Redbooks Paper

Ruzhu Chen

Performance Modeling and Characterization (PMaC) Benchmarking on POWER4+ Platforms (II)

Problem

Performance evaluation of parallel applications on an HPC machine is complicated. However, modeling the main factors of single processor performance and communication performance among processors is usually sufficient for application performance evaluation (for more details about this subject, refer to “Modeling application performance by convolving machine signatures with application profiles” by A. Snavely, et al).

This paper describes the performance measurements of our SMP POWER4™ machines p690 and p655 using the PMaC benchmark suite.

Solution

To model the single processor performance of an application, the following are measured: bandwidth and latency of data storing and loading into the memory hierarchy (main memory, caches), and the performance of total floating points and memory operations. Communication performance is evaluated by measuring the communication bandwidth and latency on and off nodes, and the percentage of communication required to distribute and gather data in a parallel application.

Description of benchmark system and tests

The PMaC HPC Benchmark Suite is a set of orthogonal benchmarks used to enable performance evaluation of essential HPC hardware and system features such as I/O, CPU, Network, and Memory (for more detailed information, refer to “A framework for application performance prediction to enable scalability understanding” by L. Carrington, et al).

Because total floating points and memory operations of an application are the same, the comparison of performance of different systems can be decided mainly by machine profiles (the load/store bandwidths of a single processor, and communication bandwidths between processors on and/or off nodes). The Benchmark Suite is derived from the synthetic

benchmarks of DoD HPCMO and Pallas MPI Benchmarks (PMB benchmarks). The suite consists of six benchmarks:

- MAPS or memory access pattern signature (to test memory load and store bandwidth)
- MAPS-CG
- MAPS-PING (to test MPI communications)
- EFF-BW (to test effective bandwidth)
- PEAK (to test floating point and memory operations)
- I/O Bench (to test I/O)

The purpose of PMaC benchmark testing is to meaningfully compare and predict machine performance on certain applications. The benchmarks are described in detail at the PMaC Web site:

http://www.sdsc.edu/PMaC

System configuration

The IBM® POWER4+ platforms p690+ and p655+ were tested for performance on the PMaC benchmarks; the details of the configurations used in this benchmarking suite are listed in Table 1.

Table 1 System and hardware configurations

Configuration		P690+	P655+
Processor		1.7 GHz POWER4+	1.5GHz POWER4+
Processors/node		32	8
Memory/node		128 GB (8-card)	16 GB (2-card)
Mem (GB)/processor		4	2
Caches	L1	64/32 KB (1-/2-way)	64/32 KB (1-/2-way)
	L2	1.5 MB/card (4-way)	1.5 MB/card (4-way)
	L3	128 MB	128 MB
OS		AIX® 5.1.0.0	AIX 5.1.0.0
AIX Kernel		64-bit	64-bit
File System(s)		Local or gpfs	Local or gpfs
FORTRAN compiler		XLF 8.1	XLF 8.1
C/C++ compiler		VAC 6.0	VAC 6.0

Note: Two processors per card

Results

In the following sections, we present the results of all six benchmarks.

MAPS

To obtain the memory performance of a single processor, a MAPS (memory access pattern signature) benchmark was executed to collect memory access information (bandwidth) for all levels of memory, using various-sized working sets and memory access patterns (stride one and random access).

Thus, the MAPS benchmark results give sustainable rates of memory load or store, depending on the access pattern and the size of the problem.

Example 1 MAPS benchmark

```
Version:[x] generic

Date test run(MM/DD/YY): 08/26/2003
File System(NFS,GPFS,local,etc): local file system
Compiler:
    Version: XLF 8.1
    Flags:  -qarch=pwr4 -qtune=pwr4 -qcache=auto -qalign=4k -O3 -lmass
-lmassvp4 bmaxdata:0x70000000
```

Figure 1 and Figure 2 on page 4 show the memory bandwidth versus size for load and store, displaying a three-stair pattern on all three systems. The memory bandwidth is highly influenced by the memory access pattern. Problems with a random access pattern ran much slower than those accessing memory sequentially (stride one). Small problems yielded significant cache reuse

To view the output results, refer to “MAPS data” on page 8.

