real world applications. For instance, traditional methods highlight the performance of network by latency using zero byte message sizes and peak bandwidth for a very large message sizes ranging from 1 MB to 4 MB for small systems (typically 32 to 64 processors.) Yet, real world applications tend to send messages ranging from 10 KB to 2 MB using not only point-to-point communication but often with a variety of communication patterns including collective and reduction operations.

The recently renamed Intel MPI Benchmarks (IMB, formerly the Pallas MPI Benchmarks) attempt to provide more information than simple tests by including a variety of MPI specific operations [3,4]. In this paper, we have used a subset of these IMB benchmarks that we consider important based on our application workload and report the performance results for the five computing systems. Since the systems tested vary in age and cost, our goal is not to characterize one as "better" than

other, but rather to identify strength and weakness of the underlying hardware and interconnect networks for particular operations.

To meet our goal of testing a variety of architectures, we analyze performance on five specific systems: SGI Altix BX2, Cray X1, Cray Opteron Cluster, Dell Xeon cluster, and NEC SX-8 [5, 12]. These five systems use five different networks (SGI NUMALINK4, Cray network, Myrinet, InfiniBand, and NEC IXS). The complete set of HPC benchmarks are run on each of these systems. Additionally, we present IMB 2.3 benchmark results to study the performance of 11 MPI communication functions for various message sizes. However, in this paper we present results only for the 1 MB message size as average size of the message is about 1 MB in many real world applications.

2. High End Computing Platforms:

In Table 1 is given the system characteristics of these 5 systems. Computing systems we have studied have three types of networks namely, flat-tree, multi-stage crossbar and 4-dimensional hypercube.

Table 1: System characteristics of the five computing platforms.

Platform	Type	CPUs/ node	Clock (GHz)	Peak/node (Gflop/s)	Network	Network Topology	Operating System	Location	Processor Vendor	System Vendor
SGI Altix BX2	Scalar	2	1.6	12.8	NUMALINK 4	Fat-tree	Linux (Suse)	NASA (USA)	Intel	SGI
Cray X1	Vector	4	0.800	12.8	Proprietary	4D-Hypercube	UNICOS	NASA (USA)	Cray	Cray
Cray Opteron Cluster	Scalar	2	2.0	8.0	Myrinet	Flat-tree	Linux (Redhat)	NASA (USA)	AMD	Cray
Dell Xeon Cluster	Scalar	2	3.6	14.4	InfiniBand	Flat-tree	Linux (Redhat)	NCSA (USA)	Intel	Dell
NEC SX-8	Vector	8	2.0	16.0	IXS	Multi-stage Crossbar	Super-UX	HLRS (Germany)	NEC	NEC

3. 0 Benchmark Used

We use HPCC Benchmark [1, 2] and Intel MPI Benchmark Version 2.3 (IMB 2.3) as described below

3.1 HPC Challenge Benchmarks

We have used full HPC Challenge [1,2] Benchmarks on SGI Altix BX2, Cray X1, Cray Opteron Cluster, Dell Cluster and NEC SX-8. HPC Challenge benchmarks are multi-faceted and provide comprehensive insight into the performance of modern high-end computing systems. They are intended to test various attributes that can contribute significantly to understanding the performance of high-end computing systems. These benchmarks stress not only the processors, but also the memory subsystem and system interconnects. They provide a better understanding of an application's performance on the computing systems and are better indicators of how high-end computing systems will perform across a wide spectrum of real-world applications.

3.2 Intel MPI Benchmarks

IMB 2.3 is a successor of PALLAS PAM from Pallas GmbH 2.2 [9]. In September 2003, the HPC division of Pallas merged with Intel Corp. IMB 2.3 suite is very popular among high performance computing community to measure the performance of important MPI functions. Benchmarks are written in ANSI C using message-passing paradigm comprising 10,000 lines of code. The IMB 2.0 version has three parts (a) IMB for MPI-1, (b) MPI-2 one sided communication, and (c) MPI-2 I/O. In standard mode, size of messages can be 0,1, 2, 4, 8, ... 4194304 bytes. There are three classes of benchmarks, namely single transfer, parallel transfer and collective benchmarks.

4.0 Results

In this section we present results of HPC Challenge and IMB benchmarks for five supercomputers.

