Cyd Burrows-Schilling

UC San Diego

- *Professional:*
- Position: Research Facilitator serving all disciplines
- Campus Champion for UC San Diego
    - In support of XSEDE
- No previous supercomputing experience

- *Personal:*
- Research interests involve GIS, mapping, analysis and visualization of spatial data

Cyd Burrows-Schilling

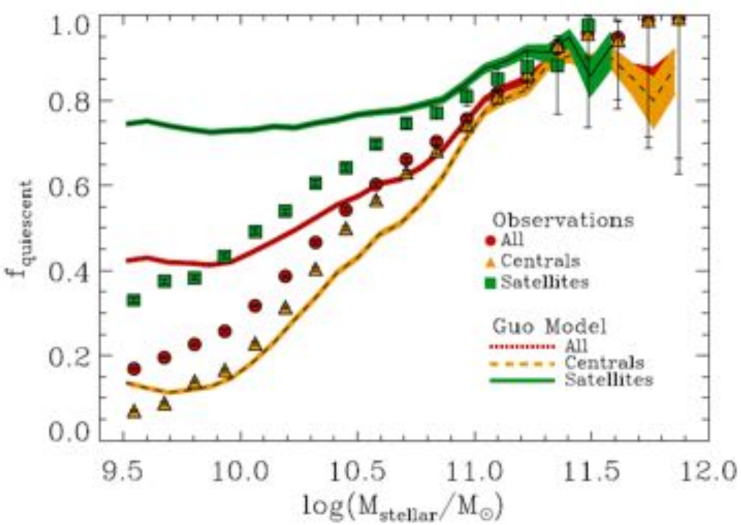
UC San Diego

- My goals in attending this workshop:
  - Improve my technical abilities in order to support new / basic users of the XSEDE platform
  - Increase my general knowledge about HPC
  - Learn about visualization techniques / concepts / tools / softwares
  - Discover and explore what I don't know
  - Learn from other attendees

Devontae Baxter

Physics Graduate at UC Irvine

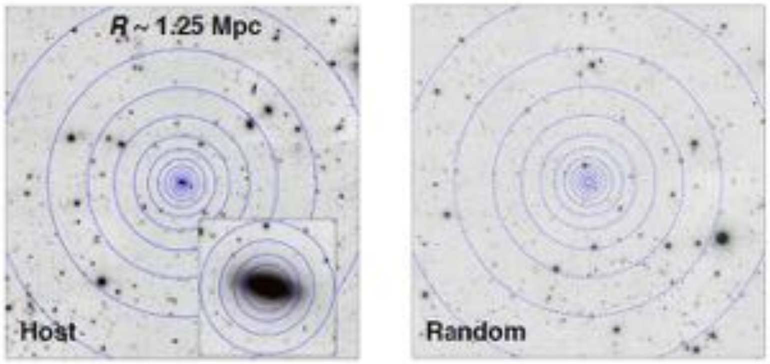
Research: I study the suppression of star-formation in satellite galaxies in the local universe and at higher redshifts (earlier times) using observational and simulated data.



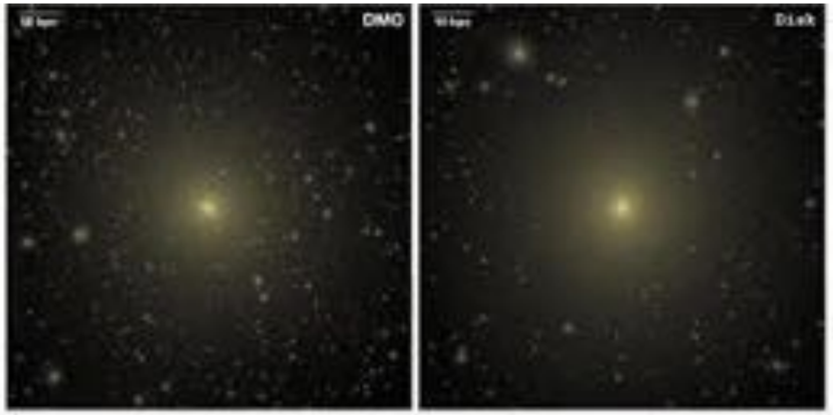
Guo et al.

Questions to answer: Which physical processes are involved in shutting down star-formation in satellite galaxies? Are the processes internal or external? What is the quiescent fraction for low-mass satellite galaxies?

Observational (imaging) data



Simulated data



Kelly et al.

Devontae Baxter

The University of California, Irvine

Some of my computational challenges include handling and analyzing large observational and simulated datasets.

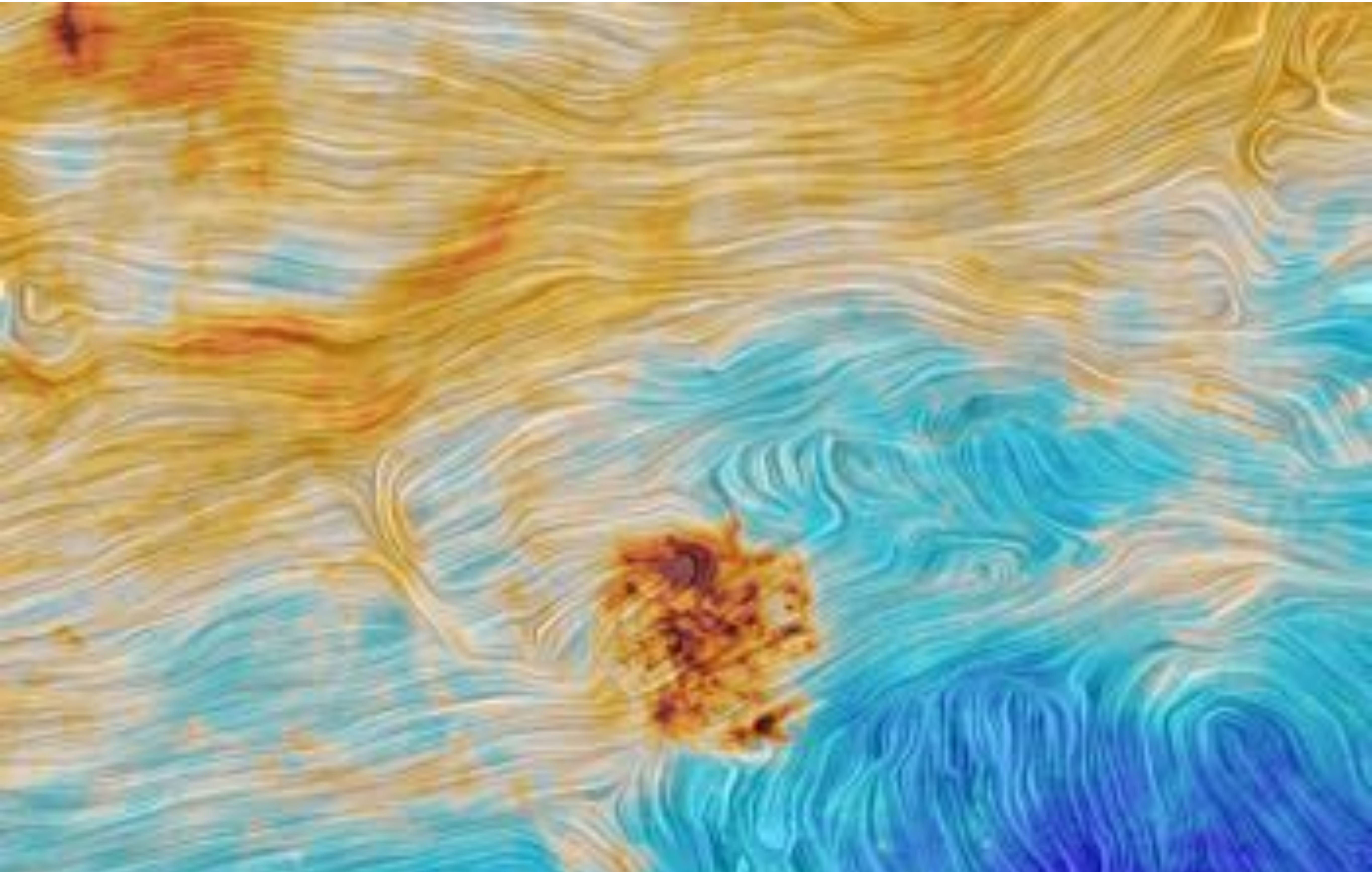
Classification: Classify photometrically observed galaxies as star-forming or quenched to ultimately explore the quiescent fraction of low-mass satellites galaxies.

Regression: Use observable quantities from simulations in conjunction with ML models obtain difficult to derive quantities relevant to the suppression of star-formation (e.g. infall-time).

SDSC goals: Learn more about machine learning, performance optimization and supercomputing with the goal of eventually running my own simulations.

Diego F Gonzalez-Casanova

UW-Madison



Diego F Gonzalez-Casanova

UW-Madison

3D

Then is important understand the “observables” (2D variables) and find relationships with the real variables 3D. For this part I use regressions, classifiers, PCA, …

In order to understand the interstellar medium, we need computer simulations. This done by solving the magneto hydrodynamic equations (with turbulence, gravity and chemistry). For this I need high-performance computing, on simulation that take about a week.

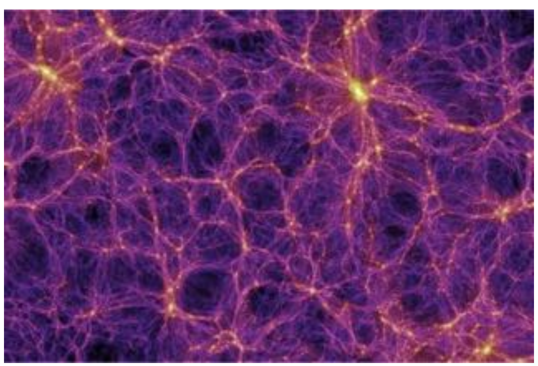
2D

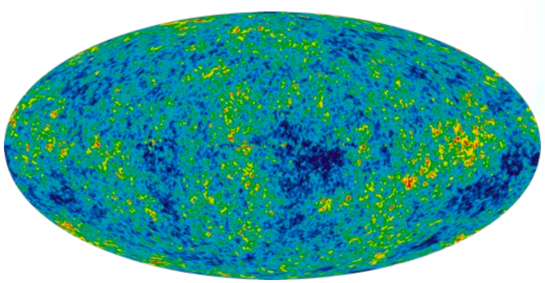
Ellie Kitanidis, PhD Candidate

UC Berkeley Physics Department

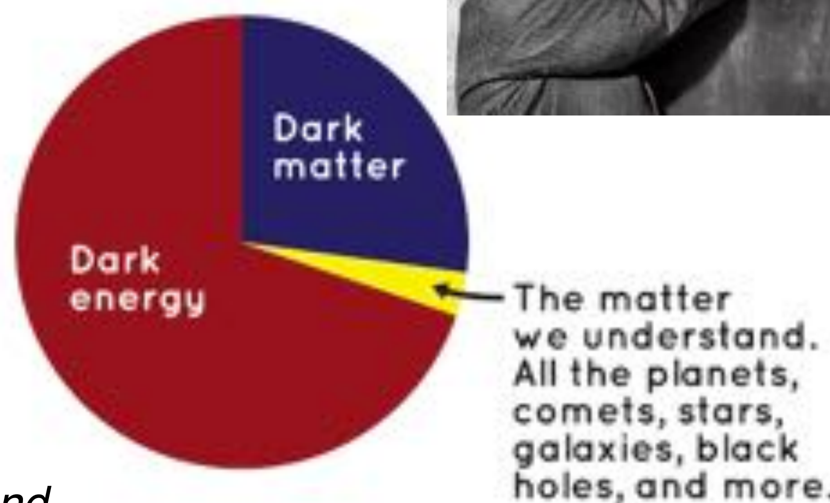
The statistical distributions of matter and light in the universe are rich with clues about dark energy, dark matter, general relativity, and other cosmic mysteries!

