USING RECYCLED COMPUTERS TO CONSTRUT A BEOWULF CLUSTER FOR MOLECULAR MODELING

Creston Flemming & Timothy Urness
Department of Mathematics and Computer Science
Drake University
2507 University Ave.
Des Moines, IA 50311
creston.flemming@drake.edu
timothy.urness@drake.edu

Maria Bohorquez
Department of Chemistry
Drake University
2507 University Ave.
Des Moines, IA 50311
maria.bohorquez@drake.edu

Abstract

The process of using recycled computers to construct a Beowulf cluster in order to simulate the molecular dynamics of a chemical system is explored. The cluster was assembled using used computers taken out of service due to upgrades, and would otherwise be discarded. When assembled in a cluster, the computer performed the numerically-intensive process of molecular modeling roughly four times faster than a current machine using synthetic benchmarks. The cluster was used to run a molecular modeling program NAMD. We used the cluster and NAMD to analyze an experiment to study a biologically active lipid called DOPC in the form of a monolayer. The monolayer was constructed using VMD (Visual Molecular Dynamics). It consists of DOPC molecules on the air/water interface. An energy minimization process is carried out before molecular dynamics can be calculated in order to render the system in its most probable physical state.

1 Introduction and Background

The Beowulf Cluster is a form of high performance computing utilizing standard personal computers and Ethernet networking in order to achieve high throughput in parallelizeable computing applications at a low cost compared to traditional supercomputing [1]. The goal was to use computers removed from service in the Mathematics and Computer Science Department at Drake University and reuse them in the Beowulf Cluster. This approach has several unique advantages and challenges. Using recycled computers has the advantage of low cost and environmental impact, as they would otherwise be discarded. This is useful for investigators with limited budgets, especially at smaller colleges and universities. Disadvantages of this approach are reduced speed per node compared with newer hardware.

The program used for the molecular dynamics simulations was NAMD [2]. NAMD was selected because of its open source nature and ease of parallelization. It is ideal for simulating molecular systems with high speed and precision [2]. Several other teams have investigated lipid monolayers and bilayers using NAMD [3][4][5].

A monolayer of DOPC [6] (1,2-Dioleoyl-sn-Glycero-3-phosphocholine) on the air/water interface was modeled. This was chosen due to the use of DOPC in wet experiments. An advantage of using this system is that we can compare the computational results to the actual results of an experiment obtained in a chemistry wet lab.

2 Beowulf cluster

The Beowulf cluster platform consists of hardware and software.

Figure 1: Cluster

2.1 Hardware

Recycled workstations from the Drake University Mathematics and Computer Science Department computer labs were obtained. The hardware is modest by today's standards, with each node consisting of a Intel Celeron 2.0GHz CPU, Intel motherboard with Fast Ethernet, 40GB hard drive, 512MB of DDR266 RAM and a 350W power supply. There are currently a total of 18 nodes. Each node connects via Fast Ethernet to a 26 port switch with gigabit uplink to the cluster server. The server consists of an AMD Athlon 3200+ CPU, 80GB hard drive, 1 GB of RAM, Fast Ethernet card, and Gigabit Ethernet card. The server links into the campus network for remote access.

2.2 Software

Sourcemage Linux was used for the cluster. This distribution was selected because of its extremely modular, source-based system. A minimal installation was made with only the required software for the nodes in order to free the maximum amount of resources for running the simulation. Each node was programmed for network booting from the server to ensure the proper software was available each run. NAMD was compiled and installed. The TCP/ssh version was selected over the MPI and clustermatic versions due to the limited number of nodes and age of the hardware, using these versions with additional overhead over Fast Ethernet would have undoubtedly reduced performance.

