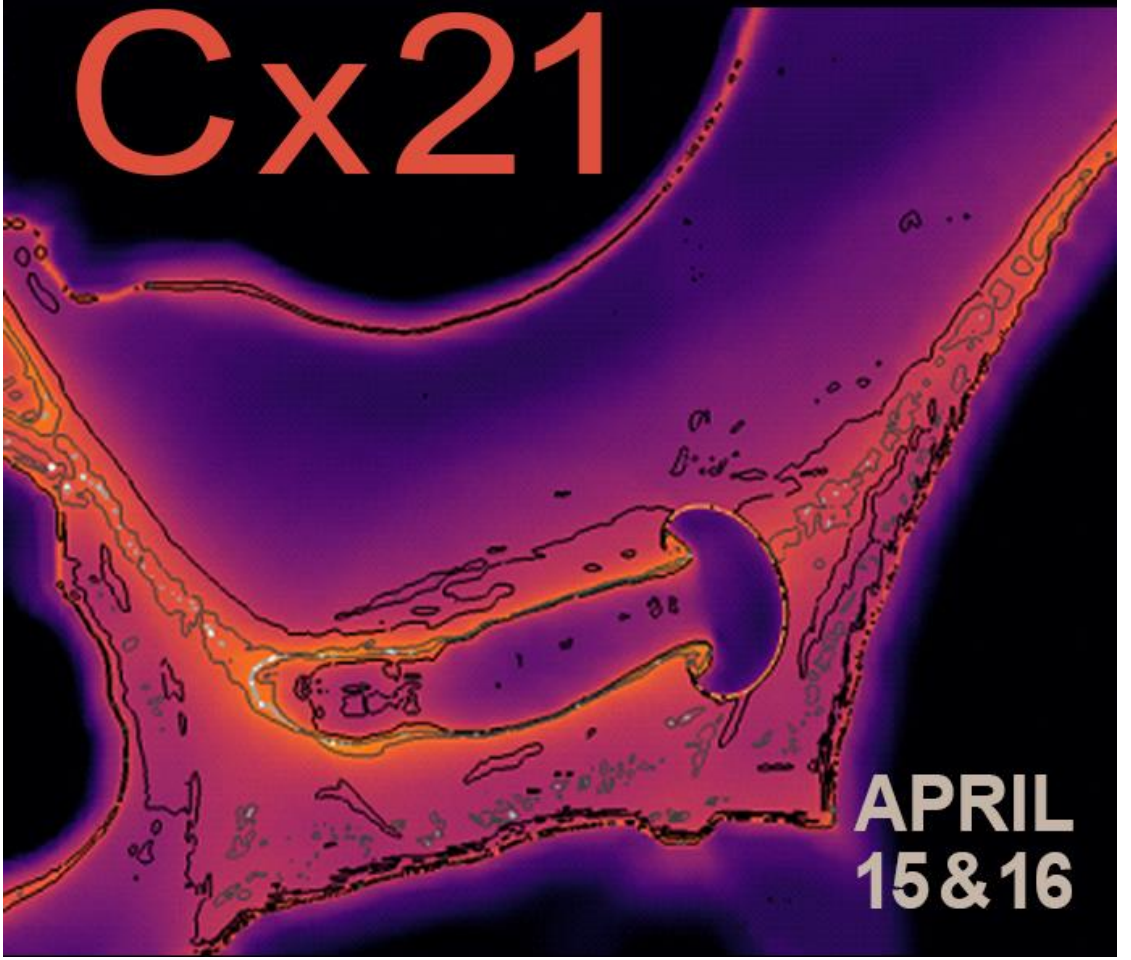
The Department of Scientific Computing
Florida State University

2021 Computational Xposition

Cx21



APRIL
15 & 16

Welcome to Computational Exposition 2021, a yearly event where students in the Department of Scientific Computing (DSC) showcase the results of their research in the last year. Higher education was confronted with singularly difficult circumstances this year due to a deadly pandemic, with most students and professors working from home, attending classes online, and having to conduct research and communicate remotely. This research covers a broad spectrum of disciplines, but shares a common thread: they concentrate on algorithm development and blend the computational and the mathematical to solve problems in the applied sciences. The innovation displayed is broad and remarkable. Our students make us proud!

The student posters reflect the breadth and depth of the research carried out in the DSC. They are the direct result of a fulfillment of our two most important missions: providing world-class interdisciplinary research and training in computational science.

Our graduate degree programs and the success of our students bolsters our confidence that we are a premier institution for the training of the next generation of computational scientists. Indeed, looking at what our current students have achieved over the past several years attests to this claim!

So, enjoy the presentation, interact with the students, challenge them, learn from them, and reflect on the fruits of their intelligence, skills, dedication and labor, and join us in thanking them for their contributions to the DSC, to FSU, and to science.

Gordon Erlebacher
Chair, Department of Scientific Computing

Front Cover: Carbon deflagration wave transitioning into a carbon detonation wave, courtesy of Ezra Brooker.
Back Cover: A single attention map of the mean activity and AD subjects from the validation dataset, adamantine, courtesy of Kevin Mueller.

