

国产万亿次机群系统NPB性能测试分析*

袁伟^{1,2,3}, 张云泉^{1,2}, 孙家昶^{1,2}, 李玉成¹

¹(中科院软件所并行计算实验室 北京 100080)

²(中科院计算机科学重点实验室 北京 100080)

³(中国科学院研究生院 北京 100039)

{yw, zyq, sun, lyc}@mail.rdcps.ac.cn

摘要 峰值计算速度达万亿次、十万亿次的国产大规模机群(Cluster)系统已相继研制成功并投入使用,我国高性能计算的硬件平台又上了一个新台阶,但高性能计算机的实际性能和应用水平却亟需提高。本文对三个国产万亿次机群系统,曙光4000A, Deepcomp 6800 和 Deepcomp 1800,进行了NPB性能测试分析,重点研究在进行大规模并行处理时(处理器数目达到上千个)的性能特点和趋势,分析了不同的处理器、互连网络等系统配置对NPB性能的影响,我们发现NPB的8个程序在三个万亿次机器上的性能特点和表现并不一致,表明国产高性能机群在设计上正在逐渐走出同质化的趋势,向多样化发展。用户需要以自己的实际应用性能为标准选择最合适的系统。进一步分析表明,目前NPB程序的可扩展性可以达到几百个处理器,但尚不能达到上千个处理器;NPB程序能发挥出的系统峰值的百分比仍然徘徊在10%左右,机群系统的并行可扩展性和应用程序对机器运算潜能的利用还需要进一步提高。对于处理器数目达到上千个的万亿次机群系统来说,对集合通信和细粒度通信能力的支持亟需提高。这也提醒并行程序设计者在万亿次机群系统上进行软件开发时,要注意减少集合通信和细粒度通信操作的数量。处理器数目达到上千个的国产万亿次机群系统的出现,在系统效率,可扩展性,并行I/O,并行程序设计和优化,系统维护管理,和系统可靠性等方面对国内高性能计算产业界,学术界和用户提出了前所未有的挑战。

关键词 万亿次机群; 性能评测; NPB

Performance Analysis of NPB Benchmark on Domestic Tera-Scale Cluster Systems

YUAN Wei^{1,2,3+}, ZHANG Yun-Quan^{1,2}, SUN Jia-Chang^{1,2}, LI Yu-Cheng¹

¹(Laboratory of Parallel Computing, Institute of Software, Chinese Academy of Sciences, Beijing 100080)

²(Laboratory of Computer Science, Chinese Academy of Sciences, Beijing 100080)

³(Graduate School of the Chinese Academy of Sciences, Beijing 100039)

Abstract With the successful development and application of domestic tera-scale cluster systems, the high performance computing hardware platforms of China reached another new peak performance level; however, the sustained application performance and utilization of these

* 基金资助: 国家自然科学基金(NO.60303020)、国家重点基础研究发展规划项目(973)《大规模科学计算研究》项目子课题《大规模计算工程软件系统的基础理论和实施》(G1999032805)、国家863“高性能计算机及其核心软件”重大专项课题《高性能计算机性能测试技术及方法研究》(No. 2004AA104020), 中科院软件所培育项目CXK25628。

systems need urgent improvement. In this paper, we performed NPB benchmarking on three domestic tera-scale cluster systems, our research focused on the performance characteristics and trends when carrying out tera-scale parallel computing on systems with thousands of processors. We analyzed the effects of different system configurations(processor, interconnection network, etc.) on final NPB performance, and found out that the programs in NPB suites got their best performance on different clusters, this phenomenon indicates that domestic HPC manufacturers begin to adopt different cluster system design and integration strategy and technologies. Users should choose the most suitable system based on their own real HPC application performance. Through further analysis, we found out that the scalability of NPB programs can reach hundreds of processors, but can't reach thousands of processors; most of NPB program can only exploit around 10% of system peak performance, the scalability of cluster systems and real application performance on tera-scale cluster systems need further improvement. For manufactures of tera-scale cluster systems with thousands of processors, the performance of collective communication and fine-grained message passing needs further improvement. For parallel programmers, they should pay more attention on the usage of collective communication and fine-grained message passing when designing large scale parallel program for tera-scale cluster systems with thousands of processors. In summary, The emergence of tera-scale cluster systems with thousands of processors become an unprecedented challenge to domestic HPC industry, academy and users from system efficiency, scalability, parallel I/O, parallel program design and optimization, system maintenance and administration, system reliability, etc.

