

Modeling Advanced Materials and Systems Biology: Building Capabilities and Collaborations for Cyber- Enabled Discovery

CONFERENCE
September 20-22, 2010

**Clemson University,
Madren Center**





EPSCoR RII Track-2 "An EPSCoR Desktop to Teragrid Ecosystem"

MONDAY SEPT 20

1:00 PM	2:00 PM	Pre-Conference Hands-On Training Session: Using the Palmetto Cluster – Intro to Unix commands (Barre Hall Room 105, Clemson Campus)	Barr von Oehsen, Galen Collier
2:00 PM	5:00 PM	Pre-Conference Hands-On Training Session: Using the Palmetto Cluster—compiling and running jobs; parallelize code; using storage (Barre Hall Room 105, Clemson Campus)	Barr von Oehsen, Galen Collier
6:00 PM	8:00 PM	WELCOME RECEPTION (Madren Center, North Lobby)	

TUESDAY SEPT 21

9:00 AM	9:05 AM	WELCOME All presentations in BellSouth Auditorium, Madren Conference Center	 Jill Gemmill Executive Director Cyberinfrastructure Integration, Clemson University
9:05 AM	9:45 AM	"Shaping the Future with CI" Jim Bottum will discuss the vital structure of funding and strategic partnerships as a solid foundation for growing cyberinfrastructure, to leverage and enable science and education in South Carolina. He will also touch on current activities and future strategic direction among leading organizations in the CI field.	 Jim Bottum Vice Provost and Chief Information Officer, Clemson University

9:45 AM	10:20 AM	Research Collaboration opportunities with the UTK Joint Institute for Computational Sciences (JICS) at Oak Ridge National Laboratory	 <p>A. Barney Maccabe Director, Computer Science and Math Division Oak Ridge National Laboratory</p>
MORNING BREAK – 30 minutes			
10:50 AM	11:30 AM	<p>“Investment in High Performance Computing as a Predictor of Research Competitiveness in U.S. Academic Institutions”</p> <p>Cyberinfrastructure is the information technology that, by definition, enables scientific inquiry. High Performance Computing (HPC) is one of the key cyberinfrastructure components. This talk traces the development of HPC in U.S. academic institutions, and describes how the Top 500 list can be used to measure the investment that institutions have made in HPC. In any given Top 500 list, roughly 30 U.S. academic institutions have an entry on the Top 500 list, and this number has been dropping in recent years. With support of a grant from the National Science Foundation to the University of Arkansas and RENCi (UNC-CH), we have studied how investments in HPC can be shown to be leading indicators of the research productivity of an institution. Results show that investment in HPC can increase the average NSF funding to an institution by a statistically significant amount. We also discuss the various costs to invest in HPC, and suggest supporting factors in making an HPC investment have a full impact.</p>	 <p>Amy Apon Professor, Computer Science and Computer Engineering; Arkansas High Performance Computing Center, University of Arkansas</p>
11:30 AM	12:10 PM	<p>“Computational Nanoscience and the Interplay Between Experiment and Theory”</p> <p>Theory and simulation have played, and continue to play, a central role in nanoscience. In fact, it can be argued that theory and simulation play a greater role in nanoscience than in the traditional, macroscopic materials and chemical sciences for at least three reasons: first, many experiments performed at the nanoscale can only be interpreted through theory; second, theory and simulation can provide a convenient framework to isolate effects and phenomena in a way that may be difficult or impossible to achieve in an experiment (i.e., in theory and/or simulation, the boundary and initial conditions are under complete control, which may be impossible to achieve in an experiment), thus making theory and simulation a crucial tool in understanding emergent phenomena in nanoscale systems; and, finally, theory and simulation can be used to design new nanostructured materials, as well as systems based on nanoscale phenomena.</p> <p>In this presentation, these roles of theory and simulation will be illustrated through examples from the published literature, from the speaker’s own research program, from user research projects at the Center for Nanophase Materials Sciences (CNMS), and from the CNMS internal scientific research program. The latter has three major themes: imaging nanoscale functionality, synthesis and dynamics of nanostructured polymeric and hybrid materials, and emergent behavior in nanoscale systems. Theory and simulation play prominent roles in all three themes, in addition to leading the emergent behavior theme.</p>	 <p>Peter T. Cummings John R. Hall Professor, Department of Chemical and Biomolecular Engineering, Vanderbilt University</p> <p>Principal Scientist, Center for Nanophase Materials Sciences, Oak Ridge National Laboratory</p>

