CHE I399 – Final Report

Man Yi “Ariel” Yeung

September 3, 2019

Abstract:

The performance of two Molecular Dynamics simulations software, NAMD and Gromacs, on the supercomputer Stampede2 is studied. Scaling analysis of a solvated ApoA1 system were performed on both software, and the processing speeds are compared. TopoGromacs is used to convert the NAMD input file format for the ApoA1 system to Gromacs-compatible file types. Gromacs versions 2018.3 and 2016.4 show faster processing speed among all versions of Gromacs and NAMD installed on Stampede2. This demonstrates an advantage of using Gromacs to conduct simulations on Stampede2, because it would save money and costs, especially when running large and complex systems, as long as the system is compatible with the TopoGromacs plugin.

Introduction

The objective of the project is to compare the performance of two Molecular Dynamics (MD) simulations software packages, Gromacs and NAMD, on Stampede2 supercomputer. MD simulations are used to study the movement of atoms in a macromolecule. Macromolecules, such as proteins, are solvated in a solvent system, where atoms of the macromolecule interact with each other and the solvent atoms. The entire system is relaxed through energy minimization to remove any inappropriate geometry [2]. During the MD simulation process, at each small time step, typically around 1 to 2 fs, the coordinates, trajectories and energy of each atom are recorded. The force and energy of each atom are calculated using force-field parameters, which features bonded energies, based on bond length, bond angles and bond dihedrals, as well as nonbonded energies, from Van der Waals and electrostatic interaction. Using the force values, velocity, acceleration and coordinates can be calculated with Newton's laws of motion [1].

Gromacs and NAMD are two software packages that can perform MD simulations. NAMD can easily utilize CHARMM force fields, which facilitates the simulations when ran on complex biological molecules, such as carbohydrates, proteins and glycoproteins. Gromacs excels at the analysis stage after the system has been processed [3]. Whether Gromacs or NAMD will run faster for the same system is tested to compare the processing speed of the two software. The test is conducted on Stampede2 supercomputer, located at the Texas Advanced Computing Center (TACC), University of Texas at Austin.

Methods

NAMD and Gromacs require different input file formats, to specify the system topology, force field parameters and simulation parameters. The TopoGromacs plugin for VMD is used to

convert NAMD's to those compatible with Gromacs. The test molecule used to compare the performance of NAMD and Gromacs is ApoA1. This molecule is chosen because it is used in the published paper "TopoGromacs: Automated Topology Conversion from CHARMM to GROMACS within VMD", to demonstrate the application of TopoGromacs to large biomolecular system. A protein data bank (PDB) file, featuring the structure of an ApoA1 solvated in water, is provided with the published paper. The solvated system contains 92,224 atoms, 70,660 bonds, and a total of 21,458 water molecules. Accompanied scripts, included in the published paper, is used to convert the NAMD input files, already provided, into Gromacs input files, thus preparing all the files necessary to run the simulation on both software for comparison [3].

The comparison is conducted on the supercomputer Stampede2 by running the simulation as Message Passing Interface (MPI) programs. The tasks that need to be executed are distributed to multiple nodes to be ran in parallel for faster run time. The nodes also communicate to get information from other nodes to perform their individual tasks. Therefore, the run time for a simulation depends on the number of nodes assigned and the number of MPI tasks assigned per node. The total number of nodes to be used and the total number of MPI tasks are specified in the job submission script to stampede2. The number of MPI tasks per node is calculated by dividing the total MPI tasks by the total number of nodes. Multiple jobs are submitted with different numbers of nodes and numbers of MPI tasks to conduct scaling analysis on NAMD and on Gromacs in order to find and compare their processing speed. All simulations were ran on Stampede2's Knight's Landing (KNL) compute nodes, which each have 68 cores[4].

Results and Discussion

For NAMD, the number of nodes specified in the SBATCH job script submitted to Stampede2 is equivalent to the number of physical cores used by NAMD, stated in the output file after the simulation is completed. The total number of MPI tasks is equal to the number of cores used by NAMD displayed in its output file. For each trial, the number of nodes is varied to conduct scaling analysis on the solvated ApoA1 system. The total number of MPI tasks is specified such that the number of MPI tasks per node is fixed at 13, as recommended by TACC's user guide for running NAMD on Stampede2 [4].