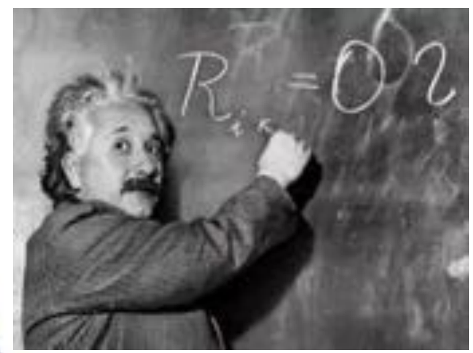
Large-scale structure

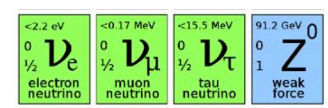


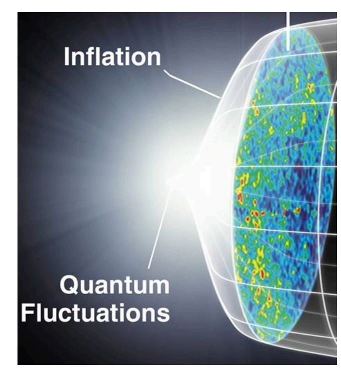


Cosmic microwave background









Ellie Kitanidis, PhD Candidate

UC Berkeley Physics Department

HPC / Big Data challenges:

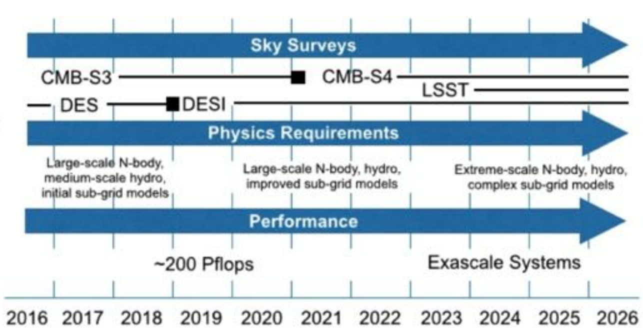
- Experiments measure hundreds of millions of galaxies; each involves many different types of data, sources, etc.
- A typical cosmological simulation requires minimum of tens of millions of CPU hours, petabytes of snapshot data
- Calculations of statistical quantities of interest (e.g. *n*-point correlation functions) scale as N^*n* or worse
- Not enough human power to keep up with data demands, need scalable AI to automate tasks

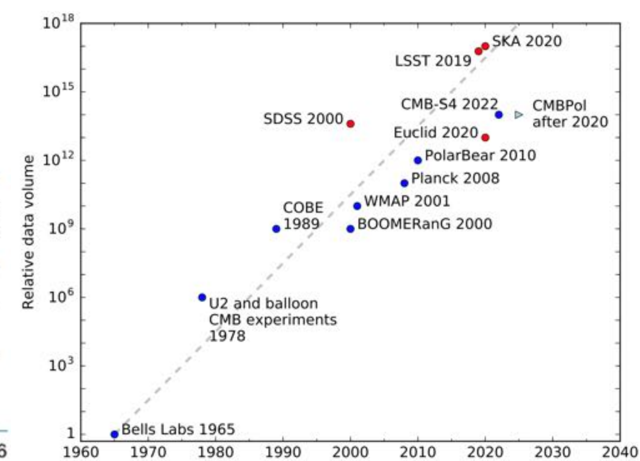
Argonne Team Breaks Record with 2.9 Petabytes Globus Data Transfer

July 9, 2019 by staff   Leave a Comment

globus

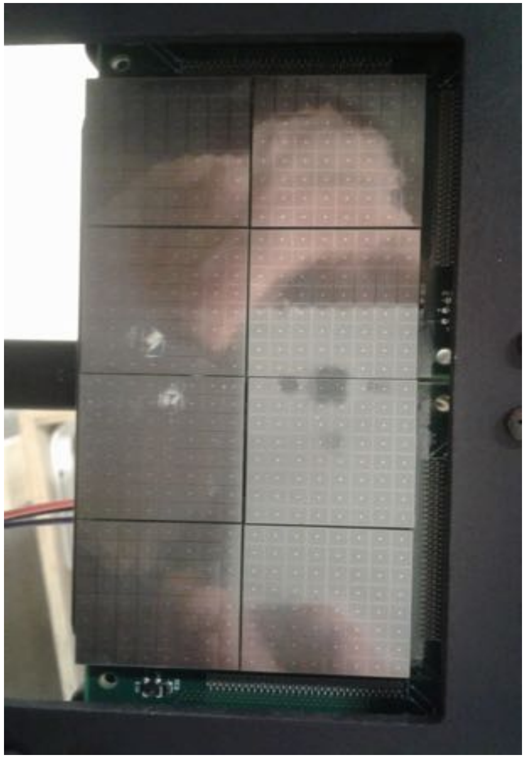
Today the Globus research data management service announced the largest single file transfer in its history: a team led by Argonne National Laboratory scientists moved 2.9 petabytes of data as part of a research project involving three of the largest cosmological simulations to date. "With exascale imminent, AI on

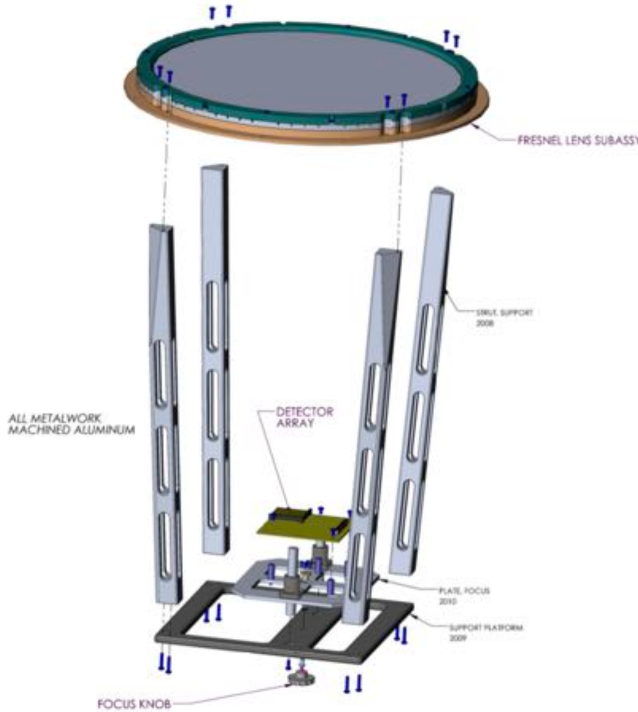


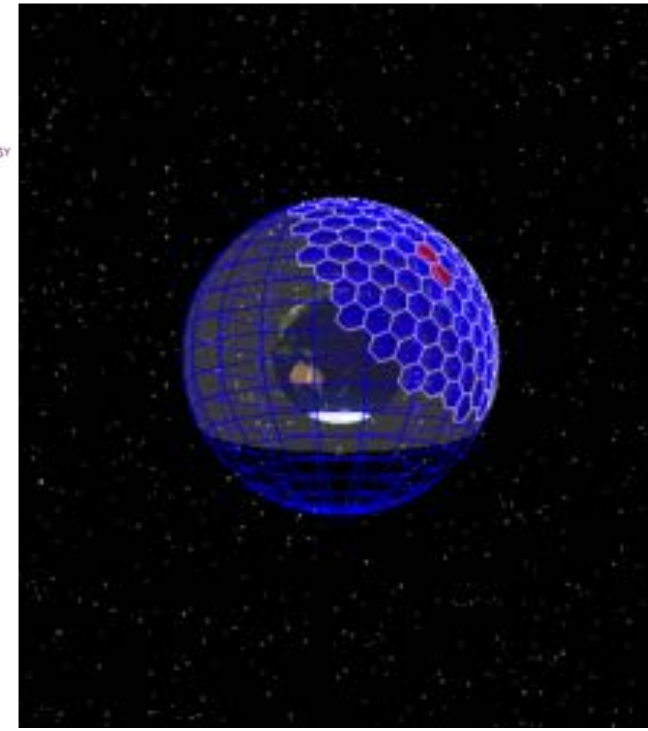


Guillaume Shippee

UC Berkeley







PANOSETI Goals: Interested in designing a telescope that can scan all the sky, all the time, for short time scale events such as pulsars, supernova, gamma ray bursts, and pulsed lasers

Guillaume Shippee
UC Berkeley

Challenges:

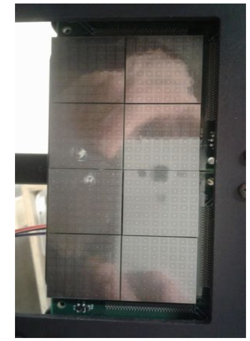
How do we quickly collect, analyze, and record data from >100,000 pixels with the conditions that:

-the pixels may be several kilometers apart

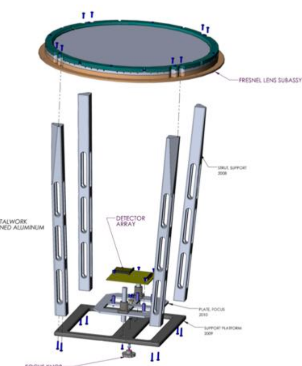
-the pixels are adapting to the environment and are well calibrated

- ‘noisy’ data can be identified and eliminated in real-time

1024 pix/mod



100 mod/dome



Jagath Vithanage

Florida International University

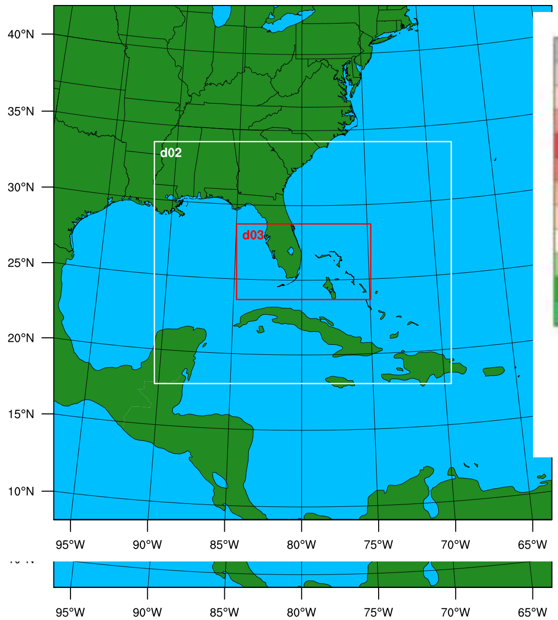
Coupling storm surge models with climate models, and I work as a hydrology and climate modeler. Uses WRF and ParFlow. Current storm surge models are not using sea wind when simulating tropical storms and has a profound effect on sea water intrusion and flooding. Land fall and low ground water table increase the complexicity of modeling. My research is an effort to include this compound effect using Cheyenne super computer and FIU HPC. WRF has components of serial and parallel, which may need to change the code to customize the output varibles/format. Input data are RCM and GCM with observed. Output data are in netCDF format and a single file can be of size of 10-100 GB

