The Department of Scientific Computing
Florida State University

Cx20

2020 Computational Xposition

April 13 & 15, 2020

Welcome to Computational Exposition 2020, a yearly event where students in the Department of Scientific Computing (DSC) showcase the results of their research in the past year. 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 fulfilment 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 becoming 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 serves as evidence that we are already there!

In this year, we are in the midst of a word-wide pandemic, our students and our faculty confined to their homes. Rather than cancel this event, we have moved it to the virtual world. Each student will present their work in a morning Zoom session, with opportunity for one-on-one interactions in the afternoon. This virtual event will take place over two days, and will be the first of its kind.

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

Faculty Mentors

Bryan Quaife, Faculty Organizer

Peter Beerli

Gordon Erlebacher

Chen Huang

Alan Lemmon

Anke Meyer-Baese

Tomek Plewa

Sachin Shanbhag

Xiaoqiang Wang

Presenting Researchers

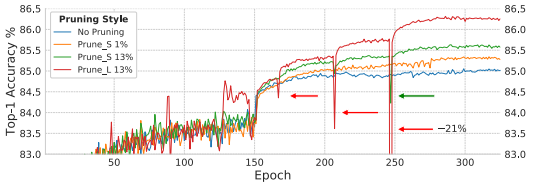
R E S E A R C H

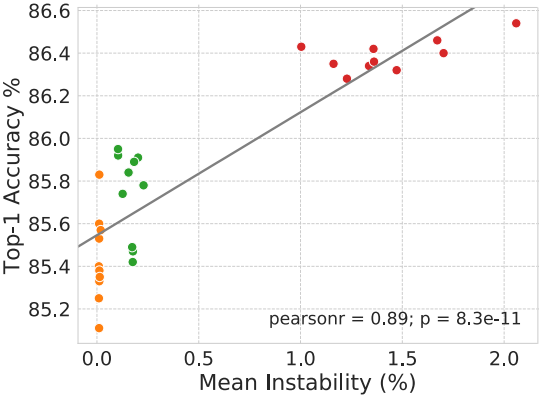
Abstracts & Graphics

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

The Generalization-Stability Tradeoff in Neural Network Pruning

Pruning neural network parameters is often viewed as a means to compress models, but pruning has also been motivated by the desire to prevent overfitting. This motivation is particularly relevant given the perhaps surprising observation that a wide variety of pruning approaches increase test accuracy despite sometimes massive reductions in parameter counts. To better understand this phenomenon, we analyze the behavior of pruning over the course of training, finding that pruning's effect on generalization relies more on the instability it generates (defined as the drops in test accuracy immediately following pruning) than on the final size of the pruned model. We demonstrate that even the pruning of unimportant parameters can lead to such instability, and show similarities between pruning and regularizing by injecting noise, suggesting a mechanism for pruning-based generalization improvements that is compatible with the strong generalization recently observed in over-parameterized networks.





Images Above and Right: Pruning instability improves generalization of VGG11 when training on CIFAR-10 (10 runs per configuration). (Left) Test accuracy during training of several models illustrates how adaptation to more unstable pruning leads to better generalization. (Right) We observe strong correlation between mean instability and accuracy, means reduce along the epoch dimension (creating one point per run-configuration combination).

Siddhartha Bishnu
Ph.D. in Computational Science
Advisor: *Bryan Quaife*

Stability Analysis of Spectral Deferred Correction Method for Time-Stepping Differential Equations

Over the years, considerable efforts have been made in designing spatial discretizations of climate models which are either high order accurate or conserve desired physical quantities. In contrast, much less attention has been devoted to the existing low-order temporal discretizations which can accumulate significant amounts of error over long simulation times. This motivates implementing a high order spectral deferred correction (SDC) method for time-stepping the fast depth-averaged barotropic equations of a stratified ocean model. SDC repeatedly applies a low-order solver to systematically increase the order of accuracy. Before applying SDC to solve the barotropic equations of a stratified ocean model equations with Coriolis force and bathymetry, it is worthwhile to perform a linear stability analysis to, equivalent to the non-linear shallow water a gain deeper insight into its underlying properties. In this poster, we apply SDC to a linear ODE and a linear system of coupled ODEs satisfying a parameter constraint. Interestingly, after undergoing Fourier Transform, the linear advection and shallow water equations, continuous or discretized, reduces to the same form as the above-mentioned ODE and the coupled system of ODEs. After verifying spectral convergence, we study the behavior of the growth rates i.e. the amplification factor for the ODE and the spectral radius of the amplification matrix for the coupled system of ODEs within the region of absolute stability. For the corresponding PDEs, this stability analysis provides us with an upper bound on the Courant number and the hence the maximum permissible time-step for a given spatial mesh and SDC parameters.