4.1 HPC Challenge Benchmarks:

4.1.1 Balance of Communication to Computation:

For multi-purpose HPC systems, the balance of processor speed, along with memory, communication, and I/O bandwidth is important. In this section, we analyze the ratio of inter-node communication bandwidth to the computational speed. To characterize the communication bandwidth between SMP nodes, we use the random ring bandwidth, because for a large number of SMP nodes, most MPI processes will communicate with MPI processes on other SMP nodes. This means, with 8 or more SMP nodes, the random ring bandwidth reports the available inter-node communication bandwidth per MPI process. Although

the balance is calculated based on MPI processes, its value should be in principle independent of the programming model, i.e., whether each SMP node is used with several single-threaded MPI processes, or some (or even one process) multi-threaded MPI processes, as long as the number of MPI processes on each SMP node is large enough that they altogether are able to saturate the inter-node network [5]. Fig.1 shows the scaling of the accumulated random ring performance with the computational speed. To compare measurements with different numbers of CPUs and on different architectures, all data is presented based on the computational performance expressed by the Linpack HPL value. The HPCC random ring bandwidth was multiplied by the number of MPI processes. The computational speed is benchmarked with HPL.

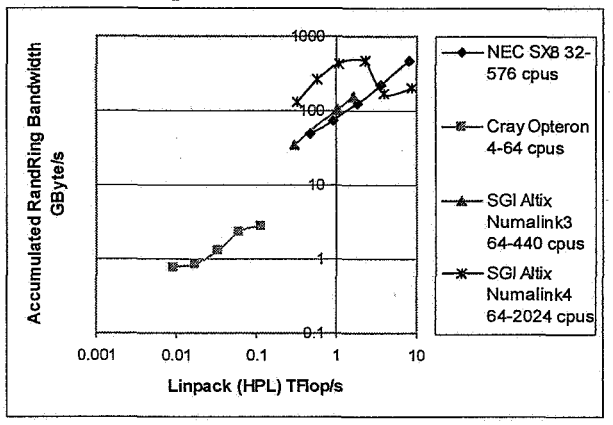

Figure 1: Accumulated Random Ring Bandwidth versus HPL performance.

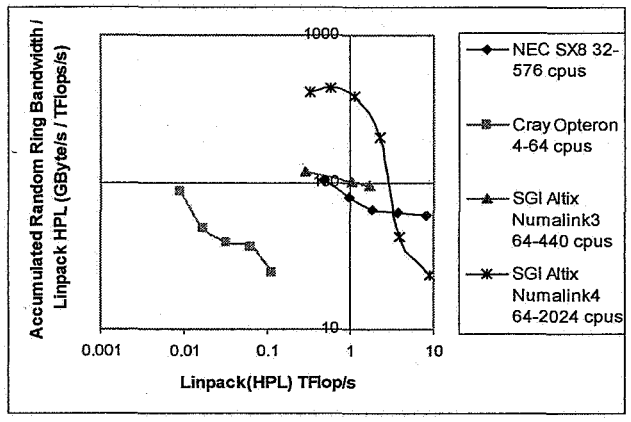

Figure 2: Accumulated Random Ring Bandwidth ratio versus HPL performance.

The diagram in Fig. 1 shows absolute communication bandwidth, whereas the diagram in Fig. 2 plots the ratio of communication to computation speed. Better scaling with the size of the system is expressed by less decreasing of the ratio plotted in Fig. 2. A strong decrease can be observed in the case of Cray Opteron, especially between 32 CPUs and 64 CPUs. NEC SX-8 system scales well which can be noted by only a slight inclination of the curve. In case of SGI Altix, it is worth noting the difference in the ratio between Numalink3 and Numalink4 interconnects within the same box (512 CPUs). Though the theoretical peak bandwidth between Numalink3 and Numalink4 has only doubled, Random Ring performance improves by a factor of 4 for runs up to 256 processors. A steep decrease in the B/KFlop value for SGI Altix with Numalink4 is observed above 512 CPUS runs (203.12 B/KFlop for 506 CPUs to 23.18 B/KFlop for 2024 CPUs). This can also be noticed from the cross over of the ratio curves between Altix and the NEC SX-8. Whereas with Numalink3 it is 93.81 (440 CPUs) when run within the same box. For the NEC SX-8, B/Kflop is 59.64 (576 CPUs), which is consistent between 128 and 576 CPUs runs. For the Cray Opteron it is 24.41 (64 CPUs).