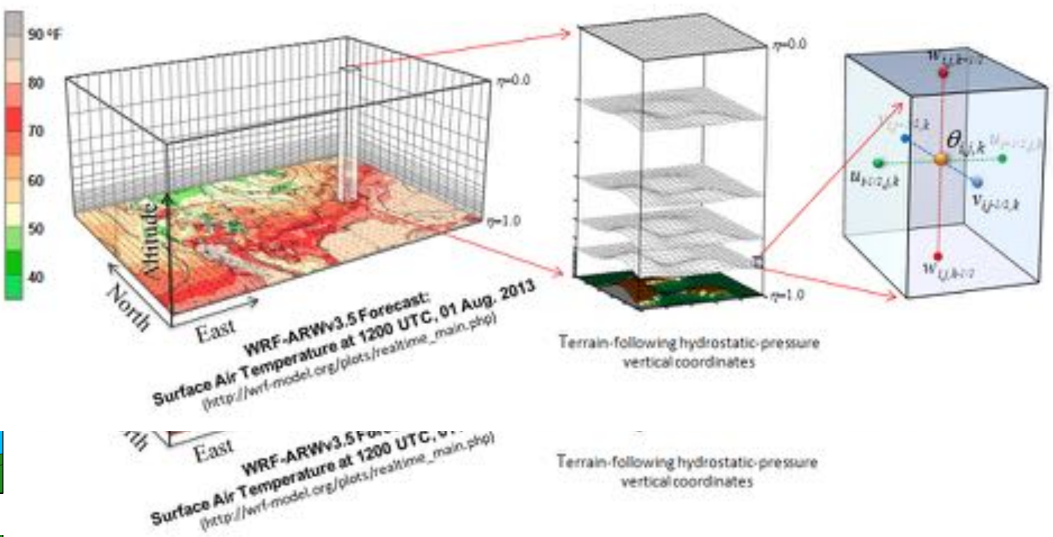
Jagath Vithanage

Florida International University

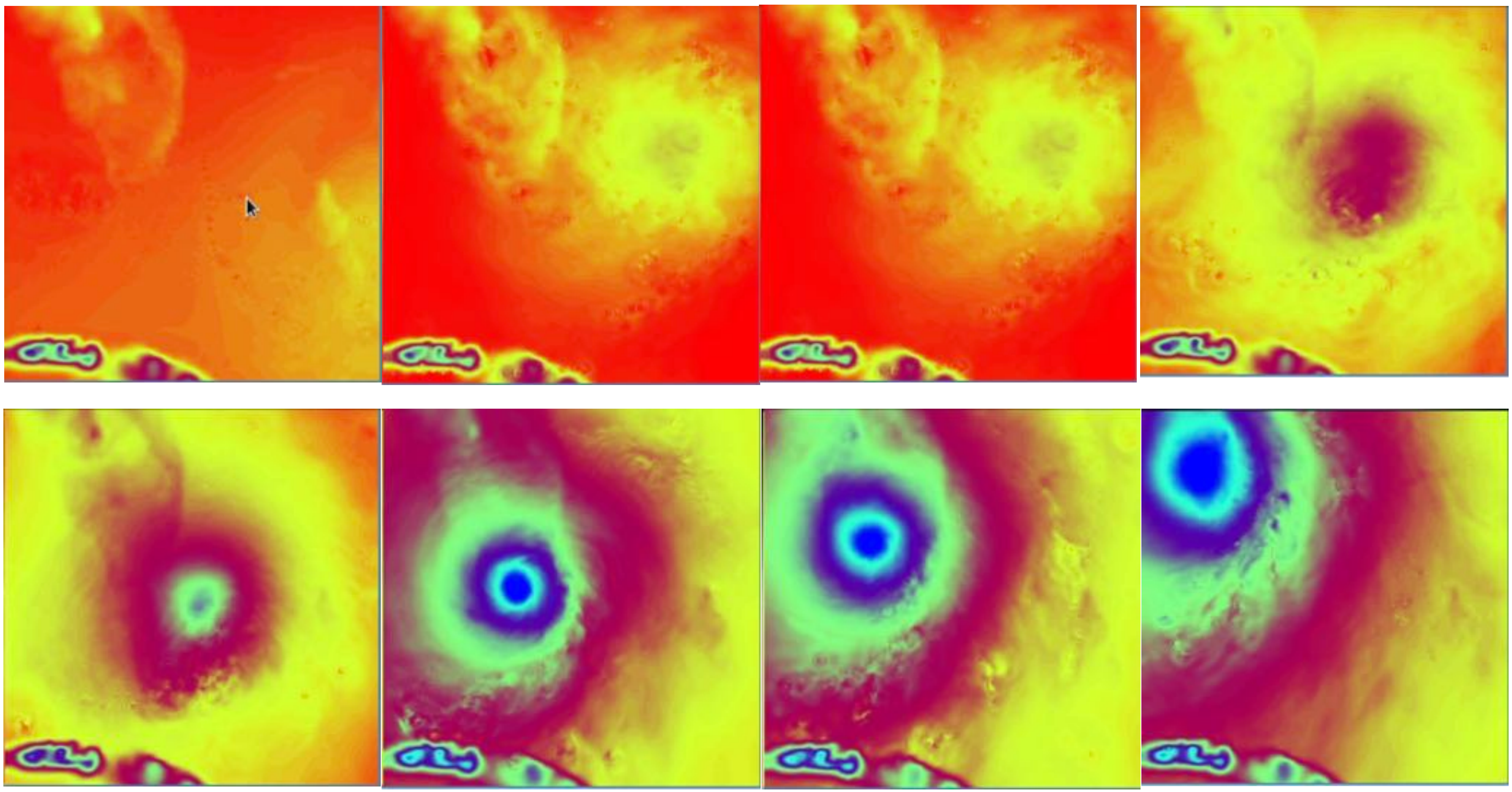
Main goals are optimization of the model efficiency after coupling the climate models with storm surge model.

WPS Domain Configuration





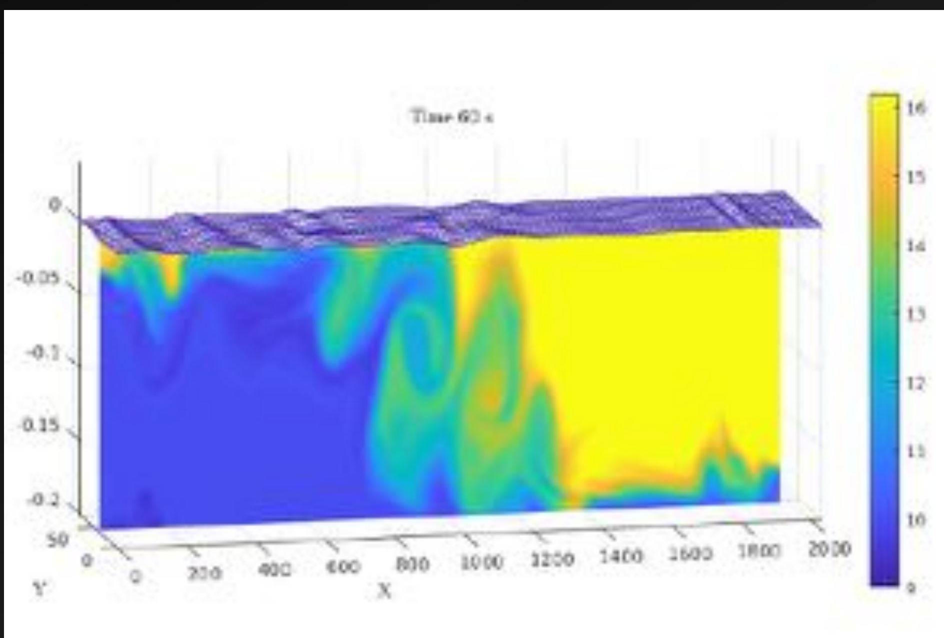
Jagath Vithanage
Florida International University

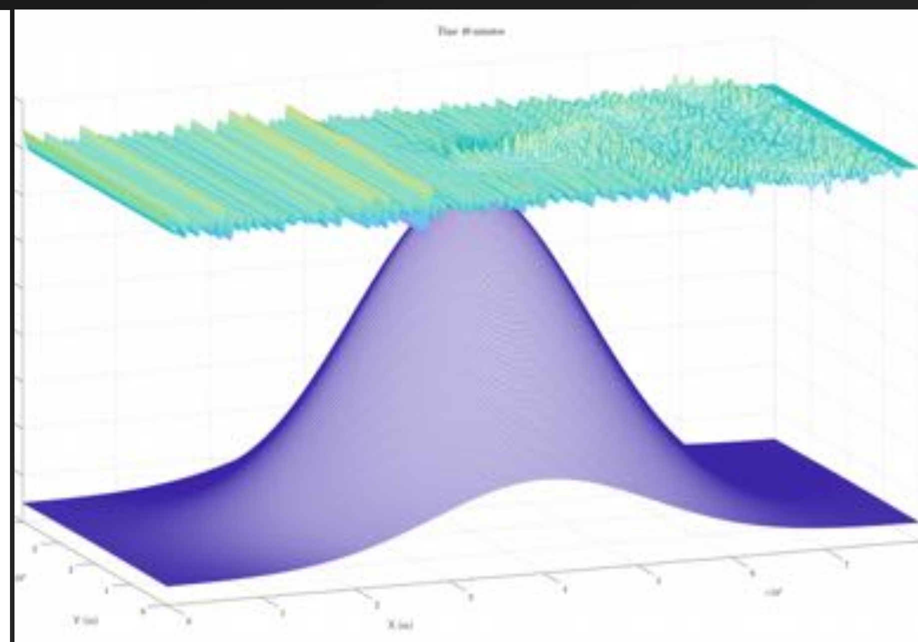


Jared Brzenski

San Diego State University

Ocean Modeling, General Curvilinear Coastal Ocean Model (GCCOM) built at SDSU. 4th order accurate. Masters Project: Coupling GCCOM with shallow water model to render surface features.





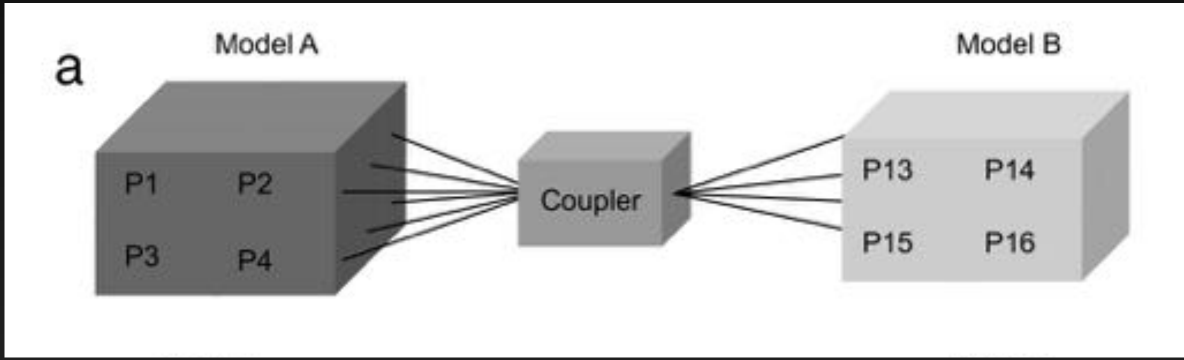
Jared Brzenski

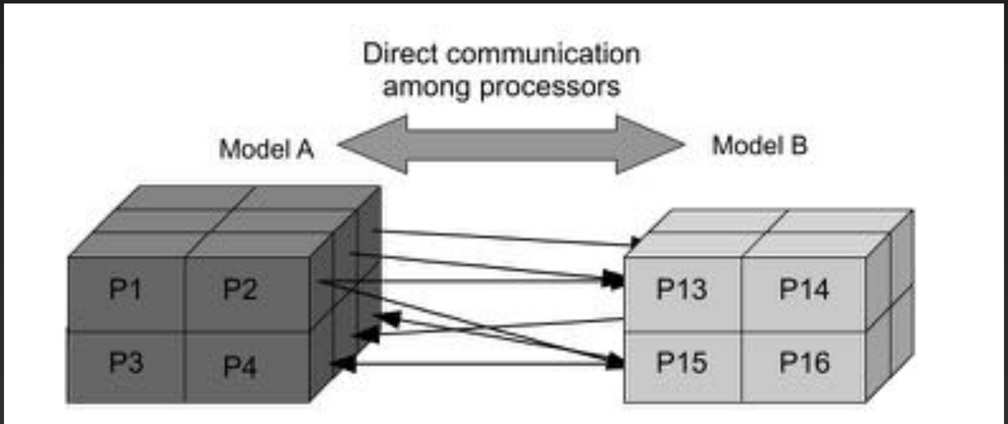
GOALS for Summer Institute

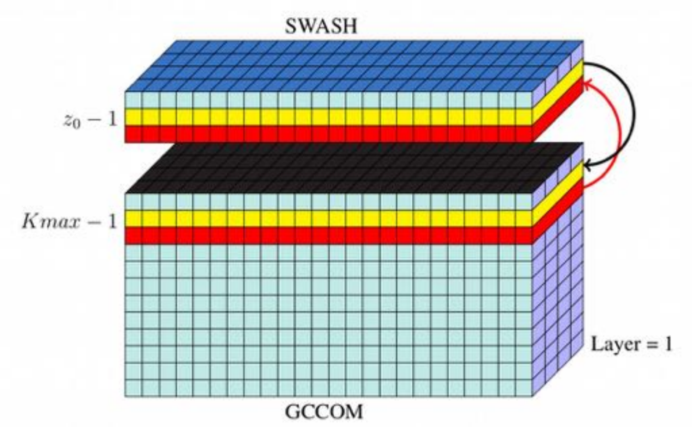
Optimize coupling of models, parallllelize output!!

Optimize code for HPC (FORTRAN, PETsC, MPI, others...)

Use containers for easier deployment







Source: Design of a Distributed Coupling Toolkit for High performance Computing Environment [9, §Figure 4]

Jin Soung Yoo

Purdue University Fort Wayne

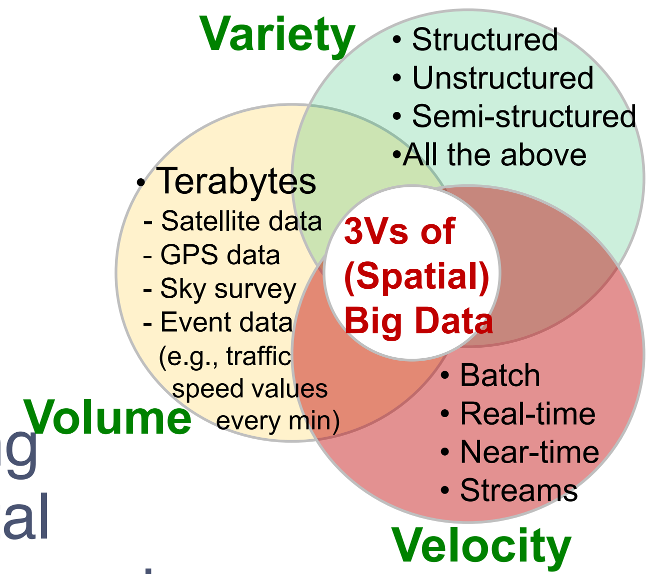
Research Area: ***Data and Knowledge Engineering*** - *Data Mining, Machine Learning, and Databases*

- **Data-driven knowledge discovery, data processing and management for non-traditional data** such as
  - Spatial/Temporal/Spatio-temporal data
  - Graph/Network data, Scientific data
- **Location-based computing, services, and Information**
- **Data Science applications** - public safe, public health, medical, life science, education, business, and so on.
- **Big Data analytics and management**

Jin Soung Yoo
Purdue University Fort Wayne

Recent Works on Big Data Analytics

- Spatial association pattern mining methods for Spatial Big Data on
  - Hadoop, MapReduce, NoSQL
  - Hadoop, Spark
- Frequent pattern mining from evolving spatial data – Parallel and Incremental Approach in cloud computing environment



Near Future Works

- Spatial/Spatio-temporal pattern mining on GPU
- Data Mining and Machine Learning for streaming data
- Exploration of Polystore systems

John E. Daniels

University of California, Davis Computational Research Service

The Computational Research Service provides quantitative computing, consulting and data services for the social sciences in College of Letters and Science.

