UNIVERSITY OF DELAWARE
DATA SCIENCE INSTITUTE

DELAWARE Data Science

DARWIN COMPUTING SYMPOSIUM

MAR 24, 2022

#DARWINcomputing2022

AGENDA

Registration	8:30-9:00am
Welcome & Introduction	9:00-9:15am
Opening Remarks: Alejandro Suarez, NSF Associate Program Director, OAC	9:15-9:30am
Advancing Research Computing at UD	9:30-9:45am
Session I: Science on DARWIN	9:45-10:45am
Coffee Break	10:45-11:00am
Session II: Democratizing DARWIN	11:00am-12:00pm
Lunch *(Poster Lightning Talks Start at 12:40pm)*	12:00-1:00pm
Poster Session	1:00-2:00pm
Session III: Science on DARWIN	2:00-3:00pm
Session IV: DARWIN for Education	3:00-3:40pm
Coffee Break	3:40-3:55pm
Accessing DARWIN	3:55-4:15pm
Panel and Discussion: DARWIN User Experiences and Accomplishments	4:15-4:55pm
Closing Remarks and Best Poster Award	4:55-5:00pm

INTRODUCTION

Jodi Hadden-Perilla

Co-Chair
DARWIN Computing Symposium Planning Committee
Assistant Professor, Chemistry & Biochemistry
University of Delaware

REMARKS FROM NSF

Alejandro Suarez

Associate Program Director
Office of Advanced Cyberinfrastructure (OAC)
National Science Foundation

ADVANCING RESEARCH COMPUTING AT UD

Sharon Pitt

Vice President for Information Technologies and CIO
University of Delaware

Sharon P. Pitt serves as the Vice President for Information Technologies and Chief Information Officer at the University of Delaware. In this role, Sharon is charged with developing strategic IT initiatives that forward the University's mission. She oversees Information Technologies in support of teaching, learning, research and the student experience at the University.

Sharon came to the University of Delaware after serving as CIO at Binghamton University, a State University of New York. Before that, Sharon served 9 years at George Mason University, most recently as its Deputy Chief Information Officer. Previous to her work at George Mason University, Sharon served as Associate Vice Provost for Learning Technologies at North Carolina State University, Director of Instructional Technology and Distance Education at James Madison University, and in the Educational Technologies division of Virginia Tech. She also was an Executive Director of 4VA (4VA.org).

Sharon has served on advisory, governance and/or strategy groups for Blackboard, Inc., E-LEARN, E&I Cooperative, the Electronic Campus of Virginia, the IBM Cloud Academy, Internet2, NYSERNet, and SUNY. In addition, Sharon has served as a co-chair of the Higher Education Information Security Council (HEISC) of EDUCAUSE and the EDUCAUSE Diversity, Equity and Inclusion Task Force.

Sharon is currently Vice-Chair of the EDUCAUSE Board of Directors.

Sharon has served as a faculty mentor of the NY CIO Leadership Academy as well as a mentor for the Leading Change Institute, Vantage's Leadership Lounge, the EDUCAUSE Senior Leadership Roundtable and the Hawkins Leadership Roundtable.

Sharon is a 2004 Frye Fellow.

SESSION I:

SCIENCE ON DARWIN

Alexei Kananenka *(Moderator)*

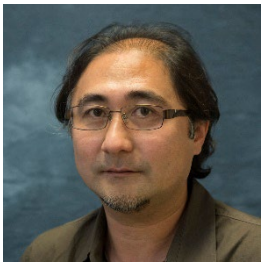
Assistant Professor, Physics & Astronomy
University of Delaware
https://web.physics.udel.edu/about/directory/faculty/alexei-kananenka

Shanshan Ding

Associate Professor, Applied Economics & Statistics
University of Delaware
https://www.udel.edu/academics/colleges/canr/departments/applied-economics-and-statistics/faculty-staff/shanshan-ding/

"High performance computing and statistical learning for tensor data, multi-source data, and time series forecasting"

Abstract: High performance computing (HPC) plays an important role in statistical machine learning. In this talk, I present some of my research work on developing new statistical learning methods for tensor data, multi-source data, and time series forecasting with implementations on HPC. Some software packages and toolboxes are also developed for performing supervised dimension reduction on these complex data structures.



Tom Hsu

Professor, Civil & Environmental Engineering
University of Delaware
https://sites.udel.edu/thsu/group-members/

"Multi-scale numerical modeling of coastal storm and flooding impact to infrastructures and environments"

Abstract: Regional-scale coastal modeling and computational fluid dynamics (CFD) for multiphase flow are applied in this study to investigate multi-scale, complex physical processes in the coastal zones. Coastal flooding is an intensifying natural hazards in the low-lying coastal areas due to the growing coastal urbanization and the accelerating climate change. This is an outstanding multi-scale and multi-physics problem because an accurate numerical model must cover the regional-scale oceanic and nearshore hydrodynamics, while in the meantime resolve the overland flows

down to the street/channel level. Researchers have long been challenged with the necessity to capture the essential physics while maintain a reasonable computational cost for a timely prediction that can benefit the coastal communities.