Presenting Researchers

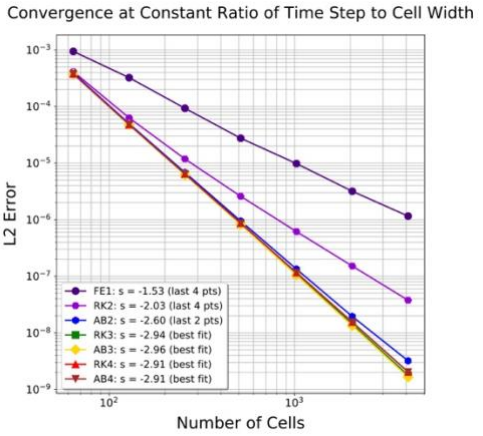
RESEARCH

Abstracts & Graphics

Siddhartha Bishnu
Ph.D. in Computational Science
Advisors: *Bryan Quaife & Mark Petersen*

On the Spatial and Temporal Order of Convergence of Hyperbolic PDEs

I will discuss the leading order terms of the local truncation error of hyperbolic partial differential equations (PDEs) on a uniform mesh. When employing a stable numerical scheme, we make the following observations in the asymptotic regime, where the truncation error is dominated by the powers of the grid spacing and the time step rather than their coefficients: (a) the order of convergence of stable numerical solutions of hyperbolic PDEs at constant ratio of time step to grid spacing is governed by the minimum of the orders of the spatial and temporal discretizations, and (b) convergence cannot be attained under only spatial or temporal refinement. We test our theory against numerical methods employing Method of Lines for any hyperbolic PDE, be it linear or non-linear, and employing finite difference, finite volume, or finite element discretization in space, and advanced in time with forward Euler, predictor-corrector, and multistep methods. If the hyperbolic PDE is reduced to an ordinary differential equation (ODE) by specifying the spatial gradient(s) of the dependent variable and the coefficients and the source terms to be zero, then the standard local truncation error of the ODE is recovered. We have performed the mathematical analysis with the generic as well as specific hyperbolic PDEs by hand and using SymPy, the symbolic algebra package of Python, and conducted a number of numerical experiments to demonstrate our theoretical findings. Future work includes considering schemes that treat space and time together and an investigation of the reduction in the spatial and temporal orders of accuracy resulting from monotonicity-preserving strategies commonly applied to finite volume methods.

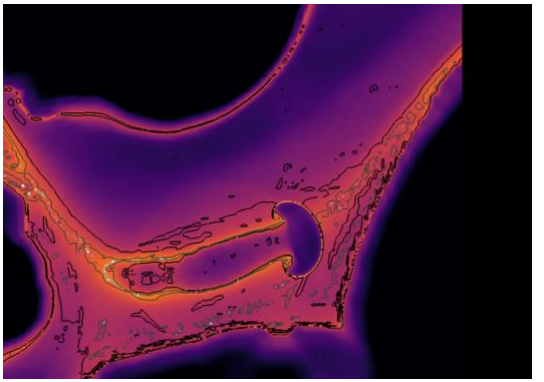


Figure, left: Convergence plot of a linear variable-coefficient advection equation using fourth-order accurate piecewise parabolic reconstruction in space, and first-order forward Euler (FE1) method, explicit midpoint -- a form of second-order Runge-Kutta method (RK2), second-order Adams-Bashforth (AB2) method, low-storage third-order Runge-Kutta (RK3) method, third-order Adams-Bashforth (AB3) method, low-storage fourth-order Runge-Kutta (RK4) method, and fourth-order Adams-Bashforth (AB4) method in time, for refinement in both space and time.

Ezra S. Brooker
Ph.D. in Computational Science
Advisor: *Tomasz Plewa*

SN Ia DDT Explosions Powered by the Zel'dovich Reactivity Gradient Mechanism

The aim of this work is to identify and explain the necessary conditions required for an energetic explosion of a Chandrasekhar-mass white dwarf. We construct and analyze weakly compressible turbulence models with nuclear burning effects for carbon/oxygen plasma at a density expected for deflagration-to-detonation transition (DDT) to occur. We observe formation of carbon deflagrations and transient carbon detonations at early times. As turbulence becomes increasingly inhomogeneous, sustained carbon detonations are initiated by the Zel'dovich reactivity gradient mechanism. The fuel is suitably preconditioned by the action of compressive turbulent modes with wavelength comparable to the size of resolved turbulent eddies; no acoustic wave is involved in this process. Oxygen detonations are initiated either aided by reactivity gradients or by collisions of carbon detonations. The observed evolutionary timescales are found sufficiently short for the above process to occur in the expanding, centrally ignited massive white dwarf. The inhomogeneous conditions produced prior to DDT might be of consequence for the chemical composition of the outer ejecta regions of SN Ia from the single degenerate channel and offer potential for validation of the proposed model.