12:10 PM

12:45 PM

LUNCH – MEETING ROOMS 3 & 4

12:45 PM

1:25 PM

LUNCHTIME SPEAKER:**"Cyberinfrastructure for the Humanities, Arts, and Social Sciences (HASS)"**

This talk will introduce a framework for cyberinfrastructure for the humanities, arts, and social science (HASS) disciplines that is rooted in the history and challenges of those disciplines, as well as the working practices of faculty scholarship. It also describes the work of The Institute for Computing in Humanities, Arts, and Social Science (I-CHASS) at the University of Illinois at Urbana-Champaign, which serves the national research and education communities by making tools available for high-performance computing, communication and collaboration, data collection and analysis, geospatial inquiry, and visualization.



Vernon Burton
Professor of History
Associate Director for
Humanities, Arts, and
Social Sciences
Clemson Cyber Institute
Clemson University

1:30 PM

3:00 PM

**Special “Birds of a Feather” Session on
Cyberinfrastructure for the Humanities, Arts, and Social
Sciences**, led by Vernon Burton (remain in dining area)

Topic: What can we do in HASS computing at Clemson? Brainstorm ideas for regional collaborations through the CU Cyber Institute. If you are interested in HASS computing, we would love your attendance.

1:30 PM

3:00 PM

PANEL: CyberEnabled Applications in Structural Biology
Nicholas Panasik, Session Chair

**“Entropic Stabilization in Thermostability; Enumerating
Conformations in the Unfolded State”**

Using LINUS, a hard sphere sterics, ab initio protein modeling and fold-prediction algorithm, a library of allowed conformations for N-Acetyl-(Ala)₄-Methyl-ester has been created by iteratively progressing through all possible allowed phi/psi dihedral angle combinations and discarding only those conformations which had steric clash or which had one or more polar backbone atoms that could not be hydrogen bonded to another atom of the peptide backbone or with a water molecule. This method generated a library of 53 million allowed polyalanyl backbone conformations that were fully hydrogen bond satisfied and demonstrated no steric clash. All possible side chain rotamer conformations from a mesophilic and a thermophilic related, super-imposable crystal structure are then modeled on these backbone conformations to enumerate all possible conformations for both the mesophilic and thermophilic sequences. These libraries are compared for their ability to rule out non-native-like conformations in the unfolded state while retaining conformational space in the folded state and provide a guide to thermodynamic origins of thermostability for certain mutations.



Nicholas Panasik
Assistant Professor, Dept.
of Biology & Chemistry,
Claflin University

**“Simulation of Protein-Surface Interactions: Problems,
Solutions, and Future Directions”**

The adsorption behavior of proteins on material surfaces is of fundamental importance in essentially any technology that involves the interaction between protein-containing fluids and synthetic materials. Examples include the design of biocompatible medical implants and substrates for tissue engineering and regenerative medicine, biosensors, nonfouling surfaces for food processing and marine applications, and the development of improved systems for bioanalytical assays, bioseparations, biocatalysis, and biodefense. Because of the great complexity of protein adsorption behavior,



Robert A. Latour
McQueen-Quattlebaum
Professor
Department of

experimental methods are extremely limited in their ability to provide an atomic-level understanding of protein adsorption behavior. Molecular simulation methods, on the other hand, are very well suited to address this type of problem, with these capabilities continuing to expand each year with increasing computational resources. These types of methods, however, require careful development, evaluation, and validation before they can be confidently applied to predict protein adsorption behavior. In this talk, I will provide an overview of the main problems that must be overcome to accurately simulate protein adsorption behavior, the solutions that we have found to address these problems, and directions for further development. Once these methods are adequately developed, they have the potential to serve as a powerful tool for the design of surfaces to directly control protein adsorption behavior to provide improved capabilities for a wide range of biotechnologies.