Resources include mixed environments of UNIX and Windows research servers desktops and access to a modest HPC cluster (120+ nodes).

Primarily consultations deal with: systems support, statistical and mathematical packages, high-level programming languages, data acquisition and usage.

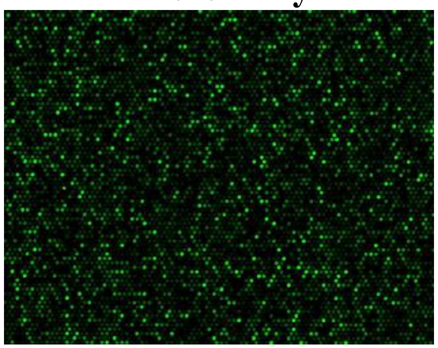
John E. Daniels
Key UCD facts

- 40k students across 102 majors
- Main campus in Davis, Medical Center in Sacramento, research centers from the Pacific to the Sierra (Bodega Bay to Tahoe Research Center) and and as far south as Tulare (VetMed Tulare Research Center)
- 4000 monkeys (rhesus)
- 650 head of cattle
- 300 sheep
- 150 goats
- 73 pigs (HR based in 100 year old pig barn)
- 40 horses
- 1-2 families of wild river otters, coyotes, and the occasional California black bear and mountain lion
- Annual campus open house Picnic Day includes maggot painting, dueling marching bands, and dachshund races!

Jose Martinez Lomeli
University of California, Riverside

Prediction of Transcription factors (proteins) binding sites

Protein Binding Microarray



Identify SNPs that alter Transcription Factor DNA binding

- Most of the transcription factors have a preference for a particular DNA sequence to bind
- Disruptions like mutations in a DNA sequence will affect the ability of transcription factor to bind
- The *in vivo* method made use of protein binding microarrays (PBMs)
- A PBM experiment allows to test hundreds of different DNA sequences for several proteins
- We trained a machine learning model base on the PBM data to predict new DNA binding sites

Jose Martinez Lomeli

University of California, Riverside

In the field of bioinformatics, we always look to improve the way we visualize the data

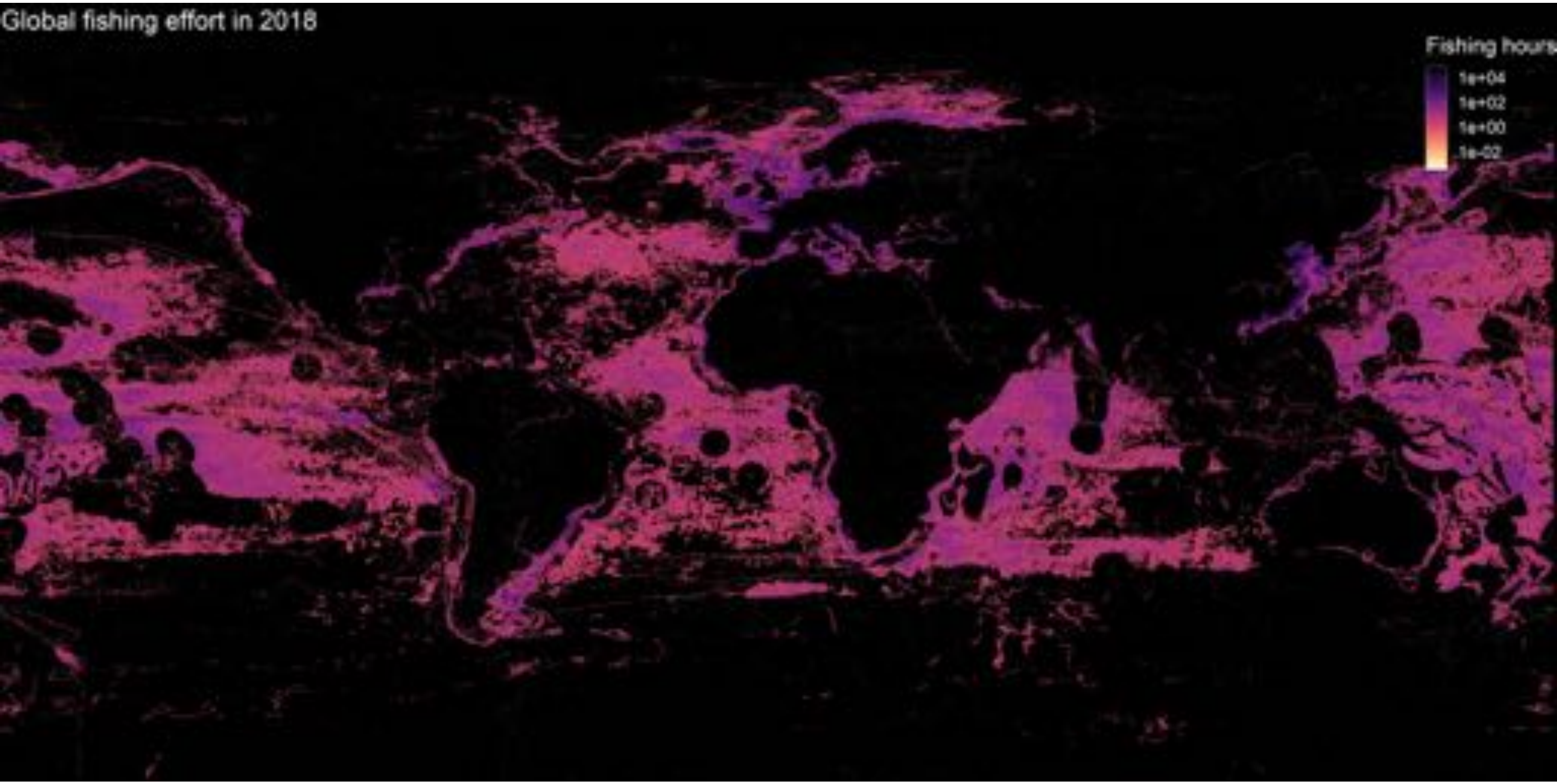
We look for more efficient ways to employ machine learning to analyze PBM, ChIP-seq, RNA-seq and other data

We would like to improve some of the current pipelines to reduce the large amount of resources that are required at different steps

We would like to use a state of the art deep learning models to improve transcription factor binding prediction

The current published models that predicts transcription factors binding sites made use of CUDA and we would like to port or create a more friendly tool using TensorFlow

Juan Carlos Villaseñor-Derbez
UC Santa Barbara



A modern approach to marine conservation

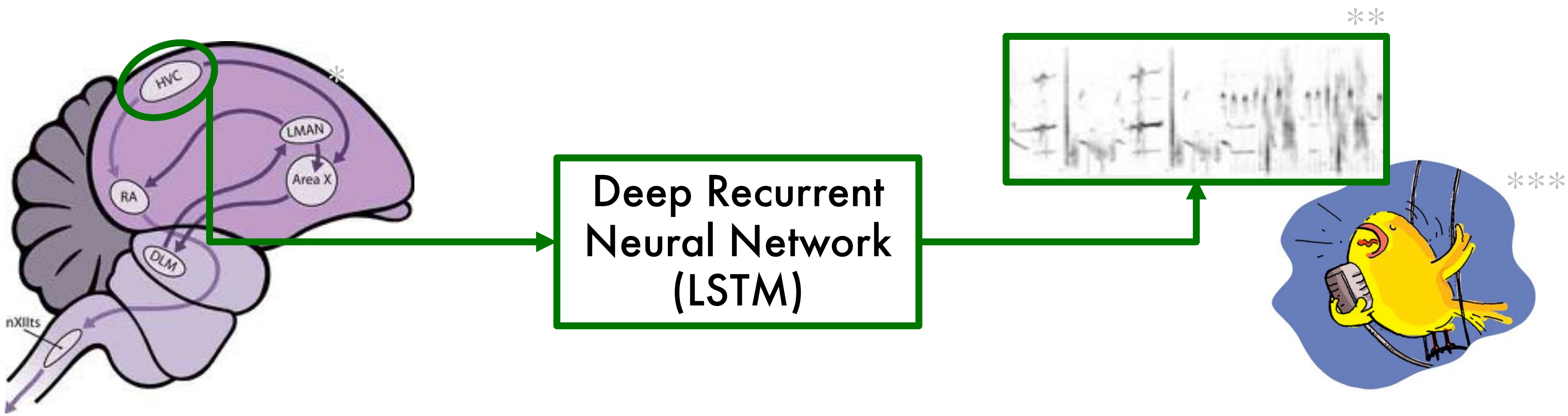
Juan Carlos Villaseñor-Derbez

UC Santa Barbara

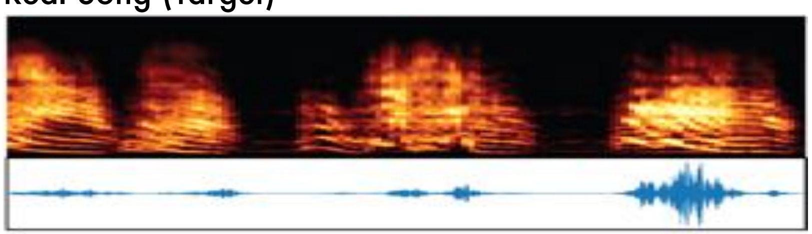
- Theoretical approaches
    - Long-running multi-scenario dynamic optimization problems
    - Embarrassingly parallel
    - No data are needed, small data are generated
- Empirical approaches
    - Often simple operations, but RAM-heavy (spatial)
    - Applied predictive modelling
    - Combination of multiple data sources (timeseries, remote sensing, fish catches, political boundaries)

A BMI to Decode Songbird Vocal Outputs from Neural Activities

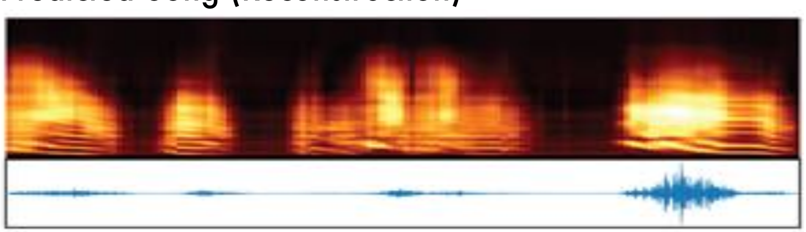
Kai Chen
UC San Diego



Real Song (Target)



Predicted Song (Reconstruction)



Future Directions:

1. Define a perceptual error function
2. Design new architecture to have longer memory
3. Optimize network for real time analysis

*Nottebohm F (2005) The neural basis of birdsong. PLoS Biol 3(5): e164
**Gentner, T. Q. (2008). Temporal scales of auditory objects underlying birdsong vocal recognition. The Journal of the Acoustical Society of America, 124(2), 1350–1359. http://doi.org/10.1121/1.2945705
*** https://www.toonpool.com/cartoons/singer_42583

Goals at the Summer Institute

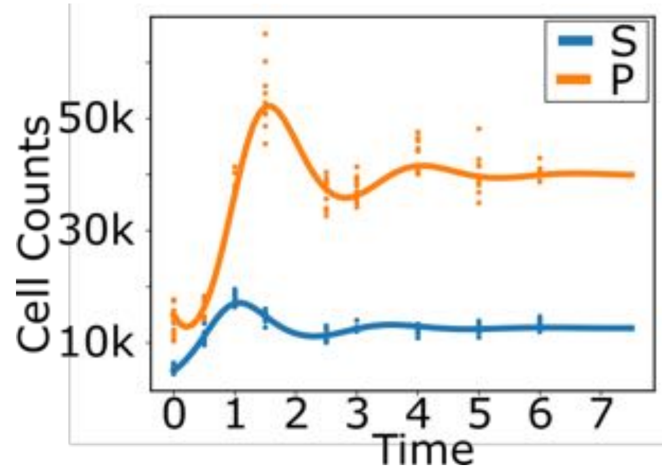
Kai Chen
UC San Diego

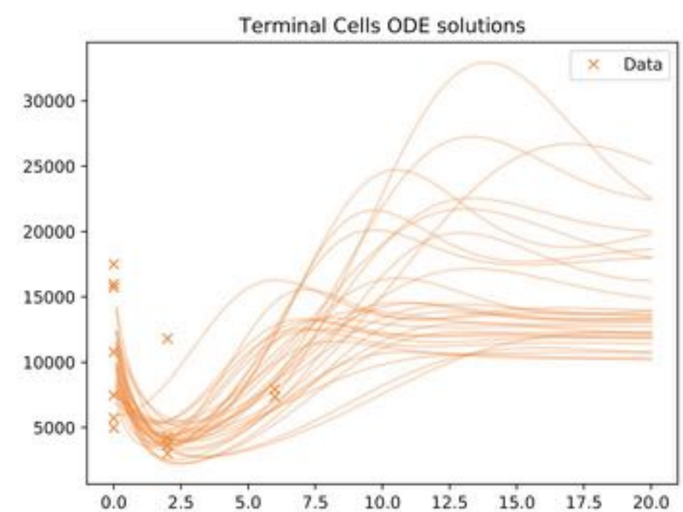
1. Learn to optimize complicated processes (model fitting!) and be more efficient
2. Have a better understanding of deep learning models
3. Improve my data management habits and curation skills
4. Meet people working on interesting projects

Bayesian Optimal Experimental Design: How to design informative experiments?