Keywords tera-scale cluster system; performance evaluation; NAS Parallel Benchmarks (NPB)

1 引言

科学计算特别是高性能计算为现代科学研究提供了一种全新的手段。作为提供高性能计算环境的基础设施，高性能计算机的研制进展迅速。处理器性能仍然按照摩尔定律在提高，而体系结构技术的发展和创新的使得越来越多的处理器互连在一起，提供了每秒数万亿次乃至数十万亿次的计算能力。近年来，机群系统以其较高的性价比推动了高性能计算的普及，很短的时间内峰值计算速度达到万亿次、十万亿次的国产高性能机群相继研制成功并投入使用，国内高性能机群的装机量快速增长，高性能计算得到更广泛应用的条件已经成熟[7, 8]。

推动高性能计算的深化应用不仅是研制峰值更高的机器，还包括对机器的客观公正评价，机器实际性能，系统管理和维护水平，用户使用水平，并行软件的质量等其它问题。正因如此，高性能计算机的设计者和用户都很关注性能评测，性能评测是高性能计算机构建和应用的重要一环，其意义在于：

首先，高性能计算机硬件性能指标往往并不代表用户可获得的实际性能，系统瓶颈严重影响整体性能，复杂的并行系统无论是硬件或软件在性能、可靠性、可用性等方面都可能存在问题。性能评测有助于分析系统瓶颈，检验系统的性能、可用性、可靠性，并做出改进；

其次，不同的应用领域对高性能计算机的具体要求存在差别，用户对系统的浮点计算性能、通信性能、存储访问性能、I/O性能、系统可扩展性等有点或多点的特殊要求。对基于不同架构和配置的高性能计算机进行性能评测，有助于不同应用领域用户选择更适用的系统配置，并对用户开发并行软件提供建议。

进行性能评测可以使用标准测试程序[1, 2, 3, 4]，也可以根据需要自己设计测试程序。我国性能评测正经历从以峰值为主要评价指标到以 Linpack 性能为主要评价指标的转变

阶段。国际 TOP500[7]和中国 TOP100[8]排行榜对促进这一转变起到了很好的推动作用。但仅仅以 Linpack 性能对系统进行评价,并不能使厂商和广大用户满意,在某些情况下,也会出现偏差。因此,开发行业测试标准软件包,以行业认可的应用软件对机器系统进行面向行业的性能评测,成为一个急需解决的问题。目前国际上已经发布了 HPC Challenge Benchmark[15],正在试图解决和弥补 Linpack 测试的不足,更全面反映系统的真实性能。国内也正在积极进行这方面的探索。

2 NPB 简介

采用基准测试程序 (Benchmark) 对系统进行评测是常用的方法,如果使用得当也是最有效的。为了满足对单点性能或多点性能的不同测试需要,已经开发出了许多基准测试程序。选择用于分析的基准程序要参照应用的特性。求解稠密线性代数方程组的 Linpack 可以测试出并行机可发挥的最佳浮点计算能力,是最经常使用的一个基准程序 [3, 5, 6],为国际上 TOP500 和中国 TOP100 排行榜所采用。Linpack 测试值可达峰值性能的 50%—90%,并不能代表一般应用程序的实际性能。

NPB[9, 16]中的 8 个程序来自计算流体力学应用领域软件,可以表现出一般应用程序的实际性能,已经被普遍接受作为高性能计算机的性能测试标准。NPB 由 NASA Ames Research Center 开发,它制定了 NPB 测试程序的规范[10],可根据不同机器平台的特点在规范允许范围内进行优化和变动。

NPB 程序包有 5 个核心程序,它们是应用较频繁的一些算法[16, 17],下面是其主要测试特点:

- IS (Integer Sort, 整数排序), 主要测试整数运算性能和集合通信 (Collective Communication) 性能,对通信延迟很敏感;
- EP (Embarrassingly Parallel, 密集并行), 通信很少, 主要测试数学函数的浮点运算性能;
- MG (3-D Multigrid, 三维多重网格), 采用多重网格算法求解三维 Poisson 方程, 要求处理器数目必须为 2 的幂次; 主要测试规则 (structured) 的非连续存储访问集合通信和点到点通信;
- CG (Conjugate Gradient, 共轭斜量法), 主要测试不规则 (unstructured) 的集合通信和点到点通信;
- FT (Fast Fourier Transform, 富氏变换), 用 FFT 求解三维偏微分方程, 主要测试集合通信;

另外 3 个是拟应用程序, 它们来自计算流体力学[17]:

- LU (Lower-Upper Triangular, 对称超松弛法 (SSOR) 求解块稀疏方程组), 要求处理器数目必须为 2 的幂次; 主要测试细粒度的非连续存储访问点到点阻塞通信,对通信延迟很敏感;
- BT (Block Tridiagonal, 解 5x5 三对角块方程组), 要求处理器网格是方形的, 主要测试通信和计算的平衡, 以非连续存储访问点到点长消息通信为主。
- SP (Scalar Pentadiagonal, 解五对角线方程组), 要求处理器网格是方形的, 主要测试通信和计算的平衡, 以非连续存储访问点到点长消息通信为主, 与 BT 的通信模式很相似, 但通信强度比 BT 高。

