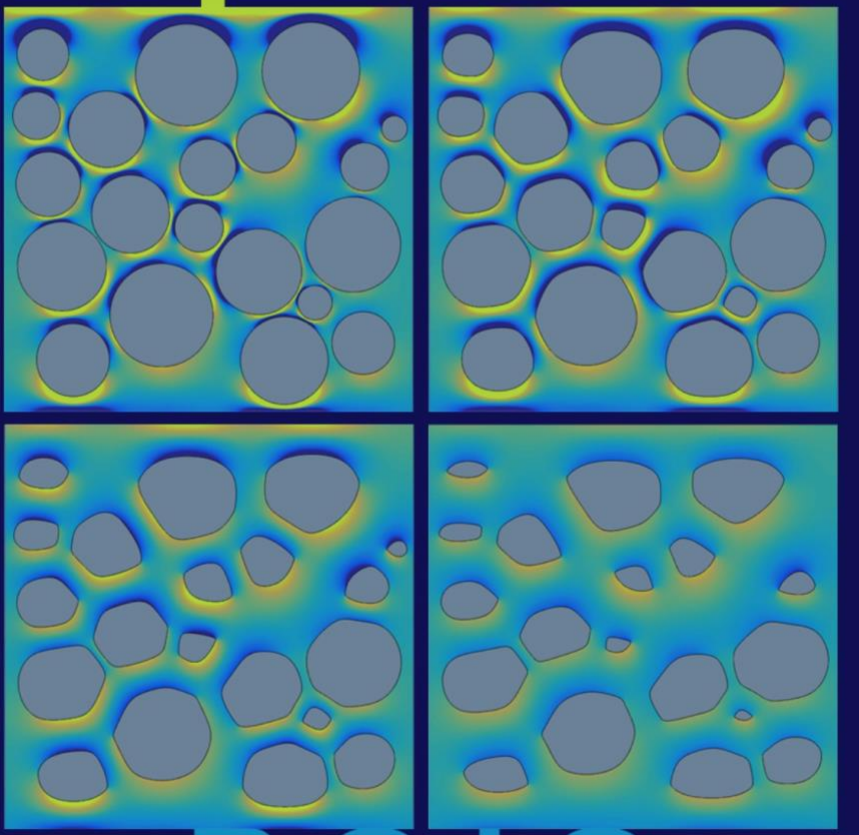
SCIENTIFIC COMPUTING

Computational Exposition



2019

Welcome to Computational Exposition 2019, 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 bolster 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!

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

Cover: Dense packing of eroding bodies, courtesy of Shang-Huan Chiu.

Presenting Researchers

RESEARCH

Abstracts & Graphics

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

When Can Neural Network Pruning Improve Generalization?

We investigate the generalization improvements conferred by several neural network pruning algorithms, including our own novel ones. We show that pruning during training, especially when pruning targets more important information pathways, can significantly increase the test accuracy of baseline models. Our investigation highlights the conditions with which pruning can offer such generalization improvements. In addition, we present preliminary results on our algorithms' competitiveness with state-of-the-art pruning strategies, illustrating that we can improve the accuracy of neural networks trained on various datasets while pruning large percentages of both convolutional-filter and linear layers.

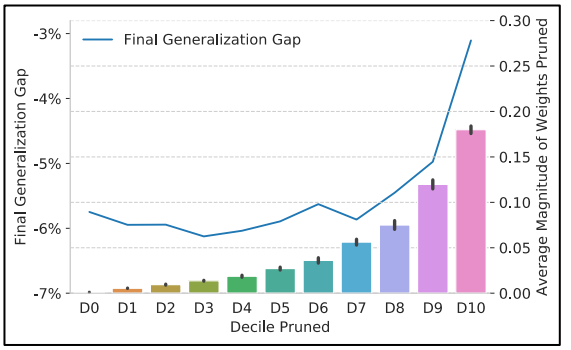


Figure 1. Decile Figure: We trained a 6-layer neural network on CIFAR10 with various pruning setups: each setup pruned the weights belonging to a particular weight-magnitude decile until all of the weights in that decile were pruned (thus, a total of 10% of all weights was pruned in each setup). We found that the final generalization gap (the difference between test and train accuracy) was smaller in the setups that pruned larger weights. The error bars are 90% confidence intervals for the mean, calculated by a bootstrap procedure operating on 10 distinct runs of each experiment.

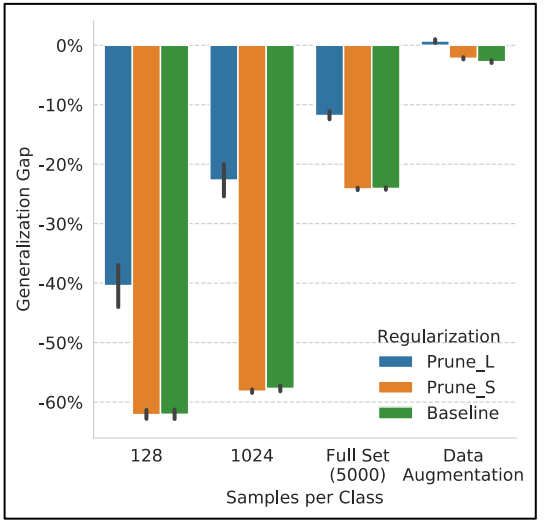
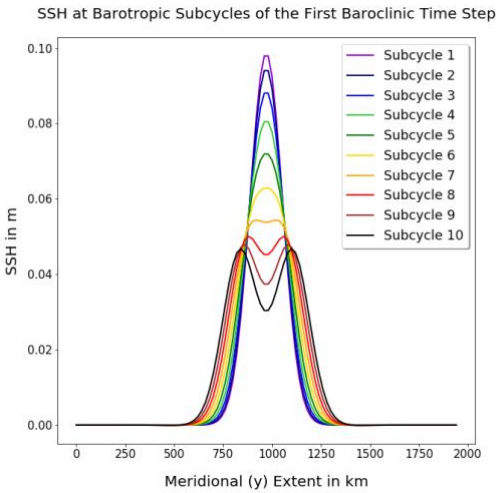


Figure 2: Subsample Figure: We trained a 6-layer neural network on CIFAR10 using four levels of data. Until the level of data rises to include augmentation with horizontal flips and random crops, the generalization-gap-shrinking (i.e., regularization) effect of pruning can significantly improve the baseline's test accuracy because the baseline's generalization gap (the difference between test and train accuracy) is responsible for a large amount of test error.

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

Efficiency of Different Filters in Split-Explicit Time-Stepping Algorithm

Ocean models permit a wide range of time scales with the speed of surface gravity waves being 100 ms^{-1} with that of the surface currents and internal gravity waves being two orders of magnitude less. It is not practical to advance the full 3D momentum equations with the smallest time step corresponding to the fastest waves due to the computational weight of the problem. So, a traditional approach is to split the momentum equations into two parts, a barotropic part for solving the depth independent fast 2D barotropic waves and a baroclinic part for solving the much slower 3D baroclinic waves. My research involves improving this barotropic-baroclinic splitting of the time-stepping algorithm of the Model for Prediction Across Scales − Ocean (MPAS-O) with a view to improving the stability and solution accuracy without compromising the computational time. More specifically, I have been studying (a) different filters for time-averaging the intermediate instantaneous barotropic modes and including the 'mean' solution in the time derivative for the next baroclinic (large) time step, and (b) different time-stepping algorithms for advancing these barotropic modes. I have programmed a one-dimensional shallow water equation solver in object-oriented Python for simulating the propagation of a surface gravity wave, where I have tested some of these filters and time- stepping schemes. In this poster, I will compare the efficiency of these filters and time-stepping algorithms with respect to the solution error norm and choose the optimum ones to be incorporated in the MPAS-O code base.



