

# 13<sup>th</sup> Annual ACSESS



Friday, April 15, 2016  
12:00 pm - 5:00 pm  
SDSU's Parma Payne Goodall  
Alumni Center

For Applied Computational Sciences and Engineering & Computational Science Curriculum Development



SAN DIEGO STATE  
UNIVERSITY

12:00 pm

## Registration / Buffet Lunch

**Ezra Bejar, Industry Projects Coordinator** Moderator  
Computational Science Research Center  
San Diego State University

1:00 pm

**Stanley Maloy, Dean** Welcome / Opening Remarks  
College of Sciences  
San Diego State University

1:10 pm

**Jose E. Castillo, Director** Computational Science at SDSU  
Computational Science Research Center and ACSESS Program  
San Diego State University

1:30 pm

**Satnam Alag** Keynote Speaker  
Vice President, Software Engineering  
Enterprise Informatics  
Illumina, Inc.  
Introduction by Victor Pereyra  
Member, CSRC Advisory Board  
“Enabling the Vision of Precision  
Medicine via Big Data”

2:30 pm

## Blitz Oral Presentations

3:00 pm

## Poster Presentations / Reception

4:00 pm

**Paul Paolini, Satchi Venkataraman,  
and Andrew Cooksy** Poster Presentation Awards  
and  
Closing Remarks  
Associate Directors  
Computational Science Research Center  
San Diego State University

## Sponsors of ACSESS 2016



INTELLISIS CORPORATION



## *W*elcome to the 13<sup>th</sup> Annual ACSESS Event

We cordially welcome our local tech and biotech industry executives, board members and CSRC faculty, and students to our Applied Computational Science and Engineering Student Support (ACSESS) Annual Event.

ACSESS was created thirteen years ago to provide graduate students an opportunity to work on real world problems. Also, with the vision to create stronger links between academic computational scientists, industry professionals and technology officers. This year's program is a reflection of that idea to continue the support and aspiration to foster CSRC educational and research missions; instilling graduate students interest in industry problems and pursuing careers in technology and biotechnology sectors.

We appreciate the great interest of our authorities, faculty and student participation in our event, and our special thanks go to our CSRC board members for their continuous advice, as well as our company sponsor's for the financial commitment to our teaching and research missions.

## *M*ission Statement

The mission of the Computational Science Research Center (CSRC), located at San Diego State University, is to promote development and advancement of the interdisciplinary subject of computational science. This is accomplished by fostering research, developing educational programs, and promoting industrial interaction, outreach, and partnership activities.

The Computational Science Research Center provides an excellent environment for scientific research at SDSU. The center facilitates the interaction between applied mathematics, computer science, and the sciences by providing the necessary infrastructure for productive research efforts.

Real world applications are the focus of the projects undertaken by the faculty and students of the center. Such projects provide a significant educational opportunity for our students to hone their industrially relevant computational skills.

## *E*xecutive Board

### **Program Director:**

Jose E. Castillo

### **Associate Directors:**

Andrew Cooksy • Paul Paolini • Satchi Venkataraman

### **Industry Projects Coordinator:**

Ezra Bejar

### **Computer Support Coordinator:**

James Otto

## *C*SRC Advisory Board

**Chair:** John M. Newsam  
CEO  
Tioga Research, Inc.

Gary Fogel  
CEO  
Natural Selection, Inc.

Victor Pereyra  
Consulting Professor  
Stanford University

Bob Parker  
Deputy CTO  
SPAWAR

Mark E. Pflieger  
Senior Scientist  
Cortech Solutions, Inc.

Antonio Redondo  
Division Leader  
Los Alamos National Laboratory

Bill Bartling  
Deputy  
California Department of Conservation

Scott Kahn  
Vice President Commercial  
Enterprise Informatics  
Illumina, Inc.

Anton Zajac  
President  
Eset Foundation

## Featured Posters

---

### *From DNA to FBA: How to Build Your Own Genome-Scale Metabolic Model*

**By: Daniel A. Cuevas, Taylor G. O'Connell, Chris S. Henry, and Robert A. Edwards**

*Advisor: Robert A. Edwards*

Microbiological studies are increasingly relying on in silico methods to perform exploration and rapid analysis of genomic data, and functional genomics studies are supplemented by the new perspectives that genome-scale metabolic models offer. A mathematical model consisting of a microbe's entire metabolic map can be rapidly determined from whole-genome sequencing and annotating the genomic material encoded in its DNA. Flux-balance analysis (FBA), a linear programming technique that uses metabolic models to predict the phenotypic responses imposed by environmental elements and factors, is the leading method to simulate and manipulate cellular growth in silico. However, the process of creating an accurate model to use in FBA consists of a series of steps involving a multitude of connections between bioinformatics databases, enzyme resources, and metabolic pathways. We present the methodology and procedure to obtain a metabolic model using PyFBA, an extensible Python-based open-source software package aimed to provide a platform where functional annotations are used to build metabolic models (<http://linsalrob.github.io/PyFBA>). Backed by the Model SEED biochemistry database, PyFBA contains methods to reconstruct a microbe's metabolic map, run FBA upon different media conditions, and gap-fill its metabolism. The extensibility of PyFBA facilitates novel techniques in creating accurate genome-scale metabolic models.

To date, fifty-eight diverse bacteria have been sequenced and annotated to generate metabolic models using PyFBA. Additionally, these bacteria have been grown on up to 192 different minimal media compositions and analyzed using the PMAalyzer pipeline (growth curves available from <http://edwards.sdsu.edu/dbbp>). These experiments supplement FBA predictions by creating a feedback system that optimizes and reconciles each metabolic model. By linking a microbe's genotype to its phenotype, a new methodology combining phenomics with genomic annotations enables predictions of gene-protein-function connections not yet explored by traditional bioinformatics approaches.

***This research is supported by National Science Foundation funds (MCB-1330800, DUE-1323809, DUE-1259951) and the Computational Science Research Center at San Diego State University.***

### *Exploration of Surrogate Models for Inverse Identification of Delamination Damage in Composites using Electrical Resistance Change*

**By: Paulina Diaz-Montiel and Satchi Venkataraman**

*Advisor: Satchi Venkataraman*

Carbon Fiber Reinforced Polymer (CFRP) materials are used in aerospace structures due to its superior mechanical properties and reduced weight. Non-destructive evaluation (NDE) techniques to detect and measure internal inter-ply delamination and intra-ply matrix cracking damage are needed for such materials. The electrical resistance change (ERC) provides a NDE technique that uses the inherent changes in conductive properties of the composite to characterize damage.