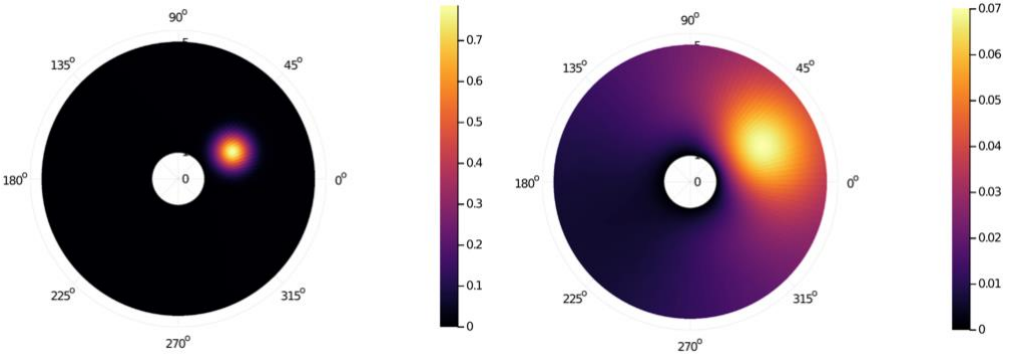
Figure, above: Carbon deflagration wave transitioning into a carbon detonation wave via the Zel'dovich Reactivity Gradient mechanism in a channel of carbon/oxygen fuel compressed by two plumes of thermonuclear ash formed by prior deflagration waves in a box of turbulent white dwarf plasma.

Jesse Cherry
Ph.D. in Computational Science
Advisor: *Bryan Quaife*

A novel solver for diffusive processes in exterior domains

To describe transport of quantities such as heat or chemical concentrations, the diffusion equation must be solved in complex unbounded geometries. Since closed-form solutions rarely exist, we must resort to numerical methods. Three major challenges that numerical methods must address include (i) computing high-order accurate solutions in both space and time, (ii) accurately capturing far-field conditions, and (iii) computing long-time behaviors. I will describe a numerical method that recasts the time-dependent heat equation into an elliptic PDE by applying a Laplace transform. By using the Laplace transform, the solution of the PDE can be found at any time without the need to time step. This elliptic PDE is solved with high-order methods that accurately capture the far-field conditions. After solving the elliptic PDE, an inverse Laplace transform must be performed. This is done by carefully choosing a Bromwich contour that results in an integrand that is not oscillatory and rapidly decays to zero. By combining these techniques, high-order solutions in both space and time are achieved.

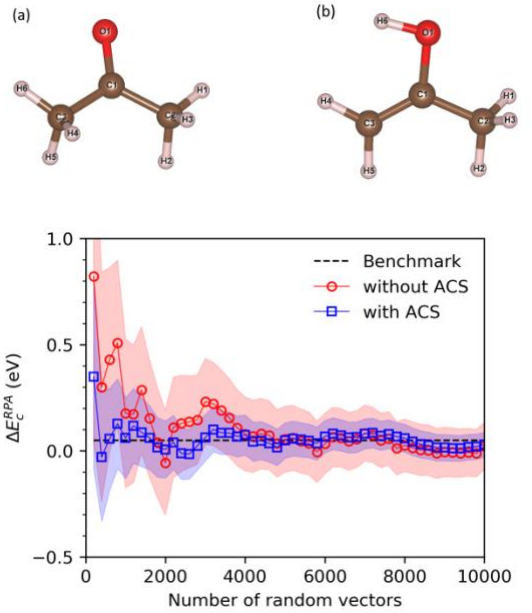
Figures: Diffusion of a point source on the domain exterior to the unit disc, with Dirichlet boundary condition, at time 0.1 (left) and 1.0 (right).



Yu-Chieh Chi
Ph.D. in Computational Science
Advisor: *Chen Huang*

Accelerate stochastic calculation of random-phase approximation correlation energy difference with atom-based correlated sampling

A kernel polynomial method (KPM) is developed to calculate the random phase approximation (RPA) correlation energy. In this method, the RPA correlation energy is formulated in terms of the density of states (DOS) for the eigenvalues of the matrix obtained as the product between the Coulomb operator and Kohn-Sham (KS) linear response function. The integration over the eigenvalues is then calculated using KPM. Since energy differences between similar systems are of much interest in practice, we also develop a scheme, termed atom-based correlated sampling (ACS), to accelerate the convergence of energy-difference calculations. ACS is used to calculate the isomerization energy between acetone and 2-propanol and the energy of water-gas shift reaction. For both cases, the convergence is much accelerated by ACS. The KPM developed in this work, together with the ACS scheme, is expected to be useful for calculating RPA correlation energy difference between two systems that only differ in a local region, for example, calculation of adsorption energies of molecules for surface catalysis. For these types of calculations, the computational cost of our method scales quadratically with the system's size.



Figures Above: Convergence of the RPA correlation energy difference between Acetone and 2-propenol with respect to the number of random vectors, calculated with and without ACS. Standard errors are represented by the red and blue bands. Benchmark is denoted by the dashed line.

