



Life Sciences Molecular Dynamics Applications on the IBM System Blue Gene Solution: Performance Overview

Version 1.0

Carlos P. Sosa, Tom Milledge, Jeffrey
McAllister, Gaolin Zheng Milledge
IBM
Minneapolis, Minnesota

Abstract: *This IBM® System Blue Gene® Solution Applications Guide provides a brief overview of the performance characteristics of a selected set of computational chemistry applications. Classical Molecular Mechanics/Molecular Dynamics (MM/MD) are commonly used to simulate large biomolecules that cannot be treated with more accurate methods. This guide is intended to help applications users in computational chemistry area evaluate(?) applications that are running on Blue Gene.*

1. Introduction

There are a variety of methods for studying molecules of interest in Life Sciences and related areas of chemistry. One of such methods is Molecular Mechanics / Molecular Dynamics where atoms are treated as points with mass and charge, governed by the laws of classical physics. Advances in biomolecular modeling using molecular dynamics necessitate simulations of larger systems of atoms on longer time-scales, which in turn, require robust and powerful computing platforms.

2. Hardware

The IBM® System Blue Gene® Solution is a massively parallel supercomputer from IBM. At the core of the system is the IBM Power PC® (PPC440) processor with the addition of two floating-point units (FPU). This system uses a distributed memory, message-passing programming model. To achieve a high level of integration and quantity of micro-processors with low power consumption, the machine was developed based on a processor with moderate frequency. Blue Gene utilizes system-on-a-chip (SoC) technology to allow a high level of integration, low power, and low design cost. Each processor core runs at a frequency of 700 MHz giving a theoretical peak performance of 2.8 Gigaflops/core or 5.6 Gigaflops/chip. The chip constitutes the compute node. The next building block is the compute card. Two compute nodes attached to a processor card with memory (RAM) create a compute card. The memory for each chip is soldered on the other side of the processor card; the amount of RAM per card is 2 GB (1 GB per compute node). The I/O card is the next building block. This card is very similar to the compute card. The I/O card has an integrated ethernet enabled for communicating with the outside world. Compute cards and I/O cards are plugged into a node card. There are two rows of eight compute cards on the node card. A Midplane consists of 16 node cards stacked in a rack. A rack holds two midplanes, for a total of 32 node cards. The largest system currently consists of 64 racks consisting of 131,072 processor cores [1].

3. Applications

In general, software packages based on molecular dynamics have been considered good candidates for the Blue Gene architecture [2]. Classical MD simulations

compute atomic trajectories by solving equations of motion numerically using empirical force fields. The overall MD energy equation is broken into three components: bonded, van der Waals, and electrostatic. The first two of these are local in nature and therefore do not make a significant contribution to the overall running time. The quadratic scaling of the electrostatics force terms, however, requires a high level of optimization of the MD application [2]. To improve performance on simulations in which the solvent is modeled at the atomic level (i.e., explicit solvent modeling), the three, AMBER, LAMMPS, and NAMD, Blue Gene MD applications employ a reciprocal-space technique called “Ewald sums” which allows the evaluation of long-range electrostatic forces to a pre-selected level of accuracy. In addition to the particle mesh Ewald (PME) method, LAMMPS also offers the particle particle/particle-mesh (PPPM) technique with characteristics that make it scale well on massively parallel processing (MPP) machines such as the Blue Gene system.

4. AMBER

AMBER [3] is the collective name for a suite of programs developed by the Scripps Research Institute that allow users to carry out molecular dynamics simulations, particularly on biomolecules. The primary AMBER module, called sander was designed to run on parallel systems and provides direct support for several force fields for proteins and nucleic acids. AMBER includes an extensively-modified version of sander, called PMEMD (Particle Mesh Complete information on AMBER can be found at their official website [4]).

For implicit solvent (continuum) models which rely on variations of the Poisson equation of classical electrostatics, AMBER offers the Generalized Born (GB)

method which uses an approximation to the Poisson equation that that can be solved analytically and allows for very good scaling [3] (Figure 1).