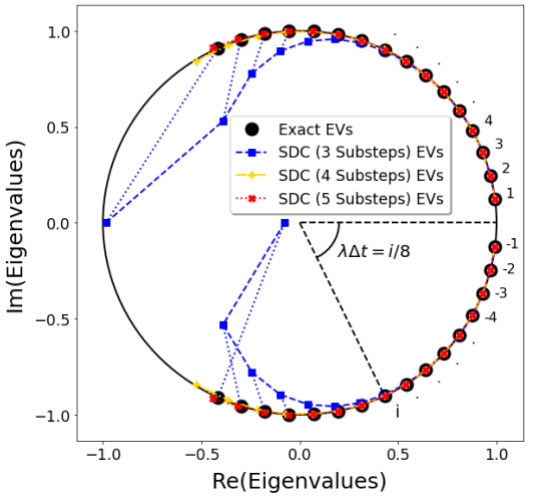


Figure 1: Eigenvalues of Amplification Matrix for Linear Coupled ODEs

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

Detonability of white dwarf plasma: Turbulence models at low densities

The deflagration-to-detonation transition (DDT) mechanism has been long studied but remains one of the major unsolved problems of theoretical combustion. It has been directly observed in a number of laboratory experiments and has been extensively studied through numerous numerical simulations. Also, astrophysicists have suspected for almost 40 years that it is directly responsible for at least a subclass of white dwarf explosions powering Type Ia supernovae (SN Ia). Astrophysical observational evidence for the DDT is, however, only indirect, which hinders progress in understanding the SN Ia explosion mechanism. Regardless of whether or not the deflagration initially exists, the viable explosion mechanism requires conditions in which the flow is energized and sizeable parcels of fuel attain a critical burning temperature. The question then is, what is the source of that energy? One possibility might be turbulence existing in the white dwarf plasma. Recent studies indicate that the sudden compression of viscous turbulence may lead to the rapid release of turbulent kinetic energy. In the SN Ia context this could result in the initiation of a detonation if the released energy is sufficiently large. Additionally, the importance of compressibility in formation of detonations for SN Ia conditions was studied recently by our group. We found that in order to initiate detonations, the turbulence has to be strongly compressively-driven. In the case that the nuclear fuel was already partially processed by deflagration, in a recently proposed unified mechanism, interaction between a highly subsonic flame with a highly subsonic turbulence produces a shock, which in turn leads to a detonation. This turbulence-induced DDT mechanism also depends on a number of parameters, such as the degree of compressibility of turbulence and properties of the flame. In this work we build upon our previous study, and consider the sensitivity of nuclear burning to the intensity of turbulence. We perform a series of well-resolved turbulence simulations of carbon-oxygen mixtures for a plasma density of 1×10^7 g/cm^3 in both 2D and 3D, systematically changing the amount of energy used to drive turbulence. Depending on this factor, we find a range of behaviors from non-explosive scenarios to cases involving partially burned fuel as well as situations when complete burning occurs. In our future studies, we will explore additional sources of plasma heating, besides kinetic energy dissipation.

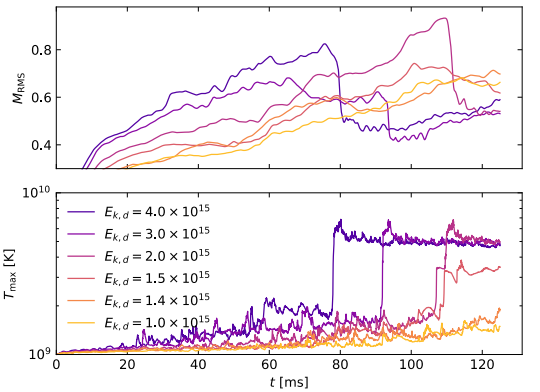
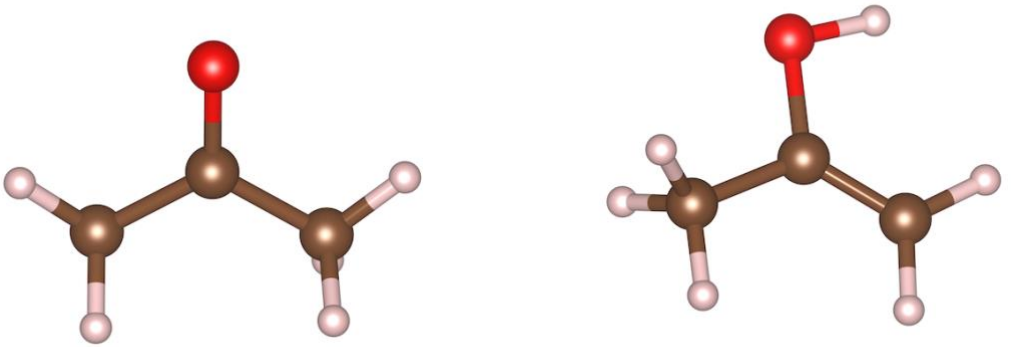


Figure 1.: Turbulent combustion models in 2D.

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

Directly calculating the RPA correlation energy difference between two similar systems

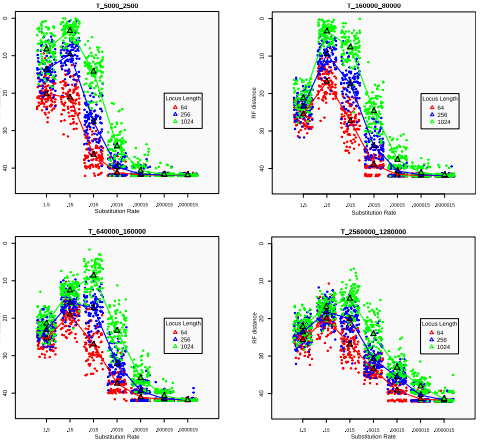
Random Phase approximation (RPA) correlation provides a route to improve the accuracy of Kohn-Sham (KS) density functional theory (DFT). In our previous project, we developed an efficient stochastic method to calculate the RPA correlation energy differences between two similar large systems. In this work, we develop a method to calculate RPA correlation energy directly. The RPA correlation energy can be formulated in terms of the trace of a large matrix which is the product between the Coulomb matrix and the KS linear response function. The trace is then expressed in terms of the eigenvalues of that large matrix. Fortunately, most of the eigenvalues are very close to zero and do not contribute much to the RPA correlation energy. We then develop a method to solve the lowest eigenvalues of this large matrix. The conjugate gradient (CG) method is used to solve for these low eigenvalues one by one. The Gram-Schmidt procedure is used to ensure that the current eigenvector is orthogonal to all lower eigenvectors. This method allows us to calculate the RPA correlation energy by solving the lowest eigenvalues of the large matrix directly.



Figures Above: The figures show our method will be applied to calculate the RPA correlation energy difference between these two similar systems.

Michael Conry
Ph.D. in Computational Science
Advisor: *Alan Lemmon*

Gene tree estimation accuracy under varying locus lengths, divergence times, and substitution rates



One of the central goals of evolutionary biology is to understand the evolutionary relationships among organisms by constructing phylogenetic estimates, commonly known as evolutionary trees. The accuracy of phylogenetic estimates can be strongly affected by the particular evolutionary processes that are taken into account during an analysis. One important process, genetic recombination, has been shown to lead to inaccurate phylogenetic estimates when ignored. In this novel simulation study we explore a subset of important parameters (locus length, species divergence time, and substitution rate) that require careful consideration during the preparation of DNA alignments used to estimate gene trees. Gene trees resulting from this simulation are compared to the true trees using the Robinson-Foulds (RF) metric to determine the extent of the discordance caused by short loci that do not contain sufficient information about the evolutionary history of a set of taxa, shallow divergence times that lead to incomplete lineage sorting (ILS), and extreme DNA substitution rates that exemplify species relationships that are difficult to resolve. This research is part of a larger study that examines the extent to which neglecting to accommodate for recombination in data partitioning approaches can lead to inaccurate species trees estimates. We intend to provide recommendations to empirical researchers as to when it is most beneficial to treat exons as independent evolutionary units in phylogenetic analyses.

Figure 1: The above four graphs demonstrate the changes locus length and substitution rate transmit to gene tree estimation accuracy. The title of each graph refers to the speciation times in a 3 species context (e.g. T_5000_2500 means species A split off from the population 5,000 generations ago and the remaining population split again to become species B and C 2,500 generations ago). The RF distances between the estimated trees and the true trees are minimized when the simulated loci have a substitution rate of .15 for shallow trees and .015 for deep coalescent times. This variance is intuitive because a short coalescence time will restrict the chance for the necessary ancestral information to arise in the form of substitution events. Consequently, the longer locus lengths both permit a lower substitution rate and help to increase gene tree accuracy by providing more ancestral information content. The question then becomes: What analysis approach can systematists employ to acquire longer sequences while still accommodating empirical assumptions like the presence of recombination?