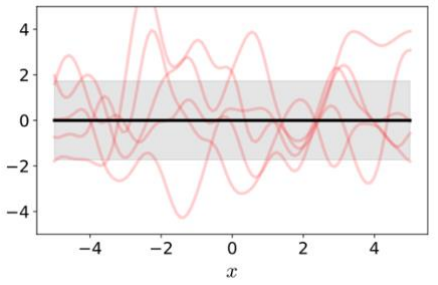
Pankaj Chouhan
Ph.D. in Computational Science
Advisor: *Sachin Shanbhag*

Spatial Modeling of Intracellular Calcium Dynamics in Branched Astrocyte Processes

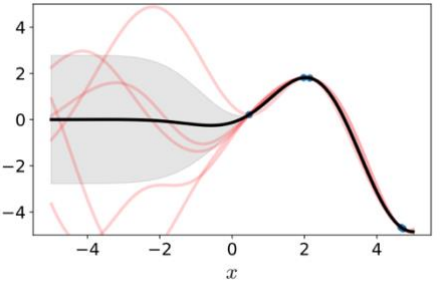
Molecular models of polymer rheology take molecular structure of a polymer mixture as input, and yield the linear or nonlinear viscoelasticity of the sample as output. In industry, usage of predictive models is an integral part, processing method can be rationally engineered to shape polymer melts into final products. Different class of molecular simulation model exist to achieve various level of accuracy, and follows the trade- off law between accuracy and speed. Certain application requires highly accurate results of molecular simulation, but can't bear the underlying computational cost, this is where surrogate models comes to rescue. Surrogate models are statistical Model that seek to mimic the input-output relationship, x → y, of computationally expensive simulations. For this presentation, a SM, y = mˆ (x), where m : Rd → RN , is considered with x representing the structure of the polymer mixture, and the output y representing the rheology.

The use of Gaussian process (GP) regression is a popular surrogate modeling technique. It provides a systematic and flexible framework to emulate simulators with multivariate input and output variables. When used within a regression framework, training data is first accumulated by simulating the true model at n training points x1, · · · , xn. The parameters of the GP are inferred, typically using maximum-likelihood (MLE). At this point the SM is trained, and can be used to make predictions at a test input point, x∗.

GPs have two features that make them particularly attractive as SMs: (i) they interpolate the true model at the training points; thus yi = m(xi) = E(mˆ (xi)) for i = 1, · · · , n and (ii) they provide uncertainty estimates σˆ∗2 = V(mˆ (x∗)), where V(.) denotes the variance in the SM pre- diction at input point x∗. Thus, the SM provides a prediction that takes the form yˆ∗ ± σˆ∗2 =, where yˆ∗ = E(mˆ (x∗)) represents the expected value or mean prediction of the SM, and σˆ∗2 represents the expected variance. Thus, we can visualize GP-based SMs as sophisticated interpolants with built-in uncertainty quantification.



(a) GP prior



(b) GP posterior

Figures above: Solid black line depicts the mean and the shaded region shows an interval of ±σ around this mean. The red lines depict 5 samples drawn from this Gaussian process. In (a) the kernel is given by the RBF with σ2 = 3 and l = 0.5. In (b), the blue circles denote 5 observations. The optimal hyper-parameters after maximizing the marginal likelihood of observing the data are σˆ2 = 7.73 and ˆl = 1.57. In posterior distribution using these MLE hyper-parameters and observed data are shown in (b).

Ashley Gannon
Ph.D. in Computational Science
Advisor: *Bryan Quaife*

Semipermeable Multicomponent Vesicles in Stokes Flow

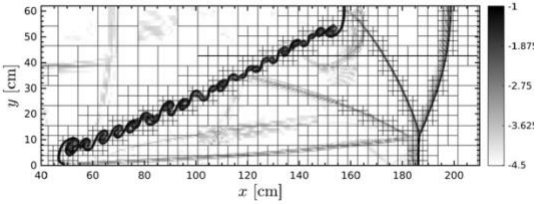
We apply a high-order boundary integral equation method to simulate multi-component, semi-permeable membranes in various Stokes flows. Our multicomponent vesicles are representative of cell membranes containing aquaporins and are only permeable to water. This semipermeability is important for many biophysical processes including cell migration and cell rupture. Our semipermeability model depends on the membrane forces where the fluid flux is proportional to the pressure drop. The multi-component model uses the Cahn-Hilliard equation to allow for phase separation between different lipid species. I will discuss our results for vesicles in quiescent flow, shear flow, extensional flow, and Poiseuille flow.

Brandon Gusto
Ph.D. in Computational Science
Advisor: *Tomasz Plewa*

A Hybrid Adaptive Multiresolution Approach for the Efficient Simulation of Reactive Flows

Computational studies that use block-structured adaptive mesh refinement (AMR) approaches suffer from unnecessarily high mesh resolution in regions immediately adjacent to important solution features. This deficiency may be a major factor limiting the performance of AMR codes. In this work a novel hybrid adaptive multiresolution (HAMR) approach to AMR-based calculations is introduced to address this issue. The multiresolution (MR) smoothness indicators are used not only to adapt the mesh, but also to decrease the computational cost of individual physics solvers in regions identified as smooth by replacing direct calculations with MR interpolation according to prespecified accuracy constraints. The accuracy of this procedure is shown to be consistent with that of the MR-driven AMR. The performance of the HAMR scheme is demonstrated for a range of test problems, from one-dimensional hydrodynamics to multidimensional turbulent combustion.