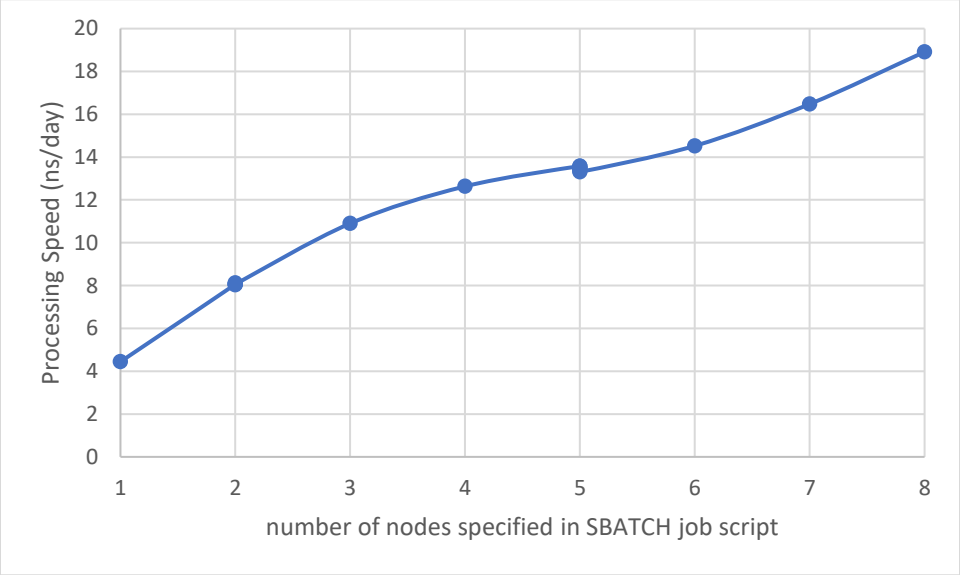

Figure 1: Scaling Analysis for ApoA1 system on NAMD version 2017_12_05

On Figure 1, the results are shown for the scaling analysis conducted on NAMD version 2017_12_05, which is the recommended and default version to be ran on Stampede2. The processing speed in nanoseconds per day is calculated by taking the inverse of days per nanosecond, recorded in the output log for each simulation runs. The processing speed increases with increasing number of nodes specified in the SBATCH job script, as tasks are more

distributed, and more tasks are ran in parallel. The graph does not show any plateauing when using less than eight nodes for the ApoA1 system of 92,224 atoms. This suggests it has not reached the stage when tasks are overly distributed and message passing between nodes takes up more computing time than processing tasks within each node.

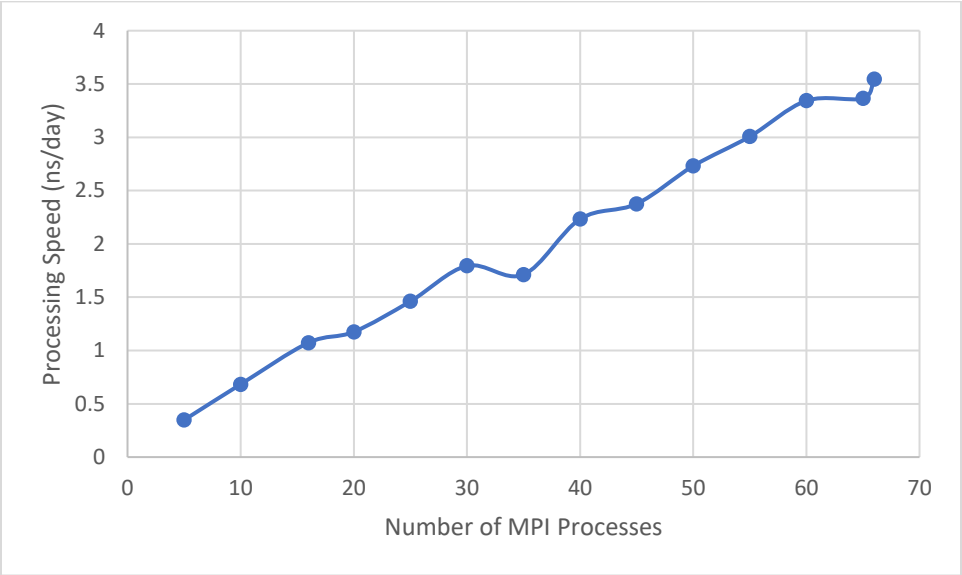

Figure 2: Scaling Analysis of ApoA1 system on Gromacs Version 5.1.2 using 1 KNL Node on Stampede2

The same scaling analysis is performed on Gromacs. However, for Gromacs, the number of nodes specified in SBATCH job script corresponds to the number of nodes used by Gromacs, not the number of physical cores like NAMD. The total number of MPI tasks also corresponds to the number of MPI processes stated in the output file. On Stampede2, there are three versions of Gromacs installed, version 5.1.2, the recommended version 2016.4, and version 2018.3. Scaling analysis was conducted on all three versions for comparison. In contrast with NAMD, there was no recommended number of MPI tasks per node stated in Stampede2's user guide. Tests were conducted to observe the change in processing speed with the total number of MPI processes when the total number of nodes is kept constant at one. As can be seen from the result plotted in

Figure 2 for Gromacs version 5.1.2, there is an improvement in processing speed when the number of MPI processes increases. It is assumed the same relationship will hold for all three versions. Therefore, future tests for the relationship of processing speed and number of nodes are conducted with the maximum total number of MPI processes that can be obtained on Stampede2 Knight's Landing computing system, which is 68 MPI processes per node.