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

Self Organizing Heirarchical Kernal Density Estimation

All observed data are samples produced from unknown stochastic processes. Determining the underlying distributions giving rise to these observations is an important task in many disciplines such as machine learning and Bayesian non-parametrics. Often times, these distributions are high in dimensionality, multimodal and complex with discontinuities such as jumps and edges. It is difficult for traditional parametric models to capture the structure of these distributions. Non-parametric techniques such as Kernel Density Estimators (KDE) provide more expressive power, though still suffer from the curse of dimensionality. Here, we demonstrate a non-parametric kernel-based method that side-steps the curse of dimensionality by approximating marginals of the true distribution and combining them hierarchically to reconstruct the full joint. We start by showing how a local learning rule from neuroscience called Hebbian Plasticity minimizes the reverse Kullback-Leibler divergence between a single kernel and an unknown distribution. This leads to each kernel finding modes in the marginal distributions. We then show how subsequent kernels listening to input from the approximate marginals learn to approximate the joint on a lower dimensional manifold.

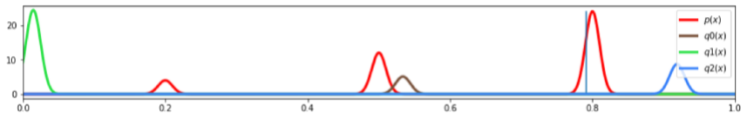


Figure 1: Three randomly initialized kernels $q_i(x)$ listen to observations (blue line) drawn from an unknown distribution $p(x)$. The parameters of each kernel are locally updated according to simple Hebbian plasticity rules with each observation. These learning rules can be shown to minimize the reverse Kullback-Leibler divergence between the kernel and the data generating distribution.

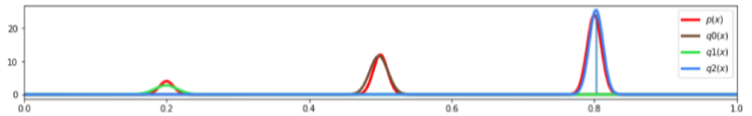


Figure 2: After a few hundred observations the kernels have all converged to local minima of the upper bound. Due to the initial conditions being spread out they each converged to a different mode of $p(x)$. If two kernels were initialized closer together they may have converged to the same mode.

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

Semi-Permeable Deformable Vesicles in a Viscous Fluid

Aquaporins are channels located on cell membranes that facilitate the movement of water into and out of a cell at much higher rates than osmosis. Studies have demonstrated that this transport across cell membranes plays a critical role in cell movement. We apply a high-order boundary integral equation method to simulate the motion of a single vesicle with a semi-permeable deformable membrane in a variety of Stokes flows. The dynamics will be compared with impermeable vesicles.

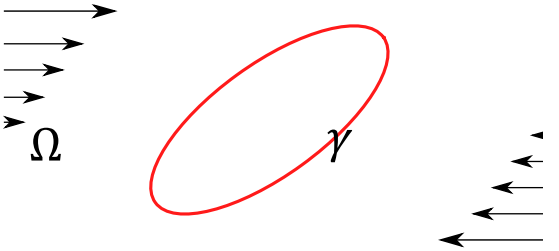


Figure 1, Above: Omega is the unbounded fluid domain and gamma is the vesicle membrane. In addition to the vesicle induced flow, a shear flow is imposed in the far-field.

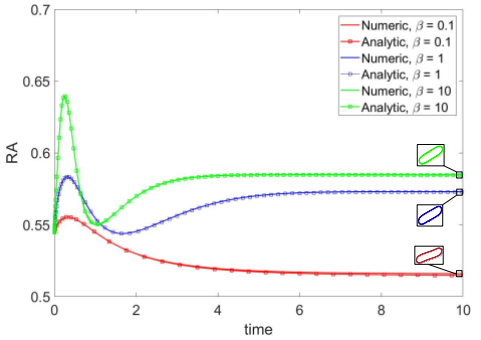


Figure 2: For each beta value, there is an asymptotic reduced area (RA). The final RA depends on the water flux coefficient, beta . In this example, the initial RA is 0.55. The analytic expression (*) is used to predict the RA values of each curve with first order accuracy.

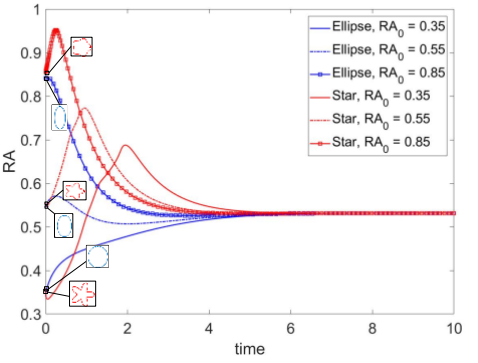


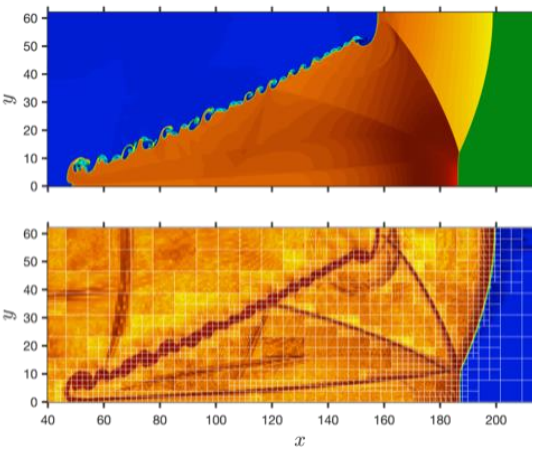
Figure 3: The final RA does not depend on the initial RA or vesicle shape. In this figure, = 1 and all vesicles have the same length.

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

A multi-resolution hybrid adaptive approach for turbulent reactive flows

A multiresolution framework for the simulation of compressible, turbulent reactive flows on adaptive quadtree/octree block-structured meshes is presented. Multiresolution indicators are used not only to adaptively refine the mesh, but also to identify where direct flux, equation of state, or reactive source term computations may be replaced with interpolation, once the mesh is fixed. The scheme addresses a major shortcoming of the tree-based adaptive mesh refinement (AMR) approach: that the graded nature of the mesh, in the presence of multi-scale fluid features, allows for the existence of potentially many blocks whose cells are resolved beyond the desired error tolerance.