根据统计[16], NPB 软件包中主要调用了 5 个集合通信函数 (MPI_Allreduce, MPI_Alltoall, MPI_Alltoallv, MPI_Barrier 和 MPI_Bcast)和四个点到点通信函数 (MPI_Send, MPI_Isend, MPI_Irecv 和 MPI_Recv)。

本文实验所用的是 2002 年 11 月发布的 MPI 版 NPB2.4[11], 另外也有采用 OpenMP 和

HPF 实现的 NPB[9, 12]。

3 测试环境

本文的实验平台为以下 3 个国产高性能机群系统，其中两台是 64 位系统。曙光 4000A 上的优化选项是 -O2, 其它是 -O3, 随机数发生器调用 randi8.

3.1 曙光 4000A

在 2004 年 6 月份发布的最新 TOP500 上排名第 10, 是目前国内峰值浮点计算速度最高的机器, 为 11.2 Tflops, 本文以下简称为 DAWNING4000A。目前安装在上海市超级计算中心。计算节点 640 个, 每节点为 4 路 64 位 Opteron CPU, 主频 2.2GHz, Myrinet 2000 网络连接。编译器为 pgi C 和 Fortran 编译器。

3.2 深腾 6800 超级机群系统

2003 年中国 TOP100 排名第一位的联想深腾 6800, 峰值浮点运算速度达到 5.324 Tflops, 本文以下简称之为 Deepcomp 6800。目前安装在中科院网络信息中心超级计算中心。计算节点 256 个, 每节点为 4 路 Itanium 2 CPU, 主频 1.3GHz, 3MB L3 Cache. 网络为 256 口全交换 full bisection 的 QsNet 网络, 网卡型号为 QM-400, elan3 3.5, 64 位的 mpich for QsNet 1.2.4, 所用编译器为 Intel C/C++/Fortran Compiler 7.0.

3.3 深腾 1800 超级机群系统

2002 年中国 TOP100 排名第一名, 2003 年中国 TOP100 排名第二位, 使用 SSE2 时峰值浮点运算速度达到 2.048 Tflops, 本文以下简称之为 Deepcomp 1800T。目前安装在中科院科学与工程计算国家重点实验室。计算节点 256 个, 每节点为 2 路 Xeon 2GHz CPU, Myrinet 2000 网络连接, mpich-GM 1.2.1, GM-1.5.2.1, 所用编译器为 GNU gcc/g++ 2.96, 缺省 Fortran 编译器为 32 位 Intel Fortran Compiler (ifc) 6.0.

4 性能分析

NPB 测试可以选择不同的类 (问题规模) 和调用不同处理器个数, 我们的测试根据研究目的制定, 测试数据尽量完整。少量数据没有测到, 主要原因是程序只能调用特定数目的处理器或受可用处理器资源和机时限制。

4.1 系统性能比较

机群系统单个节点的计算能力主要受处理器限制, 而整体性能的发挥还取决于节点之间的通信性能 [13, 14], 通信性能对机群计算性能的影响很大 [6]。为研究实际性能是否和这两方面的性能成正比, 在两个联想机群系统上, 用 64 个处理器采用 D 类问题规模对四个程序进行测试, 结果对比如图 1。其它四个程序由于两个机器上没有匹配的数据进行比较, 没有给出。

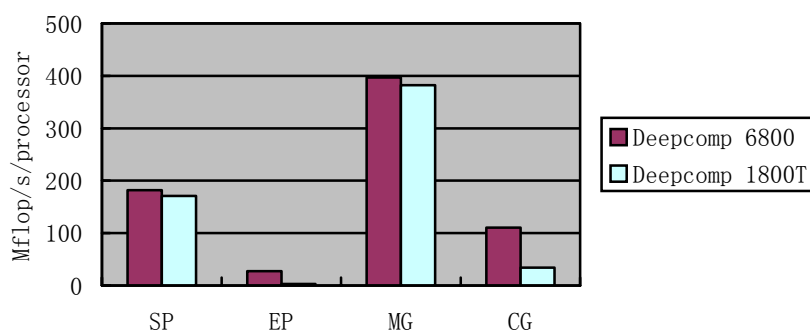


图 1 Deepcomp 6800 和 Deepcomp 1800T 上 64 个处理器 Class D 测试对比
具体比较结果见表 1。

表 1: 两个机群系统测得的 Mflop/s/processor 比值

Program	SP	EP	MG	CG
Deepcomp 6800/Deepcomp 1800T	1.06	8.94	1.04	3.25

四个程序的值都大于 1, 表明 Deepcomp 6800(Itanium + QsNet)的综合性能优于 Deepcomp 1800T (Xeon+Myrinet)。两个平台的计算节点主要性能指标和通信网络性能指标分别如表 2 和表 3 所示。