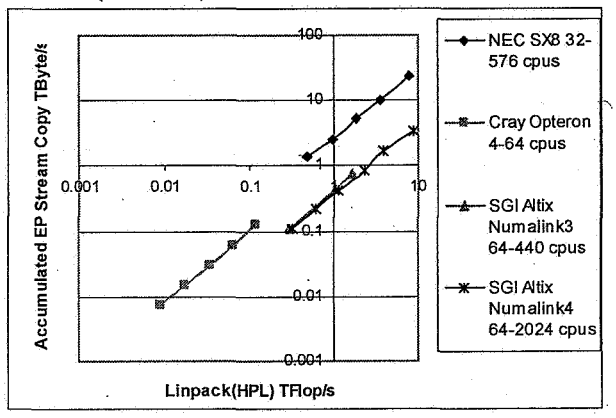

Figure 3: Accumulated EP Stream Copy versus HPL performance.

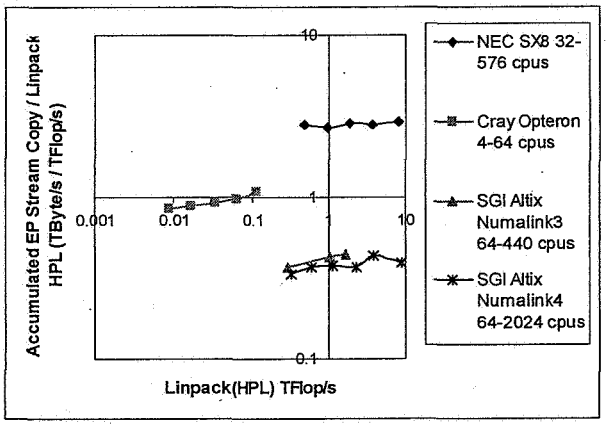

Figure 4: Accumulated EP Stream Copy ratio versus HPL performance.

Fig. 3 and Fig. 4 compare the memory bandwidth with the computational speed analog to Fig. 1 and Fig. 2 respectively. Fig. 3 shows absolute values whereas Fig. 4 plots the ratio of STREAM Copy to HPL on the vertical

axis. The accumulated memory bandwidth is calculated as the product of the number of MPI processes with the embarrassingly parallel STREAM Copy result. In Fig. 4, as the number of processors increase, the slight improvement in the ratio curves is due to the fact that the HPL efficiency decreases. In the case of CRAY Opteron HPL efficiency decreases down around 20% between 4 CPU and 64 CPU runs. The high memory bandwidth available on the NEC SX-8 can clearly be seen with the stream benchmark. The Byte/Flop for NEC SX-8 is consistently above 2.67 Byte/Flop, for SGI Altix (Numalink3 and Numalink4) it is above 0.36 and for the Cray Opteron is between 0.84 and 1.07. The performance of memory intensive applications heavily depends on this value.

4.1.2 Ratio based analysis of all benchmark

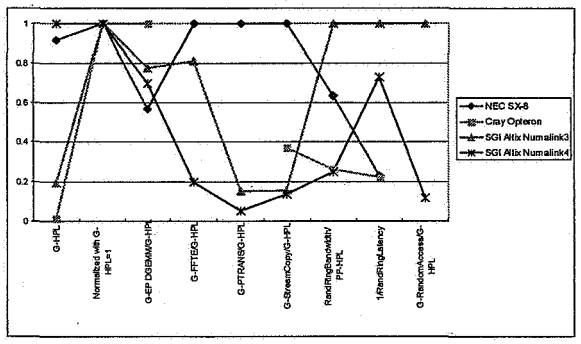

Figure 5: Comparison of all the benchmarks normalized with HPL value.

It should be noted that Random Access benchmark between HPCC versions 0.8 and 1.0 has been significantly modified. Only values based on HPCC version 1.0 are shown

Ratio	Maximum value
G-HPL	8.729 TF/s
G-EP DGEMM/G-HPL	1.925
G-FFTE/G-HPL	0.020
G-Ptrans/G-HPL	0.039 B/F
G-StreamCopy/G-HPL	2.893 B/F
RandRingBW/PP-HPL	0.094 B/F
1/RandRingLatency	0.197 1/μs
G-RandomAccess/G-HPL	4.9×10^{-5} Update/F

Table 2: Ratio values corresponding to Figure 1 in Figure 5.