This research investigates efficient numerical modeling techniques for inverse identification of delamination damage location and size in composite laminates using ERC based NDE. Identification of damage requires performing an optimization that minimizes the difference between the model predicted and the measured change in resistance at specified electrode locations. Using numerical finite element model of the laminate directly in the optimization is computationally expensive and requires the development of an accurate surrogate model. The objective is to achieve a high precision detection model for delamination cracks in a composite laminate and to quantify a measurement that could characterize the damage state of the material.

This paper presents a comparison of different surrogate modeling approaches and model reduction techniques for the ERC based damagedetection. Surrogatemodelsevaluatedincludepolynomial response surfaces and kriging. The use of model order reduction through singular value decomposition (SVD) and the accuracy of using SVD components to fit surrogate is also explored.

***This research is self supported.***

## Featured Posters

### *Optimization of Sensing Configuration Selection for Electrical Resistance Change Based Damage Detection in Composites Using an Effective Independence Measure*

**By: Luis Waldo Escalona Galvis, Paulina Diaz-Montiel, and Satchi Venkataraman**

*Advisor: Satchi Venkataraman*

Laminated carbon fiber reinforced polymer (CFRP) composite materials have multiple applications in aerospace structures due to their advantage in terms of mechanical properties and reduced weight. Evaluating the integrity of these structures requires using non-destructive evaluation (NDE) techniques to detect and measure the presence of composite interlaminar delamination and intralaminar matrix cracking damage. Amongst the known NDE techniques electrical resistance change (ERC) uses the inherent changes in conductive properties of the composite to characterize internal damage. Research on the ERC based technique has been focused mainly on thin cross-ply laminates with simple linear or circular sensing configurations. This work investigates a method of optimum selection of electrode configurations for delamination detection in thick cross-ply laminates using ERC. The ERC based inverse identification of damage requires numerical optimization of the measured response with a model predicted response. In this regard, finite element analysis allows for the computation of the electrical potential field in a CFRP composite laminate under damage after electric current is applied on specific locations defined in the model. These locations correspond to electrode locations in a real case and therefore optimizing the number of sensing combinations and locations reduces the number of required measurements and thus the hardware requirements and computational effort needed for inverse identification. This research explores the use of sensor configuration ranking based on the computation of effective independence (EI) values originally proposed for sensor location optimization in experimental vibration modal analysis. The EI is used for selecting the minimum set of resistance measurements among the ones resulting from all valid and possible electrode pair combinations. To effectively apply the EI approach on the ERC problem this research uses singular value decomposition SVD to obtain a spectral representation of the resistance measurements in the laminate. The effectiveness of EI measure in eliminating redundant electrode pairs is demonstrated by performing inverse identification of damage using the full set of resistance measurements and the reduced set of measurements. The investigation shows that the EI measure is promising for optimally selecting the electrode pairs needed for resistance measurements in ERC based damage detection.

***This research is supported by the Computational Science Research Center at San Diego State University.***

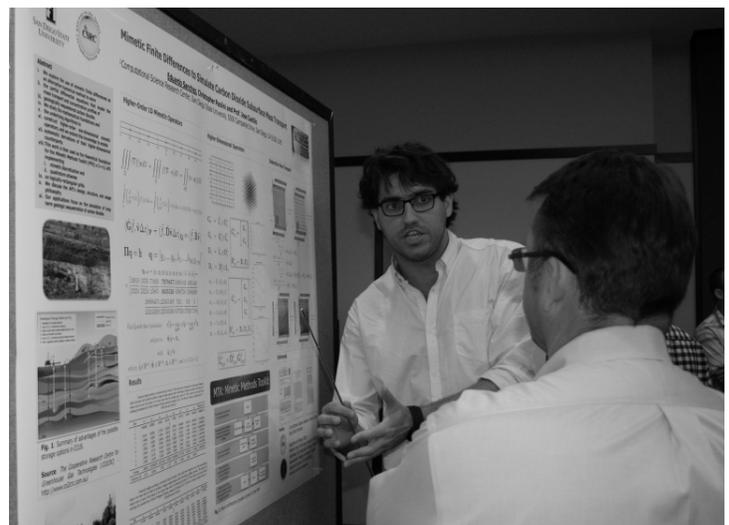
### *A Blended Product for Predicting Snow Water Equivalent in the Rocky Mountain Region*

**By: Colette Smirniotis**

*Advisor: Barbara Bailey*

Accurately modeling and predicting snow cover presents a scientific challenge, in part because it is difficult to measure this quantity directly. Thus, indirect measurements are also used to infer snow cover. Given the choice of different indirect measurements and also different algorithms to retrieve snow cover from these data, there is no single, standard gridded snow data set that can be used for climate model evaluation. The goal of this project was to use a Bayesian statistical model to create a blended product of snow water equivalent from multiple data sets and in the process also quantify the uncertainty of this across different locations. To combine the data sets, a new version of the R package LatticeKrig was implemented that can handle the “change of support” problem common in spatial statistics and inherent in this project. LatticeKrig models the spatial process with a multiresolution basis function representation in a computationally efficient manner. Monthly snow predictions from the blended product were calculated along with the standard errors of the predictors. Since the version of the LatticeKrig package was under development, several features were tested and improved as a result of this project.

***This research is supported by National Center for Atmospheric Research, the Computational Science Research Center at San Diego State University and the Department of Mathematics and Statistics at San Diego State University.***



### *Investigation of the Influence of the Subgrid-Scale Stress on Non-Intrusive Spatial Pressure Measurement Using Direct Numerical Simulation Turbulence Database*

