

Performance Analysis of Different Computational Architectures: Molecular Dynamics in Application to Protein Assemblies, Illustrated by Microtubule and Electron Transfer Proteins

Vladimir A. Fedorov¹, Ekaterina G. Kholina¹, Ilya B. Kovalenko^{1,2,3,4},
Nikita B. Gudimchuk^{1,5}

© The Authors 2018. This paper is published with open access at SuperFri.org

All-atom molecular dynamics simulation represents a computationally challenging, but powerful approach for studying conformational changes and interactions of biomolecules and their assemblies of different kinds. Usually, the numbers of simulated particles in modern molecular dynamics studies range from thousands to tens of millions, while the simulated timescales span from nanoseconds to microseconds. For cost and computation efficiency, it is important to determine the optimal computer hardware for simulations of biomolecular systems of different sizes and timescales. Here we compare performance and scalability of 17 commercially available computational architectures, using molecular dynamics simulations of water and two different protein systems in GROMACS-5 package as computing benchmarks. We report typical single-node performance of various combinations of modern CPUs and GPUs, as well as multiple-node performance of “Lomonosov-2” supercomputer in molecular dynamics simulations of different protein systems in nanoseconds per day. These data can be used as practical guidelines for selection of optimal computer hardware for various molecular dynamics simulation tasks.

Keywords: molecular dynamics, tubulin, microtubule, plastocyanin-cytochrome f.

Introduction

Molecular dynamics (MD) is a powerful method to study conformational dynamics and interactions of biomolecules, including protein assemblies. Because of a big number of particles, which make up protein systems, and multiple computational steps, usually required to achieve meaningful results, MD simulations of proteins represent a major computational challenge. Therefore, identification and use of optimal hardware for high efficiency calculations is important. In this work we systematically compare multiple currently available computer architectures in their MD simulation performance, using two types of biomolecular systems as computing benchmarks: (i) water boxes of different sizes and (ii) two protein systems. The selected protein systems include different assemblies of tubulins, the building blocks of microtubules [5], and a photosynthetic electron-transfer complex of plastocyanin and cytochrome f proteins [4].

1. Methods

All-atom explicit solvent MD was used in all tests. Calculations were performed with the use of software package GROMACS-5 [3], which allows parallel computing on hybrid architecture with the CHARMM27 force field. All benchmarks were run for 15 minutes. TIP3P water model was employed. The protein structures were obtained from the Protein Data Bank. We

¹Lomonosov Moscow State University, Moscow, Russia

²Federal Research and Clinical Center of Specialized Medical Care and Medical Technologies, Federal Medical and Biological Agency of Russia, Moscow, Russia

³Astrakhan State University, Astrakhan, Russia

⁴Scientific and Technological Center of Unique Instrumentation of the RAS, Moscow, Russia

⁵Center for Theoretical Problems of Physicochemical Pharmacology, RAS, Moscow, Russia

used higher plant plastocyanin-cytochrome f complex (PDB id 2PCF) and GMPCPP-bound tubulin structure (PDB id 3J6E). The size of the virtual cell was chosen in such a way that the distance from the protein surface to the nearest box boundary was no less than two nanometers. The particle mesh Ewald method was used for the long-range electrostatics. All-bond PLINKS constraints and mass rescaling were applied to the tested protein systems. Coulomb and Lennard-Jones cut-offs were both set to 1.25 nm. Specifications of MD systems used for benchmarking are summarized in (Tab. 1).

Table 1. Molecular dynamics systems used for benchmarking

MD systems	MD system name	Number of atoms	Box type	System size (nm)	Time step (fs)
Water box(WB)	WB-10	10,206	cube	4.7x 4.7x 4.7	1
	WB-80	80,232	cube	9.3x 9.3x 9.3	1
	WB-120	121,527	cube	10.7x10.7x10.7	1
	WB-160	159,780	cube	11.7x11.7x11.7	1
	WB-200	203,415	cube	12.7x12.7x12.7	1
Plastocyanin-cytochrome f	Pc-cyt	93,085	dodecahedron	11x11x11	5
Tubulin tetramer	Tub-4	315,718	cube	9.9x13.9x23.3	4
Tubulin 18-mer	Tub-18	1,119,458	cube	22.8x15.0x33.3	4

2. Results and Discussion

To begin with, we used our first benchmark, the water box, in order to examine performance of MD simulations as a function of the number of particles in the molecular system. We conducted MD simulations of water boxes of various sizes using 17 different single-node computer systems with various CPU/GPU architectures. Data summarized in (Tab. 2) suggest that an increase of MD system size leads to an unproportional decrease of computer performance. However, the relative extent of the decrease is almost equal for different computer systems. Not surprisingly, 2×Intel Xeon E5-2695 with four Tesla K80 GPUs shows the highest performance for all the tested MD systems. However, Intel Core i7-5930K with GTX 980 has the most optimal performance-price combination out of all hardware configurations we tested, consistent with conclusions of a previous study [2].

To further address the question of scalability, we used our second type of computing benchmark and established the dependence of the supercomputer “Lomonosov-2” performance in MD simulations on the number of computer nodes used. As expected, we could clearly see that for all three tested protein systems the performance grew as a function of the number of supercomputer nodes (Tab. 3). The relative rate of that growth did not significantly depend on the type and size of the biological system and slowed down gradually, roughly following Amdahl’s law.