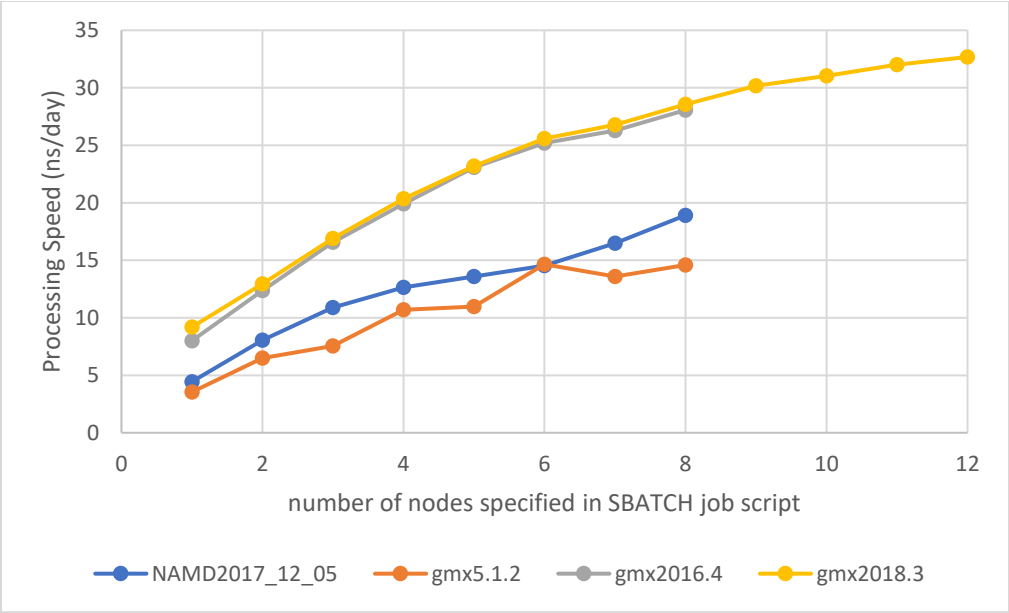

Figure 3: Scaling Analysis Comparison of ApoA1 system on NAMD version 2017_12_05, Gromacs version 5.1.2, Gromacs version 2016.4, and Gromacs version 2018.3 (gmx stands for Gromacs)

Results for scaling analysis of all three versions of Gromacs installed on Stampede2 are displayed in Figure 3. It also includes NAMD's scaling analysis data from Figure 1 for comparison purposes. The three versions of Gromacs all shows good scaling up to 8 nodes or 12 nodes for version 2018.3. Larger number of nodes was tested for version 2018.3, because it showed the fastest processing speed. It appears none of the three versions has reached a plateau for the number of nodes tested. However, the increase in processing speed slows down gradually for Gromacs version 2018.3 after 8 nodes. This suggests message passing between nodes is taking a relatively large proportion of the computing time. Comparing among the three versions

of Gromacs and NAMD on Stampede2, Gromacs version 2018.3 appears to have better processing speed with the same number of nodes, as displayed in Figure 3. Gromacs version 2016.4 also shows good result. These two versions of Gromacs has 1.5 times the processing speed of NAMD version 2017_12_05 when running the ApoA1 system with 8 nodes.

Conclusion

On Stampede2, NAMD version 2017_12_05, Gromacs version 5.1.2, Gromacs version 2016.4, and Gromacs version 2018.3 all demonstrate good scaling ability up to 8 nodes for the ApoA1 system. Gromacs version 2018.3 and 2016.4 show better performance on the solvated ApoA1 system in terms of processing speed on Stampede2. The advantage being 1.5 times over NAMD version 2017_12_05 when running the simulation with 8 nodes. This would save a significant amount of time and cost when running large complex systems, as long as the system's NAMD input files can be converted to Gromacs input files using the TopoGromacs plugin. Additional testing can be conducted to see if different types of TopoGromacs can be used for different macromolecular systems, and observe if the advantage of faster processing time applies to systems other than ApoA1.

References

[1] J. Gelpi, A. Hospital, R. Goñi, and M. Orozco, “Molecular dynamics simulations: advances and applications,” *Advances and Applications in Bioinformatics and Chemistry*, p. 37, 2015.

[2] “GROMACS Tutorial,” *Bevan Lab*. [Online]. Available: http://www.bevanlab.biochem.vt.edu/Pages/Personal/justin/gmx-tutorials/lysozyme/05_EM.html.

[3] J. V. Vermaas, D. J. Hardy, J. E. Stone, E. Tajkhorshid, and A. Kohlmeyer, “TopoGromacs: Automated Topology Conversion from CHARMM to GROMACS within VMD,” *Journal of Chemical Information and Modeling*, vol. 56, no. 6, pp. 1112–1116, 2016.

[4] “User Portal,” *Go to TACC User Portal*. [Online]. Available: https://portal.tacc.utexas.edu/user-guides/stampede2#job-scripts.