Optimization of Molecular Dynamics Core Program on the K computer

Yousuke Ohno[1,2], Rio Yokota[3], Hiroshi Koyama[4,2], Gentaro Morimoto[1,2,5], Aki Hasegawa[4,2], Gen Masumoto[4,2],
Tetsu Narumi[6], and Makoto Taiji[1,2,4,5]

[1]Laboratory for Computational Molecular Design, Computational Biology Core, RIKEN Quantitative Biology Center
[2]Highperformance Computing Team, RIKEN Computational Science Research Program
[3]Strategic Initiative for Extreme Computing, King Abdullah University of Science and Technology
[4]High Performance Computing Development Team, HPCI Program for Computational Life Sciences, RIKEN
[5]Processor Research Team, RIKEN Advanced Institute for Computational Science
[6]Department of Computer Science, The University of Electro-Communications

1 Introduction

As part of the "Integrated Simulation of Living Matter(ISLiM) Group", we are developing "Grand Challenge Applications" in the life sciences using the K computer and future super computers[1]. The "Grand Challenge Applications" are categorized into five scales/layers of life sciences, molecular scale, cell scale, organ and body scale, gene and genome, brain and neural system. Five teams; Molecular Scale Team, Cell Scale Team, Organ and Body scale Team, Data Analysis Team and Brain, and Neural System Team are assigned to each scale/layer. The High-performance Computing Team supports the optimization of these applications on the K Computer and other next-generation supercomputers.

The number of processors in supercomputers is increasing, while each processor is parallelized by multi-core and/or SIMD. The K computer system has 82,944 processors, 8 cores per processor, two 2-way SIMD floating-point multiply-and-add units per core[1]. The theoretical peak performance is 10.6 PFLOPS. Therefore we needed to optimize applications by hybrid parallel, 80,000 MPI processes, 8 threads and SIMD. We must reconsider the data structure, algorithm, etc., at deep layers of the applications to achieve such high-parallelization.

We have developed a classical Molecular Dynamics (MD) core program as a hybrid core code optimized on the K computer. We implemented cutoff methods and fast multipole method (FMM) and measured the performance using early access on the K computer[2].

2 Molecular Dynamics Simulation

In classical MD simulations, atoms are treated as a point mass, and their orbits are calculated using Newton's equation of motion. The force between atoms are composed by three forces, covalent bond, van der Waals force and Coulomb force.

The calculation cost of the covalent bond is $\mathcal{O}(N_{\text{atom}})$ because it is short range forces. The van der Waals interaction is a long range force but damps rapidly. Therefore it is accurate enough to use a cutoff distance of $R_c \sim 12$ Å, and its calculation cost is $\mathcal{O}(N_{\text{atom}})$. The Coulomb interaction is a long range interaction with a slow decay rate. The straightforward calculation cost is $\mathcal{O}(N_{\text{atom}}^2)$. Algorithms reduce the calculation cost of Coulomb interaction are used in actual MD simulation.

Particle mesh Ewald (PME) [2] and its variants are efficient methods for solving Coulomb interaction. In these methods, charges of atoms are mapped to mesh points and the Coulomb potential is calculated by FFT in reciprocal space. The calculation cost is reduced to $\mathcal{O}(N_{\text{atom}} \log N_{\text{atom}})$, the cost of FFT. They are widely used for simulation of protein because of their low calculation cost and high accuracy for periodic boundary conditions. Though, PME has two disadvantages. First, they suffer from long-range correlation artifacts and anisotropy effects for non-periodic systems such as the macromolecular crowding in cellular environments[3, 4, 5]. Second, FFT requires a global, long distance communication on highly paralleled systems.

To avoid long-range correlation artifacts, variations of cutoff methods that include long-range effects have been proposed. For example, in isotropic periodic sums (IPS), it is assumed that charge distribution in cutoff length is repeated isotropically outside the original domain. The pairwise forces are summed up from isotropic image particles[6]. The calculation cost of the cutoff method is $\mathcal{O}(N_{\text{atom}})$ when the cutoff length is constant. The cutoff methods require only short range communication because long range interactions are omitted. The accuracy of the cutoff method depends on uniformity of the charge distribution. It means that cutoff length must be longer than the scale of target structure and the calculation cost will increase for large scale simulation.

3 Performance

We measured the performance of the cutoff method and FMM on the K computer.

3.1 Performance of cutoff

Table 1 shows the performance of the cutoff method. Within the kernel loop, we achieved high efficiency and SIMD ratio. The SIMD ratio in table 1 includes all instructions. The SIMD ratio of floating point operation was over 99.9%.