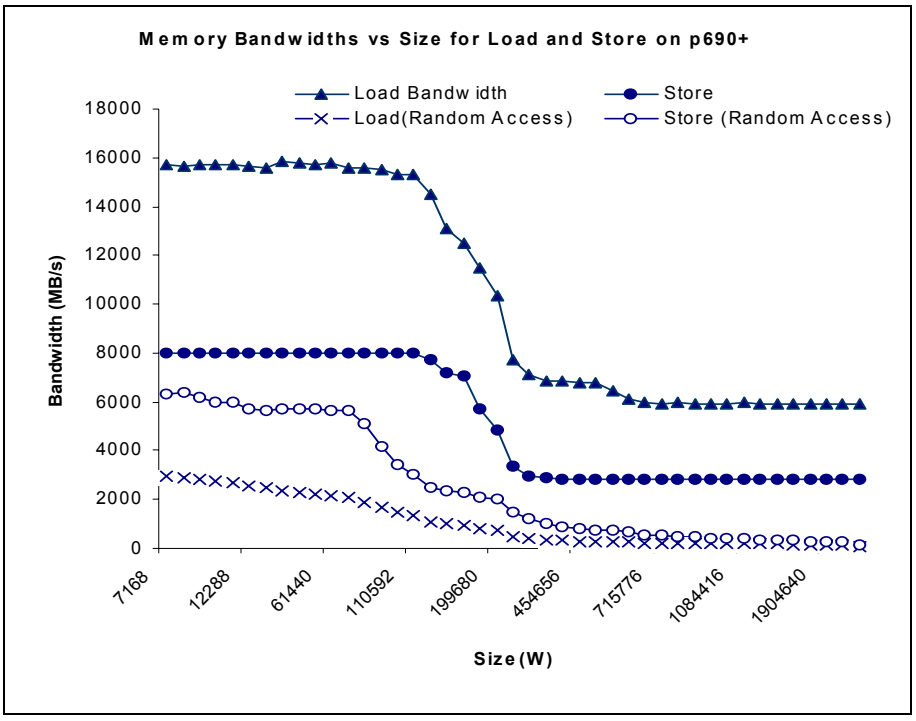

Figure 1 Memory bandwidths vs size for Load and Store on p690+

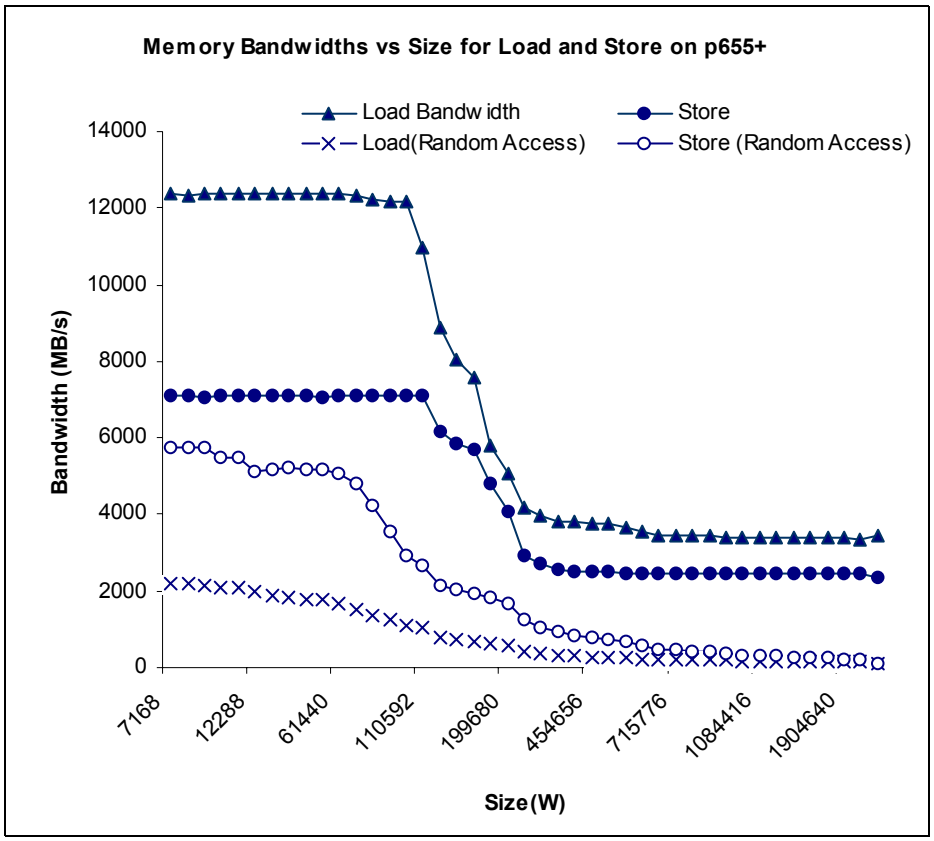

Figure 2 Memory bandwidths vs size for Load and Store on p655+

MAPS-CG

Using MPI all-to-all blocking sends and receives, the MAPS-CG benchmark tests communication bandwidths and latency among processors.

Example 2 MAPS-CG benchmark

```
Date test run(MM/DD/YY): 08/26/2003
File System (NFS,GPFS,local,etc)   : local
Compiler:
  Version: XLF 8.1 (mpxlf)
  Flags: -O3 -qtune=pwr4 -qarch=pwr4 -qcache=auto
         -bmaxdata:0x70000000
```

Table 2 lists the benchmark results.

Table 2 MAPS-CG benchmark results

Platform	Number of nodes	CPUs/Node	Bandwidth (MB/sec)/CPU	Latency (µsec)
p690+	1	4	825.05	3.80
	1	16	436.9	4.50
	1	32	310.73	4.90
p655+	1	4	676.48	3.50
	4	4	665.68	3.90
	4	8	393.60	4.20

MAPS-PING

The MAPS-PING benchmark was used to evaluate the performance of communication between processors. The test measures the bandwidths and latency between two processors on a node or cross node.

Example 3 MAPS-PING benchmark

```
Date test run(MM/DD/YY): 08/26/2003
File System (NFS,GPFS,local,etc)   : local
Compiler:
  Version: XLF 8.1 (mpxlf)
  Flags: -O3 -qtune=pwr4 -qarch=pwr4 -qcache=auto -bmaxdata:0x70000000
        (-lMPIbind # for cpu bind)
```

The resulting data on two processors, either randomly chosen or close together in one chip, is listed in Table 3.