To overcome this issue we use an additional hierarchical multiresolution data structure on a block to perform these calculations at a more coarse level if the flow is locally smooth enough. Then the multiresolution basis is used to prolong the resultant values to the block's native grid level. The scheme achieves a worthwhile decrease in simulation time for problems with multiscale features. Most importantly, we show that the error introduced by this approximation procedure is shown to be of the same order introduced by the adaptive mesh refinement procedure itself, therefore the savings come at virtually no cost to overall solution accuracy. The efficiency and accuracy of the scheme is demonstrated for problems with reaction-driven, compressible turbulence relevant to astrophysical applications.



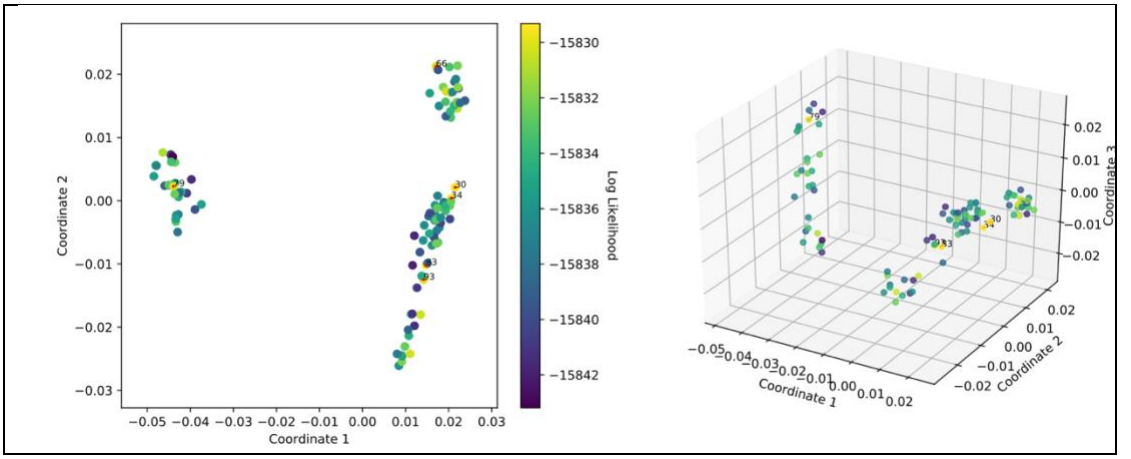
Figures: A shock wave passes through an inclined, density-stratified interface, inducing an unstable vortex layer (top). Multiresolution coefficients are used to identify the shock waves, vortices, and weak acoustic signals, and guide the mesh refinement accordingly (bottom).

Marzieh Khodaei-Gheshlah
Ph.D. in Computational Science
Advisors: *Peter Beerli*

Explorations in Tree Space: Where are the good trees?

Tree space has many islands containing phylogenies with high likelihoods. We expected that trees on the same tree-islands defined by a distance metric among trees may have similar likelihoods. This assumption seems incorrect. Finding the best tree in this space is difficult!

We generated a sample of 100,000 trees using the program REVBAYES (Höna et al. 2016) and their tutorial dataset primates_cytb_JC. From this sample, we extracted 100 rooted trees that were 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 (MDS – Trevor et al. 2000). We marked the six trees with the highest likelihoods: they are labeled 30, 34, 66, 79, 83, and 93. It turns out that these six best trees fell into several different group.



Figures Above: Visualization of tree space in the first two dimensions (left) and three (right) dimensions of the MDS plot. Each dot is a tree, the lighter the dot the higher the likelihood of the tree, the 6 best trees are marked with red. They are in different tree clouds.

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

Phase Field Modeling of a Sickled RBC

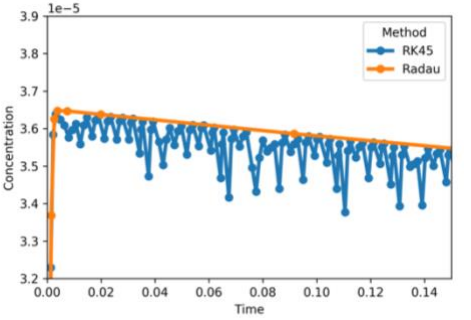
Sickle cell anemia is a genetic illness caused by an HBB gene mutation that produces an abnormal version of beta-globin known as hemoglobin S (HbS). It affects millions of people worldwide, and some of the symptoms are fatigue, acute and chronic pain, bone pain, dactylitis (swelling and inflammation of the hands and/or feet), and pulmonary hypertension. The average life expectancy of sickle cell anemia patients is 40-50 years. The only potential cure is bone marrow or stem cell transplant, which have high risks and can cause severe side effects.

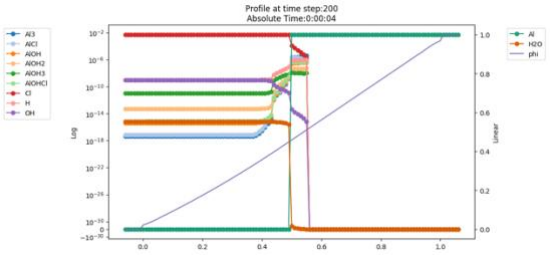
Many studies and much modeling of the dynamics of HbS polymer fibers have been done to better understand sickled red blood cells (RBC). However, integrating the dynamics and the interactions of the fibers and RBC membrane is challenging as they occur at multispatial scales, ranging from nanometers to micrometers, and models introduced so far are computationally expensive and require substantial CPU hours. In this poster, we discuss the capability of phase field method in modeling vesicle dynamics and introduce a potential HbS fiber - RBC membrane interaction model.

Eitan Lees
Ph.D. in Computational Science
Advisor: *Sachin Shanbhag*

A Simplified Nonlocal Multiphysics Model for Local Corrosion

Corrosion adversely affects the performance of metal alloys that are widely used as structural materials in the automobile, naval, and aircraft industries. In this work, I present a simplified multiphase (solid, liquid, and porous interphase) nonlocal model for pitting corrosion in aluminum alloys. I propose and solve nonlocal reaction diffusion equations that (i) explicitly account for multiple dissolved species, and (ii) enforce a nonlocal electroneutrality constraint. Initial results are presented, and shown to be qualitatively consistent with experimental observations. Future work is outlined, including model reduction on both homogeneous and heterogeneous model parameters, as well as combining results with a peridynamic failure model for the bulk material.





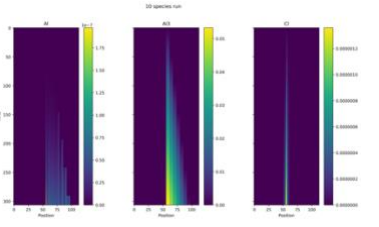


Figure 1, Top Right: Integration of chemical reaction networks often relay on processes happening at very different time scales. This leads to a set of stiff ordinary differential equations. This image compares and explicit and implicit method used in solving a set of stiff ODEs.