Ezra S. Brooker

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

The Inverse Problem: Connecting Core-Collapse Supernova Observations to the Explosion

A star much more massive than the Sun evolves rapidly and dies violently when its core can no longer produce a sufficient amount of energy to support it against the force of gravity. The resulting core collapse leads to the formation of a proto-neutron star (PNS) and, with the help of neutrinos emitted by the PNS, a successful supernova explosion.

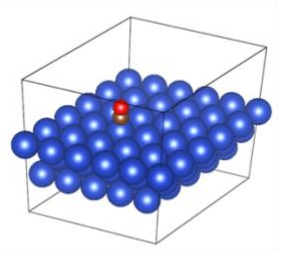
In this project, we study the evolution of core-collapse supernovae (ccSN) in search of progenitor-remnant connections. We simulate ccSN evolution, starting from the initial collapse, for about 1000 years when the young supernova remnant (ySNR) is formed and the simulation results can be directly compared to observations. We adapt existing code to efficiently include effects of various physics processes that dominate the evolution at different times. One of the proposed model modifications is to use a simplified neutrino-matter interaction model in which the neutrino heating of the post-shock region is described by a source term. This new model will use a dual discretization in which the energy source term (computed on a spherical mesh) is appropriately coupled to the hydrodynamic background (computed on the Cartesian-like mesh). The second required model modification calls for replacing the PNS with a point mass. This modification allows the code to overcome the restrictive time step limit due to the presence of PNS material characterized by extremely high sound speeds (extremely short cell sound crossing times).

Simulation results will allow us to produce model observables of young SNR and enable comparison to observations with respect to the ejecta structure and presence of compositional inhomogeneities presumably due to Rayleigh-Taylor instabilities. These instabilities develop early on during the explosion and provide information about the explosion mechanism and the structure of the progenitor star.

Fig. 1, Above: This image shows the northwest quadrant of the Vela supernova remnant in the southern constellation Vela.

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

Efficiently Calculating the RPA Energy Difference between Two Similar Systems Based on Atom-Centered Correlated Sampling



Random Phase approximation (RPA) correlation functional provides a route to improve the accuracy in Kohn-Sham (KS) density functional theory (DFT). However, the computational cost of RPA is high, and it is fourth power of system sizes $O(n^4)$. RPA correlation energy can be rewritten in terms of the density of states (DOS) of the product between Coulomb operator and the system's Kohn-Sham linear response function. The DOS can be stochastically evaluated by using the kernel polynomial method (KPM) by expanding it with Chebyshev polynomials. Therefore, the RPA energy difference can be reduced to computing the difference of the moments between the two systems. The purpose of this project is to develop an efficiently converged stochastic method to directly compute the RPA correlation energy differences between two similar large systems. Two independent series random vectors are sampled for two different systems in order to obtain the difference between the moments of the two systems. However, the convergence of the moment difference is not very fast. We develop an approach that is called atom-centered correlated sampling to improve the convergence of the moment difference. To achieve correlated sampling for the atom j, we define its Voronoi cell. For both system I and system II, the same random number sequence is used to sample the points in the Voronoi cell of the atom j. The moment difference between the two systems can be sampled using random vectors that are correlated in each atom's region. Thus, the efficient of convergence of the difference moments can be improved. We are in the progress of implementing this new method in the ABINIT program, an open-source, plane-wave based on DFT program. This method will be employed to obtain sufficiently accurate surface adsorption energies of molecules on solids to help us gain reliable understanding of heterogeneous catalysis with the atomic resolution.

Figure 1: This figure shows our method will be applied to obtain sufficiently accurate adsorption energies.

Shang-Huan Chiu
Postdoctoral Associate in Computational Science
Advisor: *Bryan Quaife*

Dense Packings of Eroding Bodies

Erosion is a fluid-mechanical process that is responsible for many naturally occurring geological phenomena. We present a boundary integral equation formulation to simulate the two-dimensional erosion of many bodies. A particular challenge of the formulation is accurately resolving interactions near an eroding body. We present a quadrature method to resolve these interactions by using a Barycentric-type formula applied to a complex variable formulation of the integral equation. The quadrature method is compared with the standard trapezoidal rule.

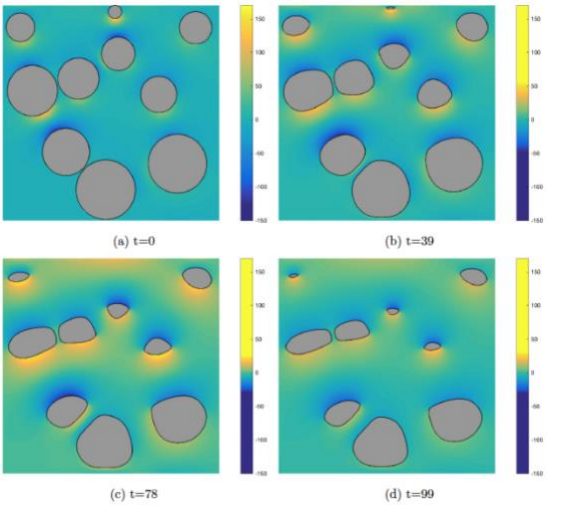


Figure 1: This figure shows our method will be applied to obtain sufficiently accurate adsorption energies.

Evan Cresswell-Clay
Ph.D. in Computational Science
Advisor: *Gordon Erlebacher*

Compartmental Modeling of Calcium Dynamics in Astrocytes

Several contemporary studies show that astrocytes, a type of glial cell, are fundamental to a variety of neural functions ranging from metabolic support to higher cognition such as recollection memory. This has led to the introduction of astrocytic dynamics into neural modeling. Most cellular functions in astrocytes are triggered by an increase or decrease in calcium concentration within the cytosol. Previous work treated astrocytic dynamics by representing calcium concentration as a point source or a completely spatial model in the cell. We now know that the role of the astrocyte takes many different perspectives. This work, which is inspired by in vivo recordings of astrocytes in the ferret visual cortex, models the different levels of intracellular calcium activity in the astrocyte.

In the model, we create a framework to enable the exploration of spatial calcium dynamics in astrocytes. We consider five astrocyte processes, modelled as five spatially partitioned compartments representing branches. Over each compartment we solve a system of differential equations for intracellular calcium dynamics. A branching structure, while not as general as a full 2D or 3D spatial simulation, allows the study of astrocyte's cellular properties over extended regions of space. Studying a spatial representation of the processes will help investigate the role of astrocyte morphology and spatially varying physiology in calcium signal propagation as well as developing intuition on the functional relationship between different levels of activity in the cell.

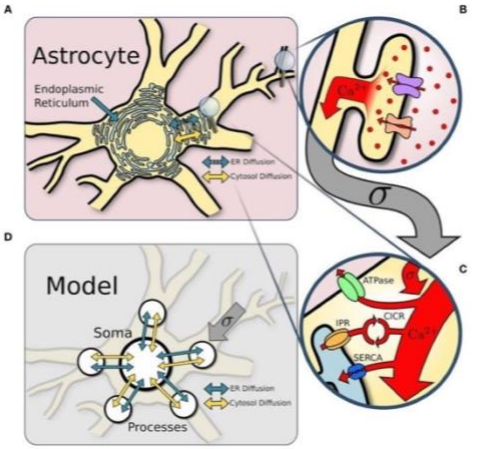


Figure 1, Modeling calcium signaling.