表 2: 计算节点主要性能指标

Processor	Cache	Memory	Peak Performance
Itanium 1.3GHz	256KB L2; 3MB L3	8GB	5.2Gflops
Xeon 2.0GHz	512KB L2	1GB	4.0Gflops

表 3: 网络性能指标

Interconnection	Bandwidth(point to point)	Latency
Quadrics	$\geq 300\text{MB/s}$	$< 7 \mu\text{s}$
Myrinet	250MB/s	$< 10 \mu\text{s}$

在处理器峰值上, Itanium2 的峰值高于 Xeon, 特别是 Itanium2 处理器有 3MB 的 L3 Cache, 对浮点计算性能很重要; 在通信网络上, Quadrics 在带宽和延迟方面都比 Myrinet 网络有优势。仅仅从硬件峰值性能来看, Deepcomp 6800 的整体浮点计算能力应该优于 Deepcomp 1800T 的整体浮点计算能力。但 SP 和 MG 两个测试程序的性能比接近 1, 说明提高系统的配置, 并不能保证提高所有应用程序的性能, 需要针对特定的应用程序特点和用户特点对系统进行配置。

MG 和 CG 主要测试通信性能, CG 测试的比值较大, 而 MG 测试的比值较小。说明 Deepcomp 6800 在处理不规则通信方面优势明显, 而在处理高度结构化的通信方面优势相对不如处理不规则通信;

EP 测试是密集并行计算, 通信量很少, 主要测试的是处理器的浮点计算能力, EP 测试的比值是 8.94, 说明 Itanium2 1.3GHz 处理器比 Xeon 2.0GHz 处理器可以发挥更高的浮点运算性能。

DAWNING4000A 和 Deepcomp 6800 的性能比较调用了 1024 个处理器, 采用 Class D, 结果如图 2。

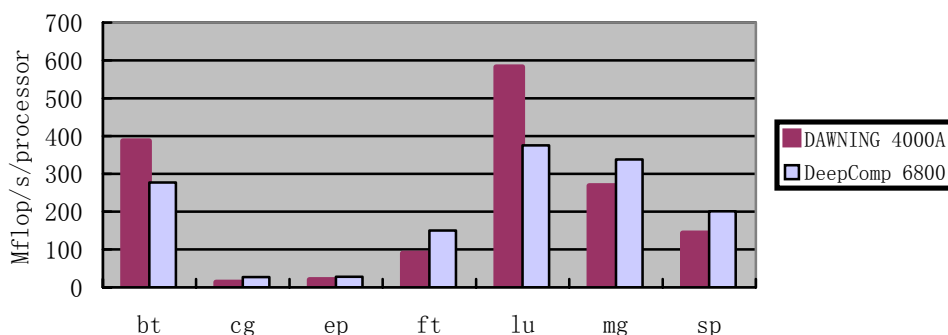


图 2 DAWNING 4000A 和 Deepcomp 6800 上 1024 个处理器 Class D 性能测试对比

可以看出各个程序在两个平台上的性能表现并不一致, 分别在不同的机器上达到最佳性能。DAWNING4000A 上 BT 和 LU 的性能高于 Deepcomp 6800, 而其它几个程序则低于 Deepcomp

6800; LU, BT 和 MG 在两个平台上都是性能比较好的。上述结果表明不同的平台适合运行不同的程序, 同一个计算平台并不能最好的满足所有类型的计算需要, 用户需要通过自己的应用软件选择最适合自己程序的并行计算平台。

4.2 系统效率

在处理器规模达到数百个、上千个的情况下, 系统的性能发挥水平很重要。如果实际性能与峰值性能相差太悬殊, 就要重新考虑单纯通过增加处理器数目来提高系统峰值的可行性。

为了分析 NPB 中各个程序在大规模并行时的性能状况, 在 DAWNING 4000A 上调用 1024 个处理器对 D 类问题规模进行测试的结果如图 2; 在 Deepcomp 6800 上分别调用 256 个和 1024 个处理器, 对 D 类问题规模的测试结果如图 3。