Figure 2, Lower Left. This is a snapshot of a 1-D model of corrosion. You can see the build up of corrosion byproducts on the surface of the bulk aluminum. Over time the electrolyte solution dissolves the Aluminum and corrodes more of the metal.

Figure 3, Lower Right: One of my research goals is to develop a reduced model of the corrosion system. This image shows the difference between the full model and one of the reduced models over time.

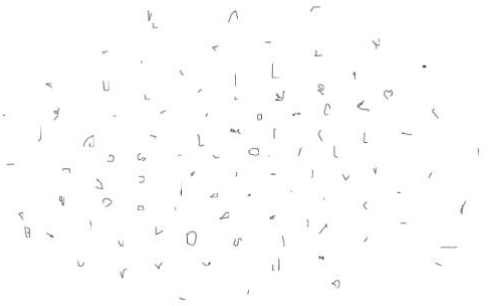
Behshad Mohebali
Ph.D. in Computational Science
Advisor: *Gordon Erlebacher*

Role of Sensitivity Analysis in Stress Testing Power System Controllers

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. Such simulations provide a closer to reality framework for testing new power hardware or newly designed controller by matching the speed of the model execution with the "wall clock" and using physical and/or data interfaces to incorporate the devices-under-test (DUT) into a simulation model. Due to their real-time characteristics, these simulations cannot be accelerated by using more powerful hardware. This makes the comprehensive evaluation of a given DUT more challenging. Usually the conditions and circumstances of the simulation (known as simulation scenario) are dictated by experts to observe the behavior of the defined performance metrics when the DUT is put under stress. Although this approach is effective in establishing whether a DUT passes the requirements of the design, it cannot push the DUT to its absolute limit. Doing so needs a systematic way to determine and manipulate a usually vast number of model parameters to check all possible scenarios and see which ones push the metrics beyond their acceptable range. However, not all the model variable have the same influence on the performance metrics. Knowing the extent of influence that each variable has on a given metric can help analysts reduce the dimension of the search space drastically and therefore reduce the processing time needed to carry the analysis. Here, we are using a Power Generation Module (PGM) model used on next generation Navy ships as a case study. Different sensitivity analysis approaches are tested on the model.

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

Learning Compositional and Causal Inductive Biases for Character Generation from Sketches



In recent years, the performance of handwriting recognition algorithms has increased dramatically, enabling many artificial intelligence (AI) applications that are now increasingly becoming part of our everyday lives. However, despite their impressive performance, current algorithms neglect to utilize the causal processes that underlie handwritten character generation, forcing them to learn every character from scratch. One way to better capture these processes is by building generative models that combine sketch and image data. Here, we show preliminary steps towards this goal by applying a state of the art sketch generative model to the omniglot dataset, a popular dataset for handwritten character generation, and show that generative models that utilize sketches have more expressive latent spaces than models that exclusively use images. Furthermore, we apply this model to individual strokes to explore how well current generative models for sketches capture the degrees of variation across strokes.

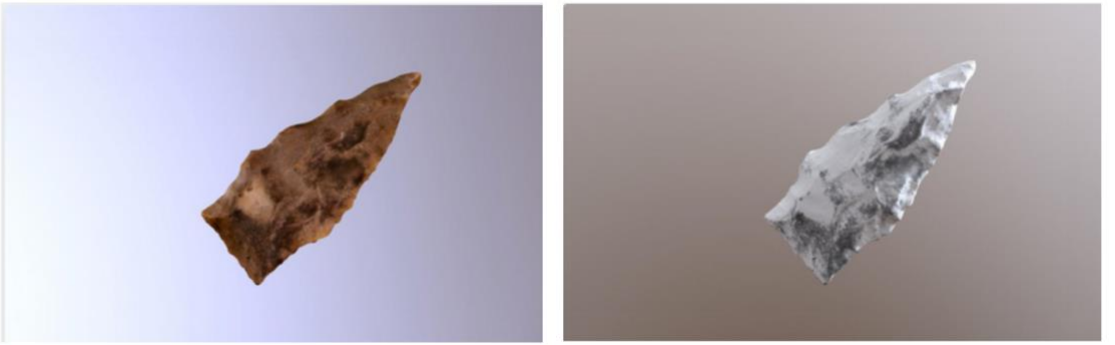
Figure 1, Left: Examples of characters from the omniglot dataset, displayed as blocks of 15 characters from 28 alphabets (Blake et. al, 2015).

Figure 2, Above, Right: A t-SNE projection (with perplexity = 40.0) of the embedding space for 100 single strokes randomly sampled from the omniglot dataset. The projection illustrates that the model is able to capture properties (i.e sharpness, roundness, length) of strokes that repeatedly appear in different characters across alphabets.

Marcelina Nagales
Masters in Computational Science
Advisor: *Bryan Quaife*

A Comparison of Three-Dimensional Scanning Methodologies in Digital Reconstruction of Lithics

We compare different methodologies used to digitize lithics projectile points/knives. North American projectile point/knife (PPK) typology and classification varies widely across cultural groups, time period, geographic locale, and archaeological interpretation. The Florida Bureau of Archaeological Research is compiling a digital archive of 3D PPKs to facilitate virtual research that would otherwise be confined to the curatorial facility. Comprising of different types of PPKs from the Southeastern United States, various scanning methods were used to create highly accurate 3D renderings, using high-resolution 3D scanners and photogrammetry. The HDI LMI 109A scanner with FlexScan software uses blue light scanning to create black and white 3D models that were rasterized with images taken of the physical object using Meshlab. The NextEngine Ultra HD scanner with ScanStudio Software uses stereographic cameras and LED light to create 3D color models. Photogrammetry---using a DSLR camera, Agisoft software, and photograph editing software---allowing for model modification in response to environmental and light exposure factors. These 3D renderings from different 3D creation methods were compared using Cloud Compare software and Ben Pomidor's Generalized Procrustes Surface Analysis (GPSA) software. Cloud Compare uses the meshes of the same object from each digitizing methodology and compares them by calculating the distance between landmarks. GPSA software also compares the meshes of the same object from each digitizing methodology by calculating distances but without using landmarks. The two methods of comparison can reinforce and improve our understanding of the difference between 3D scanners and photogrammetry. This research project discusses the differences between methods specifically for scanning a variety of PPKs and provides a set of instructions for researchers who will digitize PPKs in the future. Additionally, we continue to build the archive to include a broader representation of Southeastern lithic technology and apply this methodology to other artifact types and typologies.



FiguresThese two 3D models are of the same Jackson-type projectile point/knife (PPK). The left model was created using the Nextengine Ultra HD 3D scanner and the right model was created using the LMI HDI 109A 3D scanner.