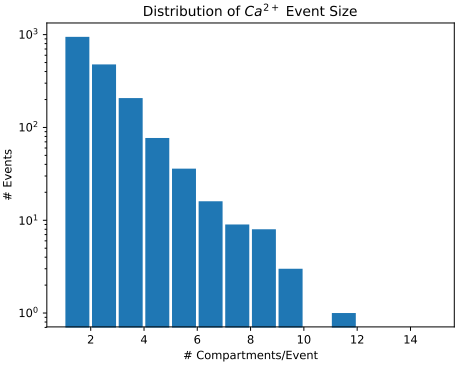


Figure 2. Distribution of calcium event sizes.

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

Self Organizing Hierarchical 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. Oftentimes, 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.

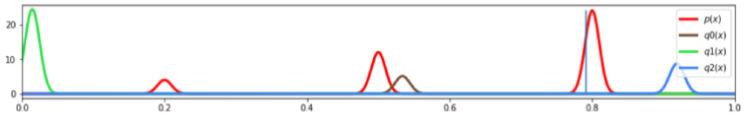


Figure 1: Three randomly initialized kernels $q_i(x)$ listen to observations (blue line) drawn from an unknown distribution $p(x)$. The parameters of each kernel are locally updated according to simple Hebbian plasticity rules with each observation. These learning rules can be shown to minimize the reverse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

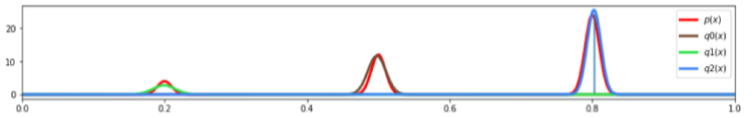


Figure 2: After a few hundred observations the kernels have all converged to local minima of the upper bound. Due to the initial conditions being spread out they each converged to a different mode of $p(x)$. If two kernels were initialized closer together they may have converged to the same mode.

Eric Dolores & Nathan Crock
Postdoctoral Associate & Ph.D, in Computational Science
Advisors: *Jose Mendoza Cortes, Gordon Erlebacher*

Cascades

We aim to develop an algebra that describes interaction between neurons. An activation diagram is a spatio temporal representation of the neurons and the path of possible signals among them, see Figure 1. External activations are induced by environmental signals, and we consider those initial conditions. Internal activations are the consequence of signals sent by neurons previously activated. We are interested in cascades, which are activation diagrams with the property that the number of internal activations is larger than the number of external activations.

Following the recent algebraic[1,2,3] and algebraic topology[4] models of neural data, we introduce operations on cascades of neurons with one self-activation and one external activation. We describe cascades that are self-sustained, and we also give an algorithm to create finite cascades.

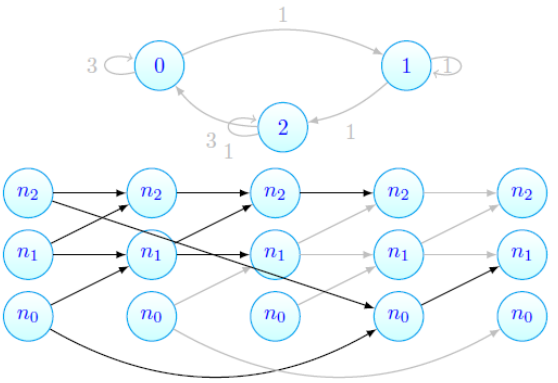


Figure 1: Above - a neural network. Below - an activation diagram with neurons externally activated at time 0.

1. C. Curto, A. Veliz-Cuba, N. Youngs. Analysis of combinatorial neural codes: an algebraic approach. Book chapter in Algebraic and Combinatorial Computational Biology. R. Robeva, M. Macaulay (Eds), 2018.
2. C. Curto, V. Itskov. Combinatorial neural codes. Handbook of Discrete and Combinatorial Mathematics, Second Edition, edited by Kenneth H. Rosen, CRC Press, 2018.
3. C. Curto, V. Itskov, A. Veliz-Cuba, N. Youngs. The neural ring: an algebraic tool for analyzing the intrinsic structure of neural codes. Bulletin of Mathematical Biology, Volume 75, Issue 9, pp. 1571-1611, 2013.
4. Giusti, C., Ghrist, R. and Bassett, D. (2016) Two's company, three (or more) is a simplex: Algebraic-topological tools for understanding higher-order structure in neural data. J. Comput. Neurosci.

Austin Eovito
M.S. in Computational Science

Advisor: *Gordon Erlebacher*

Analysis of Fraudulent Behavior with Vader Sentiment Analysis

As the needs of constituents' change, government at the local, state, and federal level must find ways to properly represent the needs of their jurisdiction. Compounding this issue is the need for government transparency and the ability to hold government accountable. In some cases, a lack of transparency leads to fraudulent behavior, which we address through sentiment analysis. Our team hypothesized that governmental communication would be business-oriented, i.e., overwhelmingly neutral. With the VADER package from NLTK, a corpus of approximately 30,000 emails was analyzed for positive, negative, and neutral sentiments in order to determine times, persons, and places that affected the overall sentiment. The results obtained from this analysis are preliminary; work in the domains topic analysis and vector space must be done in order to verify our findings.

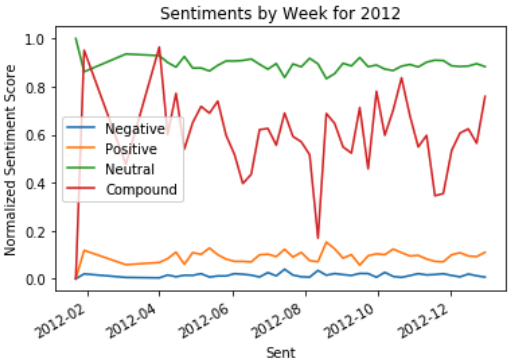


Figure 1: This is the beginning of the email, and the graph shows that there are large fluctuations in the compound valence. Specifically, the months of March and August in the year 2012, as well as November should be further investigated by breaking down the sentiment per day for the affected months. Analysis could also be done on strong upward motion in the compound sentiment, as it serves as the best indicator for changes in valence for sentiments in the corpus, as there are not many strong fluctuations in the negative or positive sentiments. This can also be interpreted as small fluctuations in negative and positive sentiments leading to an exponentially large change in the compound sentiment.
.

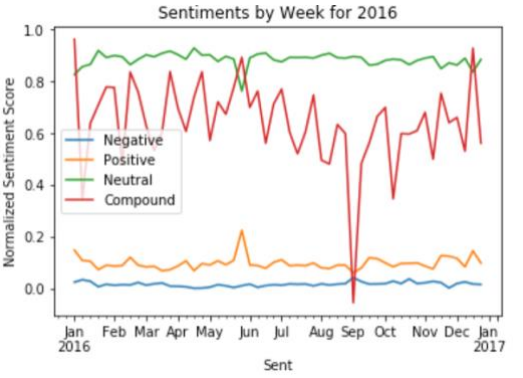


Figure 2: This graphic is the first one with a negative compound score, which happens in the month of September. There is also a large increase in the positive sentiment in the month of May that should be further broken down and analyzed. It can also be determined that an increase in the negative sentiment leads to a decrease in compound, and an increase in positive sentiment leads to an increase in the compound sentiment, which means that our analyzer is properly computing the compound sentiment.
.