Luis M. Lomeli and Abdon Iniguez

- There are multiple variables that need to be adjusted when performing an experiment
- Maximize the information gain from an experiment by choosing a design varying:
    - Number of biological replicates, timing of the records and initial conditions
- We use a mechanistic model that captures important dynamics of hematopoiesis
- We use open source libraries in R and Python to solve ODEs and run Markov Chain Monte Carlo



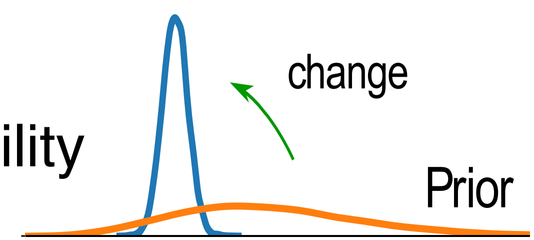


NSF grant DMS-1936833. This center was supported by a NSF grant DMS1763272 and a grant from the Simons Foundation (594598, QN).

NSF-Simons Center for
Multiscale Cell Fate Research
UC Irvine

Bayesian Optimal Experimental Design: How to design informative experiments?

Luis M. Lomeli and Abdon Iniguez



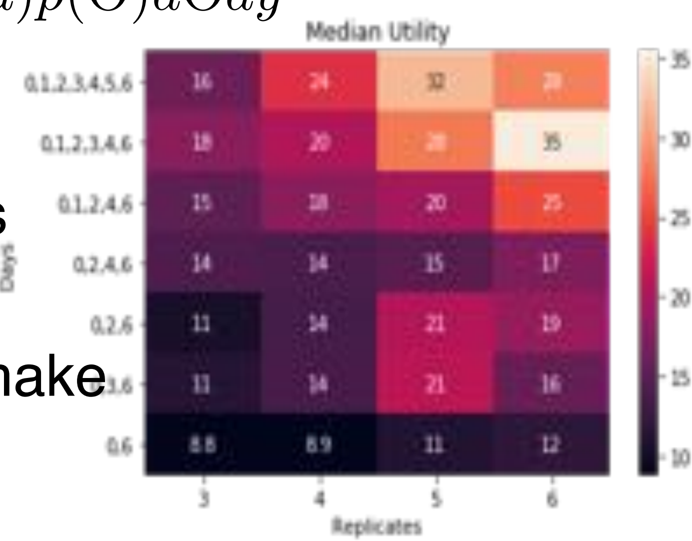
Quantify information gain

- Quantify uncertainty about the model parameters
- For each design *d*, calculate the KL-divergence utility

$$U(y,d) = \int_{\Theta} log\left(\frac{p(\Theta|y,d)}{p(\Theta)}\right) p(\Theta|y,d)d\Theta$$

- Compute the expected/median utility

$$u(d) = \mathbb{E}_{\Theta,y}\left[U(\Theta,y,d)\right] = \int_y \int_\Theta U(\Theta,y,d)p(y|\Theta,d)p(\Theta)d\Theta dy$$



- This is extremely computationally intensive since need to run tons of MCMC simulations
- Need to use HPC resources
- Also need to use Supercomputing tools to make MCMC faster and possibly more efficient

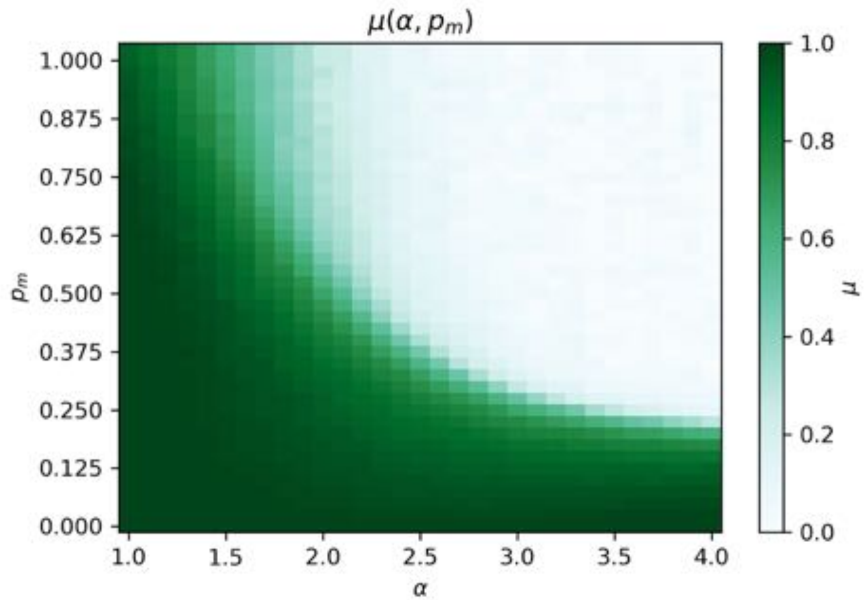
Max Block
UC Berkeley

- Physics graduate student
- Emergent phenemona in “coherent” quantum systems
  - Recent experimental advances enable control of dozens or hundreds of coherent quantum dofs (trapped neutral atoms)
  - How does classical physics (e.g. diffusion) emerge?
  - What is the role of measurement and noise?

Max Block

UC Berkeley

- Very few interesting cases are analytically tractractable
- We numerical models. Example: steady state entanglement entropy controlled by measurement rate



~50k SUs on Berkeley Savio cluster

$S_e \sim L^{\mu}$

Md Shahrier Hasan

Graduate Student, University of California San Diego

Multi-scale Modeling of the Mechanical Properties of the Nano- structured Materials

- The goal of the research project is to develop computational methods to model the mechanical properties and apply them to understand the mechanisms of deformation behavior in nanostructured materials.
- This involves molecular dynamics to describe dislocation interactions with grain boundaries and interfaces at the atomic and nanoscale.
- Atomistic to Continuum Finite Element Coupling to understand dislocation motion at the microscale

Md Shahrier Hasan

Graduate Student, University of California San Diego

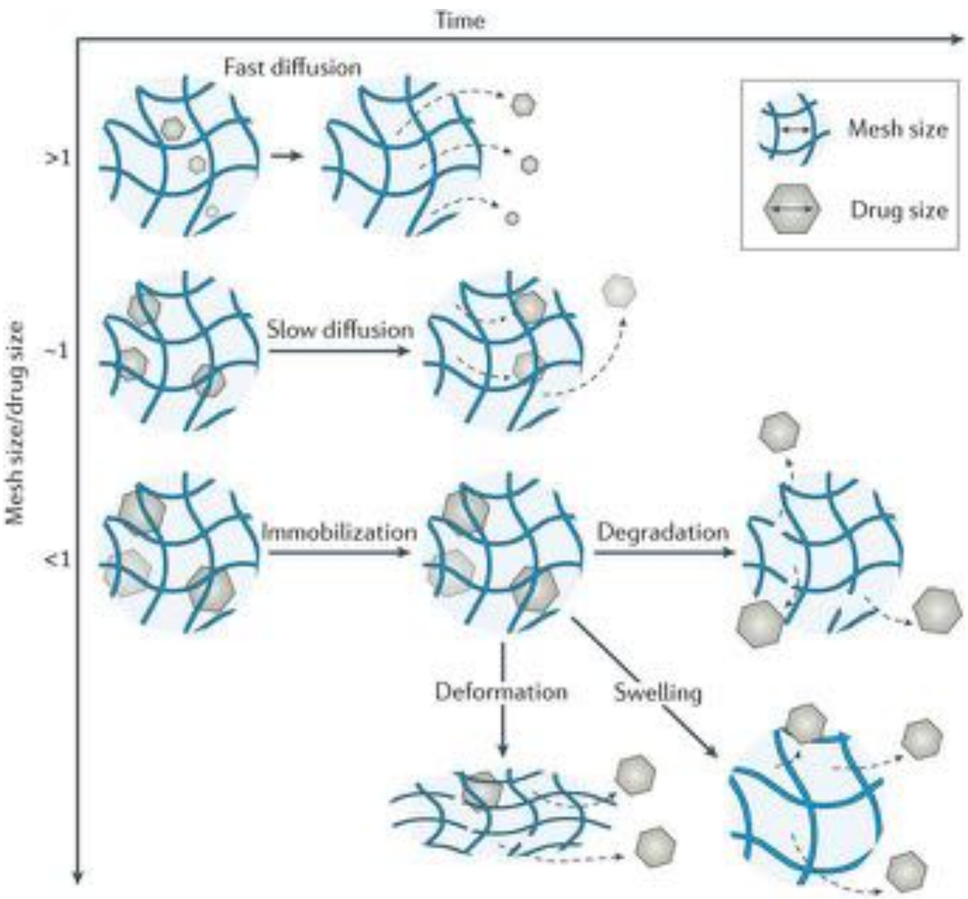
The key takeaways from SI2019:

- For Molecular Dynamics Simulation, optimize the size of the atomic system that can be modeled with effective parallelization for the available computational resource of the project
- Coupling with a suitable Continuum Finite Element software to implement Concurrent Multi-scale Modeling technique
- Effective visualization approaches to track atomic data

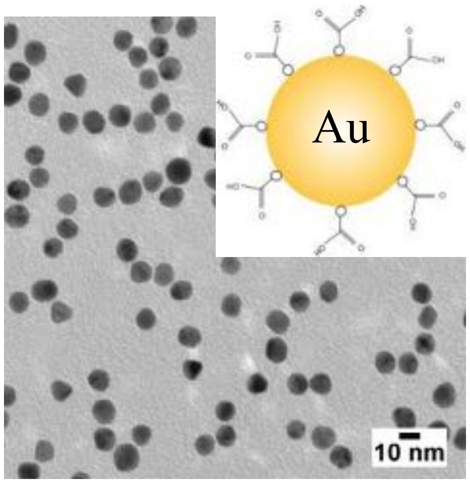
Mike Boyle

PhD Student, Materials Science and Engineering

mboy@seas.upenn.edu



Li *et al. Nature Reviews Materials* ***2016,*** *1* (12), 16071



Nanocompix.com

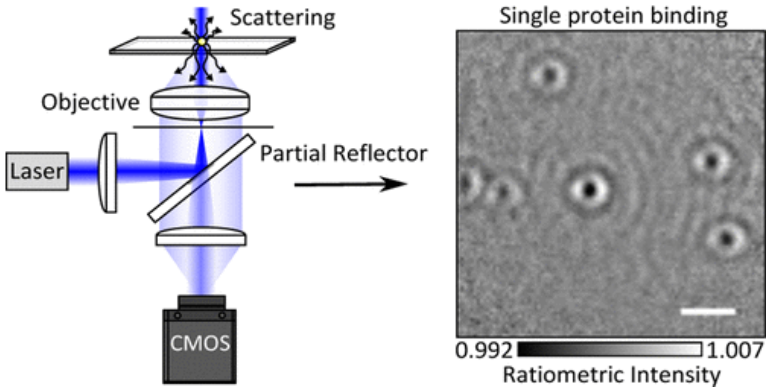
Visionshopping.com

Mike Boyle

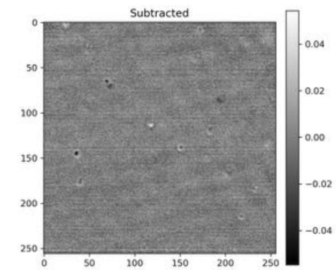
PhD Student, Materials Science and Engineering