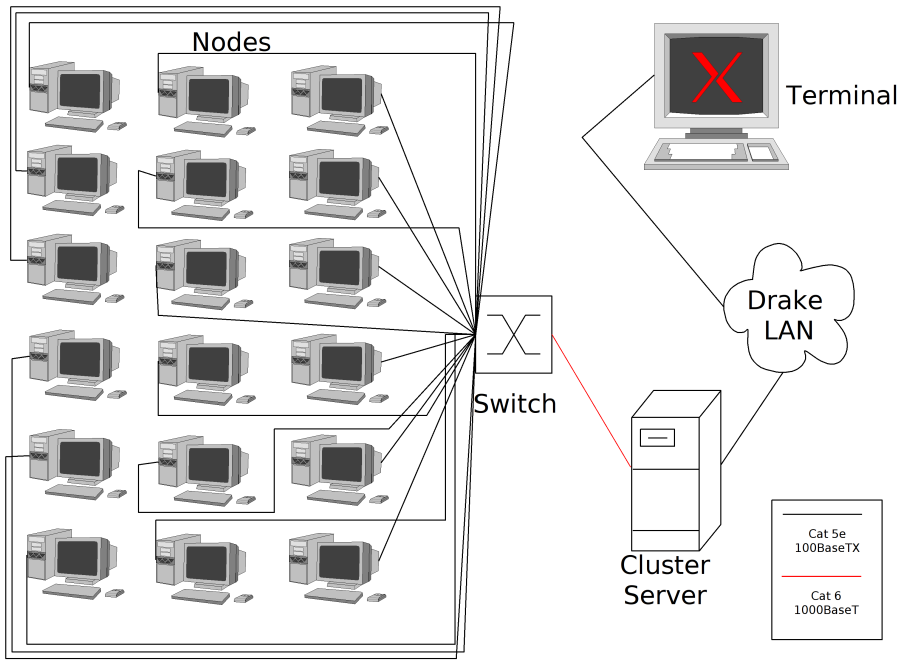

Figure 2: Cluster diagram

2.3 Cluster Performance

Performance advantages over uniprocessor machines was an important consideration

regrading the cluster. Several synthetic benchmarks and real world tests were done to gauge performance.

2.3.1 NAMD

NAMD was used as a benchmark since it was a major component of this investigation. The basic DOPC monolayer simulation was run on several computer systems to gauge performance. This is a useful measure of performance as it tests not only CPU speed, but network overhead costs.

Computer	NAMD WallClock
Intel Pentium 4 3.2GHz	245144s
Intel Core 2 Duo 6600	112778s
AMD Phenem 9850	74567s
Cluster	54791s

Table 1: NAMD benchmark

In running the NAMD benchmark, the cluster has a significant performance advantage over even the high end AMD CPU. It was over twice as fast as the Core 2 Duo, a common CPU found in many current systems.

2.3.2 Passmark CPU benchmarks

The Passmark CPU benchmark is an acclaimed CPU benchmark program. Results are available through the Passmark website [7]. This is a measure of pure CPU performance and does not take into account other factors. The result for the cluster is the multiplication of the base score by the number of nodes.

Computer	Passmark Score
Intel Pentium 4 3.2GHz	495
Intel Core 2 Duo 6600	1398
AMD Phenem 9850	2784
Cluster	4590

Table 2: Passmark benchmark

When it comes to raw CPU power, the cluster is far superior to even the high end quad core CPU. This benchmark would be useful for an application which does not use much network access.

3 Molecular Dynamics

3.1 Construction

The monolayer was constructed using VMD [8]. The monolayer consisted of 40 DOPC molecules positioned on a water box for a total of 17795 atoms. The initial conditions were 9.0 x 12.0 and subsequent simulations were reduced in size to approximate increasing concentration.

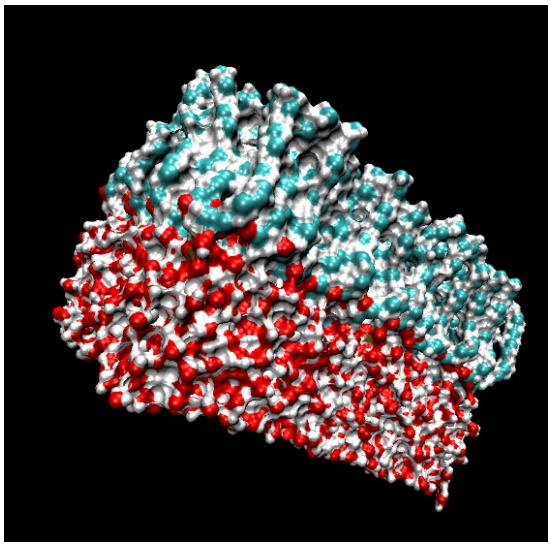

Figure 3: Monolayer

3.2 Simulation and Processing

Molecular dynamics simulations were completed using NAMD on the completed monolayers. Simulation parameters were 300 temperature, 0.5fs time step, PME enabled, and 100000 steps of minimization and 100000 steps of molecular dynamics. After each simulation, the output was processed to remove the minimization steps from the data. The surface tension was calculated by the following formula:

$$\gamma = ((P_{zz} - \frac{1}{2}(P_{xx} + P_{yy})) * L_z) \quad [9]$$

where P_{zz}, P_{xx}, and P_{yy} are pressure components in each direction and L_z is the unit cell length normal to the monolayer. The pressure components derived from the NAMD output file and the unit cell is an intrinsic property of the monolayer.

Surface pressure was then calculated by the change in surface tension.

4.0 Results

The inverse of concentration, with 1 being the baseline simulation is shown in figure 4 and the wet experiment results are shown in figure 5.

