NORTHEAST CYBERTEAM PROGRAM

Call for Participation

The Northeast Cyberteam program is a new initiative, funded by the National Science Foundation, to build a regional pool of research computing facilitators (RCFs) to support researchers at small-to-midsized institutions. We are launching our first wave of projects, and seeking a few exceptional individuals to participate!

We are looking for undergraduate or graduate students who have some research computing experience, and are interested in learning more. If matched to a project, you will be assigned a mentor, and together you will provide research computing assistance to a researcher working on a computationally intensive project. A time commitment of 20 hours per week for 3 months is required and a stipend ranging from $4500-$6000 is available. The location of each researcher is noted, but co-location is not required in most cases. We encourage qualified women and minorities to participate, as increasing diversity in the RCF community is also a program goal.

To apply for any of the projects listed below; join the mailing list for future project opportunities; or to volunteer as a mentor, please send a brief bio and your CV to: Julie Ma, jma@mghpcc.org.

Projects

Project Title	Student Profile	Target Start Date	Location
Using Genetic Algorithms & Support Vector Machines in Forest Mapping	An undergraduate student familiar with Matlab, C, and High Performance Computing	July, 2017	University of Maine
Using Station Data and Downscaled Reanalysis to Assess the Occurence of Extreme Weather	An undergraduate student with a knowledge of programming and data handling	July, 2017	University of Maine
Genetics, Cognition and Emotion	A graduate student with expertise in dealing with large data sets	July, 2017	University of Vermont
Centralizing Resources for Sea Floor Mapping	Undergraduates with an interest in system design and programming HPC devices. A lack of a fear of working with hardware and software is most desired	July, 2017	University of New Hampshire
Developing a Regional HPC Help Desk	Graduate student/Post Doc in any area of Computational Science and Engineering, with knowledge of programming and applications	August 2017	Worcester Polytechnic Institute
Developing Computational Labs for an Upper Level Physical Chemistry Course	An undergraduate/graduate student with chemistry focus that has taken physical chemistry courses; has knowledge of computational modeling and experience with Gaussian	September. 2017	Bridgewater State University

What is the Northeast Cyberteam Program?

The Northeast Cyberteam Program is a 3-year initiative funded by the National Science Foundation to build a *regional* pool of Research Computing Facilitators to support researchers at small-to-midsized institutions in Maine, Massachusetts, New Hampshire and Vermont, leveraging the work of national programs including XSEDE Campus Champions, ACI-REF and others.

Research Computimg Facilitators (RCFs) are experts at figuring out how to match the right compute resources to the task at hand, something that can stymie researchers who are, for example, sifting through billions of records to find a specific pattern of genes that correlates with a particular form of cancer; or examining massive quantities of sensor data to understand movements on the sea floor. The RCF's job is to help make use of local, regional, and national high performance computing resources when computing needs exceed the capacity of the scientist's desktop. RCFs can often be found in the research computing groups at large universities and corporations, but are scarce at smaller institutions.

Recognizing that promising research can be stopped in its tracks without high performance computing when the need arises, the Northeast Cyberteam Program was created to fill the gap. Over the next three years, the program will support 42 compute-intensive projects with RCFs-in-training, each paired with a mentor, to facilitate research computing needs for a 3 month period. RCFs-in-training will also have the opportunity to work on a live help desk with a mentor, honing their consultative skills while getting exposure to a broad range of research computing topics. As part of the program, RCFs will become part of a community of facilitators that has up-to-the-minute visibility into research computing projects and programs taking place in the region.

Leadership

The Northeast Cyberteam is a collaborative effort by led by the Massachusetts Green HIgh Performance Computing Center (MGHPCC), University of Maine, University of New Hampshire, and University of Vermont, with support from the University of Massachusetts and Worcester Polytechnic Institute. The program is modeled after other regional collaborations that have emerged since the opening of MGHPCC, a consortium created in 2013 by Harvard University, the Massachusetts Institute of Technology, Boston University, Northeastern University and the University of Massachusetts to build a high-performance data center that now houses the largest aggregation of research computing power in the Northeast.