Fig. 5 compares the systems based on several HPCC benchmarks. This analysis is similar to the current Kiviat diagram analysis on the HPCC web page [16], but it uses always parallel or embarrassingly parallel benchmark results instead of single process results, and it uses only accumulated global system values instead of per process values. Absolute HPL numbers cannot be taken as a basis for comparing the balance of systems with different total system performance. Therefore all benchmark results are normalized with the HPL system performance, i.e., divided by the HPL value. Furthermore, each of the columns is normalized with respect to the largest value of the column, i.e., the best value is always 1. Only the left column can be used to compare the absolute performance of the systems. This normalization is also indicated by normalized HPL value in Fig. 5 (column 2) which is by definition always a value of 1. For latency, the reciprocal value is shown. The corresponding absolute ratio values for 1 in Fig. 5 are provided in Table 2.

One can see from Fig. 5 that the Cray Opteron performs best in EP DGEMM because of its lower HPL efficiency when compared to the other systems. When looking at the global measurement based ratio values such as FFTE, Ptrans and RandomAccess, the small systems have an undue advantage over the larger ones because of better scaling. For this reason, the global ratios of systems with over 1 TFlop/s HPL performance are plotted. The NEC SX-8 performs better in those benchmarks where high memory bandwidth coupled with network performance is needed (Ptrans, FFTE and EP Stream Copy). On the other hand the NEC SX-8 has relatively high Random Ring latency compared to the other systems. SGI Altix with Numalink3 has better performance in Random Ring bandwidth and latency benchmarks (Numalink4 performs much better than Numalink3 within the same box). This shows the strength of its network within a box. Despite this fact the Cray Opteron performs better in RandomAccess which is heavily dependent on the network performance.

4.2 IMB Benchmarks:

On the NEC SX-8 system, memory allocation was done with MPI_Alloc_mem, which allocates global memory. The MPI library on the NEC SX-8 is optimized for global memory.

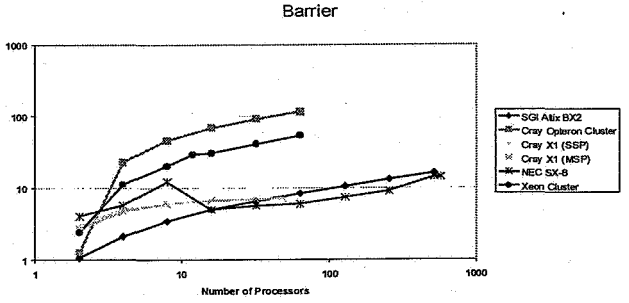

Figure 6: Execution time of Barrier benchmark on five systems in μs/call (i.e., the smaller the better).

Fig. 6 shows the performance of the Barrier benchmark from the IMB suite of benchmarks. Here we have plotted the time (in microseconds per call) for various number of processors ranging from 2 to 512 (568 on the NEC SX-8).

A barrier function is used to synchronize all processes. A process calling this function blocks until all the processes in the communicator group have called this function. This ensures that each process waits till all the other processes reach this point before proceeding further. Here, all the five computing platforms exhibit the same behavior up to 64 processors i.e. barrier time increases gradually with the increase of number of processors, except for the Cray X1 in MSP mode where barrier time increases very slowly. On NEC SX-8, the barrier time is measured using the full communicator. Varying processor count, as provided in the IMB benchmark is not used while running the barrier benchmark. In this way subset communicators are avoided and each test is done with its own full communicator (MPI_COMM_WORLD). With these runs for large CPU counts, NEC SX-8 has the best barrier time compared to other systems. For less than 16 processor runs, SGI Altix BX2 is the fastest.

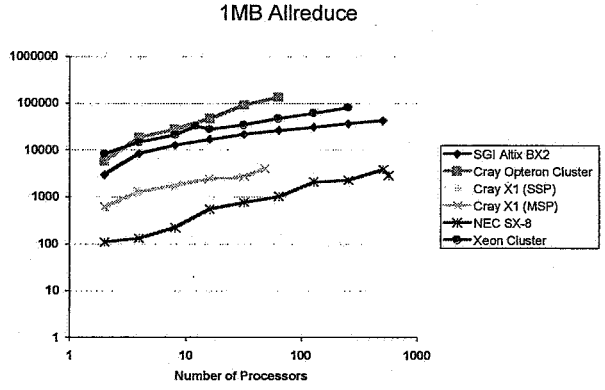

Figure 7: Execution time of Allreduce benchmark for 1 MB message for five computing systems in µs/call (i.e., the smaller the better).