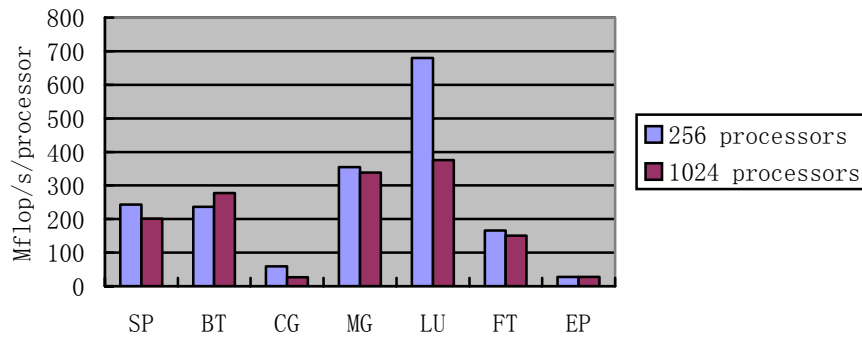


图 3 Deepcomp 6800 上 256 个和 1024 个处理器 Class D 性能测试结果

在 Deepcomp 1800T 上分别调用 64 和 256 个处理器, 对 C 类问题规模的性能测试结果如图 4。

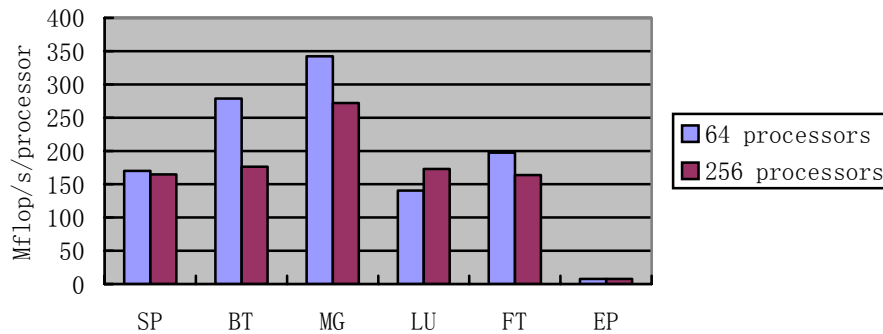


图 4 Deepcomp 1800T 上 64 个和 256 个处理器 Class C 性能测试结果

Opteron 2.2GHz CPU 的峰值浮点运算速度为 4.4Gflops; Itanium 2 1.3GHz CPU 的为 5.2Gflops; 如果使用 SSE2, Xeon 2GHz CPU 的峰值为 4Gflops, 各个程序的测试性能与处理器峰值性能的比值如表 4(na – non available)。

表 4: 单处理器上的程序性能与处理器峰值的比值 (%)

	SP	BT	CG	MG	LU	FT	EP
DAWNING4000A (1024 Processors)	3.26	8.80	0.30	6.10	13.26	2.05	0.47
Deepcomp 6800 (1024 Processors)	3.86	5.30	0.51	6.50	7.22	2.89	0.53

Deepcomp 1800T (256 Processors)	4.12	4.41	na	6.81	4.33	4.10	0.19
------------------------------------	------	------	----	------	------	------	------

在 Deepcomp 6800 上，当处理器个数从 256 个增加到 1024 个时，单处理器性能都有所下降 (BT 除外)，LU、CG 测试有显著下降。调用 1024 个处理器测试，NPB 的 7 个程序中 (不包括 IS)，实际性能最高的是 LU，为处理器峰值性能的 7.22%，而实际性能最低的 CG 仅达到峰值性能的 0.51%。

Mflop/s/processor 值随处理器规模提高而下降的现象同样出现在 Deepcomp 1800T 机群系统的测试中 (LU 除外)。调用 256 个处理器测试，NPB 的 6 个程序中 (不包括 IS 和 CG)，实际性能最高的是 MG，为峰值性能的 6.81%，而实际性能最低的 EP 仅达到峰值性能的 0.19%。

NPB 中的程序性能情况能够反映出目前一般实际应用的性能情况，在曙光 4000A 上的实际性能达到峰值性能的 0.30%—13.26%，Deepcomp 6800 上达到峰值性能的 0.51%—7.22%，Deepcomp 1800T 上达到峰值性能的 0.19%—6.8%。如果采用更加接近实际的应用程序进行测试，可能会更低。上述结果再次验证了当前大型并行应用软件仅能发挥系统峰值性能 20% 以内的事实。大规模并行处理时系统效率不高是目前所有并行计算系统面临的共同问题。高性能计算产业界，学术界和用户在提高系统效率方面还有很多工作需要做，只有在硬件，体系结构，系统软件，程序开发环境，系统维护和优化等方面不断提高和创新，才能逐步缓解这个问题。