Ashley Gannon
Ph.D. in Computational Science

Advisor: *Bryan Quaife*

Vesicle Adhesion in Constricted Geometries

In red blood cell suspensions, adhesion creates long aggregates (Figure 1) that affect the rheology of the flow. In many physiological geometries, the aggregate must be broken up so that the red blood cells can pass through a narrow constriction in single file. Therefore, hydrodynamic forces must overcome the adhesive forces. We examine the properties of a single vesicle (a proxy for red blood cells) in a constricted channel.

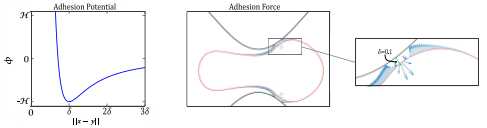
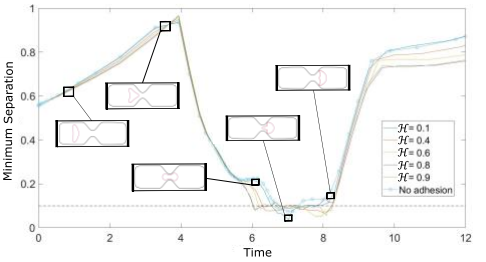


Figure 1, Above: The adhesion potential ϕ(z). Pairs of points on the vesicle and wall repel (attract) one another if they are separated by less (more) than δ. Right: The net adhesive force between the vesicle and the substrate. The green line on the magnified region has length δ = 0:1. The region of the vesicle surface that is less than δ from the substrate is repelled, indicated by the vectors pointing away from the substrate. The remaining surface is attracted, indicated by the vectors pointing towards the substrate.

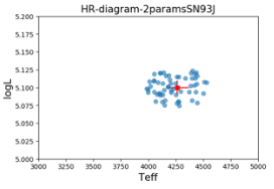
Figure 2, Below: Minimum distance between the vesicle and the substrate as a function of time. As the Hamaker constant increases, the vesicle remains close to the wall for a longer period of time after passing through the constriction. The images of the vesicle configuration correspond to the simulation with Hamaker constant H = 0:1. The dashed line is at δ = 0:1.



Antigoni Georgiadou
Ph.D. in Computational Science
Advisor: *Tomasz Plewa*

Global Optimization in Stellar Evolution Applications

Computer modeling is extensively used to probe structure and evolution of stars and planets. These computations allow astrophysicists to connect the star formation process to astronomical observations of stellar objects. Because stellar evolution is a highly complex process, matching the initial conditions to the observed objects requires substantial effort and expert knowledge. Typically probing parameter space of this problem is done using a trial-and-error approach, which is inefficient, incomplete and prone to bias. Our aim is to identify the key model parameters of the progenitor stars of SN1987A and SN1993J, their most probable values, and their observationally-constrained ranges in an automated way. I formulate a suitable constrained optimization problem, and combine select global optimization methods with the stellar evolution code, MESA. In particular I coupled the Controlled Random Search method (CRS) and the MIDACO solver to stellar evolution modules for core-collapse supernova parameter fitting. We demonstrate operation of the new stellar evolution optimization package in application to the evolution of massive stars.



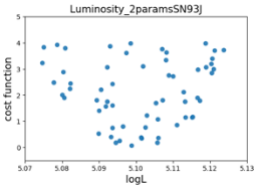


Figure 1, Far Left: Hertzsprung- Russel Diagram of the final cloud of the optimization run, encapsulating the observable constraints of the SN1993J.

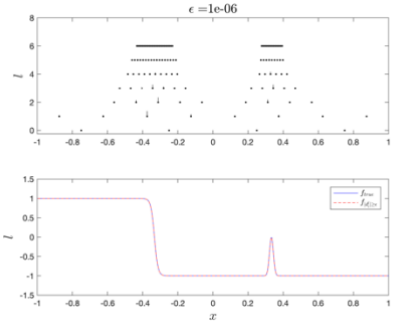
Figure 2, Left: Final cloud of the cost function values against the luminosity of the star for the SN1993J.

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

On the Efficiency of a Multiresolution Block-Adaptive Strategy for Multiphysics Problems

We present a generalization of Harten's multiresolution scheme for use with block-structured adaptive mesh refinement (AMR) codes. The method addresses a shortcoming of most existing AMR codes, which is the over-resolution introduced by block-based refinement. For problems with highly localized features and many spatial scales present this overhead can be costly. A cell-average multiresolution representation is used to indicate regions of smoothness where computationally expensive, high-order flux calculations may be avoided and instead approximated using the multiresolution basis. The flux interpolation error introduced by this approximation is related to a prescribed threshold. We find that the error does not significantly alter the rate of convergence of the solution for smooth problems. The procedure can be easily incorporated into existing parallel codes. The forward transform is applied per AMR block, obviating any need for communication of multiresolution information between blocks. The efficiency of the proposed scheme is illustrated using several one-dimensional problems.

Figure 1: An average-interpolating wavelet decomposition of a test function is shown. Non-smooth features of the function are captured by the wavelet transform. Top: The hierarchy of active cells on each refinement level is shown, with magnitudes of each respective detail coefficient plotted. Bottom: The test function f(x) is approximated accurately using the truncated wavelet basis.



Albert Iglesias
M.S. in Computational Science
Advisor: *Chen Huang*

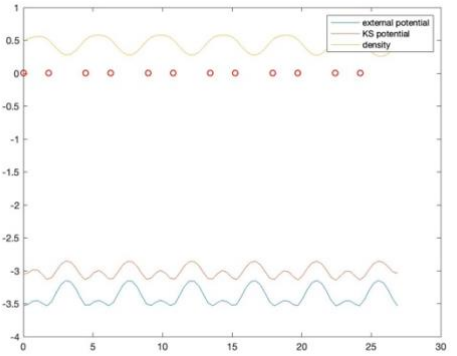
Embedded Cluster Density Approximation for Periodic Systems

Density functional theory (DFT) has been widely used to investigate the ground-state electronic structures of many-body systems, in particular atoms, molecules and the condensed phases (solid state in our case). In this project we plan to perform multiscale DFT simulations to understand and analyze Peierls' transition in 1D hydrogen chain.

We begin by taking a non-periodic Kohn-Sham DFT evaluated system with 12 hydrogen atoms in a 1D space and evaluate the density, KS potential and external potential throughout each step of the DFT calculation. We have completed turning density into a periodic function using fourth order Central Finite difference and reconstructing the Hamiltonian matrix. This allowed us to get corresponding eigenvalues, in our case these are the orbital energies of electrons in the system.

Now we are working on making the external potential periodic. To do this we are using Yukawa's potential in order to make sure the total Coulomb potential of the ions will not diverge. Once this is done we will implement k-point sampling within the first Brillouin zone in order to complete our algorithm and create a 1D fully periodic system. The goal is to demonstrate Peierls' theorem, which states that one-dimensional equally spaced chain with one electron per ion is unstable. We will test our multiscale DFT program to see how small of a cluster we can use to reasonably reproduce our benchmark.

.