**By: Seth Siddle-Mitchell**

Advisor: Xiaofeng Liu

The instantaneous pressure distribution in a turbulent flow field can be measured non-intrusively by integrating the material acceleration measured using time-resolved particle image velocimetry (PIV). However, due to the finite spatial resolution of the PIV measurement, the pressure reconstructed from PIV is actually subjected to the effect of spatial filtering. Consequently, the reconstructed pressure is effectively imbedded with the contribution of the sub-grid scale (SGS) stress, which is a term appearing in the filtered Navier-Stokes equation. To quantify the effect of the SGS stress on pressure measurements, we use box filtering to filter three dimensional velocity components in a time-varying isotropic turbulence flow field available to public from the John Hopkins University Turbulence Database (JHTDB). We then calculate the material acceleration and apply the pressure reconstruction procedures introduced in Liu and Katz (2006 and 2013) to obtain the pressure distribution. The reconstructed pressure, with and without the incorporation of the sub-grid scale stress, is compared with the pressure filtered directly from the direct numerical simulation (DNS) database, thus gauging the error due to the measurement resolution effect. A  $17 \times 17 \times 17$  box filter as a first level filtering is used to bring the DNS data resolution down to the PIV resolution level. Then a  $3 \times 3 \times 3$  box filter is further applied to the already filtered velocity data so as to establish the modeled SGS stress based on similarity SGS stress modeling. Ensemble average of 612 instantaneous realizations from the DNS database indicates that inclusion of the modeled SGS stress term in the pressure reconstruction reduces the error from 6.99% when no SGS stress term was present, to 4.78% of the r.m.s. fluctuation of the filtered pressure, indicating an advantage in using the SGS modeling in compensating the measurement error. In addition to the isotropic turbulence, the whole procedure is currently being tested with a DNS turbulent channel flow from JHUDB, which constitutes a harsher and more realistic testing case. Preliminary results show that an increase in overall error of the pressure reconstruction in the channel flow. However, by introducing the modeled SGS stress, the error was reduced from 10.92% to 8.47% of the r.m.s. fluctuation of the filtered pressure, once again indicating an improvement using the SGS modeling for resolution error compensation.

*This research is supported by UGP funds at San Diego State University and grant funds.*

### *Decoupling Task Related Cortical Activities from Multi-Channel ECoG Signals*

**By: Siddhi Tavildar**

Advisor: Ashkan Ashrafi

Electrocorticography (ECoG) is a moderately invasive method that provides high quality cortical signals with better spatio-temporal resolution compared to electroencephalography (EEG). It is the preferred choice for volitional hand movement animation after spinal cord injuries. To this end, devising an algorithm that can effectively translate task-related cortical activities of upper extremities is of utmost importance. The algorithm is divided in two major tasks: feature extraction and classification. In this paper we propose two effective feature extraction methods for classification of the task related cortical activities. First method is based on frequency transforms and uses discrete cosine transform (DCT) of the ECoG signals. The second method is based on parametric model and uses Auto-Regressive Model of the ECoG signals for feature generation. Classification results and comparison of these two methods is discussed.

*This research is supported by National Science Foundation funds (Center For Sensorimotor Neural Engineering) and the Computational Science Research Center at San Diego State University.*

### *Prescott Tests in Binary Data Under a Three Period Crossover Design*

**By: Lixia Zhu and Kung-Jong Lui**

Advisor: Kung-Jong Lui

Prescott tests are explored in binary data to compare three treatments in a three-period crossover trial. Monte Carlo simulation is employed to evaluate the performance of these test procedures in a variety of situations. Prescott tests appear to have higher power compared to other methods utilizing only those patients responding differently to treatments. The data taken from a three period crossover trial are used to illustrate the application of Prescott test on binary data under a three period crossover design.

*This research is supported by the Computational Science Research Center at San Diego State University and the Department of Mathematics and Statistics at San Diego State University.*

## Featured Posters

---

### *Designing Optical Metamaterials with Hyperbolic Dispersion Based on Al:ZnO/ZnO Nano-Layered Structure Using Atomic Layer Deposition*

**By: Priscilla Kelly, Lyuba Kuznetsova, and Mingzhao Liu**  
Advisor: Lyuba Kuznetsova

Development of the metamaterials with hyperbolic dispersion (HMM) has been a subject of recent research due to their unique optical property, namely that the dispersion of the dielectric function for HMMs exhibits a topological transition in the isofrequency surface from an ellipsoid to a hyperboloid. This important property of metamaterial with hyperbolic dispersion will offer the advantages of using this metal-dielectric HMM for applications in blue/UV LEDs and structural colors. In this work, our goal was to design and fabricate a nano-layered Al:ZnO/ZnO metamaterial using the Atomic Layer Deposition (ALD) technique. Experimental dielectric functions for Al:ZnO/ZnO structures are obtained by an ellipsometry technique in visible and near-infrared spectral range, and theoretical modeling is done using effective medium approximation. A method for analysis of spectroscopic ellipsometry data is demonstrated to extract the optical permittivity for our highly anisotropic nano-layered metamaterial. The results of the ellipsometry analysis show that Al:ZnO/ZnO structures with 1:9 ALD cycle ratio exhibit hyperbolic dispersion transition change near 1.5  $\mu\text{m}$  wavelength which pushes more towards the optical region.

*This research is supported by National Science Foundation funds (Graduate STEM) and the Computational Science Research Center at San Diego State University.*

### *Mimetic Operators and Solute Transport Modeling*

**By: Johnny Corbino**  
Advisor: Christopher Paolini

The most important equation when modeling fluid flow and species transport is the transport equation. As its name suggests, it describes how chemical species are transported through a medium as a consequence of the fluid flow and chemical interactions. In this work we use mimetic differential operators to solve transport problems.

*This research is supported by the Computational Science Research Center at San Diego State University.*

### *Comparison of Random Forest Algorithms*

**By: Peter Calhoun and Juanjuan Fan**  
Advisor: Juanjuan Fan

The world revolves around predictions, those ranging from whether it is going to rain today to how a patient will respond to a treatment. Random forests is one of the most accurate machine learning methods to make predictions and analyze datasets. They often outperform classical methods and alternative data mining algorithms and require few statistical assumptions. Advances in computing power has increased the popularity of random forests which has led to various extensions of the standard algorithm; those include randomly selecting candidate cutpoints, replacing the indicator function with a smooth sigmoid function, and adding a quality control.

The goal is to compare four random forest algorithms in terms of prediction accuracy. Several datasets were analyzed and the method that makes the best predictions was determined. The standard random forest algorithm performed well for datasets with a continuous response, but was outperformed by other proposed algorithms for classification problems. In addition, the new random forest variations also sped up computational time, a desirable property when dealing with large amounts of data.