Using DARWIN, we focus on simulating the impact of coastal storm and flooding to infrastructures and communities, particularly related to beach erosion, scour, hydrodynamic forces in built environment, and benthic exchange. As one of the Intellectual Neighborhood Accelerator project supported by College of Engineering, we use DARWIN to establish the backbone computational framework for coastal flooding prediction based on the open-source NearCoM modeling system. Numerical simulations of a realistic coastal system will be presented with a high resolution mesh that resolves small bathymetric features and a coarse resolution mesh that is typically used for an event-scale coastal flood prediction. The coarse resolution cannot resolve small bathymetric features important for morphodynamics, groundwater infiltration, and infrastructure damage assessment. High resolution simulation data is currently being used to develop sub-grid parameterizations, including the development of a physical-based machine learning algorithm. To further investigate how large waves and flows may lead to beach erosion and amplify benthic exchange via intensive sediment transport and bedform evolution, extensive numerical investigations using the open-source CFD models created in the OpenFOAM toolbox have been out on DARWIN to gain the insight into these highly nonlinear processes and improve the parameterization for regional-scale models. For example, 3D large-eddy simulations of surf zone wave breaking processes reveal the importance of onshore and offshore sand transport mechanisms that are poorly parameterized in the present regional scale morphodynamic models. Two-phase model of sediment transport reveals the importance of bedform migration in determining the benthic exchange flux; by further resolving turbulent coherent structures in the bottom boundary layer, the high resolution 3D two-phase flow simulations answer the long standing causality problem between fluid motions and bathymetric features.

Gonzalo Arce

Charles Black Evans Professor, Electrical & Computer Engineering
University of Delaware
https://www.eecis.udel.edu/~arce/

"LiDAR Imaging for Earth Science with Machine Learning"

Abstract: LiDAR imaging has emerged as a key technology for Earth Science. It can provide information on composition, structure, and function of terrestrial ecosystems. For instance, foliage and canopy distributions of forests and vegetation on Earth can be instantly and remotely attained by LiDAR imagers. The measurements can also determine elevations of glaciers and ice sheets. Current satellite-based LiDARs, such as GEDI and ICESat-2, can take LiDAR measurement at a global scale, but they are severely limited to have very low spatial resolution and lower photon count returns. Only narrow and sparse trajectory swaths can be imaged at once. Our work focuses on the development of a new generation of LiDARs technologies that will be capable of very wide-field-of-views, having the capacity of measuring Earth and planetary environments faster, with fewer swath paths, and at higher-resolution. This is being accomplished by optimal compressive LiDAR sampling, directly at the sensor, using beam-steering LiDAR technology. Scalable reconstruction from compressive measurements is being developed using advanced deep learning algorithms, leveraging NASA's high-resolution data depositories in G-LiHT. Applications on canopy and terrain characterizations in forests will be described.

SESSION II:

DEMOCRATIZING DARWIN

Benjamin Bagozzi *(Moderator)*

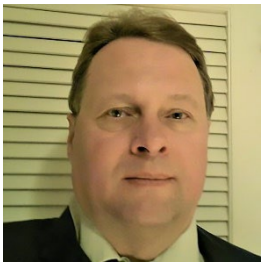
DARWIN co-PI
Associate Professor, Political Science & International Relations
University of Delaware
https://www.poscir.udel.edu/people/faculty/Bagozzib

Karthik Balasubramanian

Assistant Professor, Information Systems & Supply Chain Management
Howard University
https://profiles.howard.edu/profile/53246/karthik-balasubramanianphd

"Mobile Money: Policies for Managing Cash and Digital Currency Inventories in the Developing World"

Abstract: Mobile money systems---platforms built and managed by mobile network platform operators (MMPOs) to allow money to be stored as digital currency---connect millions of poor and "unbanked" people to the formal financial system. Unfortunately, low service levels due to the sub-optimal management of cash and digital currency (e-float) inventory impede the development of these ecosystems. Accordingly, we seek to answer the question: how should agents manage inventories of cash and e-float? This paper extends inventory theory to the mobile money context, unique in that sales of cash generate inventory of e-float and vice versa. In doing so, we address a key pain point for an emerging sector that has been demonstrated to improve lives at the base of the pyramid. We develop an analytical heuristic to determine initial stocking levels for cash and e-float, and analyze its performance on simulated and actual data. By partnering with a MMPO, we tested the performance of the heuristic inventory policy with data from more than 35 million transactions. The heuristic captured 99.9998% of the optimal profit on simulated data, and on actual data, could increase agents' profits by an average of 15.4% if they followed the recommendations. We develop a pragmatic inventory policy that performs nearly optimally. We also analyze under which conditions the performance deteriorates, and examine heterogeneity among agents with respect to the heuristic's impact on them. Thus, we equip MMPOs with guidance as to whom to target and how. By contributing to service level and profit improvements, this work can make mobile money a more effective financial inclusion tool in the developing world as well as improve the livelihoods of agents.



