Computational Molecular Modeling - Solving Quantum Mechanics Based Calculations Concerning Molecular Structure and Properties

Blue Waters Student Internship Program
Education Allocation Proposal

Team Members

PI: **Jeremy Evert** Assistant Professor Southwestern Oklahoma State University jeremy.evert@swosu.edu VOX: +1-580-774-7050 FAX: +1-580-774-7067	Student: Prabhjyot Saluja Computer Science, Junior Southwestern Oklahoma State University salujap@student.swosu.edu VOX: +1-580-302-2063

Event Abstract

This research is in partial fulfillment of the requirements for the Blue Waters Student Internship Program (BWSIP). Faculty and students at Southwestern Oklahoma State University are continuing their research in computational molecular modeling. This research will employ High Performance Computing (HPC) to solve quantum mechanics based calculations concerning molecular structures and properties. This research is under the guidance of Southwestern Oklahoma State University (SWOSU) Chemistry Faculty Dr. Doug Linder and Dr. William Kelly. The research is based on previous publications by Dr. Kelly and Dr. Linder. One component of the student internship is to port the existing code from previous research to the Blue Waters supercomputer. Previous code was written using Linda 8 and PVM. These codes are proprietary and not available on the Blue Waters system. The result of this project should be a software package which can utilize open source packages and libraries able to run on the Blue Waters supercomputer. This will significantly decrease the time to discovery for Chemistry research at SWOSU and provide a meaningful HPC experience for the student.

Project Description

Fields of study:

- Petascale computing education
- High-performance computing
- Parallel and distributed computing
- Computational science
- Molecular structure modeling
- Quantum mechanics calculations

Learning goals:

- The student will learn to collaborate with a research team in order to apply HPC optimization techniques to reduce the time to discovery in molecular modeling. The student will learn how to work on a team with faculty who are domain areas that differ from the student's background, but still desire to use HPC resources.
- The student will learn to apply core HPC concepts including: visualization; profiling; debugging; GPGPU acceleration; hybrid parallelization; and other general code performance optimization.
- The student will increase proficiency in the use of Blue Waters system with a focus on tools and procedures related to the numerical analysis of chemistry centric problems.
- The student will expand his understanding of the current state of research in computational Chemistry and molecular modeling.

How the project will allow the student to achieve these goals:

- By working with the SWOSU Chemistry faculty, the student will learn how HPC can be applied to solve a broader range of questions.
- The student will participate in the One Oklahoma Cyber Infrastructure (OneOCII) weekly phone calls. By working with SWOSU Computer Science faculty and the OneOCII team, the student will learn how to collaborate with larger teams and have professional discussions about technical questions.
- In the course of the project, the student will use the domain experts from HPC and Chemistry to reduce the time and cost to discovery. This experience as a front line HPC practitioner will enable the student to achieve the learning goals listed.

Grade level:

- The student is an undergraduate.

Instructor experience:

- Dr. Evert has been an assistant professor for the School of Business and Technology at Southwestern Oklahoma State University teaching courses in Computer Science at the undergraduate level since 2015
- Dr. Evert received $3K of travel funds for SWOSU students and faculty from the Oklahoma NASA EPSCoR space grant in 2015
- Dr. Evert received $17K of Institutional Organized Research funds for parts, labor, and travel to support the develop an HPC team at SWOSU from the Dean of Professional and Graduate Studies for fiscal year 2016
- Dr. Evert received $36K of student labor funds for SWOSU students from the Oklahoma NASA EPSCoR space grant Research Initiation Grant for the summer of 2016
- Dr. Evert is a member of OneOCII
- Dr. Evert presented at the 2015 Oklahoma Supercomputing Symposium on the topic of

Computer Science at SWOSU
- Dr. Evert presented at the 2015 Oklahoma Native American Students in Higher Education on the topic of HPC in Oklahoma
- Dr. Evert was a course facilitator for the XSEDE online course on parallel programming
- Dr. Evert earned a Ph.D. in Electrical and Computer Engineering at Oklahoma State University
- Dr. Evert completed two days of training from NVIDIA on CUDA programming
- Dr. Evert attended two days of Software Carpentry
- Dr. Evert attended four days of the XSEDE Summer Boot Camp for HPC
- Dr. Evert attended one day of training on profiling and debugging with Allinea

Dr. Evert was faculty sponsor for the SWOSU HPC team participation in the first annual Oklahoma HPC competition

Student experience:

- The student has participated in the two-week intensive Petascale Institute at NCSA. The student is well-prepared in using the tools and techniques of parallel programming (e.g. OpenMP, MPI, OpenACC, CUDA, hybrid), in addition to having hands-on experience with the Blue Waters system. The student is learning disciplines of chemistry like molecular modeling, visualization, data analysis, solving computational problems using quantum mechanics, etc.
- The student attended four days of the XSEDE Summer Boot Camp for HPC
- The student attended one day of training on profiling and debugging with Allinea
- The student attended a semester long XSEDE online course on parallel programming
- The student was programming team captain for the SWOSU HPC team in the first annual Oklahoma HPC competition

Resources Required

Number of node-hours:

- We will need ~ 25,000 node-hours. This comes from the following calculation:
  - 12 weeks during summer * 10 jobs per week * 50 nodes per job * 2 hours per job + 32 weeks during the school year * 4 jobs per week * 50 nodes per job * 2 hours per job

Balance between x86 CPUs (XE) and Kepler GPUs (XK) needed for runs:

- Roughly ~75% of the jobs will run on XE nodes.
- Roughly ~25% of the jobs will run on XK nodes.