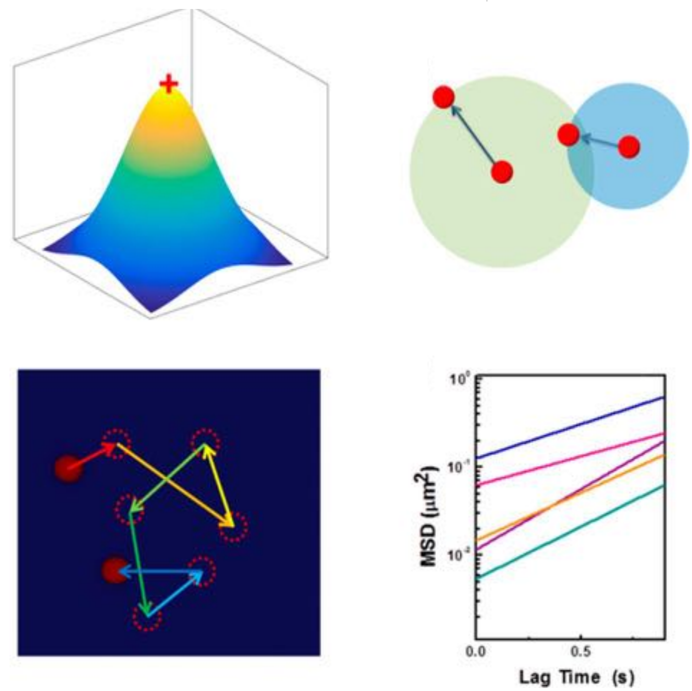
mboy@seas.upenn.edu

Single Particle Tracking with iSCAT up to 1000 Frames per second

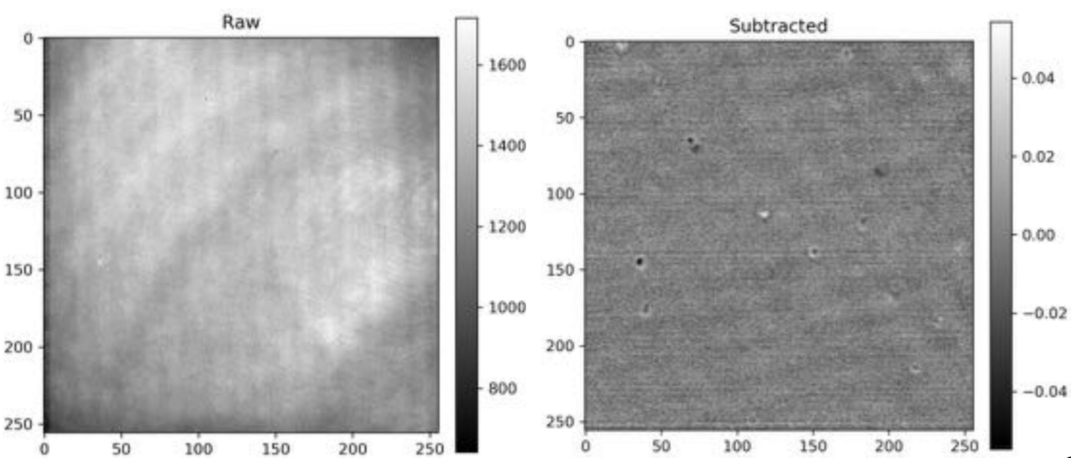


Cole *et al. ACS Photonics* ***2017***, *4*, *211-216*





Shen, H., Landes, C. F., *Chem. Rev.* **2017,** 117, 7331-7376



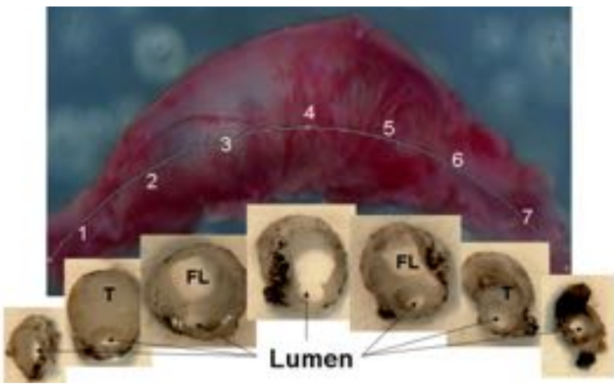
Minglang Yin
Center for Biomedical Engineering,
Crunch Group, Division of Applied Mathematics,
Brown University

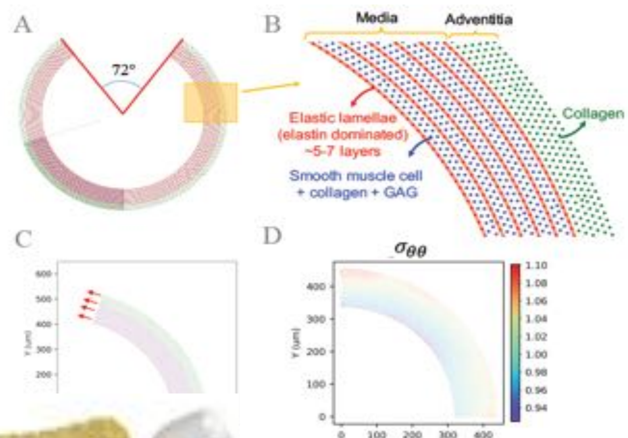
BROWN

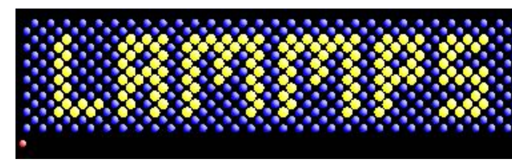
<u>Research interests</u>: Multiscale Modeling, Machine Learning +, Computational Biomechanics

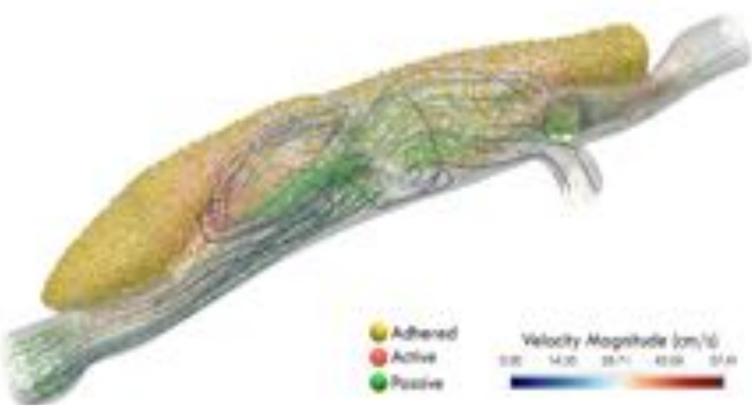
<u>Project</u>: Multimodality Imaging-driven Multifidelity Modeling of Aortic Dissection

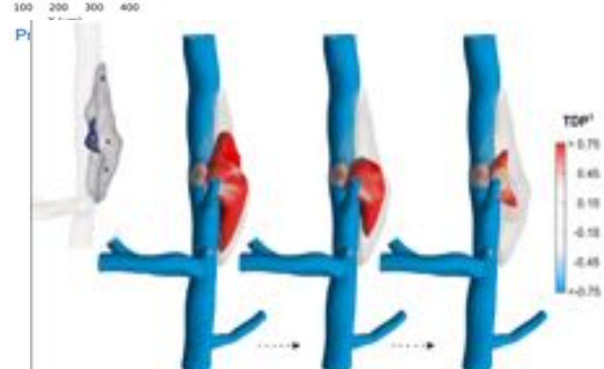
<u>Aim 1</u>: develop multifidelity deterministic models of early intramural delamination processes that lead to dissection





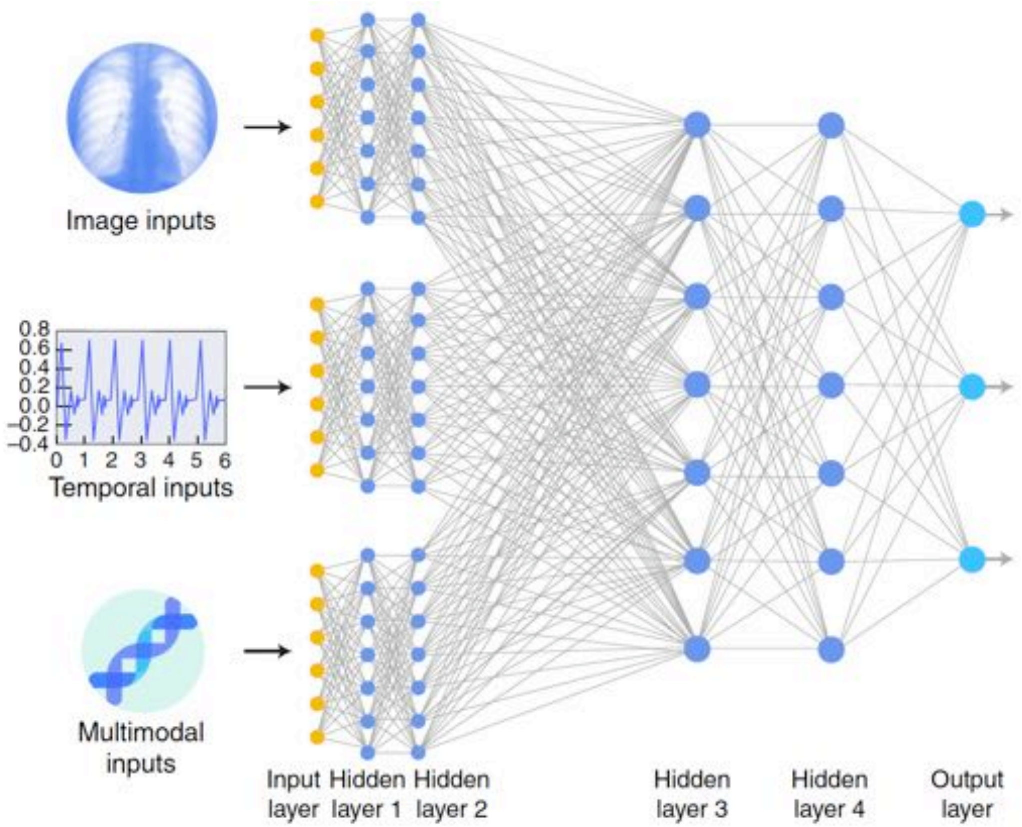






Minglang Yin
Center for Biomedical Engineering,
Division of Applied Mathematics,
Brown University

<u>Aim 2</u>: The ultimate goal is to achieve semi-supervised learning to predict outcomes by integrating findings from different animal models, computational models, and categorical data.



Esteva, Andre, et al. "A guide to deep learning in healthcare." *Nature medicine* 25.1 (2019): 24.

- **Acknowledge**

SDSC SAN DIEGO SUPERCOMPUTER CENTER

NIH National Institutes of Health

אור ותמים

LUX ET VERITAS

東京大学
THE UNIVERSITY OF TOKYO

Omar Zintan Pomona College

- **What is HPC support?** A group of student workers headed by Pomona's HPC specialist Asya Shklyar which exists to provide support to students and faculty who need to use our HPC environment.
- **Why is HPC support necessary?** The demand for HPC is on the rise, hence, the team facilitates the support that Ms Shklyar gives to faculty and students. It also gives early exposure to students interested in HPC.
- **What do we do aside HPC?** We have a lab open to all student workers to experiment with technology such as VR, 3D printing, Drones, Raspberry Pi's mainly for educational projects.

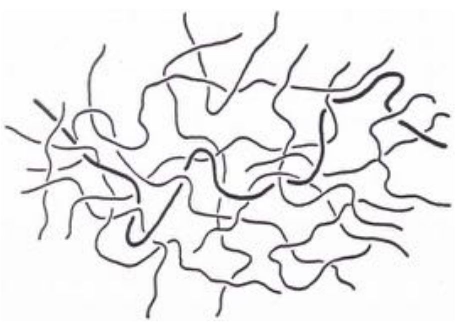
Omar Zintan Pomona College

- I partnered with a CS faculty, and one other student worker, in developing an auto grading platform on GitLab for his fall AI class
- We made a bot that automates:
  - grouping of students with the right permissions
  - forking of the class repository for each student
  - creation of branches and merge requests for each assignment
  - running pipelines for grading whenever students submit their assignment by committing to the already created merge requests
  - presenting results to students as comments in the merge requests and storing their grades securely in a grade book
- At the end of this conference I hope to have learnt enough to actively help and advise faculty and students, and even mentor my fellow student workers, on how to work with HPC.

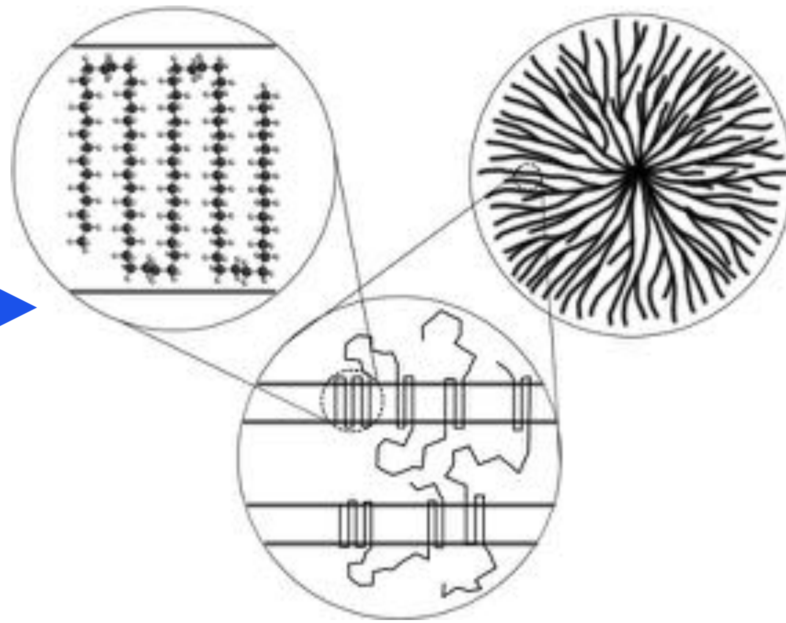
Pierre Kawak

Brigham Young University

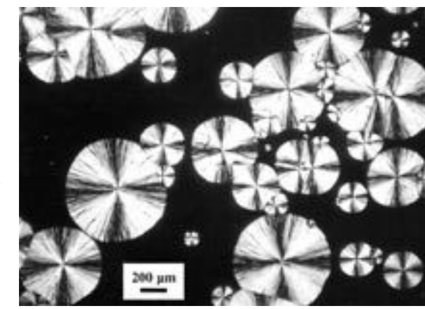
Amorphous Melt → (Primary Nucleation) → Lamellar Crystal → (Secondary Nucleation) → Spherulite → (Coarsening) → Semicrystalline Materials



https://polymerinnovationblog.com/polymers-electronic-packaging-types-polymers-used/



S. T. Milner, "Polymer crystal-melt interfaces and nucleation in polyethylene," Soft Matter, vol. 7, no. 6, pp. 2909–2917, 2011.