Andrey Semichaevsky

Associate Professor, Chemistry & Physics
D.Eng., Electrical Engineering
Lincoln University
https://www.lincoln.edu/faculty-and-staff/directory/andrey-semichaevsky

"Exploring novel Green's function-based ab initio methods for the prediction of optical properties of molecules"

Abstract: A feasibility study for a new computational physics method is presented, including analysis of assumptions, complexity, numerical algorithms, and their parallel implementation. The method uses two-particle Green's functions to solve a many-body problem in a simple molecule. Standard DFT methods are known to accurately predict structural and mechanical properties of materials but may not be enough to correctly calculate energies of excited states and electronic and optical properties of molecules. Traditional one-particle 'beyond-DFT' methods (e.g., GW approximation) are still not free of inaccuracies in bandgap energy prediction. Therefore, a principally new 'many-particle' approach to first-principles calculations was developed by a collaborating scientist. The parallel implementation of the method should include computing multiple electron-electron interaction energies using numerical Monte Carlo multidimensional integration methods. The MPI-based parallelization appears to be a simple task. More efforts are necessary to come various inputs into the model, such as unperturbed electron wavefunctions. Computer codes will be produced when time permits.

Jun Ren

Associate Professor, Physics & Engineering
Delaware State University
https://oscar.desu.edu/faculty/

"Simulations of Complex Systems and Optimization"

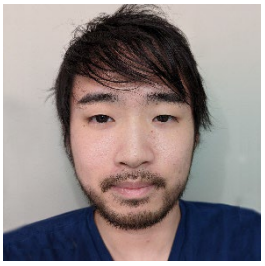
Abstract: Computational techniques and capacities have been advancing very fast in recent decades and are playing ever more essential roles at the forefront of scientific discoveries. Nowadays a significant and growing portion of practical problems can be solved only through computational methods, and very often understandings and insights are likewise apt to be gained through computed representations and visualizations. In this talk, we will show our results from using the parallel computing schemes to study complex systems in plasma science and biological structure. With the limited HPC resources available in our institution, we also pursue with studies into intrinsic parallelism methods, including (1) refining our current algorithms; (2) optimizing synchronous reading and writing I/O operations; (3) using the Tuple switched networking (TSN) over the standard MPI solutions to attain better explicit parallelism; and (4) exploring the applicability of quantum computing. This does require a modest shift in the style of programs, but it also reveals more turnabilities in the range of problems that can be processed.

SESSION III:

SCIENCE ON DARWIN

Jun Ren *(Moderator)*

Associate Professor, Physics & Engineering
Delaware State University
https://oscar.desu.edu/faculty/



Charles Cheung

Post Doctoral Researcher, Physics & Astronomy
Safronova Lab
University of Delaware
https://web.physics.udel.edu/about/directory/graduate-student/charles-cheung

"High precision atomic physics code pCI @ UD"

Abstract: High-precision atomic data and software are indispensable for experiments involving studies of fundamental interactions, development of atomic clocks, astrophysics, and many other applications. We have continued the development of a state-of-the-art parallel atomic structure code package that enables accurate computations for a wide range of systems. Our codes are also used to produce large volumes of atomic data for an online portal hosted at the University of Delaware. Efficient load-balancing is an intrinsic complication of computing matrix elements in many-electron systems. Recent progress in the code development features dynamic load-balancing, demonstrating perfectly linear speedup in our tests. Our new methods prove to be much more scalable than previous static distribution methods, allowing much larger calculations to complete in less computation time.

Alex Bryer

PhD Student, Chemistry & Biochemistry
Perilla Lab
University of Delaware
https://biophysics.chem.udel.edu/alex-bryer

"Full scale structural, mechanical and dynamical properties of HIV-1 liposomes"

Abstract: Enveloped viruses are enclosed by a lipid membrane inside of which are all of the components necessary for the virus life cycle; viral proteins, the viral genome and metabolites. Viral envelopes are lipid bilayers that adopt morphologies ranging from spheres to tubes. The envelope is derived from the host cell during viral replication. Thus,

the composition of the bilayer depends on the complex constitution of lipids from the host-cell's organelle(s) where assembly and/or budding of the viral particle occurs. Here, molecular dynamics (MD) simulations of authentic, asymmetric HIV-1 liposomes are used to derive a unique level of resolution of its full-scale structure, mechanics and dynamics. Analysis of the structural properties reveal the distribution of thicknesses of the bilayers over the entire liposome as well as its global fluctuations. Moreover, full-scale mechanical analyses are employed to derive the global bending rigidity of HIV-1 liposomes. Finally, dynamical properties of the lipid molecules reveal important relationships between their 3D diffusion, the location of lipid-rafts and the asymmetrical composition of the envelope. Overall, our simulations reveal complex relationships between the rich lipid composition of the HIV-1 liposome and its structural, mechanical and dynamical properties with critical consequences to different stages of HIV-1's life cycle.