Marzieh Khodaei-Gheshlah
Ph.D. in Computational Science

Advisor: *Peter Beerli*

Investigation of the Correlation Structure of Genetic Markers Using Measures of Distance among Derived Geneaologies

The Bayesian inference software MIGRATE calculates population genetic parameters from multiple genetic markers. These markers are currently treated as independent markers. In practice, there are reasons why this is incorrect; for example, linkage disequilibrium due to selection. A group of sampled individuals forms a genealogy. Genealogies are different when using different markers due to recombination. MIGRATE uses these genealogies as nuisance parameters and marginalizes over all possible genealogies to infer the probability distribution of the population parameters. This process discards information. Calculating a tree-distance measure among the genealogies allows investigating the correlation structure of the genetic markers because markers that are nearby will have the same genealogy, therefore small distances among their genealogies. We work on incorporating a tree distance measure (currently we use the Kendall-Colijn method) and evaluate the correlation structure, which eventually will lead to better estimates of the population parameters. This poster is a progress report.

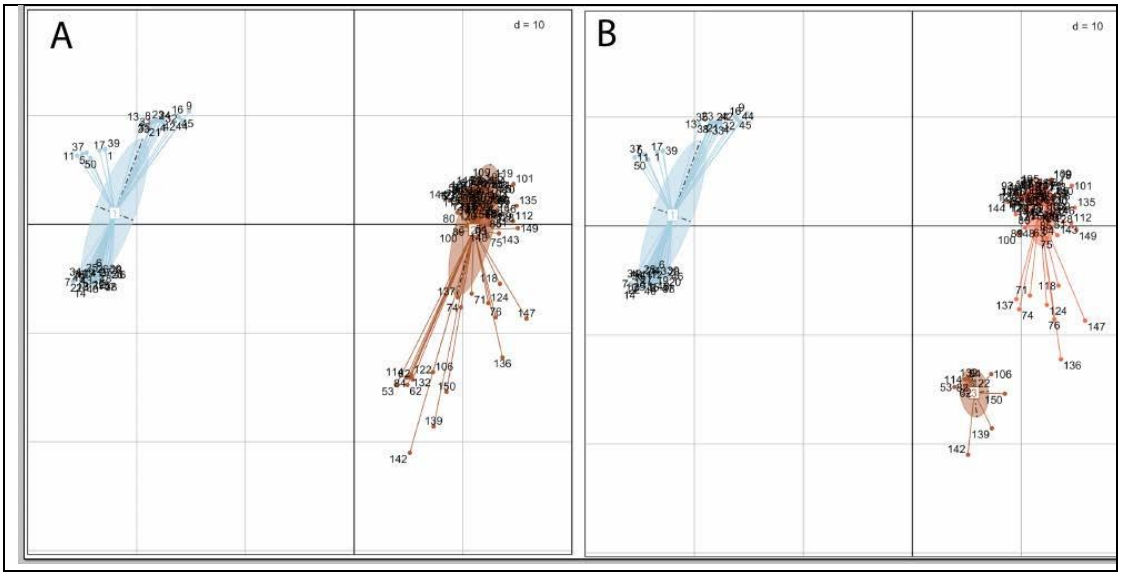


Figure 1: A comparison of 150 genealogies (trees) using the R-software *treescape* developed by Kendall and Colijn (2015) that uses the tree distance we want to build into MIGRATE.

Young Hwan Kim
Ph.D. in Computational Science

Advisor: *Xiaoqiang Wang*

Phase Field Modeling of a Simple Structure Vesicle

An average human body is estimated to contain 37.2 trillion cells, all with their own specialized functions. However, all cells share one physical feature: outer plasma membrane, which protects cells from their surroundings, allows cells to regulate the movement of ions and organic molecules in and out of cells, and allows cell communications/signaling. Modeling and simulating the dynamics of cell membranes can shed new insights to the structures and functionalities of cells as a whole.
In this poster, modeling and simulation of a simple structure vesicle using phase field coupled with bending elastic energy is demonstrated. Phase field model was chosen because it 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. Furthermore, it has been successfully applied to study vesicles.

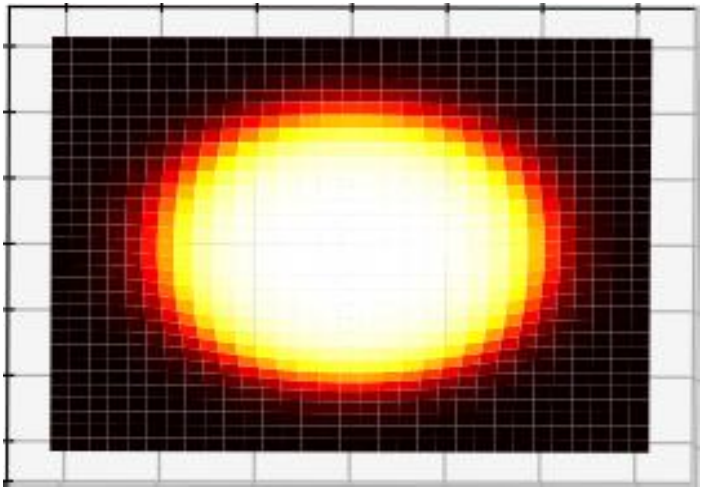


Figure 1: A cross section of a simple vesicle's conformation after bending elastic energy minimization.

Eitan Lees
Ph.D. in Computational Science
Advisors: *Sachin Shanbhag & Max Gunzburger*

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. We have developed a simplified multiphase (solid, liquid, and porous interphase) nonlocal model for pitting corrosion in aluminum alloys. We propose and solve nonlocal reaction diffusion equations that (i) explicitly account for multiple dissolved species, and (ii) enforce a nonlocal electroneutrality constraint. Results are qualitatively consistent with experimental observations. Detailed chemical kinetics are currently being explored to better capture the corrosion process but are accompanied by increased computational costs. Methods to reduce model complexity while maintaining an accurate description of the system are of interest to many areas of research as well as industrial chemical engineers. We are working on methods for the reduction of complex chemical reaction networks involving sensitivity analysis to identify important model parameters. Future work includes calibrating our model to experimental data as well as integrating the corrosion process with mechanical failure models.

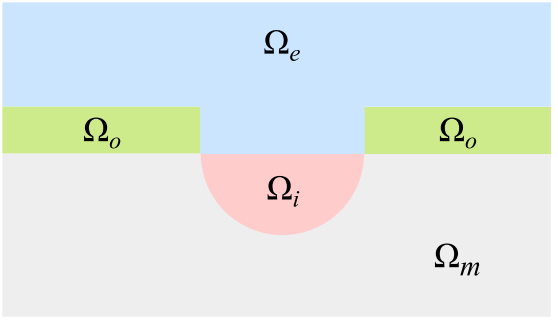


Figure 1: Schematic of the simplified pitting corrosion model.

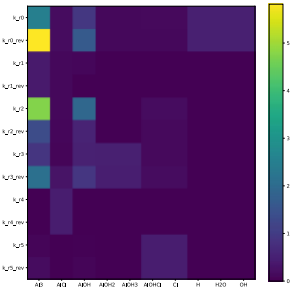
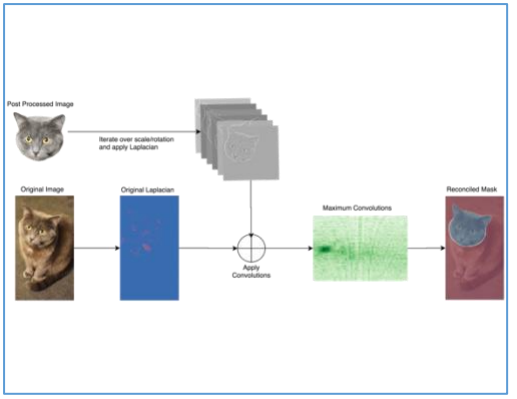


Figure 2: Sensitivity Analysis results of our chemical reaction network.