Figure, right: Log of MR detail coefficients and adaptive mesh at late time in the Hawley-Zabusky problem. At the selected MR tolerance, the algorithm allocates blocks at the finest AMR level in two main regions: along the material interface where vorticity is large, and along the passing shock waves, where pressure gradients are large. Near the important solution features there are regions where the mesh is relatively fine (due to the block-based AMR format), but the MR detail coefficients are relatively small, inviting the use of cost-saving measures.



Marzieh Khodaei
Ph.D. in Computational Science
Advisor: *Peter Beerli*

A path between trees in the phylogenetic forest

Phylogenetic trees are fundamental tools for understanding the evolutionary history of a set of species. Understanding the local neighborhoods of a phylogenetic tree is essential, but since trees are high-dimensional objects, discussing these neighborhoods is difficult. We use different distance methods to explore the phylogenetic tree landscape.

Based on the geodesic distance between pairs of trees, we developed a method to generate trees on the shortest path between two arbitrary trees. We are convinced that these paths among trees will improve searching the tree space and therefore find the best trees faster than current methods.

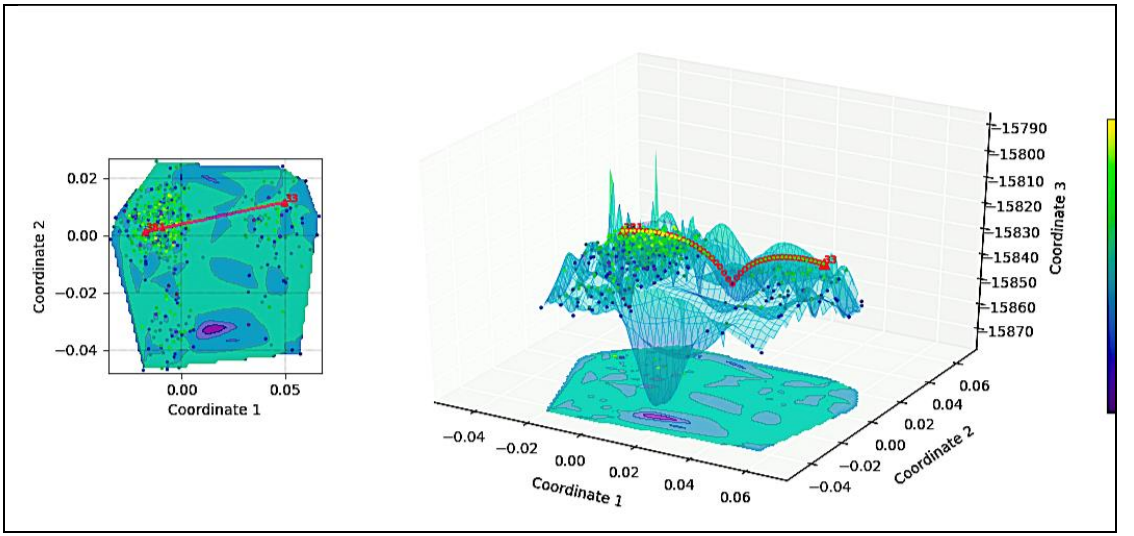


Figure: Cubic interpolation of the tree space, displaying both the contour and the surface of tree space by applying the geodesic distance method of GTP (Owen and Provan, 2010). Each dot is a tree; the lighter the dot, the higher the likelihood of the tree. The red dots are trees generated on the shortest path between two arbitrary trees. We generated a sample of 100,000 trees using the program REVBAYES and their tutorial dataset primates_cytb_JC. From this sample, we extracted 500 rooted trees collected during the MCMC run after removing half of the trees as burn-in. We then used the distance metric among all pairs to visualize the relationship among them using Multidimensional Scaling.

Young Hwan Kim
Ph.D. in Computational Science
Advisor: *Xiaoqiang Wang*

Sickle RBC Phase Field Model

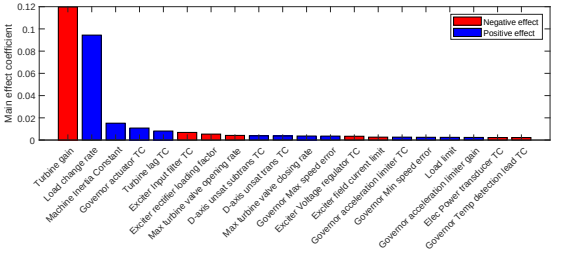
In sickle cell disease abnormal beta-globins, known as hemoglobin S (HbS), are produced, and they polymerize to form long, rigid fibers that bend red blood cells (RBCs) into sickle shapes. Many studies and much modeling of the dynamics of HbS polymer fibers have been done to better understand sickled red blood cells. However, the models are computationally expensive and do not describe the HbS fiber-RBC membrane interaction, which occurs at mult-spatial scales, ranging from nanometers to micrometers.