Table 3 MAPS-PING benchmark results

Platform	On node		Cross node	
	Bandwidth (MB/s) per CPU	Latency (µsec)	Bandwidth (MB/s) per CPU	Latency (µsec)
p690+	2470.62 *	2.45		
	2329.04	2.43		
p655+	2085.51*	2.16	432.86	16.8
	1999.20	2.29	432.86	16.6

Note: *Between logical CPUs close to each other

EFF-BW

EFF-BW evaluates the communication performance of all the participating processors by cumulating results for small and large messages.

Example 4 EFF-BW benchmark

```
Date test run (MM/DD/YY): 08/26/2003
File System (NFS,GPFS,local,etc)   : GPFS
Compiler:
Version: xlc 6.0
Flags: -qarch=auto -qtune=auto -qcache=auto
```

The results are listed in Table 4.

Table 4 Effective bandwidths

Platform	Number of nodes	PEs/Node	EFF_BW (MB/sec)	EFF_BW/pe (MB/sec)	Latency (µsec) (simple mapping)	Latency (µsec) (cross-mapping)
P690+	1	2	888.33	444.17	2.32	2.35
	1	4	1692.58	423.15	2.93	3.33
	1	8	3235.91	404.48	2.82	3.19
	1	16	5753.89	359.62	2.86	3.30
	1	24	6728.94	280.37	2.7	3.16
	1	32	7684.99	240.16	2.76	3.30
P655+	1	2	1006.42	503.21	1.99	2.00
	1	4	1776.34	444.09	2.32	2.18
	1	8	3060.73	382.59	2.16	2.37

PEAK

The PEAK benchmark measures the performance of floating and memory operations on a single processor. Four loops (division, DAXPY, DOT and 5th degree polynomial), which represent computation-intensive-to-memory-intensive applications, were executed.