*This research is supported by the Computational Science Research Center at San Diego State University and the Department of Mathematics and Statistics at San Diego State University.*



### *Experimental Characterization of Progressive Failure Damage in Countersunk Bolted Joints Tested Under Bearing Loads*

**By: Alexandru Popescu and Satchi Venkataraman**  
Advisor: Satchi Venkataraman

Composite materials are widely used in the aerospace industry due to their superior stiffness and weight-savings compared to metals. The failure of these composite laminates at fastener sites has been a long recognized as a problem and investigated by many researchers. The focus of this work is on countersunk bolted joints in composite laminates. Countersunk fasteners are preferred in aerospace to satisfy aerodynamic smoothness of airframe or wing outer surface. The ASTM D5961 procedure C is recommended to use in testing countersunk composite joints. In this paper we perform bearing tests of laminated composites, using a modified ASTM D5961 fixture designed for testing countersunk bolted joints. The progressive failure behavior and the damage resulting from the modified test fixture are quantified and compared to laminates tested using the standard ASTM D5961 Procedure A and Procedure C. Our recent work [1] finite element analysis investigation of bolted joint tests using the above mentioned three fixtures has shown that in countersunk fasteners the bearing load is transferred primarily through the non-countersunk hole region, the bending deformation of the fastener or pin leads to significant localization of stresses, and the bending of the pin and the test fixture lead to significant out of plane compressive stresses. The goal of this poster is to experimentally characterize and begin to present results of computer modeling of the progressive damage during bearing under the three test conditions and that can be used for validation of progressive failure modeling efforts. Preliminary tests on a AS4/3501-6 material 36 ply [-45<sub>2</sub>/+45<sub>2</sub>/90<sub>2</sub>/0<sub>2</sub>/±45<sub>3</sub>/90/±45/0]<sub>2</sub> laminate shows significant differences between the standard and modified tests. In standard bearing tests with straight pins the final bearing failure is characterized by a sudden load drop. Countersunk specimens do not exhibit a large load drop at the bearing failure, and continue progressively damaging. The tested specimens are inspected non-destructively (C-scan) and destructively by microscopy imaging of the specimens after sectioning to characterize and quantify the damage.

References: 1. Palwankar, Manasi, Popescu, Alexandru, and Satchi Venkataraman, "Finite Element Analysis and Testing of Countersunk Composite Bolted Joints in Double Lap Shear," Proceedings of the 57th Structures, Structural Dynamics, and Materials Conference. San Diego, CA, Jan 4-8.

***This research is supported by Office of Naval Research funds (Contract No. N00014-14-1-0297).***

### *Who is There and What are They Doing? Using FOCUS and SUPER-FOCUS for an Agile Taxonomic and Functional Analysis of Metagenomic Big Data*

**By: Genivaldo Gueiros Z. Silva, Bas E. Dutilh, and Robert A. Edwards**  
Advisor: Robert A. Edwards

Microbes are more abundant than any other cellular organism, and it is important to understand which organisms are present and what they are doing. In many environments a majority of the microbial community members cannot be cultured, and metagenomics is a powerful tool to understand the diversity of microbial communities by using only their DNA.

Analyzing the taxonomic and functional profile present in a microbial community sequences is one of the goals in metagenomics for its valuable applications in biological research. Available tools do not scale well with increasing data volumes, which is important because both the number produced by sequencing platforms keep increasing.

We developed two tools to address this problem: FOCUS, Find Organisms by Composition USage, an ultra fast model which uses k-mer abundance and non-negative least squares to profile any metagenome dataset in seconds and SUPER-FOCUS, SUBsystems Profile by databasE Reduction using FOCUS, an agile homology-based approach using a reduced SEED database to report the subsystems present in metagenomic samples and profile their abundances.

The results shows that our approaches accurately predict the taxa and subsystems present in microbial communities, and FOCUS and SUPER-FOCUS are respectively over 30,000 and 1,000 times faster than other tools.

***This research is supported by CNS-1305112, MCB-1330800, and DUE-132809 funds and the Computational Science Research Center at San Diego State University.***

## Featured Posters

---

### *Genetic and Phenotypic Analysis of Gammaproteobacteria*

**By: Blaire E. Robinson**

Advisor: Robert A. Edwards

Connecting genotypes, the genetic information in an organism, to phenotypes, the functions that an organism performs, is one of the important challenges of biology. In this study, I will identify strains of Gammaproteobacteria that are genetically similar, but display varying phenotypic profiles, and investigate the genetic source of the functional variation and gene expression that accounts for their phenotypic differences.

Bacteria were isolated from both the water column and surface/tissue of kelp species *Macrocystis pyrifera* from three locations off the coast of southern California (Catalina Island, La Jolla, and Point Loma). Isolated strains were sequenced using various Next Generation Sequencing platforms, their sequences were assembled using SPAdes and MIRA, and assembled sequences were then annotated using Rapid Annotation using Subsystem Technology (RAST). Phylogenetic association was formulated using five highly conserved housekeeping and median-level conserved genes (*ftsZ*, *gyrB*, *mreB*, *recA*, *rpoA*) in order to increase sensitivity and identify genetically similar strains. Isolate phenotypes were tested using multi-phenotype assay plates, analyzing growth curves using the high-throughput analysis pipeline PMAnalyzer to create phenotypic profiles. Phenotypic profiles were then compared to phylogenetic analysis to identify genetically similar, but phenotypically varying isolates for further analysis. Two strains were chosen for this work, ED144 and ED252. Further genetic analysis of selected strains will be performed using various bioinformatics tools to analyze individual strain structure, metabolism, proteomics, and functional subsystems in order to identify the cause of phenotypic variation. Expected sources of phenotypic variation include, but are not limited to, loss of function mutations, gain of function mutations, modulation of cellular networks, transport protein activity, and prophage activity. Identification of the basis of phenotypic and genetic variation can lead to further understanding of their source within the environment, how variation affects cell survival, and whether variation is ecologically significant to kelp health.

***This research is supported by National Science Foundation funds (MCB-1330800 and DUE-1323809).***

### *Fractal Dimension Analysis of Coral Reef-Invasive Algae Interaction Boundaries by Richardson's Algorithm and the Box Counting Method*

**By: James Mullinix**

Advisor: Antoni Luque

The ocean's coral reef population is under duress, with many coral reefs facing imminent death. To address this concerning issue, the Viral Information Institute (VII) is seeking to answer questions regarding the effects of bacteria and bacteriophage (bacterial viruses) on the survivability of the coral reefs against invasive algal colonies. To that end, the following question was raised: is the geometry of the interaction boundary between coral reefs and invasive algal colonies the result of the coral employing space filling strategies to optimize boundary size? Moreover, does this space filling strategy invoke a fractal boundary? The implication is that if a fractal boundary is found, perhaps this strategy serves to increase mucosal coverage which would increase the concentration of bacteriophage as a strategy to fend off attacking bacteria from the invasive algal colonies. We implement multiple fractal dimension analysis tools in search of the answer to this question, and in the process, encounter interesting problems to solve, such as path optimization in the presence of noise and computational efficiency of algorithms. In the end, our findings suggest that the boundary does not have fractal geometry, and, additionally, there is no significant difference in the dimensional analysis of regions in which the coral is winning the fight, losing the fight, or the fight is neutral.