Kevin Mueller & Nathan Crock
Ph.D.s in Computational Science
Advisor: *Gordon Erlebacher*

Automatic Portrait Segmentation of Cats and Dogs

With the ever-growing amount of image data available on the internet there is an increasing need to develop algorithms that can automatically detect regions of interest (ROI) in images. Currently, almost all approaches for solving this task rely on difficult to create supervised datasets that require labels for individual pixels. Here we present an end-to-end framework for automatically cropping images of cats, dogs, and humans by utilizing post manipulated images prepared by graphic designers. We accomplish this task by first applying an image reconciliation technique that creates pixel masks from the post manipulated images. We then implement a machine learning technique that learns the masks from the original images by utilizing Deep Gaze II, a deep learning model for creating saliency maps trained on human eye-tracking data, as pre-training for finding likely ROIs.



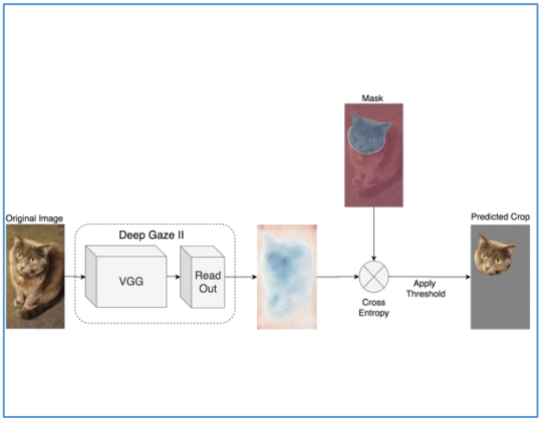


Figure 1, Left: The Laplacian is applied to the original image and all scales and rotations of the post-processed image. The kernels are then convolved with the Laplacian of the original image and the maximum convolution value is calculated. The optimal rotation/scale is found by taking the maximum of the maximum convolutions. Finally, the maximum position of the convolution for the optimal rotation/scale is taken to be the final reconciled mask location.

Figure 2, Right: The original image is first fed into the Deep Gaze II network and returns a log probability of the original image from its prior training. The mask is then used to further guide the network to the area of interest of the correctly cropped head. Finally, a threshold is applied to create a binary mask for the original image.

Jonathon Nosowitz
M.S. in Computational Science
Advisor: *Tomasz Plewa*

Detonability of White Dwarf Plasma: Survival and Growth of Detonations

The conditions created during the violent phase of the merging binary white dwarf system offer an opportunity to study and characterize the conditions for explosive burning to occur. Turbulence, a result of these merging white dwarfs, is produced in the process of collision of plasma streams when the secondary component overflows its Roche lobe and its material is accreted by the primary. Our past work provided us with a method to identify potential detonation kernels in such turbulent flow. In the ignition mechanism studies, turbulent flow is created inside a small, 3D computational box that represents a part of the boundary layer created in the merger process. Turbulence is driven using a spectral method with various contributions of compressibility and the amount of the injected energy. The initial models demonstrated the presence of plasma with short ignition times occupying regions sufficiently large formally to satisfy the ignition criteria. However, detonations were not demonstrated, and their sustained character remains unproven. We will build upon the results of the above initial study and investigate the evolution of prospective ignition kernels including the effects of nuclear burning for compressively driven turbulence. We will also consider existence of sustained detonation at various average plasma densities.

Type Ia supernovae

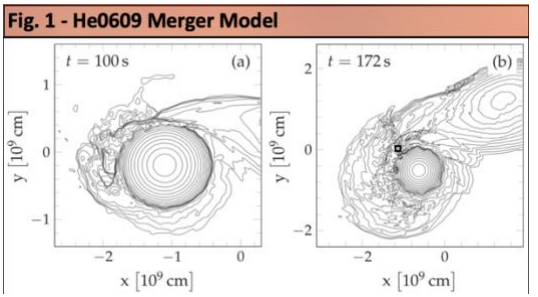


Fig. 1 - He0609 Merger Model

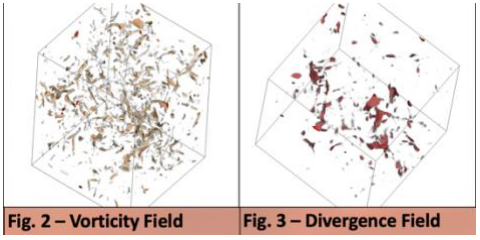


Fig. 2 – Vorticity Field

Fig. 3 – Divergence Field

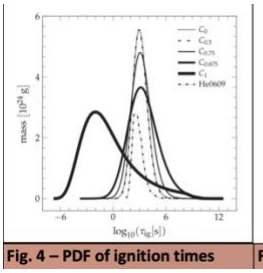


Fig. 4 – PDF of ignition times

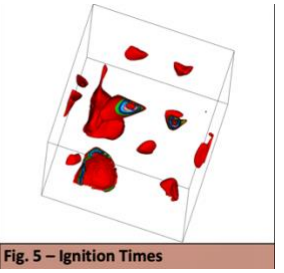
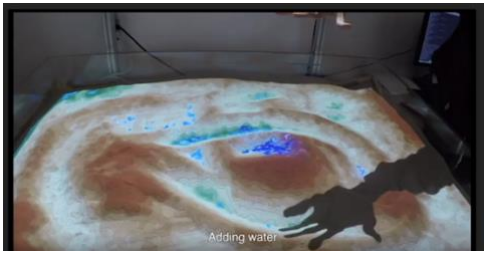


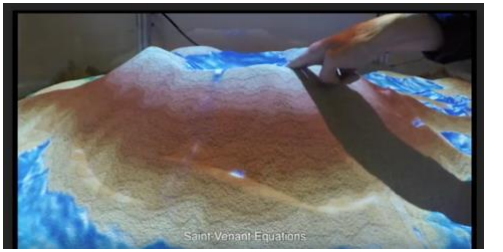
Fig. 5 – Ignition Times

David Robinson
Ph.D. in Computational Science
Advisor: *Bryan Quaife*

CAndbox: Modeling Fire Spread on Real-Time Interactive Terrain

Modeling the spread of forest fires via cellular automata is an attractive approach for its efficient computational implementation, and the surprising realism of deceptively simple algorithms in simple geometries. However, applying this method to a real wildland fire environment with complex fuel, topography, and wind requires substantial parameterization. An established approach is to use pixel-by-pixel rates of spread derived from the Rothermel model, but these models do not integrate over the contributions of multiple ignited pixels and only allow a discrete number of states. We present a new method which allows for semi-discrete states by summing over the contribution of all ignited pixels on a combustible pixel within a given neighborhood. Folding together the effects of terrain and wind speed, we employ an extended Rothermel model to form an "effective spread ellipse" defined by the rate of spread an ignited pixel contributes in any direction. Because of the computational efficiency of this method, we are able to apply it to a "sandbox", an augmented reality projection system. Initially designed to demonstrate fluid flow on an evolvable surface, this system allows the fire model to facilitate real-time adjustments to topography and burning regions. Additionally, the system handles complex initial fire shapes while remaining intuitive to the user. The outcome is a tool that is ideal for exploring the effect of topography on the spread rates of fire in a real-time environment.