Table 2. Single-node performance (ns/day) depending on various combinations of CPUs and GPUs tested by MD GROMACS simulations of water boxes of different sizes. Time step is 1 fs

Processor	GPU	System name				
		WB-10	WB-80	WB-120	WB-160	WB-200
Intel Core i7-3820	GTX 680	57.8	8.1	5.3	4.1	3.2
	GTX 780	72.9	10.5	6.6	5.1	3.9
Intel Core i5-3570K	GTX 780	64.9	9.7	6.2	4.5	3.5
	GTX 980	82.5	10.4	7.1	4.9	3.1
Intel Core i7-4790K	GTX 780	85.9	12.3	7.3	5.7	4.3
	GTX 980	106.4	13.4	8.4	6.0	4.7
2×AMD Opteron 6168	GTX 980	61.1	9.9	6.7	5.1	4.0
AMD Phenom II X6 1100T	GTX 780	56.2	7.6	4.8	3.7	2.8
Intel Core i7-5930K	GTX 960	80.5	11.2	7.5	5.7	4.5
	GTX 970	103.1	15.0	10.2	7.7	6.1
	GTX 980	116.3	17.5	11.4	8.8	6.8
2×Intel Xeon E5-2695	1×Tesla K80	88.4	13.9	9.4	7.2	5.6
	2×Tesla K80	140.4	24.3	16.6	12.6	9.9
	4×Tesla K80	162.3	36.2	24.4	19.7	15.0
	no GPU	88.9	13.3	8.7	6.7	5.1
Intel Xeon 5670 (“Lomonosov-1”)	Tesla X2070	33.4	4.7	3.2	2.4	1.9
2×Intel Xeon E5-2697 (“Lomonosov-2”)	Tesla K40	87.5	13.5	9.0	6.9	5.5

Table 3. Performance of “Lomonosov-2” supercomputer (ns/day), depending on the number of computing nodes for MD simulations of different protein systems. System size in atoms is given in brackets. Time step is 4 fs

	1	2	3	4	5	6	7	8
Pc-cyt (93085)	53.5	93.7	128.6	160.7	186.3	188.1	209.8	256.4
Tub-4 (315718)	16.1	27.2	34.1	43	48.7	51.3	53.4	61.6
Tub-18 (1119458)	4.5	5.4	8.5	12.1	16.3	18.4	21.4	22.5

Conclusion

Our comparative performance analysis suggests that for relatively small biomolecular systems, below 100,000 atoms, such as the complex of plastocyanin and cytochrome f proteins, it is quite practical to use “personal supercomputers”, i.e. single node workstations with a video accelerator. Such a computer can provide 100 ns/day performance for molecular dynamics calculation of a small biomolecular system with the size of about ten thousand atoms. For larger biomolecular systems, like a fragment of microtubule or a part of biological membrane with protein complexes, “personal supercomputers” are not currently fast enough, with typical performance of only several ns/day. Therefore, for large systems usage of modern supercomputers, like “Lomonosov-1” or “Lomonosov-2” with hybrid architecture is imperative [1]. By employing dozens of supercomputer nodes, such hardware systems are capable of accelerating calculations by an order of magnitude, providing up to 22 ns/day performance of GROMACS-5 MD simulation for a system sized more than one million particles.

Acknowledgments

This study was designed and conducted with support from the Russian Foundation for Basic Research, projects 16-34-60113 and 16-04-01862. The research was carried out using the equipment of the shared research facilities of HPC computing resources at Lomonosov Moscow State University supported by project RFMEFI62117X0011.

This paper is distributed under the terms of the Creative Commons Attribution-Non Commercial 3.0 License which permits non-commercial use, reproduction and distribution of the work without further permission provided the original work is properly cited.

References

1. Sadovnichy, V., Tikhonravov, A., Voevodin, V., Opanasenko, V.: Lomonosov : Supercomputing at Moscow State Univer. Contemporary High Performance Computing: From Petascale toward Exascale (Chapman & Hall/CRC Computational Science) pp. 283–307 (2013)
2. Kutzner, C., Páll, S., Fechner, M., Esztermann, A., de Groot, B.L., Grubmüller, H.: Best bang for your buck: GPU nodes for GROMACS biomolecular simulations. Journal of computational chemistry 36(26), 1990–2008 (2015), DOI: 10.1002/jcc.24030
3. Abraham, M.J., Murtola, T., Schulz, R., Páll, S., Smith, J.C., Hess, B., Lindahl, E.: Gromacs: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. SoftwareX 1, 19–25 (2015), DOI: 10.1016/j.softx.2015.06.001
4. Kovalenko, I., Khrushchev, S., Fedorov, V., Riznichenko, G.Y., Rubin, A.: The role of electrostatic interactions in the process of diffusional encounter and docking of electron transport proteins. In: Doklady Biochemistry and Biophysics. vol. 468, pp. 183–186. Springer (2016), DOI: 10.1134/S1607672916030066
5. Zakharov, P.N., Arzhanik, V.K., Ulyanov, E.V., Gudimchuk, N.B., Ataulakhanov, F.I.: Microtubules: dynamically unstable stochastic phase-switching polymers. Physics-Uspekhi 59(8), 773 (2016), DOI: 10.3367/UFNe.2016.04.037779