"Estimating long-distance seed dispersal using an spatially-explicit agent-based simulation"

Long-distance dispersal (LDD) is the mechanism by which organisms re-populate extinct habitats and invade new habitats. Though an understanding of this process is important for understanding the distribution of organisms, LDD is essentially impossible to observe directly. A potentially powerful technique is to infer LDD using the distribution of genetic markers in natural populations. Unfortunately most techniques make several simplifying assumptions that render traditional genetic estimates of dispersal inaccurate at best. In this presentation I show an alternative approach that implements an agent-based simulation approach to estimation of LDD parameters. These simulations are parameterized by independent ecological information and predict the distributions of genotypes across the landscape, given different dispersal scenarios. To make the inference, I try different dispersal scenarios and match the predicted genotype distribution to the distribution observed in nature. The scenario that best matches nature is taken as the best estimate of LDD. After describing the estimation framework, I apply this approach to two case studies: black cherry, one of the 100 most invasive plants in Europe, and sea rocket, one of the world's most widely dispersed plants.



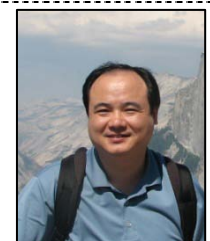
Allen E. Strand
Department of Biology
Grice Marine Laboratory
College of Charleston



Greg Peterson
University of Tennessee-
Knoxville

"Understanding Whole-genome Evolution through Petascale Simulation"

Inferring evolutionary relationships (especially from rearrangement events) is computational-intensive, and the challenge to reconstruct moderate-sized phylogenies taxes the largest computing systems available today. The availability of petascale computing systems provides a unique opportunity to accelerate the development of petascale codes for assembling the Tree of Life. We will discuss the problems involved in genome rearrangement analysis and our approaches in designing scalable algorithms that use petascale computers to infer ancestral gene orders from gene-order data and apply these analyses to several key biological problems. The newly sequenced 12 *Drosophila* genomes provides an important model for testing these new algorithms because the majority of rearrangements occur within chromosome arm and the phylogeny of the 12 species is



Jijun Tang
Associate Professor, Dept.
of Computer Science &
Engineering, University of
South Carolina

well-established. The new petacodes can then be applied to more complex sets of genomes where rearrangements occur within and among chromosomes such as mammalian genomes.

“Identify Polygenetic Disease Models from an Ontology Fingerprints Derived Gene Network for Genome Wide Association Study”

We developed a new approach to identify gene network modules as potential polygenic models for Genome wide association (GWA) study. The novel gene networks were derived by comparing ontology fingerprints of human candidate genes linking to the lipid level in plasma based on a GWA study. From this network, we were able to identify fully connected components as potential polygenic models linking to lipid concentration in plasma. Our methodology provides a novel way to reduce the search space for potential polygenic models that may contribute to human disease.



W. Jim Zheng
Assistant Professor of
Bioinformatics
Department of
Biochemistry
Medical University of
South Carolina

3:00 PM

3:30 PM

AFTERNOON BREAK – 30 minutes

3:30 PM

4:15 PM

PANEL: Speedup through hybrid computer architectures
Barr von Oehsen, Session Chair



Barr von Oehsen
Director of Computational Science
Clemson University

“GPGPU@CU”

We provide an overview of GPU-based research in modeling and rendering natural phenomena that has recently been conducted in the School of Computing at Clemson. Dramatic performance improvements over optimized (multi-threaded, SSE) CPU-based solutions are seen to be available, but such can only be achieved through careful control of memory access patterns. Examples from photon transport and ocean wave modeling are provided.



Robert Geist
Professor, School of
Computing
Clemson University

“Emerging Architectures for Computational Science”

Traditional microprocessors no longer can achieve the performance expected by computational scientists. Consequently, multicore processors, graphical processing units (GPUs), reconfigurable computers, and hybrid combinations of these technologies represent the most promising computational platforms for the future. We survey what scientific application developers and users can expect from these technologies with respect to performance, cost, ease of use, and potential challenges.



Greg Peterson
University of Tennessee
Knoxville

“Accelerators to Applications”

Insights into the use of FPGA and GPU accelerators for scientific applications.



Melissa Smith

Assistant Professor,
Computer and Electrical
Engineering
Clemson University



Jason D. Bakos
University of South
Carolina

"Heterogeneous Computing at USC"

The Heterogeneous and Reconfigurable Computing (HeRC) group at USC is exploring applications, tools, and platforms for high-performance co-processor-based computing. The group has adapted a variety of applications to heterogeneous platforms, including computational phylogenetics, genomic clustering, sparse matrix solvers, logic synthesis, and frequent itemset mining. They are also working to develop tools for application partitioning, high-level synthesis, and micro-architectural simulation. The group has worked with FPGA and GPU platforms from Annapolis Micro Systems, GiDEL, Convey Computer, and NVIDIA. This talk will give a brief overview of this work.