Figures: Screen shots from video of Robinson's research. The video is at https://drive.google.com/file/d/1hfNTWaVvR51AQhw_dLUXsqEDqwHq3qia/view.

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

Modeling Motif Occurrences with Hawkes Process

DNA sequences contain many short fragments that are special, for example they are remnants of retroviruses or mark start locations for replication and gene transcription. These fragments (Motifs) are usually short sequences. Researchers are interested in the distribution pattern of these motifs. Usually the occurrences are modeled using a Poisson process; recently Gusto and Schbath ["FADO: a statistical method to detect favored or avoided distances between occurrences of motifs using the Hawkes' model", Statistical Applications in Genetics and Molecular Biology, 4(1), 2005] have shown that the Hawkes process is a better match to discuss these patterns. We are working on a method that improves on the current implementation of the Hawkes process by exchanging the B-spline kernel that was used previously with a better fitting exponential kernel.
We show that our model will produce more accurate results.

Rachel Scarboro
Undergraduate in Computational Science
Advisor: *Dennis Slice*

Morphometric Analysis of Gait in Women Wearing High Heels and Flats

The choice of shoes may drastically impact aspects of the wearer's gait. Morris et al (2013), for instance, analyzed how walking in high-heeled shoes affects the perception of femininity, youth, and attractiveness compared to walking in flat shoes in women. This project analyzes the data from Morris' paper using Principal Component Analysis (PCA) of gait landmark data processed by Generalize Procrustes Analysis. It discusses future means of performing a comprehensive morphometric analysis to statistically characterize and compare gait between women wearing high-heeled and flat shoes.

.

Figure 1: X-ray of female foot wearing high-heeled shoes

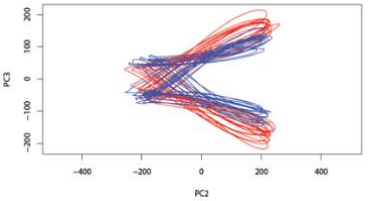
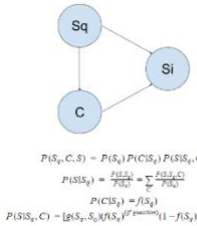


Figure 2: Gait trajectories of women wearing flat (red) and high-heeled (blue) shoes.

Eric Sharkey, Imanpreet Dhillon & Nathan Crock
Undergraduate in Computer Engineering, NewSci Labs & Ph.D. in Computational Science
Advisor: *Gordon Erlebacher*

Bag-of-Words Sentence Similarity with Synonym Augmentation

Given a target sentence, we would like to determine the most similar in a set of candidate sentences. State of the art sentence similarity methods are often very slow and require large amounts of data to perform well. Our applications require a sentence similarity algorithm that can front load most of the computational cost and yield accurate and fast results in large scale production environments. To achieve this, we model each sentence using a lightweight bag of words document frequency matrix. To improve generalization and similarity accuracy, we use WordNet to count synonyms to the words in each sentence in our frequency matrix. The candidate sentences have a label indicating whether they are interrogative or not, however, it is unknown for the target sentence. So, we train a neural network to classify the target sentence as declarative or interrogative. The output of this classifier is combined with the sentence embedding to determine the most similar sentence. This is all represented in a probabilistic graphical model.

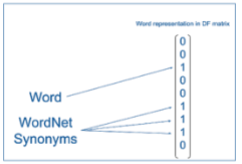


$$P(S_q, C, S_i) = P(S_q)\,P(C|S_q)\,P(S_i|S_q, C)$$

$$P(S_i|S_q) = \frac{P(S_i, S_q)}{P(S_q)} = \frac{\sum_C P(S_i, S_q, C)}{P(S_q)}$$

$$P(C|S_q) = f(S_q)$$

$$P(S_i|S_q, C) \propto \exp(S_q, S_i)\,f(S_q)^{(\text{IF question})}(1 - f(S_q))^{(\text{IF statement})}$$



.

Kyle Shaw

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

Inference of Population Parameters in Un-sampled Regions

There are many methods and programs available in population genetics for determining various population parameters in and between populations. Parameters such as population size, mutation rate, migration rate, etc. are useful numbers for field biologists and policy makers to have. However, to date, there are no methods or programs available for interpolating such parameters to populations that have not been sampled. I investigate the assumptions and the equations that are needed to infer population parameters of un-sampled populations which lie between sampled populations.

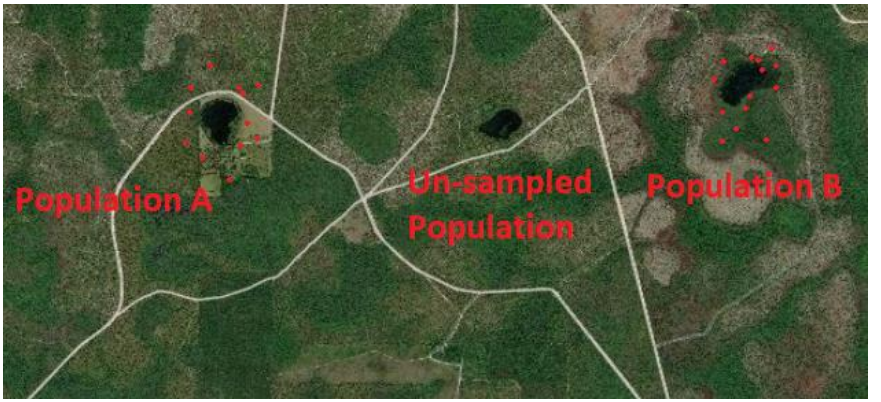
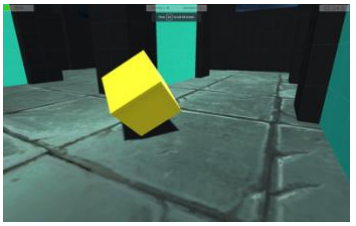


Figure 1: Field biologists often gather data from several discrete locations. In this case frogs were sampled from the areas around two ponds as shown. A third population of frogs was not sampled from the middle pond as shown. Using interpolation and migration models several interesting population parameters can be inferred about the un-sampled population.

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

Player Modeling Using Bayesian Classification to Procedurally Adapt Games

The objective of this research is to see if engagement can be increased by tailoring a game to the user's preferences. The belief is by personalizing the game, the user will have an increased play time and enjoyment. The game Preempting Path has been developed to test this hypothesis. It is an adaptive game maze game. Users will play levels and be asked how much they enjoyed the level. Along with this response, the user's gameplay metrics will be used to create the next maze in real-time. The first five levels of the game will be used to develop a baseline model for the player. At which point, the levels will begin to be generated procedurally using the model to adjust the generation parameters. Qualitative data analysis is then conducted with the observation data, generation parameters, and level ratings to investigate the engagement with the level and classify subgroups. This research will be used to further the adaptation in eRebuild. The tools used to link the Bayesian network to the procedural content generation will allow for the quick development of levels suited to the mathematical ability of each individual student. Other areas in which this may be applicable are advertising and marketing, as well as team or opponent selection and balancing.