The execution time of the Allreduce benchmark for 1 MB message size is shown in Fig. 7. All five systems scale similarly when compared to their performance on 2 processors. There is more than one order of magnitude difference between the fastest and slowest platforms. All the architectures exhibit the same behavior as the number of processors increase. Both vector systems are clearly the winner, with NEC SX-8 superior to Cray X1 in both MSP and SSP mode. Up to 16 processors, both Cray Opteron cluster and Dell Xeon cluster follow the same trend as well with almost identical performance. Here best performance is that of NEC SX-8 and worst performance is that of Cray Opteron cluster (uses Myrinet network). Performance of Altix BX2 (NUMALINK4 network) is better than Dell Xeon cluster (InfiniBand network).

Execution time of IMB Reduction benchmark for 1 MB message size on all five computing platforms is shown in Fig. 8. Here we see two clear cut performance clustering by architectures – vector systems (NEC SX-8 and Cray X1) and cache based scalar systems (SGI Altix BX2, Dell Xeon Cluster, and Cray Opteron Cluster). Performance of vector systems is an order of magnitude better than scalar systems. Between vector systems, performance of NEC SX-8 is better than that of Cray X1. Among scalar systems, performance of SGI Altix BX2 and Dell Xeon Cluster is almost the same and better than Cray Opteron Cluster.

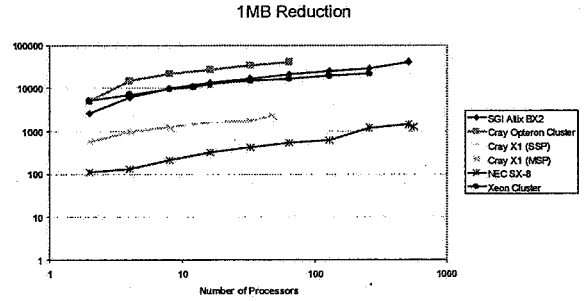

Figure 8: Execution time of Reduction benchmark on varying number of processors, using a message size of 1MB, in µs/call (i.e., the smaller the better).

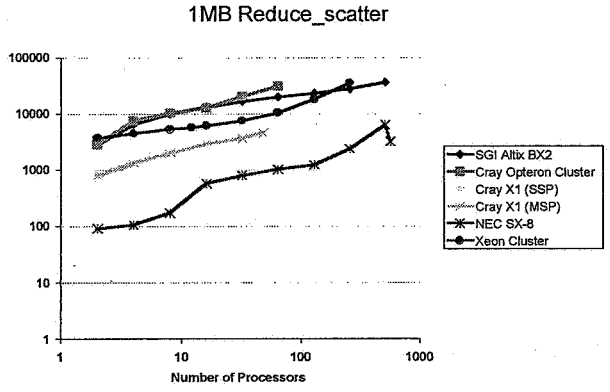

Figure 9: Execution time of Reduce_scatter benchmark on varying number of processors, using a message size of 1MB, in µs/call (i.e., the smaller the better).

Execution time of IMB Reduce Scatter benchmark for 1 MB message size on five computing platforms is shown in Figure 9. The results are similar to the results of Reduce benchmark, except that the performance advantage of Cray X1 compared to the scalar systems

is significantly worse. For large CPUs counts, NEC SX-8 shows slower results, but still better compared to the other platforms. Timings for scalar systems are an order of magnitude slower than that of NEC SX-8, a vector system.

Fig. 10 shows the execution time of IMB Allgather benchmark for 1 MB message size on five computing platforms.

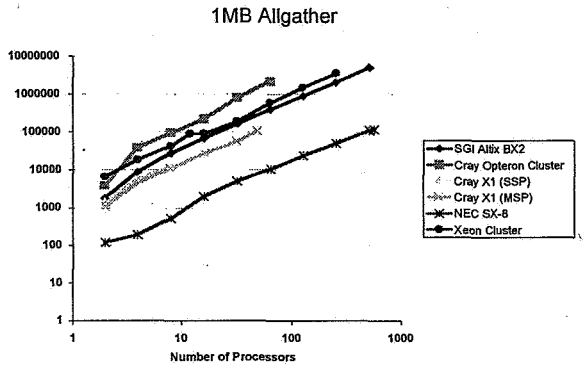

Figure 10: Execution time of Allgather benchmark on varying number of processors, using a message size of 1MB, in µs/call (i.e., the smaller the better).

Performance of vector system NEC SX-8 is much better than that of scalar systems (Altix BX2, Xeon Cluster and Cray Opteron Cluster). Cray X1 (both SSP and MSP modes) performs slightly better than the scalar systems. Between two vector systems, performance of NEC SX-8 is an order of magnitude better than Cray X1. Among the three scalar systems, performance of Altix BX2 and Dell Xeon Cluster is almost the same and is better than Cray Opteron Cluster.