4:15

5:00 PM

"The Gigantic Problem of Genomic Data Growth, and its Solutions"

The insufficient capability to translate the exponentially growing genomic data into useful knowledge is the single-most pressing grand challenge in biology. With the advent of next generation sequencing machines, genomic data is growing at a far superior rate compared to the enhancements in the computer architectures. The genomic data is currently doubling every six months when compared to Moore's law, where the processing power is doubling every eighteen months. Thus the knowledge discovery process can only be sped up by developing sensitive data mining tools along with highly scalable bioinformatics tools organized in robust, massively parallel computation infrastructures. This presentation will illustrate the importance of porting these tools to High Performance Computing architectures such as supercomputers, hybrid architectures with GPGPUs and cloud computing machines, thus enabling mining with not only the existing genomic data, but also sustain future data explosion. This process will yield substantial technology enhancements in the fields of biomedicine and bioenergy.



Bhanu Rekepalli
Research Scientist,
National Institute for
Computational Sciences,
University of Tennessee
Knoxville-Oak Ridge
National Laboratory,

5:00 PM

5:30 PM

An Introduction to the "Desktop to TeraGrid" Science Gateway

The Cyberinfrastructure Technology Integration (CITI) group at Clemson has developed a web-based collaboration and learning environment for use by participants in the EPSCoR "Desktop to TeraGrid" program. Based on the HUBzero platform, this science gateway provides easy access to a broad range of resources designed to help users of all experience levels with migrating desktop-scale computational research to more advanced systems such as Clemson's Palmetto cluster and various TeraGrid resources. Much more than a website with information and instructions, the "Desktop to TeraGrid" science gateway is a constantly evolving interface that combines advanced collaboration tools with "live" computational tools in a way that enables program participants to make immediate progress in advancing their computational research.



Galen Collier
Computational Scientist,
Clemson University

5:30 PM

6:00 PM

BREAK

6:00 PM

8:00 PM

POSTER SESSION / hors d'ouvres

7:00 PM

8:30 PM

Birds-of-a-Feather sessions

WEDNESDAY SEPT 22

8:30 AM	8:35 AM	<p>WELCOME</p> <p>All presentations in BellSouth Auditorium, Madren Conference Center</p>	 <p>Mark Fahey University of Tennessee-Knoxville</p>
8:35 AM	9:15 AM	<p>“BioVU Genomics Project Manager at Vanderbilt University”</p> <p>Vanderbilt University has created the Vanderbilt DNA Databank (BioVU), which is a DNA repository of >95,000 DNA samples (both adult and pediatric) collected from discarded blood after routine clinical testing has been completed. The samples within BioVU can be linked with clinical and demographic data within the Synthetic Derivative (SD), which is a de-identified version of electronic medical records. Thus, BioVU creates a centralized resource for making associations between genetic data and various diseases or pharmacogenetic traits across populations. Data present in the SD includes narratives, diagnostic codes, reports, lab values, and medication orders. Using this information, Vanderbilt investigators have access to a sophisticated phenotype searching mechanism for selecting their study datasets for their disease or trait of interest.</p>	 <p>Rebecca Zuvich BioVU Genomics Project Manager, Center for Human Genetics Research Vanderbilt University</p>
9:15 AM	10:00 AM	<p>"Computation aided catalyst design for water and organic oxidation, C-H bond functionalization, nitrogen fixation and drug development"</p> <p>Engineering of catalysts for a particular task requires knowledge of fundamental catalytic principles, understanding of mechanisms and factors that control rates and selectivity during the catalyst turnover. Success in this challenging field requires comprehensive and coordinated experimental and theoretical investigations. In my presentation, at first, I will briefly discuss our ongoing collaborative (experimental and computational) research projects on:</p> <ol style="list-style-type: none"> 1. Design of a visible light-driven robust nano-device for water splitting; 2. Fundamental approaches to the C-H bond functionalization (amination, alkylation, metalation and oxidation); 3. Fundamental aspects and design of late-transition-metal-oxo complexes; 4. Design of microporous catalysts that absorb, detect, and catalyze decontamination of toxic agents; 5. Se-proteins and organo-Se compounds: Search for efficient Se-based anti-cancer drugs. <p>Then, I will present details of our latest results in designing of polyoxometalate based water oxidation catalyst and nano-size device that converges solar energy to chemical energy by splitting water to O₂ and H₂.</p>	 <p>Jamal Musaev Director & Principal Scientist Cherry L. Emerson Center for Scientific Computation, Emory University</p>
10:00 AM	10:30 AM	<p>“Computational Advanced Materials Workbench Tools”</p>	 <p>Jacek Jakowski University of Tennessee-</p>