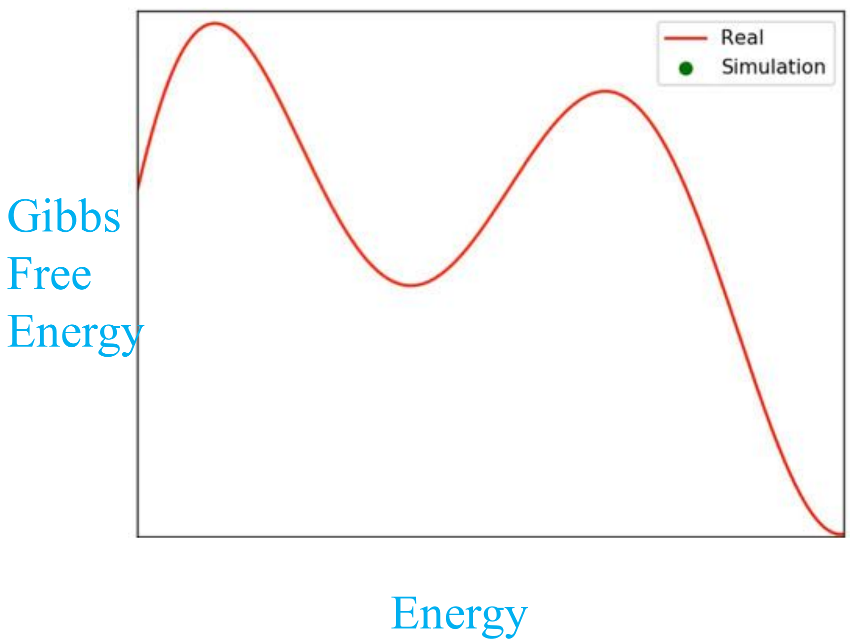
Winkler, Anders & Kloosterman, Gertjan. (2015). A critical review of fracture mechanics as a tool for multiaxial fatigue life prediction of plastics. Frattura ed Integrita Strutturale. 9.

Tree Soft Matter Theory @ BYU ChemE

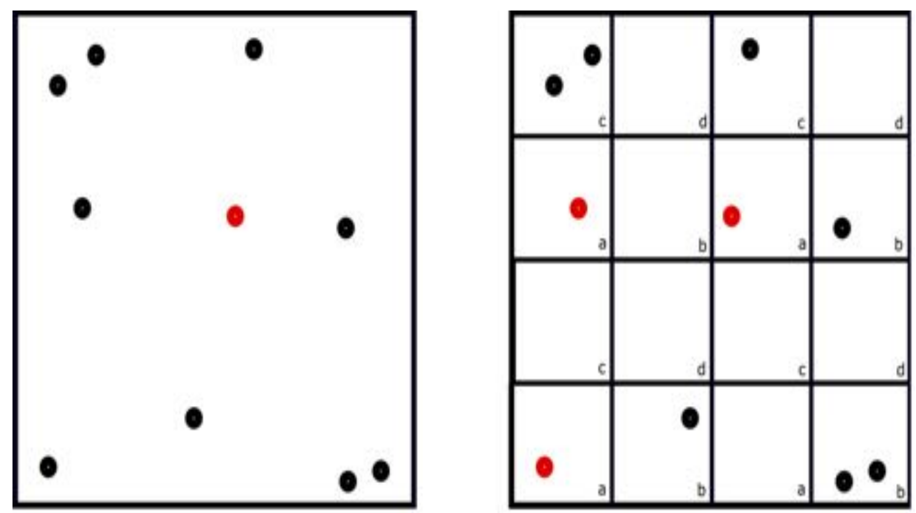
Pierre Kawak

Brigham Young University

1. Wang-Landau Sampling



2. Domain Decomposition

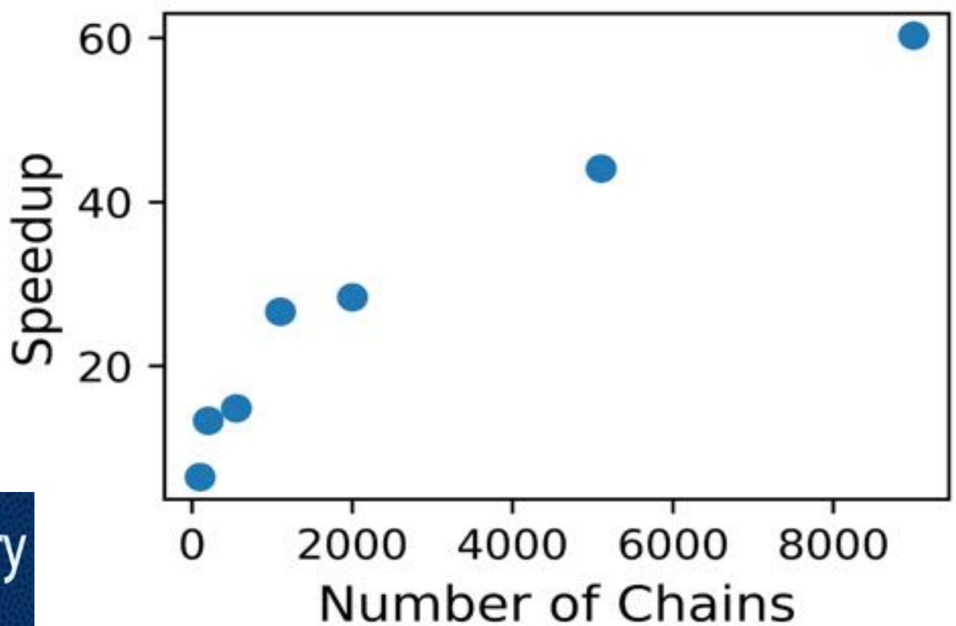


Tree Soft Matter Theory @ BYU ChemE

Pierre Kawak

Brigham Young University

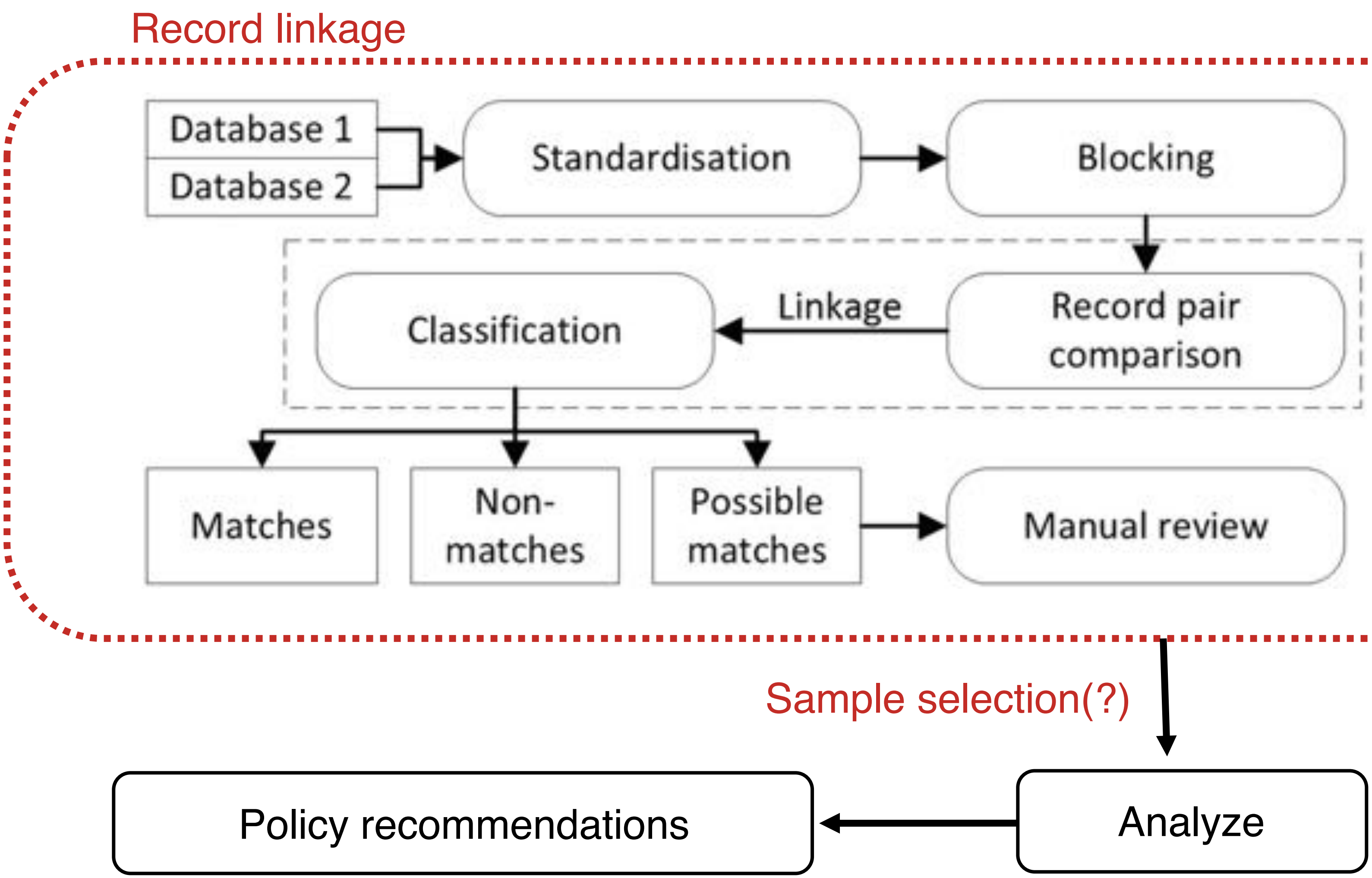
- The program is called Monte Carlo Polymer Crystallization (**MCPC**). It does **GPU-accelerated Polymer Wang-Landau Sampling Monte Carlo Simulation**.
- Preliminary studies show high achieved speedups