4.3 可扩展性

并行计算要求系统 (程序 + 机器) 有良好的可扩展性，研究不同类型程序的可扩展性特征对并行程序设计和优化具有指导意义。

一般来说，Mflop/s/processor 值与处理器规模 (N) 的函数曲线，是先单调上升，达到最大值后又单调下降，可扩展性很差的程序也可能没有上升阶段。我们把 Mflop/s/processor 值最大时对应的处理器个数定义为 Pmax，它是衡量系统可扩展性的重要指标。Pmax 越大，系统扩展性越好。程序的 Pmax 值和机器平台有关系，所以研究要在同一平台下进行。我们在 Deepcomp 6800 上研究了 NPB 各个程序的可扩展性并结合程序的特征做了分析，曙光 4000A 上由于没有测试足够的实验数据，其结果在本文没有给出。图 5 中给出的是 IS 问题规模 C (处理器数目从 1 到 256) 和 EP 问题规模 D (处理器数目从 4 到 1024) 的测试结果。图 6 中给出了其它几个程序问题规模 D 的测试结果，处理器数目从 64 到 1024。有些程序因为需要方形处理器网格，在有些情况下，没有测试结果。

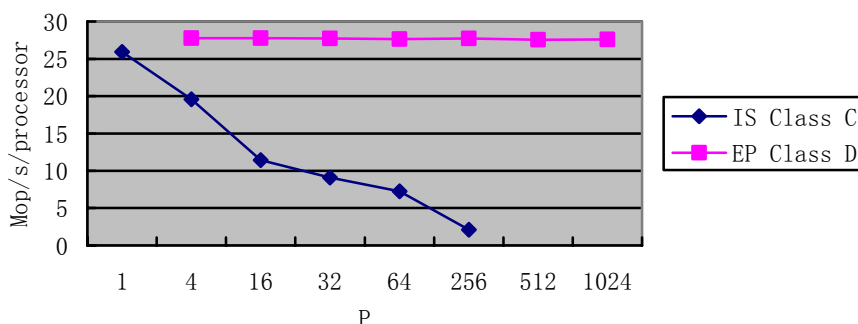


图 5 Deepcomp 6800 上 Class C IS 和 Class D EP 可扩展性测试结果

IS 做的是并行整数排序，主要调用集合通信，这类应用对通信的延迟很敏感。图 5 中给出的是采用 C 类问题规模时的 IS 测试结果，Mop/s/processor 值随着处理器规模增大下降很快，Pmax 等于 1，程序的可扩展性很差。对这类应用来说，如果通信网络性能不够高，

通信延迟大，则性能下降很快。

图 5 中 EP 测试的 Mflop/s/processor 值随着处理器数目的增加几乎没有变化。分析原因，在于 EP 测试主要是浮点运算，通信很少，Mflop/s/processor 值对处理器规模不敏感，EP 程序的可扩展性很好。

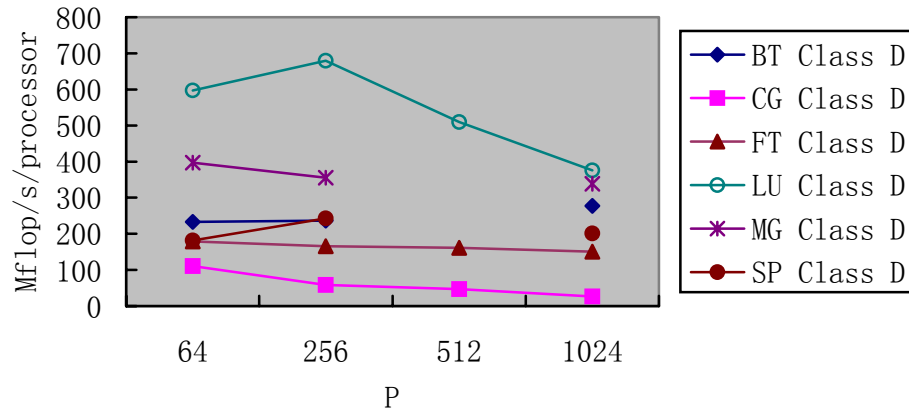


图 6 Deepcomp 6800 上 Class D 可扩展性测试结果

SP 和 BT 主要测试通信和计算的平衡，性能随着处理器数增多表现出不规则性。BT 的性能随处理器数增加而上升，Pmax 值大于或等于 1024，程序有很好的可扩展性；SP 的性能在处理器数由 64 增加到 256 时，有较大上升；在处理器增加到 1024 时又稍有下降，Pmax 介于 256 和 1024 之间。LU 主要测试细粒度的阻塞点到点通信，Pmax 在 256 附近，程序的可扩展性较好。但是在处理器数超过 256 后，性能下降很快，说明系统对细粒度阻塞点到点通信的支持还有待改进。对 NPB 软件包中的三个拟应用程序来说，在通信方面主要测试的是点到点消息传递，Deepcomp6800 上的可扩展性还是不错的，能够达到几百个处理器的水平。