***This research is supported by National Science Foundation funds (Graduate STEM) and the Computational Science Research Center at San Diego State University.***



### *Diagnostic Prediction of Autism Using RS-Functional Connectivity: A Conditional Random Forest Dimensional Reduction Approach*

**By: Afrooz Jahedi, Art Nasamran, Brian Faires, and Colleen P. Chen**

Advisor: *Ralph-Axel Müller and Juanjuan Fan*

Despite consensus on the neurological nature of autism spectrum disorders (ASD), brain biomarkers remain unknown and diagnosis continues relying on behavioral criteria. Random forest (RF) has successfully been applied to reveal complex patterns of biomarkers due to its capability in coping with small sample size and large number of predictor variables. However, RF is biased toward correlated predictor variables, which can affect variable importance measures and interpretability of the forests. In the presence of highly correlated functional connectivity MRI data, we used Conditional Random Forests (CRF) to minimize bias in the variable selection. Dealing with high dimensional parameter space, a multi-step CRF dimensional reduction algorithm were employed to maintain important information about ROI functional connectivities while significantly reducing computation time.

We used resting state (rs)-fMRI data from 462 low-motion participants (222 ASD, 240 TD) from the Autism Brain Imaging Data Exchange (ABIDE), matched on age and 0.25 mm head motion. We chose 313 ROIs from Gordon et al. (2014), 13 subcortical and 15 cerebellum ROIs from Harvard-Oxford atlas. For each ROI pair, a feature was defined as functional connectivity (time series correlation; 58311 total features). To decrease computational time, we discretized the data for each feature into 2% bins, replacing each bin by its median. A multi-step CRF dimensional reduction algorithm was used to split features into several sets of smaller chunks. Then CRF ran on each chunk in parallel and 40%-50% of top informative features were extracted from each set. These steps repeated until reducing dimension of the parameter space to 100 features.

After 8 steps of CRF dimensional reduction, the parameter space reduced from 58311 to 100 features with an OOB error of 21%. Further, the proportion of intra-network-connectivity significantly increased, while conversely the proportion of inter-network-connectivity decreased. Prominent networks were default mode, somatosensory motor-hand, ventral attention, cingulo-opercular task, cerebellar, and visual.

Using small portion of features (0.0017) can improve diagnostic prediction of autism. The prominence of sensorimotor, visual, attention and default mode ROIs is in line with previous findings. The pivotal role of ROIs in cerebellum adds novel evidence to the growing literature in ASD.

*This research is supported by National Institutes of Health funds (R01 MH081023) and the Computational Science Research Center at San Diego State University.*

### *Predicting Radiographic Osteoarthritis Grades from Cartilage Magnetic Resonance Images Using Machine Learning: Data from the Osteoarthritis Initiative*

**By: Uyen Hoang**

Advisor: *Usha Sinha*

Spin-spin relaxation time (T2) images have been shown to be sensitive to changes arising from osteoarthritis progression in articular cartilage (severity indexed by Kellgren-Lawrence (KL) grades 0 (normal) to 4 (advanced)). However, T2 mapping has low resolution and is limited by low signal-to-noise ratio, whereas three-dimensional water-excitation dual-echo steady-state (DESS-WE) morphological imaging produces high signal-to-noise ratio, high cartilage-to-fluid contrast, and near-isotropic slices. Despite advances in imaging techniques, staging osteoarthritis remains a challenge as there are only subtle changes between nearby KL grades. The purpose of this work was to evaluate the ability of machine learning to discriminate between KL grades 2 and 3 using features extracted from 3T DESS cartilage MR images. A computer-aided pattern recognition image analysis and classification tool, WND-CHRM (weighted neighbor distance using compound hierarchy of algorithm representing morphology), was used to analyze medial weight-bearing sagittal slices of cartilage MRIs. In WND-CHRM, a large number of features (e.g., texture (Tamura, Haralick), intensity histogram) are extracted from the raw images and from transforms of images (Wavelet, Fourier, Chebyshev transforms). The KL grade for all the subjects (truth data) was established based on manually scored minimal KL grade 2 and moderate KL grade 3 using radiographic images. Image features were extracted and a set of image content descriptors and image transforms identified to be informative were used to predict the KL grade. A set of 20 subjects were chosen from each KL grade and for each subject, 10 innermost slices from the medial femoral weight-bearing condyles were manually selected. This means that there were a total of 200 images from each group that were used for classification. The image classifier was trained using 180 images of KL grade 2 and 180 images of KL grade 3, and was validated using 20 images from each of the two classes. After the classifier were trained and validated, a set of independent image samples from subjects who were not part of the training and validation sets were chosen from the cohort to be classified. The accuracy of classification for KL Grade 2 was 62.5% while that for KL Grade 3 was 80%. The most informative image features were low frequency features and those computed using image transforms, which are highly non-intuitive to human perception and difficult to sense by the unaided eye. Our results demonstrate the potential of using image features from morphological images in the classification of severity of osteoarthritis. Our future work will include features extracted from T2 maps to improve the accuracy of classification and to extend to other KL grades.

*This research is supported by National Science Foundation funds (Graduate STEM) and the Computational Science Research Center at San Diego State University.*

## Featured Posters

---

### *Empirical Evaluation and Enhancement of Storage System Request Scheduling*

**By: Wen Pan, Deng Zhou, and Vania Fang**

Advisor: Tao Xie

Request scheduling for an enterprise-class storage system is crucial for it could significantly impact system performance. It is also challenging because different workloads have distinct characteristics and various types of requests have diverse timeliness constraints. Besides, the interplay between front-end and back-end requests as well as the combined effect of a spectrum of scheduling factors on performance make request scheduling even more complicated. Unfortunately, while there are many studies on block-level disk I/O scheduling, little has been reported in the literature concerning enterprise storage system file-level request scheduling. Thus, we have little knowledge about how an enterprise file system scheduler works and, even more importantly, how various scheduling factors such as request dispatching and queue weight setting affect performance. Moreover, we are in lack of a good understanding on how to enhance request scheduling to adapt to the changing characteristics of workloads and hardware resources. To answer these questions, in this paper we first build a request scheduler prototype based on WAFL, a mainstream file system running on numerous enterprise storage systems world-wide. Next, we use the prototype to quantitatively measure the impact of various scheduling configurations on performance on a NetApp's enterprise-class storage system. Several observations have been made. For example, we observe that in order to improve performance the priority of write requests and non-preempted restarted requests should be boosted in some workloads. Inspired by these observations, we further propose two scheduling enhancement heuristics called SORD and QATS. Finally, we evaluate the two heuristics by conducting a wide range of experiments using workloads generated by SPC-1, SFS2008, and SFS2014 on both HDD-based and all-flash platforms. Experimental results show that the combination of the two can noticeably reduce average request latency under some workloads.