Jonathon Nozowitz
Ph.D. in Computational Science
Advisor: *Tomasz Plewa*

Development of LAVAflow: A Large-Scale Data Analytics Package

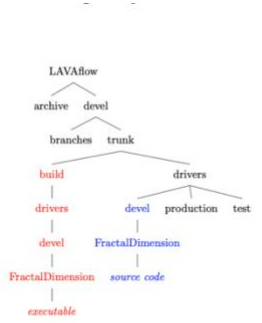


Figure 1: Organization of the LAVAflow package related to fractal dimension analysis with location of the source code (shown in blue) and executable building path (shown in red). This structure is generic to other LAVAflow analysis modules.

High resolution computer simulations can generate datasets with a size on the order of tens of terabytes. Should more physics components be added to the problem, such as magnetic fields or complex compositions, the data size requirements may substantially increase. Often, analyzing the data requires the use of a computer system of a size similar to that used to produce the actual simulation results. Thus, analysis of simulation results is frequently cumbersome, may require substantial computational resources, and motivates the development of novel data analysis methods in order to identify and understand the role of participating physics in the application.

To address this problem, we developed LAVAflow, Library for Advanced Variable Analysis. LAVAflow is a package comprising several modules for efficient analysis of computer simulation results of complex, multiphysics problems. The modules provide analysis capabilities for both grid-based and particle-based data in multidimensions. The computational mesh can be non-uniform, including block-based adaptive grids, and the data can be provided in a variety of formats including VTK and HDF5. The directory structure of the package provides information about the contents of that directory and makes navigating the package contents easier.

LAVAflow offers a high level of computational efficiency by using modular structure, dynamic memory management, and operating in parallel. The package optimizes the use of data to only the necessary subsets of solution variables to limit the required memory footprint. LAVAflow's modularity also enables continual development of the package. We are currently investigating the role of diffusive processes in magnetized turbulence. The corresponding analysis modules will be made available in the next release of the package.

Resolution	Time (s)
128^3	0.570642
256^3	3.946171
512^3	29.073812

Table 1: Times required to perform fractal dimension analysis for turbulent combustion simulation results obtained on meshes with resolutions of 128^3, 256^3, 512^3 using 6 cores on a Linux desktop workstation.

Marjan Sadeghi
Ph.D. in Computational Science
Advisor: *Peter Beerli*

Modeling Motif Occurrences with Hawkes Process

Haldane introduced the idea of using Poisson point processes to model recombination events in 1919. Recently, this process was incorporated in the genealogical history of a sample of sequences to handle recombination and coalescence, for example, Wiuf and Hein (1999). The Poisson process assumes that recombination events occur independently, and hence ignores phenomena such as interference and hotspots. We propose to replace Poisson processes with Hawkes processes to model recombination events. Poisson process has a deterministic intensity function, while a Hawkes process has stochastic intensity function making the occurrences of events depend on the entire history of the process. Therefore, the proposed algorithm will cover the dependency between the recombination events, such as interference and hotspots. We hope our algorithm improves the accuracy of recombination estimation.

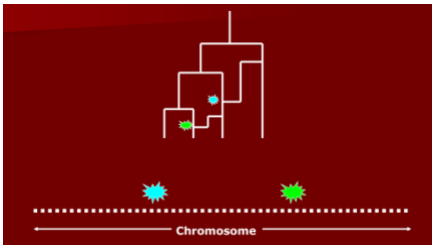


Figure 1: Wiuf and Hein algorithm. [4]

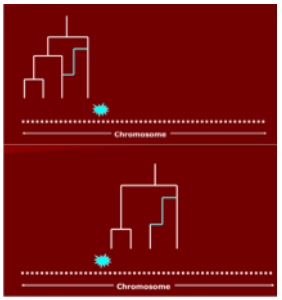
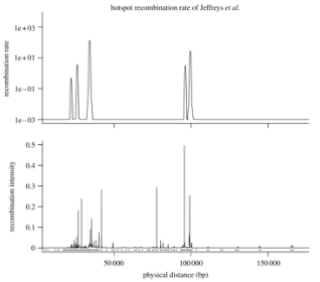


Figure2: SMC algorithm. [4]



Top figure shows recombination rates inferred from sperm-typing studies and the bottom shows the statistically inferred recombination intensities for the HLA. [2]

Daryn Sagel
Ph.D. in Computational Science
Advisor: *Bryan Quaife*

A New Way to Look at Fire: Computer Vision Applied to Fire Dynamics

Recent attention has focused on detecting wildfire activity and modeling its spread, but tools that enable us to compute the fluid and atmospheric properties from fire images are lacking. As opposed to sparse measurements obtained with expensive instrumentation, computer vision principles enable the analysis of fire, wind, and plume behavior from visual and infrared (IR) video. Data that quantifies the transport of heat and fire spread, turbulent statistical information, and plume structure can be obtained from either visual or IR images and contribute to our evolving understanding of fire behavior. Due to the visually unique environment of fires and complex turbulent nature of the dynamics, existing black-box computer vision programs are not suitable. I will describe several modifications of classical computer vision algorithms with adapted graph theory techniques that allow existing principles to be applied to diverse instances of this environment, and use them to extract data from prescribed fire videos. These data extraction experiments improve our understanding of the dynamics in complex environments and can be used to validate many variants of fire spread models.

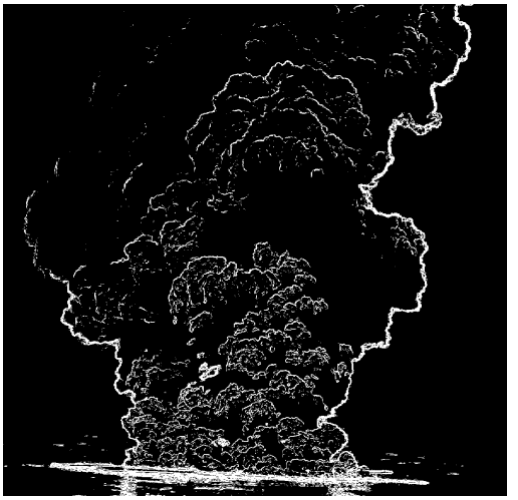


Figure 1: Structure of a smoke plume obtained through convolutions (3x3 Sobel) and thresholding

Kyle Shaw

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

Exploring effects of population density and selection coefficient on spread of new alleles

Population size is known to influence the interaction between selection and random drift. When viewed in a spatial landscape the density and size can be decoupled. We show that the density, not the overall size influences the interaction between selection and random drift. Using Messer's continuous simulation we model the introduction of an advantageous allele and its corresponding wave of advance first described by Fisher in 1937. Under low densities these advantageous alleles are influenced greatly by random drift. Within more dense populations the selection coefficient becomes much more important. We also show that as density increases the speed of the wave of advancement also increases.