Christian Heil

PhD Student, Chemical Engineering
Jayaraman Lab
University of Delaware
https://oscar.desu.edu/faculty/

"High-performance computing for studies of structure and dynamics in soft materials"

Abstract: Mixtures of polymers and nanoparticles are a class of soft materials that are used in a wide range of applications including optics, photonics, separations, sensing, batteries, automobile parts, etc. To tailor these polymer-nanoparticle mixtures to function as intended in these applications, one needs to control equilibrium structure as well as dynamics of the components during preparation of these materials. Years of experimental, computational, and theoretical research have established the structure and dynamics in polymer solutions (without nanoparticles) and (solvent-free) nanoparticles-polymer mixtures at certain conditions like small nanoparticle sizes and low volume fractions. Relative to that large body of work, much less is known about structure and dynamics in polymer solutions with nanoparticles of large diameters (~100 nm) and high-volume fractions (>0.25). Such systems comprised of solvent, polymers, and nanoparticles are important to study as they are encountered during the processing of solvent-free polymer-nanoparticle mixtures. Computational studies of systems comprised of solvent, polymers, and nanoparticles have unique challenges due to the broad range of relevant length scales (i.e., nanoparticle size (~100 nm), polymer chain size (~10 nm), and solvent size (~1 nm)) and time scales (i.e., polymer relaxation time significantly larger than the solvent translation time). We use coarse-grained molecular dynamics (CG-MD) simulations to provide a way to bridge these different length and time scales to study structure and dynamics of polymer solutions around crystalline arrangement of nanoparticles of large sizes (~100 nm). We demonstrate how the use of high-performance computing on DARWIN enabled these computationally intensive CG-MD simulations to complete in reasonable timeframe.

SESSION IV:

DARWIN FOR EDUCATION

Jodi Hadden-Perilla *(Moderator)*

Co-Chair, DARWIN Computing Symposium Planning Committee
Assistant Professor, Chemistry & Biochemistry
University of Delaware
https://sites.udel.edu/jhadden/

Arthi Jayaraman

Centennial Term Professor for Excellence in Research and Eucation, Chemical & Biomolecular Engineering
Professor, Materials Science
University of Delaware
https://cbe.udel.edu/people/faculty/arthij/

DARWIN in the Classroom: "Educating our students on application of molecular modeling, simulations, and machine learning for soft materials discovery and innovation"

Abstract: In this talk, I will highlight two elective courses that I have developed to teach undergraduate and graduate students modeling and simulation as well as machine learning tools to study soft materials in laboratories and industry. Both these elective courses require the use of high-performance computing resources, including DARWIN.

The first course is an elective that I developed for graduate students, juniors, and seniors in the departments of Chemical Engineering and Materials Science and Engineering. Students in Chemistry and Physics departments also have relevant technical background to take this course. This elective course provides an overview of molecular models (e.g., coarse-grained models, atomistic models, and multi-scale models), simulation methods (e.g., molecular dynamics and Monte Carlo simulations) in various ensembles, and teaches students on how to apply them to study complex problems in soft materials (e.g., polymers, liquid crystals, nanocomposites, peptides, and nucleic acids). When this course was offered in Spring 2019, I co-taught this with (then) 5th year graduate student from my research group, Thomas E. Gartner (now an Assistant Professor in Georgia Tech); we designed assignments to provide hands-on experiences for the students and used UD's Farber supercomputing resources for students' homework assignments and projects. I will be teaching this elective again in Fall 2022 and intend to use DARWIN supercomputing resources for the planned home-work assignments and team projects.

The second course is also an elective that I developed and currently co-teach with Sunita Chandrasekaran (UD Department of Computer and Information Sciences). In this elective course graduate students from different disciplines – Chemical Engineering, Materials Science and Engineering, Physics, Chemistry, Computer Science, and Electrical and Computer Engineering- solve (in teams) soft materials (MAT) problems submitted by researchers and engineers in industry and national lab. In Spring 2022, we received on-going projects from Procter & Gamble, Merck, Dupont, Gore,

and Brookhaven National lab that needed high-performance computing (HPC), molecular modeling and simulation and/or data science (DS) tools to be solved. Each problem is tackled by an interdisciplinary team of 3 students with 1 from HPC background, 1 from DS, and 1 from MAT background. This team composition also forces the students to learn to communicate with researchers outside of their primary expertise. The researchers in industry and national labs who provided these problems also serve as mentors to these student teams during this course. We co-instructors, Austin Brockmeier (actively involved guest lecturer for Spring 2022), and industrial/national lab mentors guide the teams through teamwork, appropriate approaches for the problem at hand, and oral & written communication exercises. The teams are using DARWIN supercomputing resources for this course.

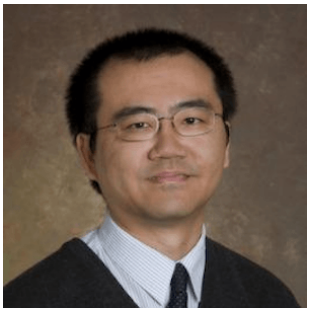
Sunita Chandrasekaran

Associate Professor and David L. & Beverly J.C. Mills Career Development Chair, Computer & Information Sciences
University of Delaware
https://www.eecis.udel.edu/~schandra/

DARWIN for Hackathons: "DARWIN for the next-generation workforce"

Abstract: This talk will highlight the use of DARWIN compute resources for training purposes along with usage of the resources for research projects in Prof. Chandrasekaran's research lab. The computing system is equipped with state of the art CPUs and GPUs that enables the infrastructure to be used as a testbed for training the next-generation workforce in High Performance Computing.

ACCESSING DARWIN



Chuming Chen

Data Scientist, Data Science Insitute
Associate Professor, Computer & Information Sciences
University of Delaware

DARWIN Allocations Overview

Abstract: Overview of the types of allocations available on DARWIN and information about using these allocation types. The presentation will contain information necessary to submit Startup, Education, and Research allocations. Additionally, information on allocation Service Units (SUs) will be provided. Statistics about current allocations will also be shared.