*This research is supported by NetApp funds and the Computational Science Research Center at San Diego State University.*

### *Simulation of Fluid Pressure and Fracturing in CO2 Sequestration*

**By: Jonathan Matthews, Christopher Paolini, and Jose Castillo**

Advisor: Christopher Paolini

CO2 sequestration in underground aquifers has the potential to reduce atmospheric CO2 emissions on the order of hundreds of gigatons. However, rock fractures, formed during injection, may release toxic species into the water table and release CO2 into the atmosphere. A model to compute the internal rock stresses induced by the injection of CO2-rich water is developed to simulate the initiation and propagation of rock fractures. A mixed finite element-finite volume model is used to calculate the fluid pressure, stresses and strains induced by injection of CO2 into a geologic saline aquifer-caprock formation.

When the effective stress exceeds the rock tensile stress at a given depth, simulated fractures are induced. The effect of the simulated fractures is upscaled to the reservoir scale by estimating an incremental permeability by means of Oda's permeability tensor.

The effect of fractures on the mineralization of CO2 is simulated by estimating the impact of fractures on the free surface area of the solid, rock phase. The free surface area impacts the dissolution and precipitation of carbonate minerals.

*This research is supported by US Department of Energy (DOE), National Energy Technology Laboratory (NETL), Web-based CO2 Subsurface Modeling, Geologic Sequestration Training and Research Funding Opportunity (DE-FOA-0000032), Simulation and Risk Assessment funds and the Computational Science Research Center at San Diego State University.*



## Featured Posters

### *Modeling Calcium Fluxes During Contraction of Neonatal Cardiac Cells*

**By: Esteban Vazquez-Hidalgo and Elesha Bartolotta**  
Advisor: Paul Paolini

Heart failure (HF) affects nearly 2 million people in the United States. A consistent finding in failing myocardium is altered calcium cycling. Calcium cycling is a complex process involving numerous proteins. We focus primarily on three proteins responsible in regulating intracellular calcium levels: RyR, SERCA, and NCX. Extracellular calcium enters the cardiocyte through sarcolemmal L-type channels, triggering systolic release of intracellular calcium stored in the sarcoplasmic reticulum (SR) via RyR channels. Diastolic calcium levels are restored by SERCA resequestering SR calcium, and NCX removing additional extracellular calcium. However, changes in normal production of these proteins alter calcium handling. Heterogeneity in SR calcium release and slowing of calcium cycling are likely to decrease transient amplitude, prolong calcium decline, and raise diastolic levels of calcium. Increased cytosolic calcium increases the likelihood of ventricular myocyte apoptosis. Reduced SERCA and increased NCX has been noted in HF animal models and in the failing human heart. Previous gene silencing experiments in our lab using neonatal rat ventricular myocytes show that inhibiting RyR increases NCX and decreases SERCA while decreasing transient amplitude and prolonging cytosolic calcium removal; inhibiting NCX increases SERCA production and increases transient amplitude. Our current goal is to use realistic mathematical models that accurately simulate cellular level physiological processes. We then perform parameter sensitivity analyses by systematically varying each parameter, observing the effects of the change, and generating a set of predictions that could be experimentally tested. The model coupled with experimental data could provide additional insight in cellular and subcellular calcium handling dynamics.

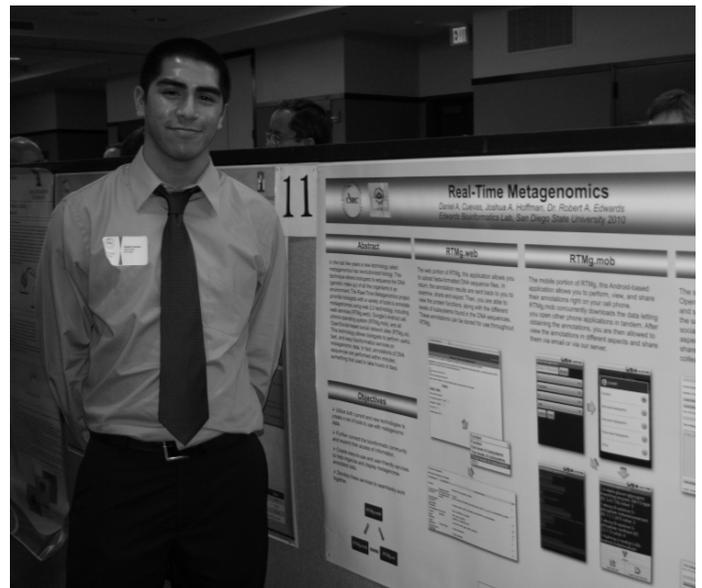
***This research is supported by National Science Foundation funds, California Metabolic funds, and the Computational Science Research Center at San Diego State University.***

### *Assessing Instructional Modalities: Individualized Treatment Effects for Personalized Learning*

**By: Joshua Beemer and Richard Levine**  
Advisor: Richard Levine

Large enrollment bottleneck courses and increased interest in distance learning options have lead to a rise in online offerings, particularly of quantitatively-oriented courses such as statistics. San Diego State University offers a Course Design Institute (CDI) to train and support instructors in the development of online courses, with a focus on student engagement, assessing student learning outcomes, and instructional technology. In this study, we compare an online offering of a second semester applied statistics course, developed through the CDI, with previous traditional offerings of the course by the same instructor. We apply an ensemble machine learning approach to compare student performance, identify factors indicative of student success, and analyze persistence into the next applied statistics course in the sequence under each instructional modality. We also introduce an “individualized treatment effect” (ITE) to evaluate individual student performance in the online and traditional offerings, though a given student enrolls in only one course offering (received only a single treatment). The ITE allows us to further characterize students that benefit from either the online or traditional offerings. We find that students performed at least as well on a common final exam and in course grade in the online class as in the face-to-face sections.

***This research is supported by ASIR, the Computational Science Research Center at San Diego State University and the Department of Mathematics and Statistics at San Diego State University.***



## Featured Posters

---

### *Which Parameters Are Important? A Sensitivity Analysis of Nuclear Interactions*

**By: Stephanie Lauber**

*Advisor: Calvin Johnson*

Many-body simulations allow us to determine behavior of nuclei, including those well beyond experimental limits, through the use of interaction parameters fit to data or derived from theory. These interaction parameters described hundreds of nuclear levels but a direct comparison between parameter sets offers very little insight into the underlying physics. The number of parameters for a given model space ranges from dozens to hundreds, but not all parameters contribute equally to resulting energy spectra. We aim to characterize these interactions based on the most important linear combinations (MILCOMs) of parameters.