MG, CG, FT 主要测试集合通信和点到点通信的混合。这三个程序测试的性能随处理器数增加而下降，Pmax 小于 64，程序的可扩展性不够好。分析原因，主要归因于程序中调用了集合通信，随着处理器的增多，集合通信的开销上升很快[18]。

不同程序表现出的可扩展性差异，首先是并行算法本身的可扩展性问题，由算法的特点和实现决定；其次是万亿次机群系统对集合通信和细粒度通信的支持和性能需要提高。

对于在万亿次机群系统上的并行程序设计，要选择可扩展性较好的并行算法，提高计算和通信比（甚至计算换通信），减少全局通信，提高并行粒度。

据我们所知，目前国内一般并行程序的并行规模都在数十个处理器，能够用好数百个处理器的应用软件很少，更不用说上千个处理器。规模增大，性能会急剧下降，有的甚至会出现错误或不能运行的情况，我们在大规模并行计算系统的可扩展性方面也面临挑战，用户需要重新设计其并行软件，甚至需要采用全新的算法和程序设计语言才能克服这一挑战。

5 结论及未来工作

本文通过 3 台国产万亿次高性能机群上的 NPB 测试结果，分析了影响万亿次机群系统性能的关键因素。对并行规模达到上千个处理器情况下的系统效率做了分析，并结合各个程序的特征做了可扩展性分析。分析表明，目前 NPB 程序的可扩展性一般可以达到几十到几百个处理器，但尚不能达到上千个处理器；NPB 程序能发挥出的系统峰值的百分比仍然徘徊在 10% 左右，机群系统的并行可扩展性和应用程序对机器运算潜能的利用还需要进一步提高。不同计算和通信特点 NPB 程序的可扩展性相差较大，对于处理器数目达到上千个的万亿次机群来说，对集合通信和细粒度通信能力的支持亟需提高。并行程序设计者在万亿次机群系统上进

行软件开发时,要注意减少集合通信和细粒度通信操作。本文在测试结果分析的基础上,进一步指出了我国高性能计算产业界,学术界和用户将在系统效率,可扩展性,并行 I/O,程序设计和优化,系统维护管理和系统可靠性等方面面临的前所未有的挑战。

NPB 基准程序主要测试的是系统的浮点运算性能和通信性能,对综合评测机群性能还是不够的,还需要其它如 I/O 性能、数据访问能力等方面的测试,设计更准确更全面的基准程序也是高性能计算机性能评测人员的努力重点。关于究竟何种 Benchmark 最有权威的争论已经持续了一段时间,其实没有任何单一的测试可以用来全面比较高性能计算机的性能,目前趋向于认同这样一个观点:对高性能计算机性能评测最有实际意义的是用户的应用程序,从这个意义上说,机器的评价并没有统一的标准,用户需要的是最适合本领域应用的计算机系统。我们还要进一步研究针对特定应用领域和系统特定性能的性能评价软件包,通过分析各种类型程序在运行时的性能特征,对高性能计算用户在购买机器,系统优化,编制应用程序等方面提供规律性的认识。

致谢: 高性能计算机的处理器和机时是很昂贵的资源,我们的测试研究得到了很多单位的支持和帮助。感谢国内高性能计算机厂商联想、曙光,感谢中科院计算所国家智能中心、中科院计算机网络信息中心超级计算中心、中科院数学与系统科学研究院科学与工程计算国家重点实验室等单位提供的机器资源和帮助。感谢中科院数学与系统科学研究院科学与工程计算国家重点实验室的张林波研究员对本文提出的宝贵修改意见。