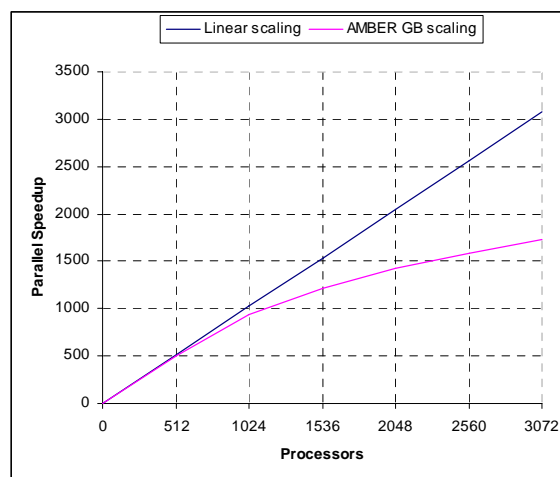


Figure 1. Parallel scaling of AMBER on Blue Gene. The experiment is with an implicit solvent (GB) model of 120,000 atoms (Aon benchmark).

AMBER also incorporates the PME algorithm, which takes the full electrostatic interactions into account and to improve the performance of electrostatic force evaluation (Figure 2).

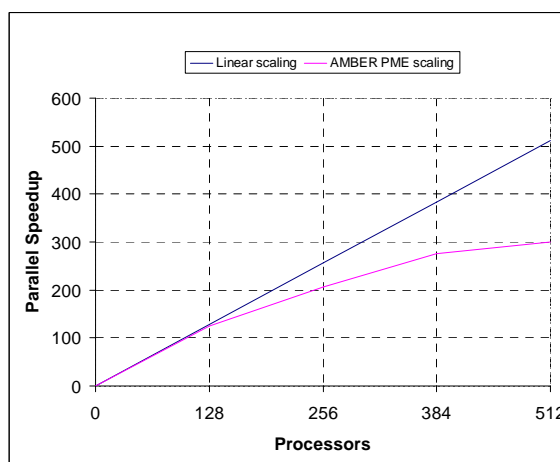


Figure 2. Parallel scaling of AMBER on Blue Gene. The experiment is with an explicit solvent (PME) model of 290,000 atoms (Rubisco).

5. LAMMPS

LAMMPS (Large-scale Atomic / Molecular Massively Parallel Simulator) [5] is an MD program from Sandia National Laboratories designed specifically for massively parallel processing. LAMMPS is implemented in C++ and is distributed freely as open-source software under the GNU Public License (GPL) [6]. LAMMPS can model atomic, polymeric, biological, metallic, or granular systems using a variety of force fields and boundary conditions. The parallel efficiency of LAMMPS varies from the size of the benchmark data and the number of steps being simulated. In general, LAMMPS can scale to more processors on larger systems (Figure 3).

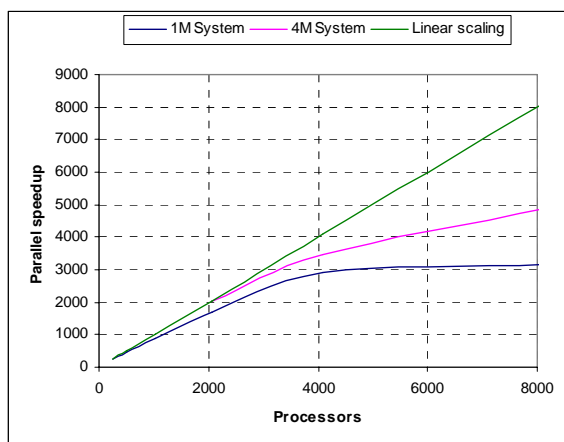


Figure 3. Parallel scaling of LAMMPS on Blue Gene (1M System: 1-million atom scaled rhodopsin, 4M System: 4-million atom scaled rhodopsin).

For a 1-million atom system, LAMMPS is able to scale up to 4096 nodes. For a larger system such as a 4-million atom system, LAMMPS is able to scale up to 4096 nodes as well. As the size of the system increases, so does the scalability.

6. NAMD

NAMD is a parallel molecular dynamics application developed for high-performance calculations of large biological molecular systems [7]. NAMD supports the force fields used by AMBER, CHARMM [8] and X-PLOR [9] and is also file compatible with these programs. This commonality allows simulations to migrate between these four programs. The C++ source for NAMD and Charm++ are freely available from UIUC. Additional information on NAMD can be found at their official website [10].