Here I present the calculations for the MILCOMs of the Brown-Richter USDB interaction in the sd-shell, which consists of 66 parameters. By performing a perturbative sensitivity analysis using nuclei from Neon-20 through Argo-36, we were able to characterize the USDB interaction for comparison to other available sd-shell interactions. In addition, we hope to use the MILCOMs to quantify the parameters derived from ab initio calculations to determine what, if any, physics is missing from these more fundamental matrix elements.

*This research is supported by DE-FG02-03ER41272 funds and the Computational Science Research Center at San Diego State University.*

### *Bacteriophage Dynamics in the Presence of Mucus Layers*

**By: Kevin Joiner and Antoni Luque**

*Advisor: Antoni Luque*

Recent studies have shown mucus, which protects human organs, offers a rich source of nutrients that attract bacteria. It is unclear, however, how these bacteria are regulated to remain at a concentration that is beneficial for the human body. Phages are viruses that prey upon, infect and ultimately destroy bacteria. Recent findings in the Viral Information Institute show that phages interacting with mucus are more effective against bacteria. For the current project we numerically study the key factors which effect phage exploitation of bacterial colonies in mucal habitats. This model will allow us to investigate under which physiological conditions the phage is more efficient. It will also guide the engineering of phages that could be used in phage therapy to regulate commensal and pathogenic bacteria.

*This research is supported by Department of Defense funds (SMART) and the Computational Science Research Center at San Diego State University.*

### *Nearshore Internal Bores Formation Along Monterey Bay Using the General Curvilinear Coastal Ocean Model (GCCOM)*

**By: Manuel Valera, Mariangel García, Paul Choboter, and Ryan Walter**

*Advisor: Jose Castillo*

Internal waves are ubiquitous features in the coastal ocean. The breaking and eventual transformation of these waves into higher-frequency waves and bores, play a significant role in the generation of turbulence and energy mixing across scales in the coastal environment. The numerical modeling of these waves is specially difficult due to the non-hydrostatic dynamics, nonlinearity of phenomena and large range of scales involved. Additionally, the interaction of eddies with realistic bathymetry is crucial in the simulation of dominant coastal dynamics features that result from its interactions with major bathymetry abutments. Moving towards finer resolution models, preserves more structures on the terrain but increases the slope interaction, giving rise to nonlinearity in the form of turbulent flows. In this work a 3D case study for the Monterey Bay is presented and validated to demonstrate the capabilities of the GCCOM model for simulating non-hydrostatic flows such as stratified flows, internal bores formations and strongly nonlinear wave processes of generation and propagation, as well as the capabilities of handling abrupt bathymetry trough the implementation of a fully 3D curvilinear mesh.

*This research is supported by COAST funds and the Computational Science Research Center at San Diego State University.*

### *Implementation of Vehicular Camera Sensors in V2V Platform for Predicting Lane-changing Behavior and Corresponding Accidents*

**By: Chenlei Zhang and Xiaobai Liu**

*Advisor: Jose Castillo*

In order to access the traffic statuses of isolated vehicles as well as connected vehicles, we will form a dynamic system based on V2V platform which can conduct vision processing and extract the observed information of perceived vehicle such as their speed, location and driving behavior.

*This research is self supported.*

## Featured Posters

---

### *MSUFS: Towards Joint Management of Memory and File on Storage Class Memory*

**By: Deng Zhou, Wei Wang, and Wen Pan**  
Advisor: Tao Xie

Emerging storage class memory (SCM) devices exhibit a huge potential to replace DRAM as they have high-capacity, low cost, low energy-consumption, and low latencies comparable to DRAM. Developing a file system on an SCM device that is directly connected to main memory bus becomes a modern trend. In this case, a memory management module runs on a DRAM system, whereas a file system manages data on an SCM device. We argue that memory and file should be jointly managed so that mutual benefits can be achieved by eliminating redundant operations and data structures. As the first step towards this goal, we design, implement, and evaluate a new file system called MSUFS (Memory Storage United File System), which not only leverages existing memory management module to accomplish space management but also simplifies memory management itself. Experimental results show that MSUFS is superior to existing memory based file systems in several aspects.

*This research is supported by National Science Foundation funds (CNS-1320738) and the Computational Science Research Center at San Diego State University.*

### *Mode-Tracking in Surrogate-Based Inverse Identification of Rotor Blade Geometry Using Campbell Diagram*

**By: Vaibhav Yadav and Satchi Venkataraman**  
Advisor: Satchi Venkataraman

This work presents a new method for tracking vibration mode shapes of blades in integrally bladed rotors used in modern turbine engines. Traditional cross-correlation type measures for modal correlation such as the modal assurance criterion (MAC) and its derivatives have been used to correlate experimentally observed modes with numerically computed modes. These traditional methods work as long as the geometry is fixed and you only need to update stiffness, mass or damping parameters. The proposed modal tracking method was developed for tracking vibration mode-shapes when the geometry is changed. The need for tracking modes between design geometry variations arises in the inverse identification of turbine blade Gregory designs using vibration test data. The method, based on statistical approach, identifies mode crossings by an iterative surrogate model fitting and reclassifying procedure. The method was applied for demonstrated for tracking modes in a ten dimensional parametric design space of a jet engine turbine blade geometry.

*This research is supported by AFOSR funds (FA9550-12-C-0047).*

### *Convergence Properties of the Randomized Extended Gauss-Seidel and Kaczmarz Methods*

**By: Anna Ma, Deanna Needell, and Aaditya Ramdas**  
Advisor: Deanna Needell

The Kaczmarz and Gauss-Seidel methods both solve a linear system  $X\beta = y$  by iteratively refining the solution estimate. Recent interest in these methods has been sparked by a proof of Strohmer and Vershynin which shows the randomized Kaczmarz method converges linearly in expectation to the solution. Lewis and Leventhal then proved a similar result for the randomized Gauss-Seidel algorithm. However, the behavior of both methods depends heavily on whether the system is under or overdetermined, and whether it is consistent or not. Here we provide a unified theory of both methods, their variants for these different settings, and draw connections between both approaches. We compare convergence properties of both methods and their extended variants in all possible system settings. In doing so, we also provide an extended version of randomized Gauss-Seidel which converges linearly to the least norm solution in the underdetermined case. With this result, a complete and rigorous theory of both methods is furnished.