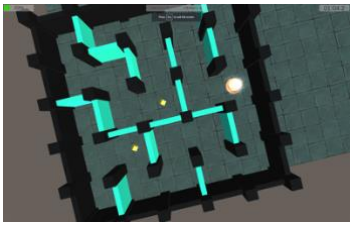


Figure 1. Bottom, Left: First-person view of the game Preempting Path. You can see a collectable power up as well as the statistics used as observations in the Bayes' net.
Figure 2. Top, Right: A top-down view allows the player to quickly route to the end or search out other paths and treasures.
Figure 3. Bottom, Right: The summary screen allows the user to rate the experience. Self-reporting allows users to influence the generation of future levels.

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

The Formation and Dynamics of Social Networks

We propose a method of inferring the opinions of individual nodes in a social network, given their initial opinions on a topic and a vector of observable characteristics. We extend the DeGroot-Friedkin model of opinion exchange and social power dynamics with a preference-based formation process recently developed by economists to explain clustering and segregation in social networks based on observables. This formation process results in a distribution over possible network topologies from the family of Exponential Random Graph Models or ERGMs.

We estimate this distribution using a Markov Chain Monte Carlo Maximum Likelihood Estimation (MCMCMLE) and discuss the results.

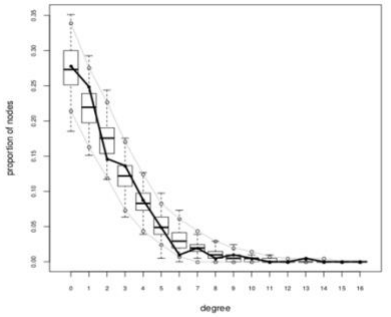


Figure 1: Goodness of fit of the MCMCMLE degree distribution

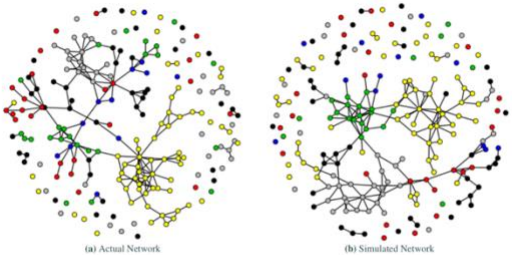


Figure 2: Results of a single simulation. The colors of the nodes represent the students' grade levels. (2a) is the Mesa High friendship network data from [5]. (2b) is drawn from the ERGM distribution as specified and using the the Mesa data with coefficients given by Table 1 and over the same node set.

Stephen Townsend
Ph.D. in Computational Science
Advisor: *Dennis Slice*

A Comparison of Classification Methods for Animal Vocalizations

Animal vocalizations are a very important point of study for scientists in the fields of ecology and biology. They have been used widely to study how animals communicate. They have also been used to assess how animal species and even populations can be acoustically differentiated. In this work, I will compare methods for acoustically differentiating individual populations of the Cope's Gray Treefrog Hyla chrysoscelis. This work is built upon much of my previous work and focuses on the use of the individual pulses of an animal vocalization as

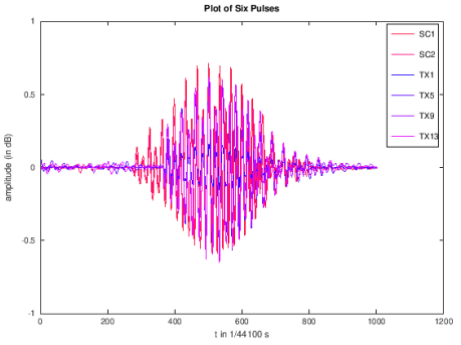
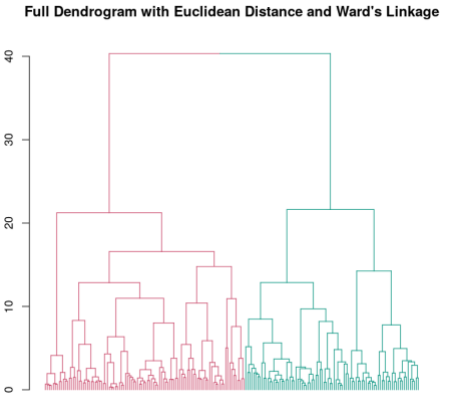


Figure 1: Six pulses from different calls after alignment..



the basis for the data that is to be analyzed. In this work, I compare standard Hierarchical Clustering Methods including Ward's Method, Single Linkage and Complete Linkage Clustering, K-Means Clustering, Fuzzy C-Means Clustering, OPTICS, DBSCAN and DIANA Clustering with modern Machine Learning Methods including Random Forests and Voting Classifiers. The objective is to determine which methods show the best overall performance on data where the correct dialect population grouping is known from classic work by Garhardt in the 1970s.

Figure 2, Top: Ward's Method Clustering with 187 Frog Calls. The Green Group represents the Western Dialect group and the Red Group represents the Eastern Dialect Group.

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

Visualization Methods in Email Topic Analysis

Topic models are powerful unsupervised generative statistical methods. They are widely used to explore huge document collections. However, visualizing the trained topic model is also a big problem since the latent parameters are embedded in high dimensional space. In this work, we implement standard Latent Dirichlet Allocation (LDA) model on a massive email corpus to calculate the topic proportions and the topics, defined as sets of words and their probabilities. We present several ways of visualizing the result with respect to time and author. Our method can intuitively show the trend of topics as well as the topic distribution for different authors, which could help to detect fraud among a large number of people over a long time period.

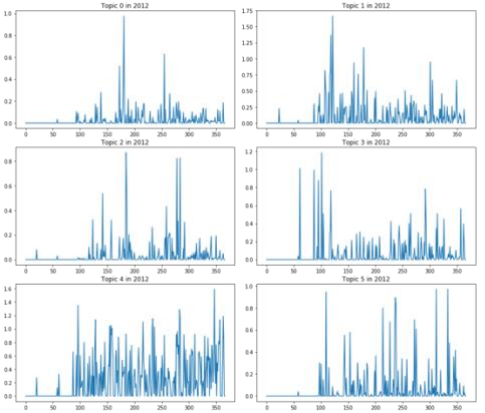


Figure 1. Topics trends of a year. Y axis shows the topic score of the first 6 topics out of 15. Higher topic score indicates more emails related with corresponding topics are sent in that day.

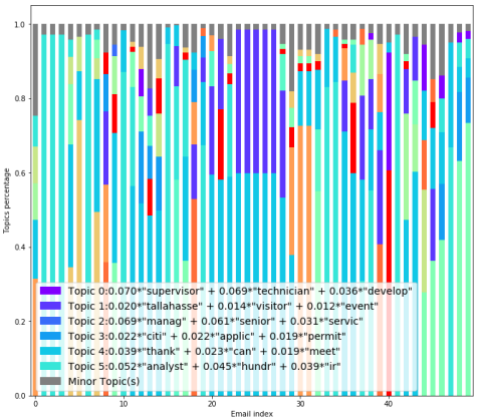


Figure 2. Topic inspection on 50 emails from a specific person. Different colors represent different topics. The figure above shows that topic 3 is likely to appear in conversation with this person.

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

The Limits of Gene and Species Tree Resolution

Many phylogenetic programs rely on correct gene trees as input. As sequencing becomes cheaper access to whole genomes is becoming more common. There are no scientific standards for selecting subsections of the genome to obtain the most accurate and well resolved gene trees, such as what region size to choose for the gene tree, whether or not recombination must be considered, and if the location of the region matters. This poster provides preliminary results from the vertebrate wide genome alignment, and presents current progress in setting up a pipeline to answer these questions.