NAMD incorporates the PME algorithm, which takes the full electrostatic interactions into account and reduces the computational complexity. To further reduce the cost of the evaluation of long-range electrostatic forces, a multiple time step scheme is employed. The local interactions (bonded, van der Waals and electrostatic interactions within a specified distance) are calculated at each time step. The longer range interactions (electrostatic interactions beyond the specified distance) are computed less often. An incremental load balancer monitors and adjusts the load during the simulation.

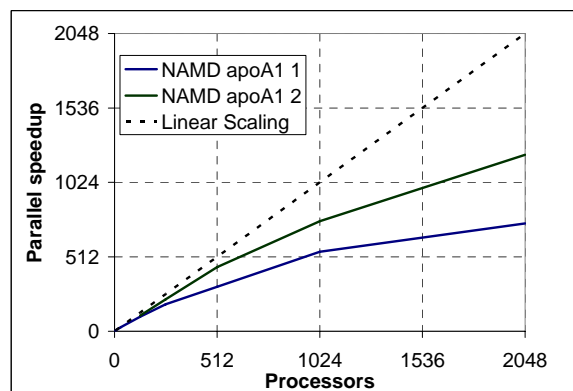


Figure 4. Parallel speedup for the NAMD standard apoA1 benchmark. The apoA1 1 line represents scaling with minimal configuration changes. Aggressive optimization, shown by the apoA1 2 line,

Due to the Blue Gene system's good balance of network and processor speed NAMD is able to scale to very large processor counts. While scalability is affected by many factors, many simulations can make use of multiple Blue Gene racks. Work by Kumar et al. [11] has reported scaling up to 8192 processors. Timing comparisons often use the "Benchmark time" metric instead of wallclock time to completion. This metric omits setup, I/O, and load balance overhead. While benchmark scaling can be considered a guide to what is possible, ideal load balance and I/O parameters for each case must be found for the wallclock time to scale similarly. Careful consideration of these parameters may be necessary to achieve the best scalability.

Summary

Although processor clock rate has traditionally been used as a predictor of application performance, the molecular dynamics applications profiled here demonstrate that the ability to scale to many thousands of processors is the key to leverage Blue Gene's massively parallel architecture. The efficiency of parallelization often depends on the overheads of communication and load distribution between processors as well as system size. These overheads vary for different application and benchmark types, but in general, larger runs show better parallel scaling than the smaller ones.

Blue Gene site: ibm.com/servers/deepcomputing/bluegene.html

References

1. Allsopp, N., et al., *Unfolding IBM eServer Blue Gene Solution*. IBM RedBook, 2005. <http://www.redbooks.ibm.com/abstracts/sg246686.html?Open>

2. Suits, F., et al., *Overview of Molecular Dynamics Techniques and Early Scientific Results from the Blue Gene Project*. , IBM Res. & Dev., 2005. **49**: p. 475-488.
3. Case, D.A., et.al., *The Amber biomolecular simulation programs*. Journal of Computational Chemistry, 2005. **26**(16): p. 1668-1688.
4. <http://amber.scripps.edu/>.
5. Plimpton, S., *Fast parallel algorithms for short-range molecular dynamics*. J. Comput. Phys., 1995. **117**(1): p. 1-19.
6. <http://lammmps.sandia.gov/>.
7. Phillips, J., et al., *Scalable molecular dynamics with NAMD*. J Comput Chem, 2005. **26**(16): p. 1781-1802.
8. Brooks, B.R., R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan, and M. Karplus., *CHARMM: A Program for Macromolecular Energy, Minimization, and Dynamics Calculations*. J. Comp. Chem., 1983. **4**: p. 187-217.
9. Brünger, A.T., *X-PLOR, Version 3.1, A System for X-ray Crystallography and NMR*. 1992: The Howard Hughes Medical Institute and Department of Molecular Biophysics and Biochemistry, Yale University. 405.
10. <http://www.ks.uiuc.edu/Research/namd/>.
11. Kumar, S., et al., *Achieving Strong Scaling with NAMD on Blue Gene/L*. Proceedings of IEEE International Parallel & Distributed Processing Symposium, 2006.

© IBM Corporation 2006

Produced in the United States of America
October 2006
All Rights Reserved

IBM, Blue Gene and PowerPC are trademarks or registered trademarks of International Business Machines Corporation in the United States or other countries or both.

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