Anita Schwartz

Scientific Application Consultant IV, IT - Research Cyberinfrastucture
University of Delaware

DARWIN in the Classroom

Abstract: An overview of the XSEDE program, how the XSEDE process integrates with DARWIN, and best practices for getting started with XSEDE. The presentation will explain the differences between startup and research allocations, along with information about what is required for a successful proposal. A preview of the upcoming replacement for XSEDE: ACCESS.

PANEL DISCUSSION: DARWIN USER EXPERIENCES / ACCOMPLISHMENTS

Juan Perilla *(Moderator)*

Assistant Professor, Chemistry & Biochemistry
University of Delaware

Alexei Kananenka

Assistant Professor, Physics & Astronomy
University of Delaware

William Totten

DARWIN co-PI
Enterprise Architect
Information Technologies, Enterprise Systems & Cloud Services
University of Delaware

Hamed Meshkin

Post Doctoral Fellow, Chemistry & Biochemistry
University of Delaware

IN-PERSON POSTER SESSION

1. **Hassan Baker, PhD Student, Electrical & Computer Engineering, UD**
   Title: Exploring networks in resting-state fMRI using voxel-to-voxel causal modeling feature selection
   Abstract: Functional networks characterize the coordinated neural activity observed by functional neuroimaging. Techniques for network estimation rely on statistical correlation or dependence between voxels. We propose a voxel-level two-stage algorithm for voxel subset selection that uses different sparsity-inducing regularization approaches to identify subject-specific causally predictive voxels. To reveal the functional networks among these voxels, we then apply independent component analysis (ICA) to model these voxels' signals as a mixture of latent sources each defining a functional network. Based on the inter-subject similarity of the sources' spatial patterns we identify independent sources that are well-matched across subjects but fail to match the independent sources from a group-based ICA. These are resting state networks, common across subjects that group ICA does not reveal. These complementary networks could help to better identify neurodegeneration, a task left for future work.
2. **James Fitzgerald, PhD Student, Physics & Astronomy, UD**
   Title: The Molecular Determinants of Membrane Viscosity
   Abstract: Lipid membrane viscosity is critical to biological function. Bacterial cells grown in different environments have been shown to alter their lipid composition in order to maintain a specific viscosity, and membrane viscosity has been linked to the rate of cellular respiration. In order to understand the factors which determine the viscosity of a membrane, we ran equilibrium all-atom simulations of single component lipid bilayers and calculated their viscosities. The viscosity was calculated via a Green-Kubo relation, with the stress tensor autocorrelation function fit to a stretched exponential by a maximum-likelihood Markov chain Monte Carlo method. By simulating a series of lipids at different temperatures, we establish the dependence of viscosity on several aspects of lipid chemistry, including hydrocarbon chain length, unsaturation and backbone structure. Sphingomyelin is found to have a remarkably high viscosity, roughly 10-20 times that of DPPC.
3. **Shweta Ghimire, MS Student, Civil & Environmental Engineering, UD**
   Title: Cost-Aware Cloudlet Placement in Edge Computing Systems otic Enteritis in broilers
   Abstract: Necrotic Enteritis (NE), caused by Clostridium perfringens (CP), induced high mortality and low growth performance in poultry. Antibiotics bacitracin lincomycin, and oxytetracycline is used in treating necrotic enteritis. Since FDA has implemented an antibiotic phase out plan for overused antibiotics in poultry feed, it created new demands for alternatives to antibiotics. We studied if our new probiotic Bacillus subtilis isolates can control necrotic enteritis (NE) and improve animal performance in broiler chickens. Isolate 103a and 62a having strong antimicrobial and antioxidant properties were chosen for research in a feeding trial. Both strains of Bacillus inhibited CP. The probiotic isolates decreased mortality, lesion score, FCR, and antioxidant level of plasma, but had significantly higher microbiota richness, compared to control groups finding. Promising positive impacts are found for both probiotic isolation in our pilot trial and will be evaluated in large feeding trial.
4. **Vineeth Gutta, PhD Student, Computer & Information Sciences, UD**
   Title: Machine learning approaches for drug response predictive models
   Abstract: The (RNAseq Count Drug Response Machine Learning) RCDML framework can be used to classify patients of rare diseases as high/low responders to drug therapy based on RNA-seq count data and drug response AUC scores. The framework has been tested using data from the BeatAML clinical trial, where over 300 (Acute Myeloid Leukemia) AML patients were treated on 133 drugs. Our results demonstrate better predictive performance when using explainable approaches such as SHAP for feature analysis instead of more traditional approaches (PCA, DGE). Another project involves CANDLE framework's deep learning models that are based on cell line data. Among them the COMBO model has a tendency to generate false positive results. The aim of this work is to test the performance of cell-line based COMBO model on real patient data. Then it will be retrained using patient data. Both of these projects focus on drug response predictive models.
5. **Peter Jones, MS Student, Chemistry & Biochemistry, UD**
   Title: AMBERff at scale: Multimillion-atom simulations with AMBER force fields in NAMD
   Abstract: The application of molecular dynamics (MD) simulations to study large-scale biomolecular systems containing millions of atoms can reveal incredible details about viral and cellular processes that are inaccessible to experimental methods. The AMBER family of force fields (AMBERff) includes the most accurate and extensively validated description of nucleic acids. Yet, it is not currently possible to use amberFF for multimillion-atom simulations. To address this, the authors take advantage of the direct cancellation between the potential energy functions to refactor several popular force fields within AMBERff into the CHARMM file format. Direct comparison of the single point energies for a comprehensive set of test systems show that this process preserves the integrity of the AMBERff force fields. Further, case studies using well characterized biological systems show that these refactored files faithfully reproduce a number of biophysical properties across several classes of biomolecules.