MASSACHUSETTS GREEN HIGH PERFORMANCE COMPUTING CENTER

Want to know more? Please contact: **Julie Ma**
NE Cyberteam Program
www.mghpcc.org/NE_Cyberteam

100 Bigelow Street
Holyoke, Massachusetts 01040
413-552-4967
jma@mghpcc.org

Detailed Project Descriptions

1) Using Genetic Algorithms and Support Vector Machines in Forest Mapping

Satellite-derived maps of forest conditions play diverse roles in research and resource management. Maps provide a basis for planning and executing field studies, developing and calibrating models, quantifying ecosystem processes or services, and evaluating environmental change. However, decisions are made within the context of map error, and methods used to produce maps generally result in patterns of error that are potentially detrimental. We have developed machine learning techniques that effectively reduce undesirable systematic error when mapping forest attributes from satellite imagery and geospatial data. Our approach is based on the optimization of support vector machines using a multi-objective genetic algorithm (GA) designed to simultaneously minimize both total and systematic error. Using this approach, we have obtained outcomes that compare well against other mapping approaches. Our algorithms are, however, computationally demanding, and large-scale applications will require more effective use of computing resources. This project will develop software that enables statewide and regional application of our algorithms through enhanced parallelization and new approaches to coordinate and accelerate the convergence of GAs. Algorithm improvements will be coupled with more efficient data handling. The primary project outcome will be software that supports locally adaptive mapping of forest resources and environmental conditions across large spatial scales through innovative algorithms and efficient use of available cyberinfrastructure.

2) Using Station Data and Downscaled Reanalysis to Assess Occurrence of Extreme Weather

Meteorological observations across North America and Europe suggest a significant increase in the frequency and intensity of extreme weather (heat waves, cold waves, precipitation events) coincident with satellite-measured major decline of Arctic sea ice over the past decade. This project will assess the occurrence and impact of extreme weather events across Northern New England using both station data and climate reanalysis models. Weather and climate are critically important across Northeast New England, owing to the heavy reliance of natural resources for its economy.

3) Genetics, Emotion and Cognition

The overarching goal of this project is to identify a predictive, quantitative framework describing individual differences in genetic, epigenetic, cognitive, and behavioral markers of emotion-cognition regulation in response to academically stressful situations. Each year, large numbers of young adults drop out of college and university due to self-sabotaging and seemingly irrational behaviors when faced with academic stressors in their young adulthood. This proposal utilizes a cross-disciplinary approach to understanding neuro-biological functionalities and resultant behaviors across a spectrum of neuro-typical and neuro-atypical young adults, the latter being identified as those with diagnosed learning disabilities, such as dyslexia, ADHD, and college-able autism. This project-partnership includes faculty and students from

the University of Vermont (sequencing data analyses), Landmark College (research subject recruitment), University of New Hampshire (research subject recruitment), University of Maine (model simulation), and Vermont Genetics Network. Some trial work has been done at MGHPCC, promising results. However, a more efficient workflow will be needed to generate full scale results -- running a single sample currently requires 2TB storage and 5 days of processing with 64GB memory and 12 cores, and the planned project has 3,000 samples.

4) Centralizing Computation Resources for Sea Floor Mapping

This project will assist in the migration of data and computational resources from departmental facilities located in the Chase OE Building to the Research Computing Center in Morse Hall using HPC resources in the Lenharth Data Center.

5) Developing a Regional Research Computing Help Desk

Participate in an effort to create/pilot a regional research computing help desk system that can respond to researchers in a timely fashion; quickly direct them to appropriate and available compute resources; and/or match them with people with relevant scientific/computing expertise to help them move their work forward when experiencing research computing issues.

6) Developing Computational Labs for an Upper Level Physical Chemistry II Course

Out of all the upper level chemistry courses, physical chemistry is the only course that provides an in-depth insight into the fundamental principles underpinning the concepts taught in various sub-disciplines of chemistry. Further, physical chemistry provides a connection between microscopic and macroscopic worlds of chemistry through mathematical models and experimental methods to test the validity of those models. Therefore, computational techniques are a perfect vehicle to teach physical chemistry to undergraduate students. Additionally, the American Chemical Society recommends computational chemistry to be incorporated into the undergraduate chemistry curriculum. At Bridgewater State University (BSU) physical chemistry is a two-semester course referred to as 'physical chemistry I' and 'physical chemistry II'. While the overarching goal is to develop computational experiments (referred to as 'dry-labs'), the project proposed here focuses on designing and developing dry labs for 'Physical Chemistry II' course at BSU. The inherently theoretical nature of this course along with its connection to wide range of spectroscopic techniques commonly used by chemists and physicists makes this course a perfect choice for assessing BSU students' reception to the idea of dry labs. Currently, there are no computational experiments in the current physical chemistry curriculum at BSU. The proposed project focuses on developing 4 – 6 computational experiments to be introduced (in spring 2018) as either stand-alone dry-lab experiments or accompany currently existing experiments. These dry labs will be developed on Gaussian 09 platform, currently installed on C3DDB server at MGHPCC. A longer term goal is to make these experiments available to other New England instructors interested in incorporating computational chemistry into their physical chemistry II courses (or equivalents).