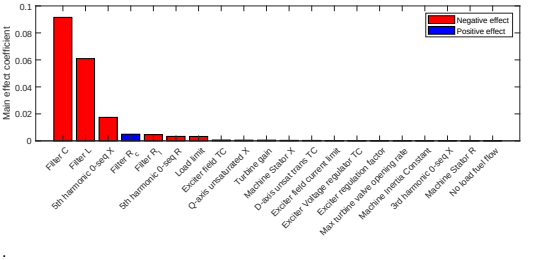
In this poster, we discuss a phase field model for HbS fiber-RBC membrane interaction. Phase field is a global method that describes complex interfaces as a relatively simple function, and it allows topological changes without the need to modify meshes for interface tracking. The RBC membrane is modeled by representing Helfrich elastic bending energy using the phase field function $\Phi(x)=\tanh(d(x,\Gamma)2\epsilon)$. The HbS fiber is modeled as a chain of particles, and since distance information is included in $\Phi(x)$, the interaction between a particle and the membrane is described by $1\Phi p(x)$, where $\Phi p(x)$ is the particle's phase field value interpolated using inverse distance weighting.

Behshad Mohebali
Ph.D. in Computational Science
Advisor: *Gordon Erlebacher*

Real-time Hardware-in-the-Loop

Real-time Hardware-in-the-Loop (HIL) simulation has been a major step in design, development, and implementation of new technologies in the field of power systems. Due to their real-time characteristics, these simulations cannot be accelerated by using more powerful hardware, making the comprehensive evaluation of a given device-under-test (DUT) more challenging. Usually the simulation scenarios (mainly events and parameter values) are orchestrated by experts to observe the behavior of the DUT under stress through calculating the performance metrics. The issue is that it cannot push the DUT to its absolute limit. Doing so requires a systematic approach to determine and manipulate a usually vast number of model parameters to check all possible scenarios and identify the ones that push the metrics beyond their acceptable range. However, model variables have different levels of influence on the performance metrics. Here, we use a Power Generation Module (PGM) model found on next generation naval vessels to implement fractional factorial design as a factor screening approach on two metrics defined for the PGM. The results confirm that a very small subset of the model variables have most of the influence on the performance metrics. Knowing so will let us reduce the dimension of the search space and therefore the computational burden of the analysis.





.

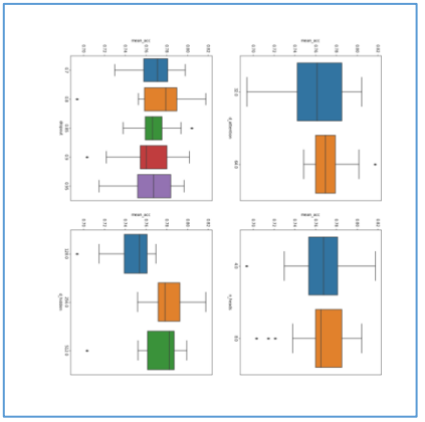
Kevin Mueller
Ph.D. in Computational Science
Advisor: *Anke Meyer-Baese*

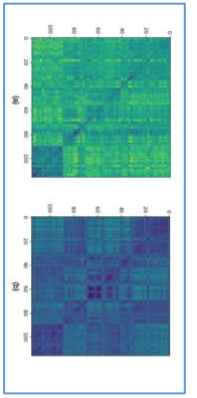
Multimodal Region-Based Transformer for the classification and prediction of Alzheimer's disease

Numerous deep learning approaches have been proposed to automatically classify Alzheimer's disease (AD) from medical images. However, common approaches such as convolutional neural networks (CNNs), lack interpretability and are prone to overfitting when trained on small datasets. As an alternative, significantly less work has explored applying deep learning approaches to region-based features that are commonly attained from atlas partitions of known regions of interest (ROI). In this work, we propose a self-attention mechanism to jointly learn a graph of connectivities between ROIs as a prior for learning meaningful features for AD prediction. We apply our method using numerous atlases and to an assortment of classification tasks to systematically compare its performance to other ML approaches for ROI-based methods. Finally, we perform exploratory analysis and analyze the interpretability properties of the learned attention graphs for AD prediction.

Figure 1, Left: Distribution of the mean accuracy attained for different hyperparameter settings.

Figure 2, Right: An example of a single attention map of the mean activity for control and AD subjects from the validation dataset.





Eric Sharkey
Masters in Computational Science
Advisor: *Gordon Erlebacher*

Transformer Text Style Transfer

Text Style Transfer (TST) is a Natural Language Generation task that attempts to transform a sentence into a new sentence with the same content but a different style. The content of a sentence dictates what the sentence is about, while the style is commonly interpreted as the sentiment and emotions used to convey the information. One may think of TST as disentangling the style and content of a sentence to allow for style control. Text Style Transfer has many applications, such as controlling the opinions in product reviews, and creating more reactive chatbots. One difficulty in training TST models is the absence of parallel corpora, containing sentences with identical content but different styles. Thus, many TST models use adversarial training methods with non-parallel corpora, which are more available. Another difficulty is the presence of long-term word dependencies present in sentences meaningful for text style transfer. Existing Recurrent Neural Network based models do not perform well because of this. This poster presents a transformer based variational autoencoder model that effectively disentangles style from content in text. Transformers have been shown to perform well in capturing long term dependencies in text using its attention mechanisms. Its architecture is also largely parallelizable, allowing for faster training. The encoder memory is separated into distinct style and content spaces using convolutional layers, and disentanglement is encouraged in these spaces using adversarial training methods. This model improves on existing Style Transfer models that use implicit disentanglement methods by leveraging Transformers' attention mechanisms to better capture important text dependencies.