6. **Lev Levintov, Postdoc, Chemistry & Biochemistry, UD**
   Title: Effect of Allosteric Inhibitors on the Integrase Conformation Probed by All-Atom Simulations
   Abstract: One of the key steps in the retroviral replication process is the integration of viral DNA (vDNA) into the host cell chromosome. The integration of vDNA is catalyzed by a viral enzyme integrase (IN) which became an important retroviral target for therapeutic agents. A new distinct class of inhibitors binds at the dimer interface of IN and disrupts the catalytic activity of IN. Although the binding pocket of ALLINIs is well-characterized using experimental methods, the exact mechanism of inhibition is still unclear. In this work, we performed a set of all-atom molecular dynamics (MD) simulations to determine the effect of two ALLINIs on the HIV-1 IN dynamics. The MD simulations revealed that ALLINIs maintain the structural integrity of the IN dimer by mediating the interactions between the catalytic core and the C-terminal domains. Further analysis of the binding pockets revealed key interactions which were formed or ruptured between amino acids of the binding pockets and inhibitors.
7. **Jorge Penaloza-Giraldo, PhD Student, Civil & Environmental Engineering, UD**
   Title: Toward modeling flocculation in turbulence-resolving simulations for cohesive sediment transport
   Abstract: Direct Numerical Simulation (DNS) is a very attractive tool to integrate with fine sediment transport equations due to its high accuracy. However, most existing models assume a constant settling velocity for the sediment by neglecting the flocculation processes. Flocculation can change the settling velocity of cohesive sediment depending on sediment properties and the aquatic system. We evaluate a loosely coupled method between a Eulerian turbulence-resolving fine sediment transport model and a floc size class-based flocculation model for a turbulent statistically-steady open channel flow in dilute conditions. This flow case has a Reynolds number 180 based on the friction velocity and channel height. Results show that the turbulent shear rate is higher close to the bed compared to the value obtained at the free surface. Since the concentration profile follows the same trend, the flocculation model reaches the equilibrium state more quickly close to the bed.
8. **Yahira Rivera, PhD Student, Physics & Astronomy, Delaware State University**
   Title: Intensity Correlation Analysis of Ficoll Raman Spectra
   Abstract: Biomacromolecular crowding is recognized to affect various biological processes such as cell signaling, protein folding and stability, binding of small molecules, and enzymatic activity. Our research interest has focused on developing and applying in-situ optical-based spectroscopy and imaging methodologies, combined with data-driven analytical tools, to probe and quantitate various chemical and physical processes in biopolymeric and cellular systems. Our overall goal is to develop data-driven (e.g. spectral and imaging) and physics-informed models that will be trained to identify and assess optical signatures associated with the structural and dynamical properties of diverse nanoparticles (e.g. proteins, amino-acids, metallic, bacteria) while dispersed in relevant samples.
9. **Miguel Romero Rosas, PhD Student, Electrical & Computer Engineering, UD**
   Title: Measuring the Impact of Automatic Program Parallelization Techniques
   Abstract: Cetus is a source-to-source translator for programs written in the C language. The primary use is as a parallelizing compiler, translating C programs to equivalent C code annotated with OpenMP parallel directives. Cetus is a research platform to study parallelization techniques and related program transformations. This poster presents an overview and evaluation of the existing analysis and transformation techniques in the Cetus source-to-source compiler infrastructure. This work seeks to measure the impact of the existing Cetus techniques on the newer versions of some of the benchmarks from the NAS Parallel benchmark suite v3.3 and PolyBench Benchmark Suite v4.2.
10. **Liam Sharp, Postdoc, Physics & Astronomy, UD**
    Title: Sampling Protein-lipid Interactions With Swarms of Trajectories and a Transition Matrix Analysis
    Abstract: Lipid-protein interactions are critical for integral membrane protein function, where lipids can play the role of both solvent and ligand. Because resolving the lipids that solvate a membrane protein is experimentally challenging, simulations have come to play a prominent role in linking lipid solvation and membrane protein function. However, converging the lipid distribution around a membrane protein can require surprisingly long simulations, especially in cholesterol-rich mixtures where lipid diffusion and exchange at the surface of the protein can be quite slow. Recently, we introduced an algorithm1 that uses trivially parallel swarms of relatively short trajectories to parametrize a transition matrix, which is then used to predict the equilibrium distribution. Here, we present a careful analysis of the convergence of the method, using Martini simulations to ensure that the expected distribution is known.
11. **Benjamin Tsai, PhD Student, Civil & Environmental Engineering, UD**
    Title: Large-Eddy Simulation of Cross-Shore Hydrodynamics under Random Waves in the Surf and Swash Zones
    Abstract: Sand dunes play an important role in minimizing the impact of beach erosion and breaching during storms. However, the understanding of the interactions between the hydrodynamics and complex geometry is limited. We carry out 3D wall-modeled large-eddy simulations to study the cross-shore wave transformation, wave-breaking, and the undertow current in the surf and swash zones. The numerical model is validated with a near prototype-scale laboratory experiment conducted in a large wave flume and the model domain size is identical to it. Model results show that free-surface elevations along the dune