Student tasks:

- The student will meet with the SWOSU Computer Science mentor twice per week to discuss his work with the research team and take notes during the meetings. The student will store his notes on a GitHub repository.
- The student will provide weekly reports on his progress. Student will store their reports on a GitHub repository.
- The student will participate in the One Oklahoma Cyber Infrastructure (OneOCII) weekly phone calls. The student will store the shared Etherpad notes on a GitHub repository.
- The student will work with the Chemistry researchers to comment and document existing proprietary code. The student will store his old code on a GitHub repository.
- The student will manage access to and maintain a GitHub repository to track the progress of his project and share his work across the research team, enabling outside members to offer suggestions on how to port the legacy proprietary code to current open source libraries.
- The student will act as cluster administrator for the current SWOSU Intel HPC cluster to prototype and test his open source code. Cluster administration notes will be maintained on a GitHub repository.
- The student will transfer his open source code to the Blue Waters supercomputer and use Git for source code management. The student will maintain working notes on his progress in a GitHub repository.
- The student will use the visualization, profiling, and debugging tools on Blue Waters to optimize the code using OpenMP to use the full potential of a single XE node.
- The student will use MPI to expand his code to use MPI to run across multiple XE nodes.
- This student will explore the use of OpenACC on a Blue Waters XK node.
- The student will pursue conference publication opportunities in the areas of HPC and molecular modeling.
- The student will explore journal publication opportunities to pursue for joint publications with his internship mentor and research mentors.
- The student will make a presentation over his experience to the SWOSU computer club.
- The student will make a presentation over his experiences at the Oklahoma Supercomputing Symposium.
- The student will make a presentation over his experiences at the SWOSU Education career and resources fair.
- The student will have a poster presentation at the Oklahoma Research Day.

Benefits of using Blue Waters over other resources:

- The Blue Waters system is a world class machine with general name recognition in the HPC community.
- The significant investment required to build the system is a fun and exciting way to begin conversations and presentations about high performance computing in Oklahoma.
- Blue Waters is a national resource available to many academic researchers. By using this

national asset, researchers at other institutions could clone the GitHub repository and continue the research without requiring additional time and effort to port the code to a different machine with a different set of compiler modules new set of optimal hardware selections.

Storage / data access requirements:

- The existing files to support research at SWOSU fill 8TB of storage. However, education and training accounts on Blue Waters are limited to less than 1 TB of spinning disk space. The team will most likely use a significant portion of the available storage space for modeling results.

Plans for using parallel programming methods:

- The code will begin as serial code using open source libraries.
- The first step will be to measure code acceleration using OpenMP to quantify the capabilities of a single XE node.
- The second step will be to rewrite the code to use MPI and OpenMP to quantify code performance over multiple XE nodes.
- The third step will be to evaluate the possibility of GPGPU acceleration by using OpenACC and XK nodes.
- If OpenACC suggests the possibility of using GPGPU acceleration would produce a significant reduction in time to discovery, the code will be rewritten to use hybrid parallelization using MPI, OpenMP, and CUDA. The expectation is the use of CUDA would allow us to get the maximum acceleration possible using the XK nodes. This is also of interest to the research mentor who is interested in the ratio of wall clock time decrease due to accelerated code to the number of hours spent writing the code to get the acceleration.

Level of scaling:

- The initial code translation from proprietary legacy code to modern open source code will be produced through iterative runs on the local SWOSU Intel clusters.
- Once an open source code has been produced locally, the code will be transitioned to Blue Waters.
- The student will measure the performance of the code from one XE node up to fifty XE nodes using OpenMP and MPI.
- The student will begin running new experiments for the Chemistry team after they have established a way to run simulations using openMP and MPI.
- The student then begin to measure performance of the code on a single XK node using OpenACC to see if GPU acceleration is a viable option.
- If the OpenACC code suggests a high return on time investment, the student will measure the code performance using OpenMP, MPI, and CUDA on one to fifty XK nodes.

- After establishing the optimal hardware configuration for this problem, the student will use this collaboration to continue to run additional simulations for the research team, generating the necessary results for the Chemistry research.

Software requirements:

- We do not have any special software requirements beyond what is currently available on Blue Waters

Project Timeline

- For June 2016, the student will interact with the SWOSU HPC team, attend additional training on HPC, and coordinate with the SWOSU Chemistry researchers.
- For July 2016, the student will work with the SWOSU Chemistry researchers to apply their HPC training to accelerate the proprietary code and document it as much as possible. The student will also install the canned software, Quantum Espresso and NWChem on the SWOSU Intel Cluster. The student will use these resources to begin the process of porting existing code and calculations to use open source libraries. The student will develop several validation experiments to show that the existing legacy proprietary code and the modern open source code provide the same results.
- In August 2016, the student should complete a functional copy of code that can run serially on the Blue Waters system using interactive calls. The student should then begin the process of optimizing the code to reduce the time to science and developing batch calls to complete validation experiments.
- In September 2016, the student will use iterative calls to optimize the code to run with OpenMP. If a complete single experiment can be run on a single node using OpenMP, the student will collaborate with the SWOSU Chemistry researchers to begin new experiments.
- In October and November 2016, the student will optimize the code using iterative calls to first run the code with MPI and then MPI with OpenMP.
- In December and over winter break, the student will work with the SWOSU Chemistry researchers to start new research experiments using MPI and OpenMP.
- In January and February 2017, the student will use batch calls to run new experiments and continue the research. At the same time, the student will use interactive calls to test the applicability of GPGPU accelerators on validation experiments. The GPGPU accelerators will be accessed using OpenACC.
- Beginning in March and April 2017, the student will examine his results using visualization software on Blue Waters, including ParaView and Quantum Espresso. The student will reduce his focus on developing new code and experiments and increase his focus on preparing conference and journal publications.