10:30 AM 11:00 AM

MORNING BREAK – 30 minutes

11:00 AM 12:45 PM

PANEL: Cyber Enabled Materials Sciences Applications

Steven Stuart, Session Chair

“Reactive Molecular Dynamics for Nanomaterials Applications”

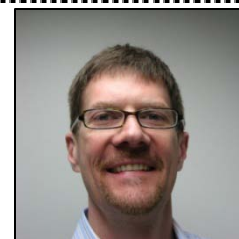
In many nanomaterials applications, computational modeling approaches require that the model be capable of breaking and forming covalent bonds. Consequently, the reactive bond-order potentials developed over the past few decades have found use as a way of treating covalent bond breaking and forming. Because these models are classical, they enable simulations of many thousands of atoms, without the computational expense of a quantum mechanical method. Nonetheless, the simulations can be quite computationally intensive, and benefit from modern high-performance computing architectures. Additionally, it is common for these models to require parameterization when extended to new systems. This parameterization process, conventionally very subjective and dependent on human input, has also begun to yield to parallel computing approaches. An automated, parallel parameterization approach will be described, in which the inherently stochastic nature of the objective function is accounted for and parallel architectures are used in dynamically optimizing potential functions. Applications will be discussed briefly, including formation of carbon nanostructures in energetic molecular beam collisions.



Steven J. Stuart
Professor, Department of
Chemistry
Clemson University

“Computer simulations of high-density quantum solids”

The computation of material properties of crystalline solids (properties such as equations of state, constitutive parameters, and phonon dispersion curves) provides sensitive tests of the underlying many-body potential energy functions used to model the solid. For quantum solids, which are characterized by large-amplitude zero point motions of the constituent atoms or molecules about their nominal crystal lattice sites, there are good reasons to expect that three-body interactions become important at moderately high densities and pressures. We will present a brief summary of our GPU-accelerated simulations of solid 4-He, a prototypical quantum solid, and will compare the properties of the solid when three-body interactions are either included or excluded from the solid's many-body potential energy function. We find that a many-body potential energy function based on recently computed ab initio two- and three-body terms gives properties for solid 4-He in excellent agreement with experiment, thus validating the underlying ab initio calculations.



Robert J. Hinde
Professor, University of
Tennessee-Knoxville

How to Control DNA?

Synthetic nanopores provide a tight confinement constriction which localize and control the motion of a biomolecule, typically submerged in aqueous electrolytic environment. The electric current of ions passing simultaneously with the molecule through the pore, or the electron tunneling current measured transversally through the molecule depends on the structural properties of the molecule, which provides a basis for a nondestructive physical recognition of a single molecule. This approach has been recently recognized as a prospective for development of a cheap and fast method for the DNA sequencing [1]. A possible recognition or sequencing device has to resolve three main challenges: (1) single-molecule or in case of DNA single-base resolution, in presence of thermal and other sources of the uncertainties, (2) electric reader matching of the speed of the molecule translocation through the pore, and (3) massive fabrication of robust and simple single-molecule readers. We are developing, in collaboration with the researchers from the Yale University a nanoscale quadrupole Paul trap in aqueous environment. The molecule is stabilized by a combined action of a radio-frequency



Predrag Krstic
Oak Ridge National
Laboratory

(RF) multipolar trapping electric fields, drag forces in the electrolytic solvent, and lateral electric field which control its translocation through the device. The main advantage of the Paul trap is relaxation of critical dimension control, which eliminates the effects of channel walls. A single-base resolution is predicted by using transverse electron tunneling from N-functionalized CNT electrodes.