profile are well predicted. The agreement between the modeled and measured velocity and undertow are highly correlated. We present refined simulations in the swash zone, including LES model validation, and insights into wave breaking, turbulence, and wave-induced currents. The refined LES results can provide comprehensive data to improve turbulence-averaged models and wave-averaged models.

12. **Ilya Tyagin, PhD Student, Computer & Information Sciences, UD**
    Title: Accelerating COVID-19 Research With Graph Mining and Transformer-based Learning
    Abstract: We present automated transformer-based hypothesis generation systems AGATHA-C and AGATHA-GP for COVID-19 research. The systems are based on the graph mining and transformer models. They are massively validated and achieve high-quality predictions across multiple domains in fast computational time and are released to the broad scientific community to accelerate biomedical research. We also show that the systems are able to discover ongoing research findings such as the relationship between COVID-19 and oxytocin hormone.
13. **Ryan Wileys, MS Student, Physics & Astronomy, Delaware State University**
    Title: Implementation of Shor's Algorithm in Prime Factorization and Discrete Logarithms
    Abstract: RSA Encryption and the Diffie-Hellman Key Exchange rely on the computational intractability of (a) factoring a product of two large prime factors and (b) the computation of discrete logarithms, respectively. With access to a sufficiently large quantum computers however, a user would be able to break these encryption schemes with relative ease. The purpose of this presentation is to introduce Shor's period-finding algorithm and demonstrate a quantum system's potential ability to solve the RSA and discrete logarithm problems by taking advantage of quantum superposition and interference. The examples of Shor's algorithm illustrated here are only proof of concept for now as quantum processors available to us are of insufficient size to prevail the classical encryption methods. Using IBM's quantum computing platform Qiskit, we will show successful implementation of Shor's Algorithm on a quantum simulator then later direct the operation on a real quantum processor.
14. **Lianxin Xin, PhD Student, Delaware State University**
    Title: Running PIC code on a small HPC cluster
    Abstract: To improve the accuracies of computing real physical systems and extend the dimensions of simulation, large-scale computing systems with advanced infrastructure designs and continuing optimization of code algorithms are required. This presentation concentrates on improving parallel code performance by optimizing synchronous reading and writing I/O operations. An NVMe protocol is applied where NVMe-oF is combined with the InfiniBand RDMA technology. Further improvement is attained by replacing the NFS with the Lustre IO so that data can be provided/distributed faster. We run the Particle-in-Cell code OSIRIS on our system. Code speed evaluation and benchmark test showed that the overall system performance is improved by nearly ten times. This newly upgraded system enables us to run 1D prototype simulations of light-plasma interactions with various complex boundary conditions and paves the way for collecting more sounding results from 2D and 3D runs when using larger computing facilities.