*This research is supported by Air Force Office of Scientific Research MURI funds (FA9550-10-1-0569) and the Computational Science Research Center at San Diego State University.*



## Featured Posters

---

### *Covariates Modulated False Discovery Rate in Hidden Markov Random Field Model*

**By: Rong W. Zablocki, Richard A. Levine, Chun C. Fan, and Wesley K. Thompson**

*Advisor: Richard Levine and Wesley Thompson*

Large-scale hypothesis testing is facing the challenge of properly controlling for type I error. Traditional multiple-comparison procedures tend to be underpowered. Procedures that control false discovery rate (fdr) are more powerful; yet most of proposed methods treat all hypothesis tests as exchangeable and independent, ignoring auxiliary covariates that may influence the distribution of the test statistics and ignore the prior knowledge of the network structure. The current work incorporates Hidden Markov Random Field (HMRF) into a Bayesian two-group mixture model for a covariate-modulated false discovery rate (cmFDR) on data that are not independent. The simulations are carried out to compare the performance in terms of sensitivity, specificity and false discovery proportion (FDP) between the new method and existing methods. The method can be applied to any field where network structure, test statistics and covariates are available. A real data application in gene-network is currently under construction.

*This research is supported by National Institutes of Health funds (R01GM104400) and the Computational Science Research Center at San Diego State University.*

### *Linear Stability Analysis of a 2D Double Shear Flow Configuration*

**By: Hareshram Natarajan**

*Advisor: Gustaaf Jacobs*

The Direct Numerical Simulation (DNS) analysis of a 2D incompressible double shear layer flow configuration was initially performed. It showed that the presence of the additional shear layer enhanced mixing properties of the shear layer. This provides a motivation to investigate the growth of linear instability modes of the double shear layer flow configuration. Linear Stability Analysis (LSA) using normal modes is performed using a double shear layer base flow. The growth of the spatial and temporal modes are obtained by numerically solving the linearized Navier-Stokes equations. Growth rates of the linear instability for different Reynolds numbers are calculated and compared with the growth rate of single shear layer flow. Additionally, a DNS of an inviscid shear layer is performed and compared with the results from LSA.

*This research is supported by AFOSR funds and the Computational Science Research Center at San Diego State University.*

### *Utilizing Interaction Trees to Inform Assessment of Pedagogical Innovations and Intervention Strategies*

**By: Kelly Spoon, Richard Levine, Juanjuan Fan**

*Advisor: Richard Levine*

In an educational setting, a common research goal is to assess the impact of an instructional method or intervention with regards to student success. Traditionally, methods used to quantify this impact are simplistic and do not account for differences due to selection bias in observational studies. This talk discusses a methodology that extends the traditional random forest framework to situations where the main interest is the treatment effect due to a particular teaching modality or intervention, specifically in observational studies. These results can be used to create overall treatment effects, individual treatment effects, marginal treatment effects for particular variables and subgroups of students who will have similar treatment effects.

Simulation studies showing the efficacy of the method and an application to a study at SDSU will be presented.

*This research is supported by the Computational Science Research Center at San Diego State University and the Department of Mathematics and Statistics at San Diego State University.*

### *A Frequency Channel Allocation Algorithm for Creating a Network of Wireless FDD Nodes*

**By: Moein Parsinia**

*Advisor: Sunil Kumar*

Many Frequency Division Duplex (FDD) systems operate on a fixed uplink and down link frequency bands where the downlink channel is used for the transmission of data from the node and an uplink is used for reception of data. Systems consisting of these nodes are designed for point to point (P2P) links in a single hop operation. In order to use these nodes in a multi-hop wireless network, these frequency channel assignment should be carried out carefully, such that the network is robust to the node mobility and meet the traffic/application demands.

In this research, we will present our channel assignment scheme so that the operation of these nodes can be extended to the multi-hop network topology. Our algorithm assumes multibeam antenna at each node and establishes a link connection such that each node can talk to multiple neighbors at a time. This design is robust to the node mobility and interference, and significantly enhances the throughput performance of the entire network.

*This research is supported by Air Force Research Lab funds and the Computational Science Research Center at San Diego State University.*

## Featured Posters

---

### *Non-Conservative Variational Approximation for Nonlinear Schrödinger Equations*

**By: Julia Rossi, Ricardo Carretero-González, Panayotis G. Kevrekidis, and Mariana Haragus**

*Advisor: Ricardo Carretero*

Recently, Galley [Phys. Rev. Lett. 110, 174301 (2013)] proposed an initial value problem formulation of Hamilton's principle applied to non-conservative systems. We explore this formulation for complex partial differential equations of the nonlinear Schrödinger (NLS) type, examining the dynamics of the coherent solitary wave structures of such models by means of a non-conservative variational approximation (NCVA). We showcase the relevance of the NCVA method by exploring test case examples within the NLS setting including combinations of linear and density dependent loss and gain. In particular, we present some examples applied to (i) exciton polariton condensates that intrinsically feature loss and a spatially dependent gain term and (ii) coherently-driven passive optical Kerr resonators that also include gain and loss terms. For (ii), we apply a NCVA and a Galerkin analysis to the spontaneous temporal symmetry instability observed by Xu and Coen [Opt. Lett. 39, 3492 (2014)] and find pitchfork and Hopf bifurcations.

*This research is supported by ARCS funds and the Computational Science Research Center at San Diego State University.*

### *Estimating Uncertainty Bounds for Blade Vibratory Stresses Under Multi-Mode Excitation*

**By: Amudha Varshini Kamaraj and Vaibhav Yadav**

*Advisor: Satchi Venkataraman*

The accurate prediction of peak vibration stresses of Integrally Bladed Rotors (IBRs) in their operating conditions is essential in engine survey testing. Stress estimation can be performed computationally by means of computer aided modeling and finite element analysis and blade vibration measurements. The advances in non-intrusive vibration measurement using blade tip timing data can help identify excited modes and the blade vibration in those modes and their relative phase. This poster presents the development of a software tool for automated analyses of the rotor geometry in ANSYS and post processing results to obtain the complex modal stresses. With the availability of modal stress real time updates of the maximum stress can be obtained using a superposition approach presented here. Estimating maximum stress requires a spatial and temporal search. The paper presents the effect uncertainty bounds of the maximum stress due to the uncertainties in the measured modal vibration amplitude and phase of the excited modes. These measurements errors are due to the limited number of sensors used in BTT measurements, small random variations in sensor sensitivity, sensor locations, and blade to blade variations.

*This research is supported by AFOSR funds (FA9550-12-C-0047).*