参考文献

- [1] Lei Hu and Ian Gorton, Performance Evaluation for Parallel Systems: A Survey, Technical report, UNSW-CSE-TR-9707,1997
- [2] Marcelo Lobosco, Vítor Santos Costa, and Claudio L. de Amorim, Performance Evaluation of Fast Ethernet, Giganet and Myrinet on a Cluster, In Proceedings of the International Conference on Computer Science (ICCS'2002), LNCS 2329, 296-305, Springer-Verlag, Amsterdam, April 2002.
- [3] Jack Dongarra, "Performance of Various Computers Using Standard Linear Equations Software (Linpack Benchmark Report)", University of Tennessee Computer Science Technical Report, CS-89-85, October 2003.
- [4] A.B. Yoo and B.R. de Supinski and F. Mueller and S.A. McKee and I. Kumar, Memory Benchmarks for SMP-based High Performance Parallel Computers, Technical Report Lawrence Livermore National Laboratory, UCRL-JC-146246, Nov. 2001.
- [5] 罗水华, 杨广文, 张林波, 石威, 郑纬民, 并行集群系统的 Linpack 性能测试分析, 数值计算与计算机应用 2003,24(4):285-292。
(LUO Shui-Hua, YANG Guang-Wen, ZHANG Lin-Bo, SHI Wei, ZHENG Wei-Min, Linpack performance test on parallel cluster system, Journal on Numerical Methods and Computer Applications, ISSN :1000-3266,2003,24(4):285-292)
- [6] 都志辉, 吴傅, 刘鹏, 陈渝, 王小鸽, 李三立, LINPACK 与机群系统的 LINPACK 测试, 计算机科学, 2002,29(5):8-10,59。
(DU Zhi-Hui, WU Fu, LIU Peng, CHEN Yu, WANG Xiao-Ge, LI San-Li, LINPACK and LINPACK Benchmark, Computer Science, ISSN :1002-137X,2002,29(5):8-10,59)
- [7] TOP500, <http://www.top500.org>.
- [8] TOP100, <http://www.samss.org.cn>.
- [9] NAS Parallel Benchmarks, <http://science.nas.nasa.gov/Software/NPB>.

- [10]D. Bailey, E. Barszcz, J. Barton, D. Browning, R. Carter, L. Dagum, R. Fatoohi, S. Fineberg, . Frederickson, T. Lasinski, R. Schreiber, H. Simon, V. Venkatakrisnan, S. Weeratunga. The NAS Parallel Benchmarks. NAS Technical Report RNR-94-007, NASA Ames Research Center, Moffett Field, CA, 1994.
- [11]Rob F. Van der Wijngaart, NAS Parallel Benchmarks Version 2.4, NAS Technical Report NAS-02-007, NASA Ames Research Center, Moffett field, CA , 2002.
- [12]Michael Frumkin, Haoqiang Jin, and Jerry Yan. Implementation of NAS parallel benchmarks in high performance fortran. Technical Report NAS-98-009, NASA Ames Research Center, Moffet Field, CA, September 1998.
- [13]孟杰, 孙彤, 李三立, MPI 网络并行计算系统通信性能及并行计算性能的研究, 小型微型计算机系统, 1997,18(1):13-18.
(MENG Jie, SUN Tong, LI San-Li, Communication Performance and Parallel Performance Research of Networked Parallel Computing System, Mini Micro Systems, 1997,18(1):13-18)
- [14]胡明昌, 史岗, 胡伟武, 唐志敏, 通信对机群并行计算性能的影响, 小型微型计算机系统, 2003,24(9):1569-1573。
(HU Ming-Chang, SHI Gang, HU Wei-Wu, TANG Zhi-Min, How Does Communication Influence Cluster's Parallel Computing Performance, Mini-Micro Systems, ISSN : 1000-1220, 2003,24(9):1569-1573)
- [15] <http://icl.cs.utk.edu/hpcc/>。
- [16] Ahmad Faraj and Xin Yuan, "Communication Characteristics in the NAS Parallel Benchmarks," In the proceedings of Fourteenth IASTED International Conference on Parallel and Distributed Computing and Systems (PDCS 2002), pp.729-734, Cambridge, MA, November 4-6, 2002.
- [17]Qingda Lu, Jiasheng Wu, Dhableswar Panda, P. Sadayappan, Applying MPI Derived Datatypes to the NAS Benchmarks: A Case Study, In the proceedings of 2004 International Conference on Parallel Processing Workshops (ICPPW'04), pp. 538-545, Montreal, Quebec, Canada, August 15 - 18, 2004.
- [18] J.S. Vetter and M.O. McCracken, Statistical Scalability Analysis of Communication Operations in Distributed Applications. In the Proceeding of ACM SIGPLAN Symposium on Principles and Practice of Parallel Programming (PPOPP'01)., pp.123-132, Snowbird, Utah, USA, June 20 - 22, 2001.

作者简介

袁 伟, 1979年4月生, 男, 河南睢县人, 硕士研究生, 主要研究领域为大型并行数值软件、并行程序设计和性能评价; Tel: 62553023

张云泉, 1973年9月生, 男, 山东聊城人, 博士, 副研究员, 主要研究领域为大型并行数值软件、并行程序设计和性能评价,并行计算模型; Tel: 62542214

孙家昶, 1942年1月生, 男, 浙江宁波人, 研究员, 博士生导师, 主要研究方向为高性能计算及并行应用软件;

李玉成, 1961年6月生, 男, 河北邯郸人, 研究员, 主要研究领域为大型并行数值软件.