[1] `http://www.csrp.riken.jp/index_e.html`

[2] The K computer is under development and the results of this work are preliminary

Table 1. latest Performance on the main loop and the kernel loop for 418,707 atom simulation by 64 process

main loop	
Time consumption (ms/step)	116.7608
Performance (MFLOPS)	3,503,227
Efficiency	0.428

kernel loop	
Time consumption (ms/step)	87.1046
Performance (MFLOPS)	4,660,130
Efficiency	0.569
SIMD ratio	0.563
FLOP counts per pair of atom	109.7

Table 2. Time spent in blocks in millisecond per step for weak scaling (6,542 atom/node)

Number of nodes (N_{node})	64	24,576
Number of atoms (N_{atom})	418,707	160,783,488
Total T(N_{node})	116.733	120.826
Force calculation	97.436	97.441
Communication	13.772	17.868
Other	5.526	5.517
Performance (TFLOPS)	3.503	1299.66
Efficiency	0.4276	0.4132

Table 2 shows weak scaling of the cutoff method. The parallel efficiency is 116.733/120.826 = 0.97.

We measured the performance of the previous version of our code using 79,872 processors on the K computer system with a nominal peak performance of 10.2 PFLOPS. A sustained performance of 3.863 PFLOPS has been achieved for the simulation with 482 million atoms.

3.2 Performance of FMM

Table 3 shows the performance of the FMM. The cost of the multipole part was smaller than the cost of other parts. We measured details of the multipole part and confirmed the cost is log(P) for P processes.

4 Conclusion

We have developed an all-atom classical MD core program that targets massive-parallel computers such as the K computer. The performance benchmarks of the cutoff method have shown good weak scalability, 0.96, up to 79,872 processors on the K computer. A sustained performance of 3.863 PFLOPS has been achieved using 79,872 processors of the K computer system with a nominal peak performance of 10.2 PFLOPS for the simulation with 482 million atoms. We have performed further optimization, and the efficiency of the cutoff kernel and main loop have reached 57 % and 40 %, respectively. We have implemented FMM for long-range Coulomb interaction and have achieved good weak scalability, 0.91, up to 24,576 processors. The efficiency was 12.3 % but the time consumption was 48 ms/step for 161 million atoms on 24,576 processors. The communication of FMM was proportional to log(P) for P processors. This shows that the FMM is more scalable than PME, which has $\mathcal{O}(P)$ asymptotic communication.

Table 3. Time consumption (ms/step) and Performance of FMM for 418,707 atoms/processor

Number of nodes	64	4,096
Node shape	4x4x4	16x16x16
Number of atoms	418,707	26,797,248
mainloop	43.421	45.355
calcforce	28.265	28.713
communication	10.938	12.387
other	4.218	4.255
Performance(TFLOPS)	1.170	71.715
Efficiency	0.143	0.137

Acknowledgement

This work was partially supported by the Integrated Simulation of Living Matter Project, commissioned by the Ministry of Education, Culture, Sports, Science and Technology, Japan. This work was partially supported in part by the Japan Scienceand Technology Agency (JST) Core Research of EvolutionalScience and Technology (CREST) research programsHighly Productive, High Performance ApplicationFrameworks for Post Petascale Computing. Part of the results is obtained by early access to the K computer at the RIKEN Advanced Institute for Computational Science. We thank the Next-Generation Supercomputer R&D Center, especially Kazuo Minami, Akiyoshi Kuroda and Masaaki Terai (in the Application Development Team) for their outstanding support. We appreciate Mikio Hondou and Hikaru Inoue, from Fujitsu Co. Ltd., for their optimization support.

References

[1] Akinori Yonezawa, Tadashi Watanabe, Mitsuo Yokokawa, Mitsuhisa Sato, and Kimihiko Hirao. Advanced Institute for Computational Science (AICS): Japanese national high-performance computing research institute and its 10-petaflops supercomputer "K". In *State of the Practice Reports*, SC '11, pages 13:1–13:8. ACM, 2011.

[2] T. Darden, D. York, and L. Pedersen. Particle mesh ewald: An n log (n) method for ewald sums in large systems. *J. Chem. Phys.*, 98:10089–10092, 1993.

[3] P.H. Hünenberger and J.A. McCammon. Ewald artifacts in computer simulations of ionic solvation and ion–ion interaction: a continuum electrostatics study. *J. Chem. Phys.*, 110:1856–1872, 1999.

[4] P.E. Smith and B.M. Pettitt. Ewald artifacts in liquid state molecular dynamics simulations. *J. Chem. Phys.*, 105:4289–4293, 1996.

[5] Francisco Figueirido, Gabriela S. Del Buono, and Ronald M. Levy. On finite-size effects in computer simulations using the ewald potential. *J. Chem. Phys.*, 103(14):6133–6142, 1995.

[6] X. Wu and B. R. Brooks. Isotropic periodic sum: a method for the calculation of long-range interactions. *J. Chem. Phys.*, 122(4):44107, Jan 2005.