Example 5 PEAK benchmark

```
Date test run (MM/DD/YY): 08/26/2003
File System (NFS,GPFS,local,etc)   : local
Compiler:
  Version: XLF 8.1 (xlf)
  Flags: -O4 -qessl -qarch=pwr4 -qcache=auto -qtune=pwr4
         -lmass -lmassvp4
  Version: VAC 6.0 (xlc)
  Flags: -O4 -lmass -lmassvp4
```

Table 5 PEAK performance results

Platform	Loop length	Division (MFlops)	DAXPY (MFlops)	DOT (MFlops)	5th degree Poly (MFlops)
p690+	1024	240.94	2007.84	2533.60	3840.00
	4096	234.05	1585.54	1831.75	3860.10
	16384	236.87	1598.43	1843.20	3855.05
	65536	231.30	1155.51	1465.40	3618.55
	262144	137.97	478.07	767.25	2711.83
	524288	129.45	448.10	676.50	2496.60
	1046576	129.02	448.53	676.82	2522.99
	4194304	128.39	436.90	645.27	2419.79
p655+	1024	211.13	1780.86	2174.86	3413.33
	4096	203.10	1412.41	1594.11	3455.23
	16384	206.95	1414.44	1615.95	3409.38
	65536	196.60	987.97	1150.87	3223.08
	262144	116.50	420.55	517.10	2194.69
	524288	112.75	391.25	471.27	2055.03
	1046576	113.48	390.83	469.00	2062.59
	4194304	110.37	381.30	457.56	2029.50

IO-Bench

IO-Bench tests a machine's I/O performance through different read/write patterns.

Example 6 IO-Bench

```
Date test run (MM/DD/YY): 08/26/2003
  File System (NFS,GPFS,local,etc)   : local
  Compiler:
    Version: xlc 6.0
    Flags: -O4 -q64 -qarch=pwr4 -qtune=pwr4 -qcache=auto
-bmaxdata:0x800000000 -bmaxstack:0x800000000
```

The results are listed in Table 6 on page 7.

Table 6 I/O benchmark test results

Platform	Test	File size (MB)	Buffer size (KB)	Bandwidth (MB/sec)		
				Max	Min	Avg
p690+	Sequential Write	10	4	333.33	200.00	266.67
	Sequential Read	10	4	1000.00	1000.00	1000.00
	Random Read Rewrite	10	4	333.33	333.33	333.33
	Random Read	10	4	1000.00	1000.00	1000.00
	Write Backwards	10	4	500.00	500.00	500.00
	Backward Read	10	4	200.00	200.00	200.00
p655+	Sequential Write	100	4	294.1	204.1	249.1
	Sequential Read	100	4	769.2	714.2	741.2
	Random Read Rewrite	100	4	250.0	250.0	250.0
	Random Read	100	4	588.2	588.2	588.2
	Write Backwards	100	4	256.4	356.4	256.4
	Backward Read	100	4	185.2	185.2	185.2

Summary

In this paper, the benchmark performance on the POWER4+ platforms p690+ and p655+ was evaluated using the PMaC Benchmark Suite, which was previously used for three POWER4 platforms. The results on memory bandwidths and latency for load and store, I/O performance, communication bandwidths and MFLOPS for loop applications were summarized.

MAPS data

Table 7 P690+ Load and Store Bandwidths (MBytes/second)