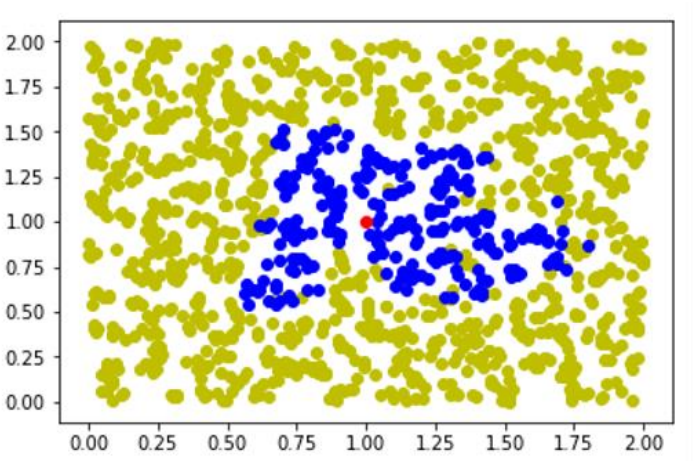


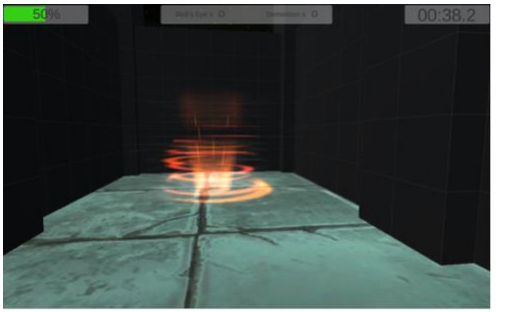
Figure 1: A simulation of a population. Each dot represents an individual with the axis representing geographic location. The Yellow dots represent individuals without the new mutation, while the blue dots are individuals with the new advantageous allele. The red dot shows the point of origination. Notice how the mutation has spread semi-radially about the point of origination. As a population becomes more dense, this process becomes less stochastic.

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

Bayes2Unity

Bayes2Unity is a package that allows for easy creation and use of Bayesian networks in Unity. The package relies on three main components: Unity, Infer.NET, and Netica. Each of these products has a unique role in the package. Netica allows for easy creation of a network. The graphical user interface allows users to simply layout and connect nodes. Initial training will be accomplished in Netica as well. Infer.NET is an inference engine that completes all the assessment, updating, and prediction. Most of the computation is completed by Infer.NET. Unity acts as the game engine and development environment. It provides a front end for displaying results and more importantly, a game world to propagate the results into. Two projects currently use the Bayes2Unity package: eRebuild and Preempting Path.

eRebuild is a first-person building game focusing on math education. A player collects materials and builds structures to home displaced families. These families then request players perform a variety of other tasks to help restore and improve the community. These include trading, painting, fencing, moving, and organizing. Each of these tasks links to a math competency with the results of the level being feed to a Bayesian network powered by



Preempting Path is a first-person exploration game that presents a series of procedural generated mazes with increasing difficulty to the player. As the player searches for the end, they may also pickup different power-ups to make it easier. The game starts with the simplest maze, a 2x1 maze, and generates a larger maze after each success. As the player completes levels, a Bayes2Unity updates a player model that represent their current play style. Preempting Path presents a simple experience that anyone can jump in and play without any explanation. We designed Preempting Path as a testbed for the experience driven procedural content generation. Initial results show a modest increase in engagement compared to players who experience a generic procedurally generated game world.

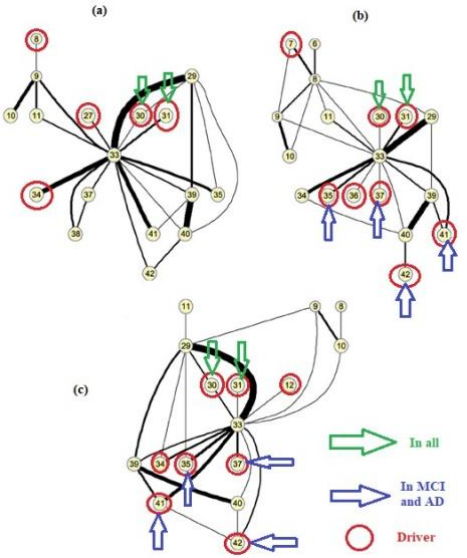
Figure 1, eRebuild. Bottom, Left: Students can login and continue the game from school or home. Teachers and parents can track progress using this login info.

Figure 2, Preempting Path. Top, Right: Levels are generated by the stats shown at the top: items collected, time, and percent complete. However, the most important factor is the rating they give each level.

Zlatko Sokolikj
Ph.D. in Computational Science
Advisor: *Anke Meyer-Baese*

Controllability of Structural Brain Networks in Dementia

The dynamics of large-scale neural circuits is known to play an important role in both aberrant and normal cognitive functioning. Describing these phenomena is extremely important when we want to get an understanding of the aging processes and for neurodegenerative disease evolution. Modern systems and control theory offer a wealth of methods and concepts that can be easily applied to facilitate insight into the dynamic processes governing disease evolution at the patient level, treatment response evaluation and revealing some central mechanism in a network that drives alterations in these diseases. Certain areas of the brain or nodes in the connectivity graph (structural or functional) can act as drivers and move the system (brain) into specific states of action. To determine these areas, we apply the novel concept of exact controllability and determine the minimum set and the location of driver nodes for dementia networks. Our results applied on structural brain networks in dementia suggest this novel technique can accurately describe the different node roles in controlling trajectories of brain networks, and show the transition of some driver nodes and the conservation of others in the course of this disease.



Figures, Left: Leader nodes for structural data. (A) Controls, (B) Mild cognitive Impairment (MCI), (C) AD subjects.

Philip Solimine
M.S. in Computational Science, Ph.D. in Economics
Advisor: *Anke Meyer-Baese*

Strategic Formation of Cooperative Networks

We hypothesize that providing information which enables direct reciprocity should increase gains from the networks, by making efficient cooperative equilibria more salient. We place subjects in groups of four, and have them play a network sharing game for 15 periods, after which they are told there will be a second part to the experiment, but not told what that will be. After 15 rounds, subjects are told the rules for the second part of the experiment. In the second part, subjects remain in the same groups but are assigned new identifiers and play another 15 rounds of the game. In treatment sessions, subjects are also shown their incoming benefits from the sharing of others in their group. We observe the effects of this intervention on several outcome variables.