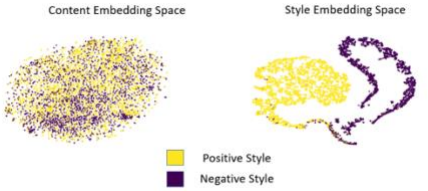


Figure 1: Disentanglement of Style and Content

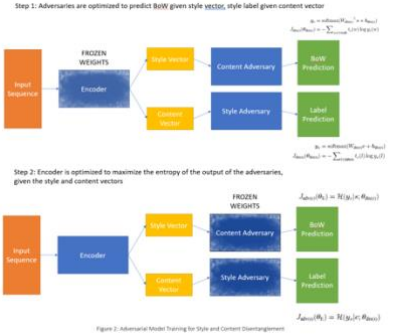
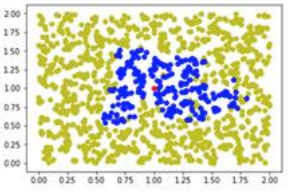


Figure 2: Adversarial Model Training for Style and Content Disentanglement

Kyle Shaw
Ph.D. in Computational Science
Advisor: *Peter Beerli*

Population Structures' effect on Genetic Patterns

Genetic data is intricately tied to the sample that was collected at a particular geographic point. Each member of a species can only travel so far; thus, genetic data can only travel so far in each generation. Thus, population structure and geographic features can have stark effects on the patterns found in genetic data. However, few modern methods include geographic data, and of those few, none do it very well. Many researchers are attempting to change this. We are developing simulation experiments to understand how geography affects genetic patterns. A simulated advantageous allele experiment shows that new mutations often spread through the population like a wave in a pond. A structured population simulation helps reveal how difficult it can be for genetic data to travel through sparse population patches. These methods will, in turn, help with the creation of new, more accurate methods.



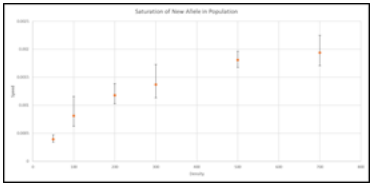


Figure 1: Snapshot of the spread of a new gene though a population (travelling wave).
Figure 2: Velocity of travelling wave dependent upon population density.

Stephen Townsend
Ph.D. in Computational Science
Advisors: *Peter Beerli & Anke Meyer-Baese, with special acknowledgement to Dennis Slice*

A Theoretical look at the Application of Shape Analysis with Acoustics and Signals

Over the past several years, a small community of researchers has arisen to develop and apply statistical Shape Analysis and Morphometrics to the study of Acoustics and other Time-Varying signals. Here, we discuss the theoretical underpinnings of Shape Analysis and Morphometrics as it applies to Acoustics and Signals. Ultimately, we show in this work that direct theoretical parallels can be drawn from the concepts and techniques of Morphometrics to generate meaningful and valuable results in the field of Bioacoustics. We discuss how this theoretical framework differs from previous work and provide new insights into Bioacoustics and acoustic ecology. Using this theoretical framework, we build a pipeline for analyzing bioacoustic data in a Morphometric context, and we show that the pipeline can be highly scalable to large problems. We also discuss algorithmic and implementational aspects of this pipeline including those for the extraction, alignment and analysis of individual pulses in a time-varying signal with pulse-like properties.

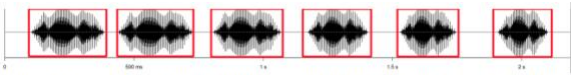


Figure 1: An example how the pipeline finds and extracts individual pulses (red boxes). The call from which these pulses were generated was simulated using the SoundGen Package in R.

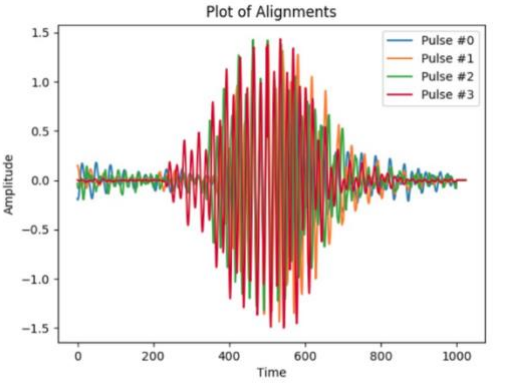


Figure 2: An example of a set of 4 maximally aligned pulses from a sample of real Frog Calls obtained from the MacAulay Library of the Cornell Lab of Ornithology.