Size	Load (Stride one)	Load (Random)	Store (Stride one)	Store (Random)
7168	15747.9	2955.46	7984.18	6328.11
7424	15661.3	2909.7	7978.36	6354.44
8192	15734.2	2844.56	7981.99	6198.8
8704	15685.1	2764.14	7989.81	5994.08
9216	15713.8	2718.14	7990.2	5967.07
12288	15680.3	2560.08	7993.87	5739.21
15360	15595.9	2454.25	8001.86	5672.85
21504	15879	2345.63	7999.19	5693.57
33792	15751.5	2250.04	7992.34	5728.8
43008	15710	2202.82	7992.41	5702.73
61440	15751.3	2142.08	7990.95	5654.42
70656	15599.1	2065.86	7988.29	5654.92
79872	15558	1867.72	7987.77	5104.86
89088	15520	1654.22	7988.86	4167.69
101376	15329.8	1468.39	7989.9	3392.15
110592	15326	1350.03	7988.73	3037.11
138240	14477.7	1098.79	7737.39	2461.91
150528	13085.5	993.078	7203.1	2327.66
156672	12480.3	938.244	7020.1	2255.07
178176	11491.8	825.769	5732.81	2099.43
199680	10332	713.034	4845.56	1985
254976	7708.5	491.766	3347.99	1487.07
297984	7150.51	414.724	2961.01	1179.23
350208	6865.21	358.805	2856.58	982.432
402432	6820.2	323.205	2829.11	873.557
454656	6792.99	299.253	2817.58	801.374
506880	6774.58	281.604	2814.07	749.218
559104	6467.34	258.944	2811.07	705.69
611328	6135.08	237.492	2806.6	641.224
663552	5963.15	222.178	2802.85	557.893
715776	5939.77	214.852	2802.7	521.733
768000	5945.29	208.858	2802.62	494.68
820224	5941.89	203.285	2803.15	470.161
921600	5942.16	194.1	2803.14	431.723
1022976	5942.08	186.362	2802.86	400.515
1084416	5944.68	182.496	2803.4	384.699
1185792	5939.88	176.267	2802.72	360.889
1299456	5943.56	170.455	2802.82	338.591
1400832	5942.15	165.748	2802.94	321.334
1566720	5940.32	159.203	2802.79	298.128
1904640	5941.18	146.486	2802.67	254.687
2036736	5943.27	142.557	2802.9	242.491
8150016	5941.01	87.5686	2799.52	108.962

Table 8 P655+ Load and Store Bandwidths (Mbytes/second)

Size	Load (Stride one)	Load (Random)	Store (Stride one)	Store (Random)
7168	12366.1	2194.7	7084.4	5730.42
7424	12346	2181.98	7085.82	5728.48
8192	12383.7	2133.81	7077.69	5731.57
8704	12356	2093.2	7080.71	5478.85
9216	12380.2	2068.66	7087.34	5491.42
12288	12393.4	1964.44	7092.5	5104.03
15360	12386.7	1903.07	7093.4	5174.83
21504	12382.1	1842.98	7087.5	5243.41
33792	12395	1784.58	7094.28	5184.71
43008	12389.5	1758.46	7072.2	5165
61440	12388	1665.99	7090.41	5051.74
70656	12314.5	1536.59	7091.53	4828.75
79872	12235.3	1362.3	7092.04	4232.81
89088	12184.4	1245.02	7092.23	3548.51
101376	12149.3	1113.84	7090.63	2949.79
110592	10963.6	1020.83	7079.65	2641.04
138240	8873.95	781.96	6179.1	2133.16
150528	8020.36	725.924	5827.64	2015.61
156672	7568.25	694.955	5689.97	1953.37
178176	5809.95	625.203	4784.28	1808.25
199680	5044.96	552.831	4089.79	1655.55
254976	4176.06	434.247	2945.34	1262.07
297984	3963.25	379.189	2697.01	1068.91
350208	3834.16	335.139	2585.43	921.071
402432	3793.32	305.973	2518.66	834.629
454656	3779.78	285.221	2498.3	772.489
506880	3768.65	269.499	2482.4	728.798
559104	3664.26	248.051	2478.87	680.331
611328	3536.56	228.508	2475.45	581.884
663552	3455.45	213.83	2472.72	488.318
715776	3440.32	207.201	2471.96	452.696
768000	3432.05	201.783	2471.5	428.559
820224	3428.92	196.134	2471.55	402.823
921600	3417.75	187.308	2470.72	366.662
1022976	3408.53	179.589	2470.74	337.85
1084416	3408.43	175.566	2470.21	323.189
1185792	3409.18	169.717	2469.8	303.297
1299456	3394.66	164.178	2470.25	285.1
1400832	3388.84	159.103	2470.02	268.494
1566720	3387.82	152.216	2469.99	247.531
1904640	3378.99	137.651	2469.75	205.833
2036736	3365.21	132.727	2469.32	192.824
8150016	3422.48	81.9893	2360.78	90.6231