VIRTUAL POSTER SESSION

1. **Parinaz Barakshan, PhD Student, Electrical & Computer Engineering, UD**
   Title: iCetus: A Semi-automatic Parallel Programming Assistant
   Abstract: The iCetus tool is a new interactive parallelizer, providing users with a range of capabilities for the source-to-source transformation of C programs using OpenMP directives in shared memory machines. While the tool can parallelize code fully automatically for non-experts, power users can steer the parallelization process in a menu-driven way. iCetus, which is still in its early stages of development, is implemented as a web application for easy access, eliminating the need for user installation and updates. The tool supports the user through all phases of the program transformation process, including program analyses, parallelization, and optimization.
2. **Abdul Rehman, PhD Student, Physics & Astronomy, UD**
   Title: Classification and Denoising of Cosmic-Ray Radio Signals using Deep Learning
   Abstract: Cosmic rays (CRs) are high-energy particles coming from outside the solar system. Cascades of secondary particles, called air showers, are produced when CRs interact with our atmosphere. One of the ways to measure the properties of primary CRs is to detect the radio emission emanating from these air showers. To mitigate the effect of the irreducible galactic and thermal background on the radio detection of CRs we employ deep learning techniques, specifically Convolutional neural networks (CNNs). Taking advantage of the early access period of the DARWIN cluster we were able to train our models using Monte Carlo simulations. The promising results encouraged us to apply these techniques to real background data acquired with antennas at the South Pole, which we do with the research allocation on DARWIN . Once trained and optimized, these models will help improve the detection threshold and increase the measurement accuracy of current and future cosmic-ray radio detectors.
3. **Juan S. Rey, PhD Student, Chemistry & Biochemistry, UD**
   Title: Deep-learning in situ classification of HIV-1 virion morphology
   Abstract: Transmission electron microscopy (TEM) has a multitude of uses in biomedical imaging due to its ability to discern ultrastructure morphology at the nanometer scale. As applied to HIV-1 research, TEM is critical to evaluate activities of inhibitors that block the maturation and morphogenesis steps of the virus lifecycle. However, both the preparation and analysis of TEM micrographs requires time consuming manual labor. Through the dedicated use of computer vision and machine learning techniques, we have developed a convolutional neural network backbone of a two-stage Region Based Convolutional Neural Network (RCNN) capable of identifying, segmenting and classifying HIV-1 virions at different stages of maturation and morphogenesis. Our results outperformed common RCNN backbones, achieving 80.0% mean Average Precision on a diverse set of micrographs comprising different experimental samples and magnifications.
4. **Diego Roa, PhD Student, Electrical & Computer Engineering, UD**
   Title: DECARD, a Distributed Runtime for Heterogeneous Architectures
   Abstract: Aiming to develop a runtime for parallel, heterogeneous and distributed systems, designed to manage communications and scheduling that does not depend on the underlying hardware, we have developed a set of protocols and mechanisms. Namely, DECARD and eDARTS. The bi-layered runtime based on the Codelet Program Execution Model (PXM) describes simple yet sufficient mechanisms to handle computation and control signals required for parallel applications. The data-centric nature of dataflow models for computation provides an ideal framework to manage the program's parallel execution based on the data's availability. The proposed runtime is implemented as a digital design with straightforward finite state machine (FSM) descriptions for each element of the system. Memory and control flow messages serve as input that define the next state of the FSM. The ability to integrate specialized units that manage communications and scheduling with CPU, GPU or custom cores within a single chip can potentially provide better performance for specific applications and increase programmability.
5. **Hamed Meshkin, Postdoc, Chemistry & Biochemistry, UD**
   Title: The impact of phosphorylation modulation in the binding of NLS peptides to importin-alpha
   Abstract: Importin proteins transport molecular cargo from cytoplasm into the nucleus. The localization of Importin to the nucleus is facilitated via Nuclear Localization Signal (NLS) peptides. Microtiter binding experiments reveal that phosphorylation of NLS peptides perturbs Importin-alpha binding affinity. To derive the NLS conformational landscape and its interaction with Importin-alpha, we performed MD simulations of NLS and four phosphorylated constructs: 88pThr 91pSer, 92pSer, and 88pThr+91pSer+92pSer. Kinetic maps yielded via Markov state modeling reveal several metastable states of NLS. Furthermore, we found that the lowest and highest mean first passage times belong (MFPT) to wild-type and poly-phosphorylated NLS, respectively. We further probed the binding energies of each construct with Importin-alpha, corroborating the dramatic reduction in affinity the phosphorylated constructs compared to wild-type NLS.

6. **Karl Franke, Postdoc Fellow, Center for Pediatric Auditory and Speech Sciences, Nemours**
   Title: Improving low allelic variant detection using error corrected sequencing coupled with a binomial error rate
   Abstract: Next generation sequencing (NGS) techniques have been extensively utilized in the medical field to help understand the association of genomic mutations and disease onset and progression. A tremendous amount of effort has been directed towards improving bioinformatic techniques for detecting genomic alterations, with a major focus on improving the types of variants detected, such as structural variants and insertions and deletions. Our team focuses on improving the detection of low allelic events via a technique called error corrected sequencing, or ECS. This technique involves the use of unique molecular indexes, UMI, that enables the bar-coding of individual molecules being prepared for NGS. The UMIs, coupled with downstream bioinformatics, allows for the ability to correct for library preparation "errors" that are introduced during the low level amplification steps, required in most library preparation techniques. We further link these results with a binomial reduction method to further remove sequencing errors.

DARWIN COMPUTING SYMPOSIUM PLANNING COMMITTEE

Rudolf Eigenmann, PI

Arthi Jayaraman, co-PI

Benjamin Bagozzi, co-PI

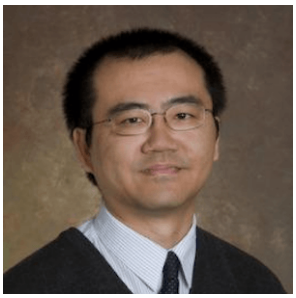
William Totten, co-PI

Cathy Wu, co-PI

DARWIN COMPUTING SYMPOSIUM PLANNING COMMITTEE (CON'T)

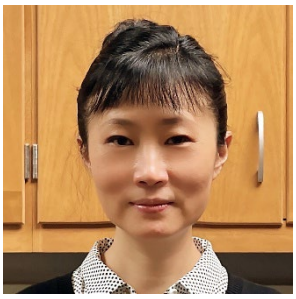
Jodi Hadden-Perilla, Co-Chair

Juan Perilla, Co-Chair



Chuming Chen

Julie Cowart



Jun Ren

Andrea Trungold

Parinaz Barakhshan

James Korman

THANK YOU TO OUR INDUSTRY SPONSORS

DELL

NEMOURS CHILDREN'S HEALTH

VISIT THE DATA SCIENCE INSTITUTE AT:

https://dsi.udel.edu/

JOIN US ON TWITTER

@dsiudel

hashtag: #DARWINcomputing2022