Toan Tran
Ph.D. in Computational Science
Advisor: *Gordon Erlebacher*

Exploration of population-level patterns in spontaneous neural activity in mice

Research interest in spontaneous brain activity has increased over past decades as the idea of these noisy patterns holding valuable information about the structural and functional architecture of the brain gained traction among neuroscientists. This kind of activity is generally characterized by the absence of stimulation and how it is defined can vary depending on the settings of experiments. In this project, we analyzed neural activity of unconstrained mice in darkness which was recorded simultaneously across several brain regions. Despite the noisy appearance, the spontaneous activity exhibits structured patterns on the population level. Specifically, there are periods of synchronized surge in the activity of most neurons in the dataset. We distinguished neurons with activity patterns representative of the population trend from their noisy counterparts using two complementary approaches: (1) use sliding window embedding to convert each spike-train/time series into a more geometric representation and characterize it topologically with persistent homology, (2) analyze the autocorrelation and partial autocorrelation functions of the spike-trains. Further investigation of the distinction between these two groups of neurons may have functional implications regarding spontaneous brain activity.

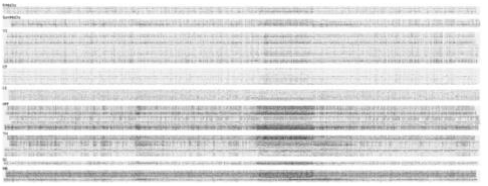


Figure 1: Raster plot of the spike-trains of neurons in one mouse separated by brain regions.

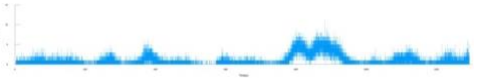


Figure 2: Activity of a “representative” neuron.

Jingze Zhang

Ph.D. in Computational Science
Advisor: *Gordon Erlebacher*

Written Style analysis using parts of speech (POS)

Stylometry studies written language style. Generally, writing style is only determined by the choice of words, sentence and paragraph structure, regardless of semantic meaning. Traditional methods operate on the frequency of function words to represent the style. However, function words, such as articles and pronouns, comprise only a small component of style. Part-of-speech (POS) vectors, which classify each word into one of 49 categories, define an alternate vocabulary, which will allow an elementary characterization of style. Starting from a collection of 10,000 emails, we first construct a frequency table of POS elements, to which we apply an SVD. Each text is represented by a vector, where each entry represents the POS frequency. We apply POS analysis to translate each entity (word, punctuation) into a POS element. We then construct a POS frequency table across the entire mail corpus. We then display the coefficients associated with the three dominant singular values for each email. As illustrated in the figures below, different people use POS elements differently, thus have different style. In future work, we plan to measure written deception through stylistic change.

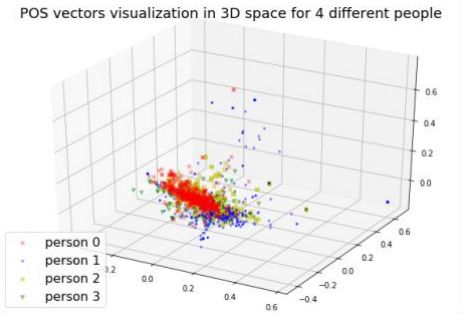


Figure 2. The high dimensional POS vectors are projected into 3D space. Each point corresponds to a single email, while color denotes the email author.

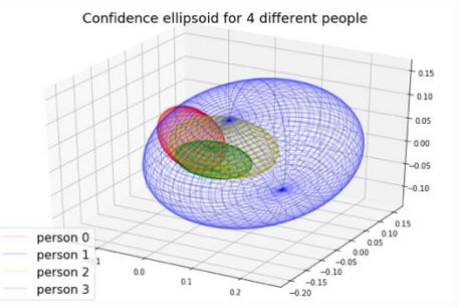


Figure 1. Confidence ellipsoids define a one-sigma region about the mean for a given author. A larger ellipsoid indicates that the author changes writing styles frequently. The different locations of ellipsoid suggest that different people have different writing styles.

Kevin Ziegler
Ph.D. in Computational Science
Advisor: *Alan Lemmon*

Estimating Accurate Gene Trees in the presence of Intra-locus Recombination; a Simulation Study

Data was simulated to test the impact of intra-locus recombination on estimating gene trees; the aim of the study is to compare feasibility of different methods and approach in dealing with recombination. The data simulated emulates a model of human evolution; 10 individuals are sampled from 6 populations, which results in a 60 taxon tree. Recombination rates, population sizes, and substitution rates are varied. A standard maximum likelihood method, Fast Tree, is tested along with newer ancestral recombination graph (ARG) like programs. Results show high levels of intra-locus recombination can completely disrupt Fast Tree's ability to accurately estimate a gene tree. Tsinfer, an ARG like program, was the most accurate in estimating trees and the most robust to changing simulation parameters.

Figure, below: Simulated C-Gene (regions not containing any recombination) lengths on a log scale color gradient. The x axis represents a population scaling parameter where 1.0 represents ancestral human population sizes. The y axis represents recombination rate scaling where 1.0 represents 1*10^-8. The size of recombination free regions ranged from almost 1 to 100,000.