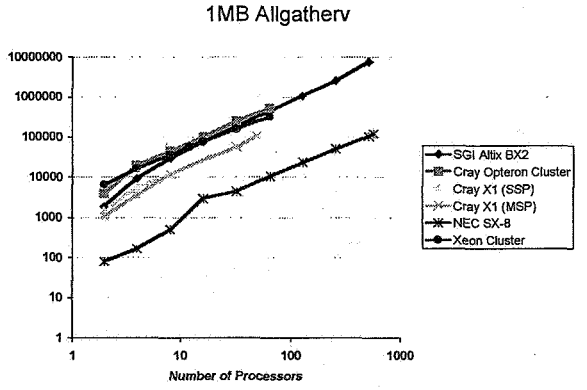

Figure 11: Execution time of Allgatherv benchmark on varying number of processors, using a message size of 1MB, in µs/call (i.e., the smaller the better).

Results shown in Figure 11 are the same as in Fig. 10, except that a version of the Allgatherv with variable message sizes was used. The performance results are similar to the results of the (symmetric) Allgather. On the NEC SX-8, the performance increase between 8 and 16 processors is based on the changeover from a single shared memory node to a multi SMP node execution. Performance of all scalar systems is almost same. Between two vector systems, the performance of NEC SX-8 is almost an order of magnitude better than Cray X1.

Fig. 12 shows the execution time of AlltoAll benchmark for a message size of 1 MB on five computing architectures. This benchmark stresses the global network bandwidth of the computing system. Performance of this benchmark is very close to the performance of global FFT and randomly ordered ring bandwidth benchmarks in the HPCC suite [12]. Clearly, NEC SX-8 out performs all other systems. Performance of Cray X1 (both SSP and MSP modes) and SGI Altix BX2 is very close. However, the performance of SGI Altix BX2 up to eight processors is better than Cray X1 as the SGI Altix BX2 (uses NUMAlink4 network) has eight Intel Itanium 2 processors in a C-Brick. Performance of Dell Xeon Cluster (uses IB network) and Cray Opteron Cluster (uses Myrinet PCI-X network) is almost same up to 8 processors, after which performance of Dell Xeon cluster is better than Cray Opteron Cluster. Performance results presented in Fig. 11 show NEC SX-8 (IXS) > Cray X1 (Cray proprietary) > SGI Altix BX2 (NUMALINK4) > Dell Xeon Cluster (Infiniband network) > Cray Opteron Cluster (Myrinet network). It is interesting to note that performance is directly proportional to the randomly ordered ring bandwidth, which is related with the cost of the global network.

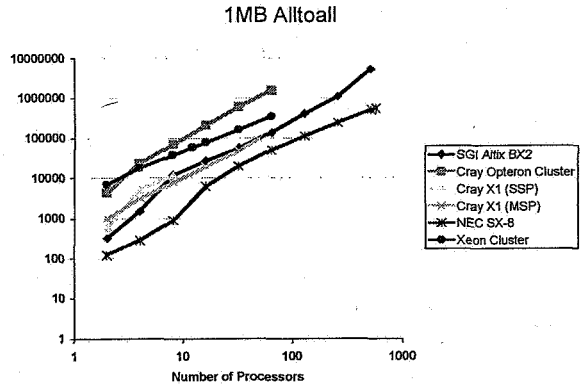

Figure 12: Execution time of AlltoAll benchmark on varying number of processors, using a message size of 1MB, in µs/call (i.e., the smaller the better).

Fig. 13 presents the bandwidth of IMB Sendrecv benchmark using 1 MB message. Clearly, performance of NEC SX-8 is the best followed by SGI Altix BX2.

Performance of Xeon cluster and Cray Opteron is almost the same. After 16 processors, the performance of all the computing system becomes almost constant. For all platforms, systems perform the best when running 2 processors. This is expected for BX2, Opteron and Xeon because all of them are dual processor nodes and also for NEC SX-8 with its 8-way SMP nodes. Therefore this Sendrecv is done using shared memory and not over the network. Here, it would be interesting to note that on the NEC SX-8 with 64 GB/s peak memory bandwidth per processor, the IMB Sendreceive bandwidth for 2 processors is 47.4 GB/s. Whereas for the Cray X1 (SSP), IMB Sendreceive bandwidth is only 7.6 GB/s.