Stephen Townsend
Ph.D. in Computational Science
Advisors: *Peter Beerli & Anke Meyer-Baese*

A Scalable Pipeline for Single-Pulse Analysis of Animal Vocalizations

The study of animal vocalizations is of central importance to the fields of ecology and population biology. In recent years, a myriad of new methods have been developed for working on animal vocalizations. Many, if not most of these, center on the use of spectrograms or summary numerical representations of the elements of the vocalization signals themselves as the basis for analysis. In previous posters, the authors introduced a new method for animal vocalizations which is termed single-pulse analysis. This method presents a robust and intuitive approach to analyzing vocalizations whose structure consists of self-similar pulses. In this poster, we present a scalable and parallelizable pipeline for performing single pulse analysis.

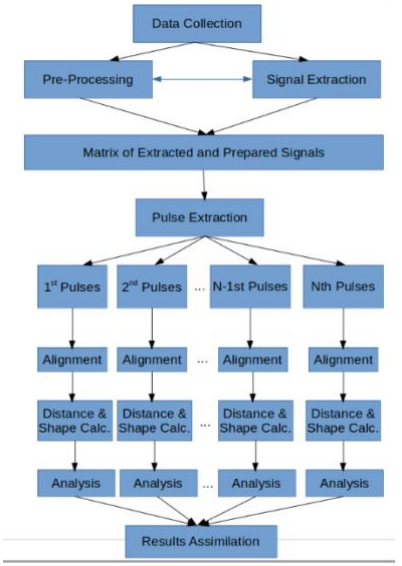


Figure 1: This figure shows a simple diagram of the entire framework being proposed with all major steps included.

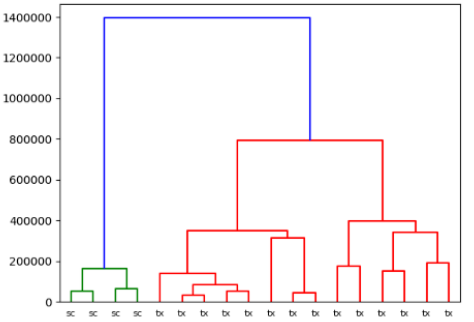


Figure 2: This dendrogram shows the results of running the framework on a single pulse from each of 18 calls from the populations investigated in Gerhardt (1974).

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

Written style analysis using Part-of-speech (POS) tags

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.

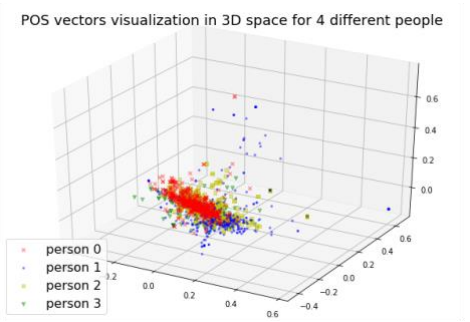


Figure 1. The high dimensional POS vectors are projected into 3D space. Each point corresponds to a single email, while color denotes the email author.

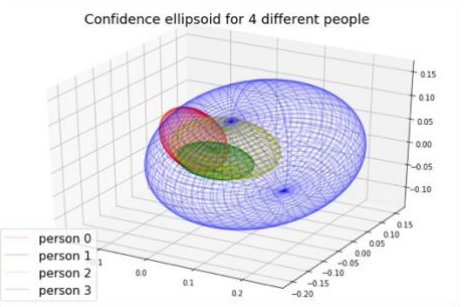
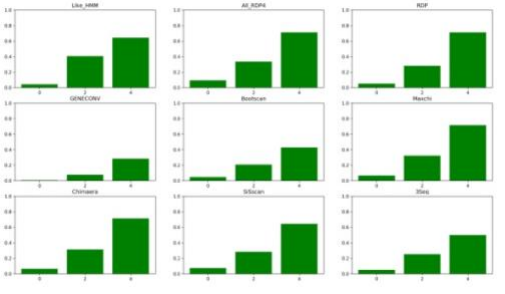


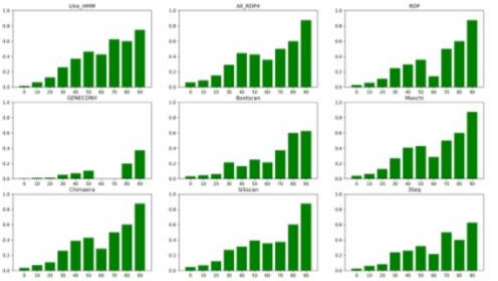
Figure 2. Confidence ellipsoids define a one-sigma region about the mean for a given author. A larger ellipsoid indicates that the author changes writing styles frequently. The different locations of ellipsoid suggest that different people have different writing styles.

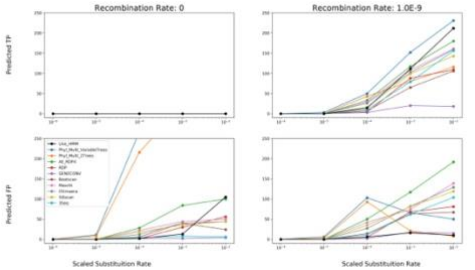
Kevin Ziegler
Ph.D. in Computational Science
Advisor: *Alan Lemmon*

Improved Recombination Breakpoint Estimation Utilizing Likelihood Ratio Tests: Like_HMM

Recombination causes genomic sites to have differences in their genealogical history. For this reason, recombination can produce error when phylogenetic trees are estimated under models that assume all sites have the same history [1][2][3]. One solution to this problem is to identify and remove loci with a history of recombination. A second solution, more amenable to genome-wide alignments, is to split loci with a history of recombination into two or more separate loci for analysis. A subset of the available methods aim to identify these recombination breakpoints. To our knowledge the accuracy of these methods have not been compared comprehensively. The method presented here automates and extends an existing method for detecting recombination breakpoints in an DNA sequence alignment using a Hidden Markov Model (phyML_Multi, [4]). The performance of our method is compared to existing methods using a simulated dataset designed to emulate Hominidae. Our program uses Likelihood Ratio Tests (LRTs) to substantially reduce the rate at which phyML_Multi falsely predicts recombination break points.







Top Left: figRFPlot: All simulated breakpoints are grouped based on the Robinson Foulds distance between the to alignments the breakpoint separates. The breakpoints separating alignments with large RF values are more likely to be detected.

Top Right: figwRFPlot.jpg : All simulated breakpoints are grouped into bins of size 10 based on the Weighted Robinson Foulds distance between the to alignments the breakpoint separates. The breakpoints separating alignments with large WRF values are more likely to be detected.

Left: test.jpg : Recombination rate 0 has 0 significant simulated breakpoints, recombination rate 10-9 has 440 significant breakpoints. The first row = true positives the second row = false positives. The left column = R0 and the