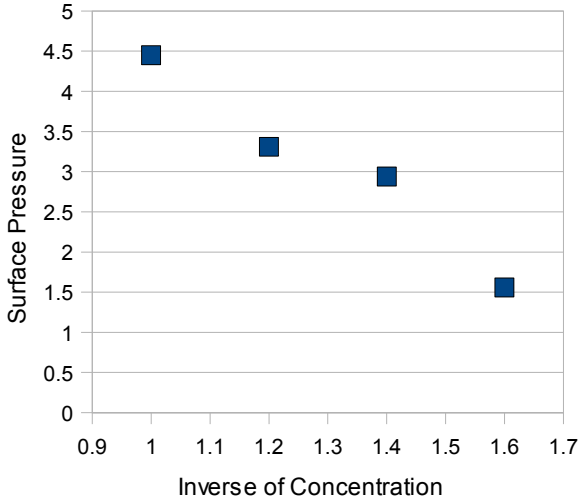

Figure 4: Surface Pressure vs Inverse of Concentration

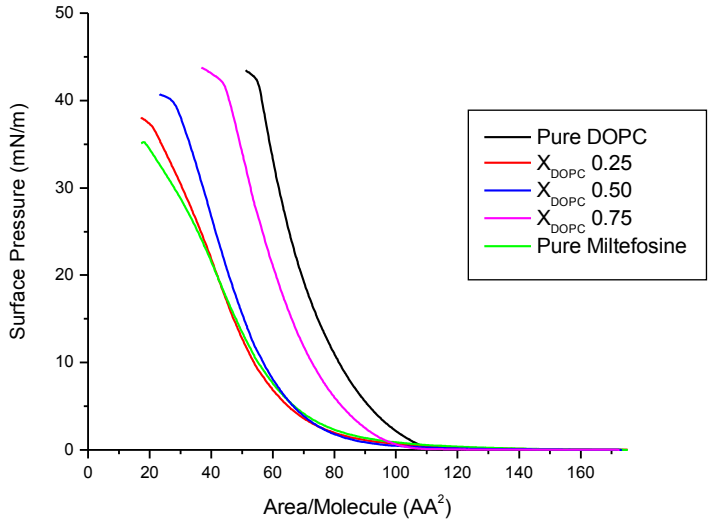

Figure 5: Original wet experiment

5.0 Discussion and Conclusions

Using recycled computers to construct a Beowulf cluster is a cost effective way to increase processing power. The use of eighteen 4-year-old computers provides a significant processing advantage over a contemporary high end computer, especially in situations where minimal data needs to be continually shared among the nodes. Gigabit Ethernet could be used to link the nodes for applications which require high network throughput. Future research could involve determining how performance and efficiency change when additional nodes are added. The effects of a mixed cluster (nodes without identical hardware) could also be investigated, as this is a very real occurrence when using recycled computers.

Initial results of the molecular dynamics simulation appear promising. The NAMD results were similar to the experimental results, with a general downward slope. The exact correspondence between the NAMD results and the wet lab data has yet to be completely calculated.

References

[1] T. Sterling, D. Becker, D. Savarese, J. Dorband, U. Ranawake, C. Packer, "Beowulf: A Parallel Workstation for Scientific Computation," *Proceedings of the 24th International Conference on Parallel Processing*, 1995.
[2] NAMD, Accessible on the Internet at World Wide Web URL http://www.ks.uiuc.edu/Research/namd/.
[3] T. Woolf, B. Roux, "Molecular dynamics simulation of the gramicidin channel in a phospholipid bilayer," *Proceedings of the National Academy of Science*, *91*, November 1994, pp. 11631-11635.
[4] Y. Kaznessis, S. Kim, R. Larson, "Simulations of Zwitterionic and Anionic Phospholipid Monolayers," *Biophysical Journal*, *82*, April 2002 pp. 1731–1742.
[5] R. Venable, Y. Zhang, B. Hardy, R. Pastor. "Molecular Dynamics Simulations of a Lipid Bilayer and of Hexadecane: An Investigation of Membrane Fluidity," *Science*, *262*, October 1993, pp. 223-226.
[6] DOPC, Accessible on the Internet at World Wide Web URL http://www.ucalgary.ca/~tieleman/download.html.
[7] Passmark, Accessible on the Internet at World Wide Web URL http://www.cpubenchmark.net/.
[8] VMD, Accessible on the Internet at World Wide Web URL http://www.ks.uiuc.edu/Research/vmd/.
[9] P. Adhangale, D. Gaver, "Equation of state for a coarse-grained DPPC monolayer at the air/water interface," *Molecular Physics*, *104*, October 2006, pp. 3011–3019.