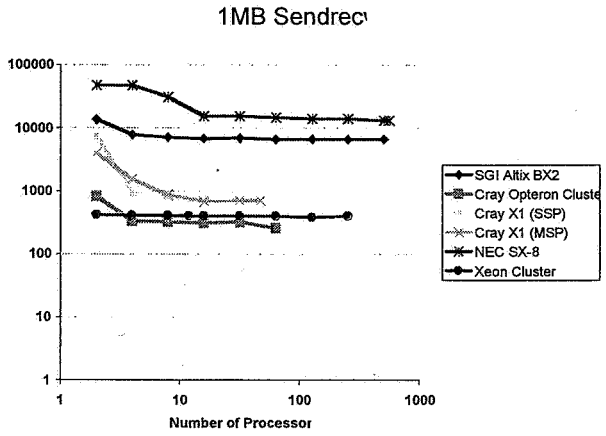

Figure 13: Bandwidth of Sendrecv benchmark on varying number of processors, using a message size of 1MB, in MB/s.

Fig. 14 shows the performance of the IMB Exchange benchmark for 1 MB message size. The NEC SX-8 is the winner but its lead over the Xeon cluster has decreased compared to the Sendrecv benchmark. The second best system is the Xeon Cluster and its performance is almost constant from 2 to 512 processors, i.e., compared to Sendrecv, the shared memory gain on 2 CPUs is lost. For a number of processors greater than or equal to 4, the performance of the Cray X1 (both SSP and MSP modes) and the Altix BX2 is almost same. For two processors, the performance of the Cray Opteron cluster is close to the BX2, and the performance of Cray Opteron cluster is the lowest.

In Fig. 15, we plot the time (in micro seconds) for various numbers of processors for 1 MB broadcast on the five computing platforms. Up to 64 processors, the broadcast time increases gradually and this trend is exhibited up to 64 processors by all computing platforms. Only 512 processor results are available for SGI Altix BX2 and NEC SX-8. For the BX2, broadcast time suddenly increases for 256 processors and then again decreases at 512 processors.

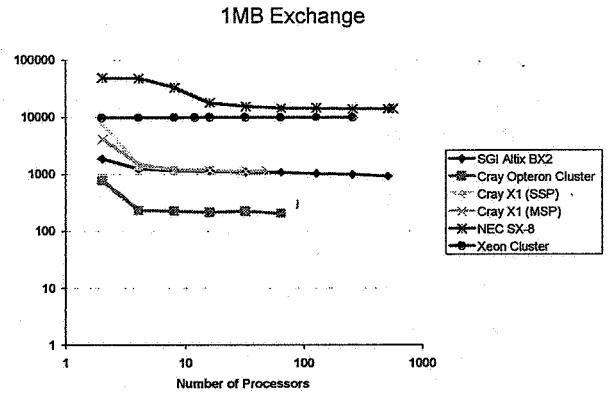

Figure 14: Bandwidth of Exchange benchmark on varying number of processors, using a message size of 1MB, in MB/s.

A similar but quite smaller behavior is seen for NEC SX-8 – increases for broadcast time up to 512 CPUs and then a decrease at 576 processors. The best systems with respect to broadcast time in decreasing order are NEC SX-8, SGI Altix BX2, Cray X1, Xeon cluster and Cray Opteron cluster. The broadcast bandwidth of NEC SX-8 is more than an order of magnitude higher than that of all other presented systems.

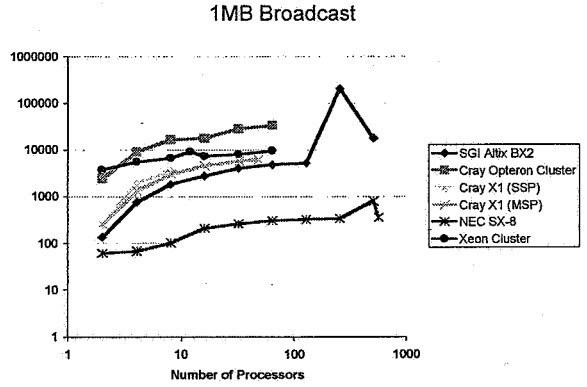

Figure 15: Execution time of Broadcast benchmark on varying number of processors, using a message size of 1MB, in μs/call (i.e., the smaller the better).

5. Conclusions

We present the results of HPCC and IMB benchmarks separaetely.