In collaboration with researchers from Arizona State and Columbia University we are also developing a control of translocation of a DNA segment using single-walled carbon nanotubes which play the role of nanopores. Unexpected results were obtained in the measured translocation current of electrolyte and biomolecules through the nanotube and these have recently reached a general agreement with theory [2]. [1] D. Branton et al, Nature Biotechnology 26, 1146 (2008). [2] H. Liu et al, Science 327, 64 (2010).

“Approximate quantum-mechanical description of nuclei in reactive molecular systems”

The broad questions we consider are when to take into account the quantum nature of nuclei in chemical processes of large molecular systems and how to estimate the magnitude of quantum effects without expensive quantum-mechanical simulation. Quantum effects on molecular dynamics are essential for understanding of many chemical processes, such as those in surface reactions, photochemistry, in interactions of molecules with electric field, reactions in enzymes and nanomaterials and in bond-selective chemistry. We develop an approximate, thus, scalable to large molecular systems, methodology based on real and imaginary-time dynamics of an ensemble of quantum trajectories representing a nuclear wavefunction. This approach is used to compute thermal reaction rate constants and low-lying energy eigenstates. The dominant quantum, inherently non-local, effects are incorporated into trajectory dynamics through a “mean-field”-type approximation to the quantum force, which can be turned off to assess its importance in a chemical process. The approximate quantum trajectory approach has been applied to reactions of oxygen with molecules and high-dimensional model systems. Large chemical systems will be investigated, when the code is parallelized and computation of electronic energies and gradients on-the-fly becomes realistic.



Sophya Garashchuk
Assistant Professor,
University of South
Carolina

“Large Scale Molecular Dynamics Simulations of Soft Matter”



Gary Grest
Distinguished Member of
Technical Staff, Sandia
National Laboratory

"Shape-Based Similarity Search in Computer-Aided Drug Design"

Molecular similarity search has become a standard virtual screening tool in the biotech and pharmaceutical industry. It premises on the simple similarity principle which states that similar molecules tend to have similar biological activities or physicochemical properties. Even though the principle is simple, it is a much harder task to devise effective computational measures to quantify the molecular similarity



between molecules. Over the past 30 years, various molecular similarity measures have been developed and applied to both chemical information retrieval and the discovery of biologically active compounds. The most commonly used methods are based on molecular fingerprints such as MACCS keys and Daylight fingerprint based Tanimoto measures. These measures are derived from the graph representation of chemical structures and afford only an overall similarity measure without detailed atom-by-atom structural alignment. To obtain such detailed information, 3D methods have been developed (such as SQ, SEAL, ROCS) to address the 3D conformational flexibility of molecules and offer superimposition of molecules as testable hypotheses for chemists and biologists to test experimentally. However, these 3D methods are often computationally expensive due to the needs to have 3D conformers. We devise a simple new scheme (named EZSim) based on Open Eye Scientifics Toolkit to calculate molecular similarity between 2D depictions of molecules. It avoids the problem of conformational analysis, and affords molecular superimposition data that are so critical to chemists' thinking and designing new molecules. We applied this method to Directory of Useful Decoys (DUD) validation dataset (Huang, N. et al. J. Med. Chem. 2006, 49, 6789-801.) and demonstrated the effectiveness of this new 2D similarity method in virtual screening experiments. In this talk, I will briefly give an overview the computational protocol implemented in EZSim, the feasibility of the approach in a distributed computing environment, as well as EZSim's superior computational efficiency when compared to other 3D-based shape matching methods.

Jerry Ebalunode
Scientific Middleware
Engineer, University of
South Carolina

"Quantum dynamics at the nanoscale"

Using the time-dependent density functional theory we study the electron dynamics in nanostructures. The computer simulation of electron dynamics is based on an efficient real space approach using Lagrange functions. The simulations include studies of electron transport, field emission, laser excitation of nanostructures.



Kalman Varga
Vanderbilt University

12:45 PM 2:00 PM

Discussion Groups w. assigned questions / box lunch
Pick up lunch in North Lobby; Discussion Tables in Meeting Rooms III and IV

CONFERENCE ENDS

1:30 PM 4:00 PM

LEADERSHIP TEAM MEETING WITH EXTERNAL ADVISORY BOARD

Seminar Room II