The team that wrote this paper

- Ruzhu Chen

  ruzhuchen@us.ibm.com

- Clarisse T. Taffe-Hedglin

  clarisse@us.ibm.com

pSeries® & HPC Benchmark Center

IBM Poughkeepsie, NY

References

- IBM Redpaper *PMaC Benchmarking on Three POWER4 Platforms*, REDP-3724-00, by Chen, R. and Ebbers, M.:

  http://w3.itso.ibm.com/itsoapps/Redbooks.nsf/RedbookAbstracts/redp3724.html?Open

- “A framework for application performance prediction to enable scalability understanding” by Carrington, L., et al, 2002, San Diego Supercomputing Center:

  http://www.sdsc.edu/pmac/Papers/papers.html

- “Modeling application performance by convolving machine signatures with application profiles” by Snavely A., et al, 2002, San Diego Supercomputing Center:

  http://www.sdsc.edu/pmac/Papers/papers.html

- For a detailed description of the PMaC Benchmark Suite:

  http://www.sdsc.edu/PMaC

Notices

This information was developed for products and services offered in the U.S.A.

IBM may not offer the products, services, or features discussed in this document in other countries. Consult your local IBM representative for information on the products and services currently available in your area. Any reference to an IBM product, program, or service is not intended to state or imply that only that IBM product, program, or service may be used. Any functionally equivalent product, program, or service that does not infringe any IBM intellectual property right may be used instead. However, it is the user's responsibility to evaluate and verify the operation of any non-IBM product, program, or service.

IBM may have patents or pending patent applications covering subject matter described in this document. The furnishing of this document does not give you any license to these patents. You can send license inquiries, in writing, to:
IBM Director of Licensing, IBM Corporation, North Castle Drive Armonk, NY 10504-1785 U.S.A.

The following paragraph does not apply to the United Kingdom or any other country where such provisions are inconsistent with local law: INTERNATIONAL BUSINESS MACHINES CORPORATION PROVIDES THIS PUBLICATION "AS IS" WITHOUT WARRANTY OF ANY KIND, EITHER EXPRESS OR IMPLIED, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF NON-INFRINGEMENT, MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE. Some states do not allow disclaimer of express or implied warranties in certain transactions, therefore, this statement may not apply to you.

This information could include technical inaccuracies or typographical errors. Changes are periodically made to the information herein; these changes will be incorporated in new editions of the publication. IBM may make improvements and/or changes in the product(s) and/or the program(s) described in this publication at any time without notice.

Any references in this information to non-IBM Web sites are provided for convenience only and do not in any manner serve as an endorsement of those Web sites. The materials at those Web sites are not part of the materials for this IBM product and use of those Web sites is at your own risk.

IBM may use or distribute any of the information you supply in any way it believes appropriate without incurring any obligation to you.

Information concerning non-IBM products was obtained from the suppliers of those products, their published announcements or other publicly available sources. IBM has not tested those products and cannot confirm the accuracy of performance, compatibility or any other claims related to non-IBM products. Questions on the capabilities of non-IBM products should be addressed to the suppliers of those products.

This information contains examples of data and reports used in daily business operations. To illustrate them as completely as possible, the examples include the names of individuals, companies, brands, and products. All of these names are fictitious and any similarity to the names and addresses used by an actual business enterprise is entirely coincidental.

COPYRIGHT LICENSE:
This information contains sample application programs in source language, which illustrates programming techniques on various operating platforms. You may copy, modify, and distribute these sample programs in any form without payment to IBM, for the purposes of developing, using, marketing or distributing application programs conforming to the application programming interface for the operating platform for which the sample programs are written. These examples have not been thoroughly tested under all conditions. IBM, therefore, cannot guarantee or imply reliability, serviceability, or function of these programs. You may copy, modify, and distribute these sample programs in any form without payment to IBM for the purposes of developing, using, marketing, or distributing application programs conforming to IBM's application programming interfaces.

Send us your comments in one of the following ways:

- Use the online **Contact us** review redbook form found at:
  ibm.com/redbooks
- Send your comments in an Internet note to:
  redbook@us.ibm.com
- Mail your comments to:
  IBM Corporation, International Technical Support Organization
  Dept. HYJ Mail Station P099
  2455 South Road
  Poughkeepsie, NY 12601-5400 U.S.A.

Trademarks

The following terms are trademarks of the International Business Machines Corporation in the United States, other countries, or both:

@server®	POWER™	Redbooks™
@server®	POWER4™	Redbooks (logo)™
AIX®	POWER4+™	Redbooks (logo) ™
IBM®	pSeries®	Sequent®

Other company, product, and service names may be trademarks or service marks of others.