5.1 HPCC Benchmark Suite

The HPCC benchmark suite highlights the importance of memory bandwidth and network performance along with HPL performance. The growing difference between

the peak and sustained performance underlines the importance of such benchmark suites. A good balance of all the above quantities should make a system perform well on a variety of application codes. In this paper, we use the benchmark analysis to see the strengths and weaknesses of the architectures considered. The ratio based analysis introduced in this paper provides a good base to compare different systems and their interconnects.

It is clear from the analysis that NEC SX-8 performs extremely well on benchmarks that stress the memory and network capabilities, like Global PTRANS and Global FFTs (G-FFT). It is worth mentioning that the Global FFT benchmark in the HPCC suite does not completely vectorize, hence on vector machines (like Cray X1 and NEC SX-8) the performance of FFTs using vendor provided optimized libraries would be much higher. The interconnect latency of SGI Altix BX2 is the best among all the platforms tested. However, a strong decrease in the sustained interconnect bandwidth is noticed when using multiple SGI Altix BX2 boxes. On SGI Altix BX2, G-FFT does not perform well beyond one box (512 CPUs) and this degradation in performance is also reflected by a decrease in the random order bandwidth benchmark of the HPCC suite. G-FFT involves all-to-all communication and therefore for it to perform well it must have very good performance on the IMB benchmark All-to-All. G-FFT is not expected to perform well on Cray X1 and NEC SX-8 as the G-FFT benchmark in the HPCC suite is not vectorized.

The scalability and performance of small machines (Cray Opteron and Cray X1) cannot be compared to that of larger machines as the complexity and cost of the interconnect grows more than linearly with the size of the machine.

5.2 IMB Benchmark Suite:

Performance of both the vector systems (NEC SX-8 and Cray X1) is consistently better than all the scalar systems (SGI Altix BX2, Cray Opteron Cluster and Dell Xeon Cluster). Between two vector systems, performance of NEC SX-8 is consistently better than Cray X1. Among scalar systems, the performance of SGI Altix BX2 is better than both Dell Xeon Cluster and Cray Opteron Cluster. We find that the performance of IXS (NEC SX-8) > Cray X1 network > SGI Altix BX2 (NUMAlink4) > Dell Xeon Cluster (Infiniband) > Cray Opteron Cluster (Myrinet).

In the future we plan to use IMB benchmark suite to study the performance as a function of varying message sizes starting from 1 byte to 4 MB for all 11 benchmarks on the same five computing systems. We also plan to include three more architectures – IBM Blue Gene, Cray XT3 and a cluster of IBM POWER5.

References

1. Luszczek, P., Dongarra, J., Koester, D., Rabenseifner, R., Lucas, B., Kepner, J., McCalpin, J., Bailey, D., Takahashi, D. "Introduction to the HPC Challenge Benchmark Suite," March, 2005. http://icl.cs.utk.edu/hpcc/pubs.
2. Rabenseifner, Rolf, Tiyyagura, Sunil, and Mueller, Matthias, Network Bandwidth Measurements and Ratio Analysis with the HPC Challenge Benchmark Suite (HPCC), EuroPVM/MPI'05, 2005.
3. http://www.intel.com/cd/software/products/asmo-na/eng/cluster/mpi/219847.htm
4. Karl Solchenbach, Benchmarking the Balance of Parallel Computers, SPEC Workshop on Benchmarking Parallel and High-Performance Computing Systems, Wuppertal, Germany, Sept. 13, 1999.
5. Saini, S., Full Day Tutorial M04, Hot Chips and Hot Interconnect for High End Computing Systems, *IEEE Supercomputing 2004*, Nov. 8, 2004, Pittsburgh.
6. http://www.ncsa.uiuc.edu/UserInfo/Resources/Hardware/XeonCluster/
7. Infiniband Trade Association, Infiniband Architecture Specifications, Release 1.0 October 24, 2000.
8. http://www.intel.com/technology/infiniband/
9. Quadrics, Quadrics Ltd. http://www.quadrics.com.
10. Myricom. Myricom Inc. http://www.myri.com.
11. J. Liu et al., Performance Comparaison of MPI Implementations over Infiniband Myrinet and Quadrics, *http://www.sc-conference.org/sc2003/paperpdfs/pap310.pdf*, SC 2003, Phoenix, Arizona, Nov. 2003.
12. Koniges, A., M. Seager, D. Eder, R. Rabenseifner, M. Resch, *Application Supercomputing on Scalable Architectures, IEEE Supercomputing 2004*, Nov. 8. 2004, Pittsburgh.
13. Intel MPI Benchmarks: Users Guide and Methodology Description, Intel GmbH, Germany, 2004.