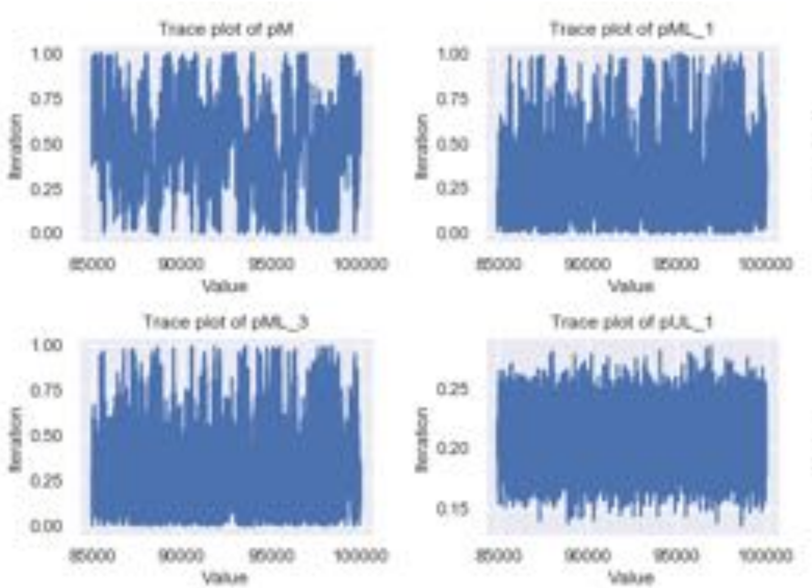
Tree Soft Matter Theory @ BYU ChemE

Rachel Anderson

PhD Student, Princeton University



Rachel Anderson
PhD Student, Princeton University



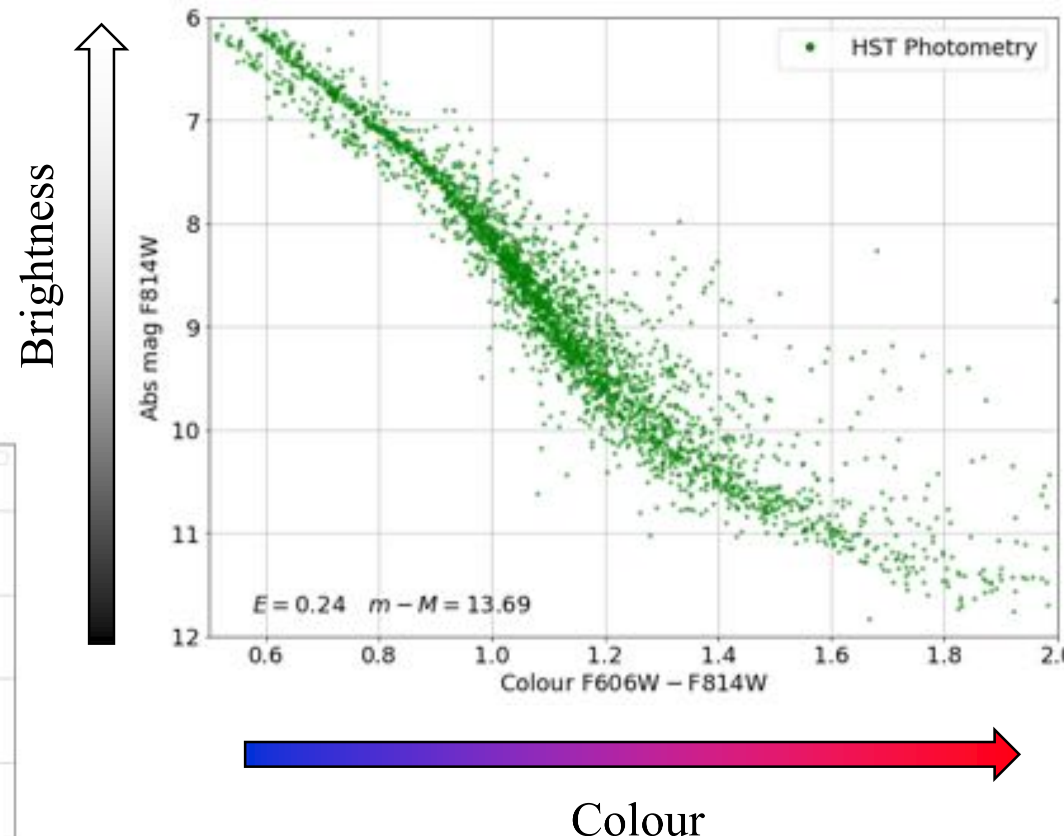
$$\hat{\theta} = \arg\min_{\theta \in \Theta} \left\{ \min_{\mathcal{Z}} \sum_{i=1}^{n} q(z_i; \theta) \right\}$$

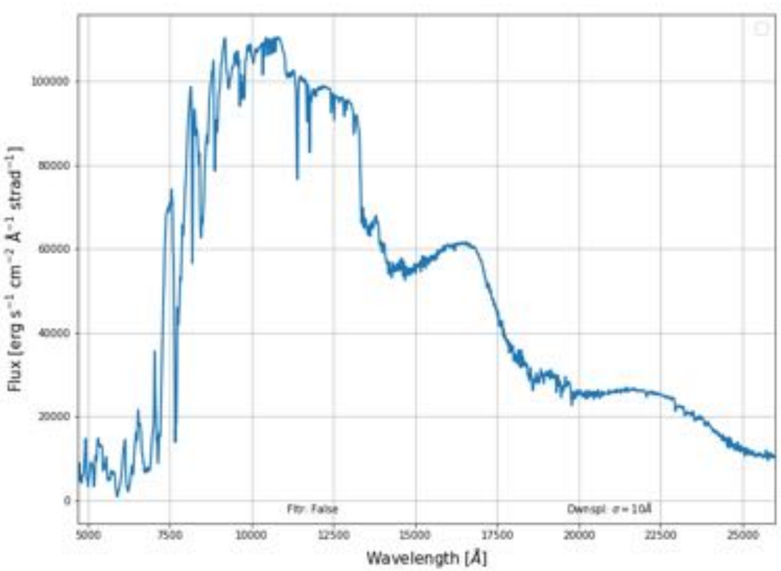
- Learn to perform computations efficiently, in parallel
- Focus on reproducibility and accessibility
- Introduction to resources and community for future projects

Roman Gerasimov
UC San Diego

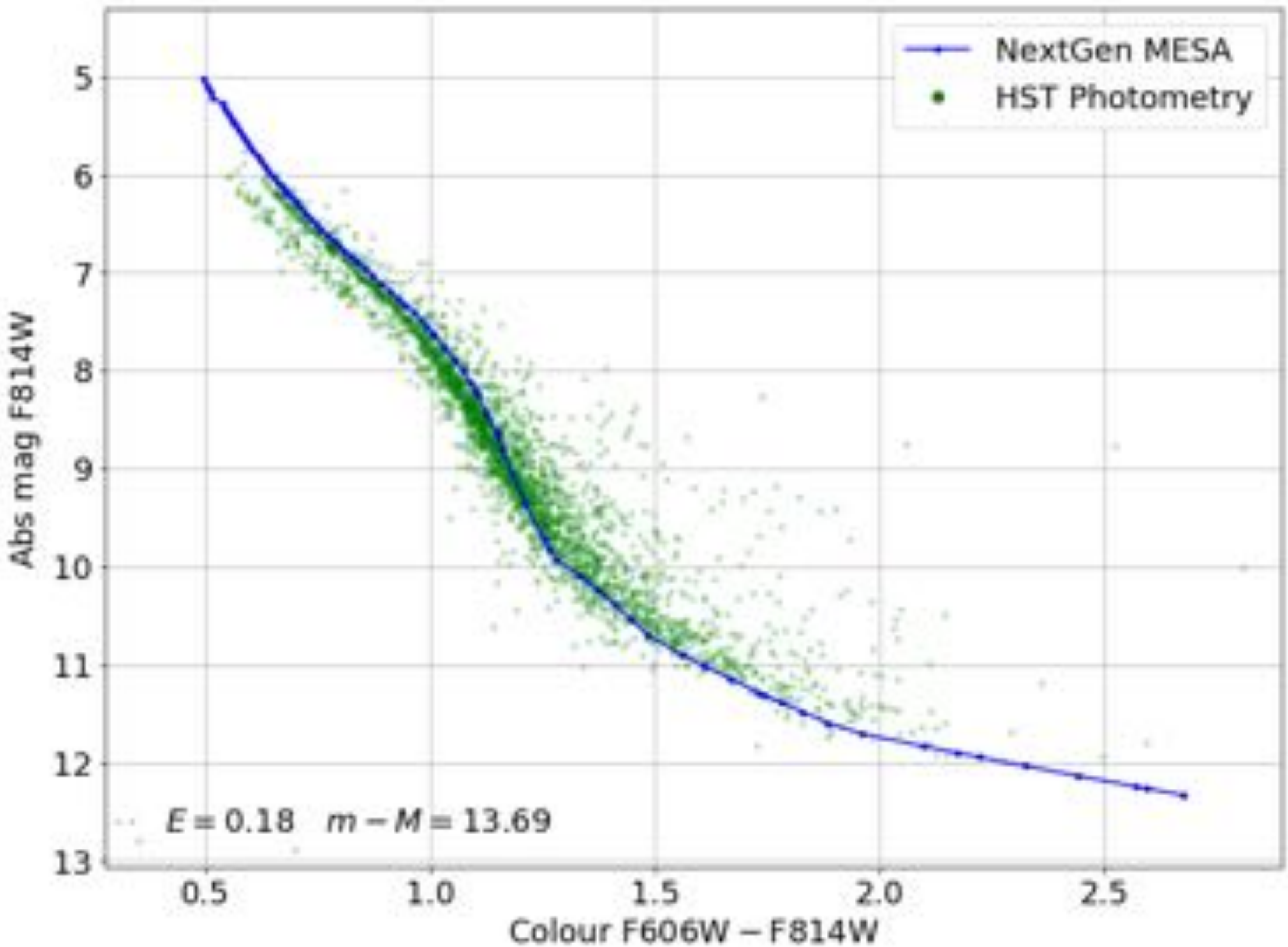
Credit: European Southern Observatory

NGC 5139 aka ω Cen





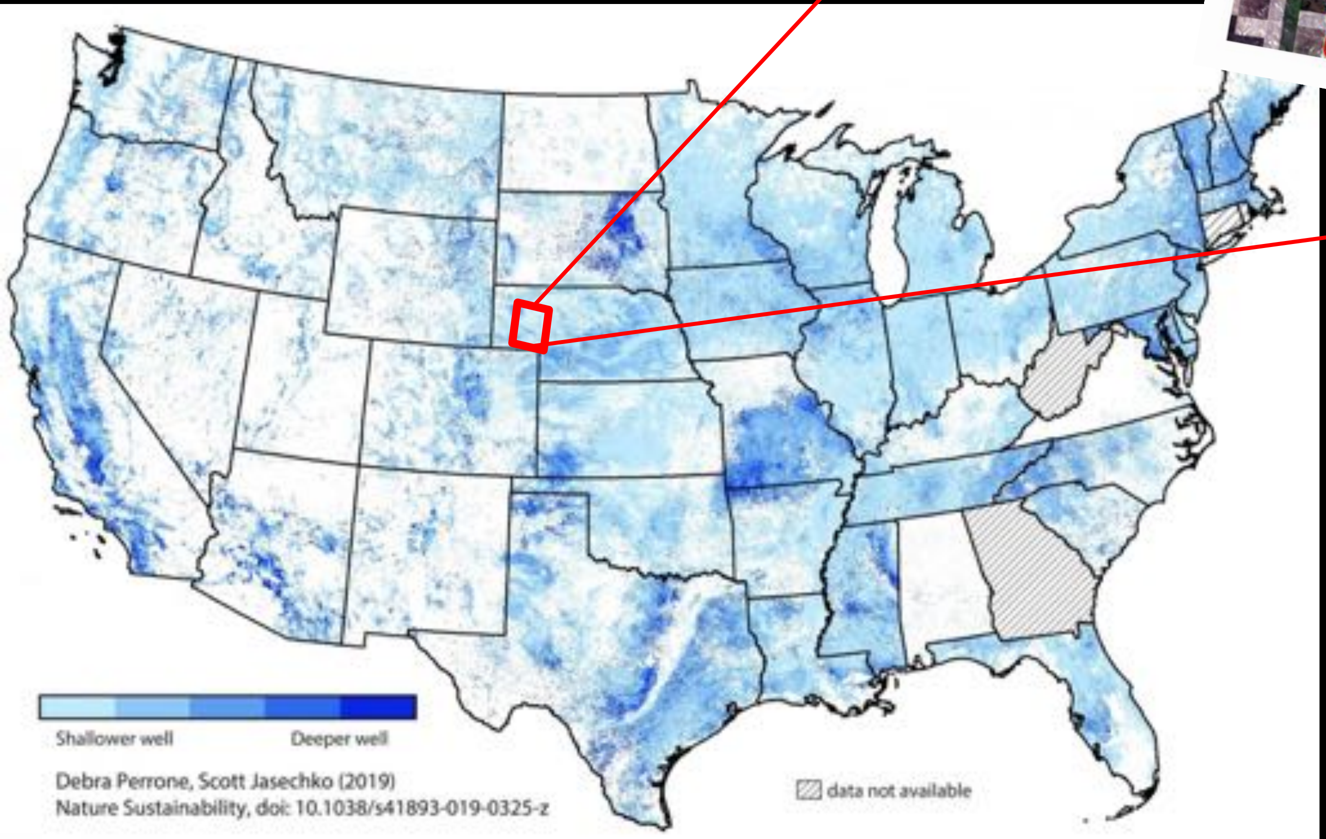
Roman Gerasimov
UC San Diego



Ryan Avery, UCSB

Problem: We know that globally, wells are being drilled deeper and deeper and water resources are increasingly strained.

But there is a lack of water use monitoring in developing regions.



Question: Can we train a model that can generalize across arid agricultural zones to locate past and future agricultural fields? Can we then locate hotspots (agricultural expansion) and cold spots (abandonment) with satellite imagery?

https://github.com/ecohydro/CropMask_RCNN

Ryan Avery, UCSB

Computational Challenges

4 scene footprints * 4 seasonal images * 1.5 Gb per image = **24 Gb of imagery for training the Keras Mask R-CNN model**

~ 40 scene footprints * ~10 seasonal images * ~30 years of Landsat imagery * 1.5 Gb per image = **18 Tb of imagery for inference**

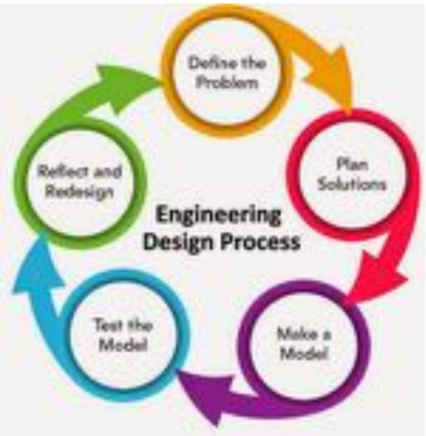
I’m here to get experience with profile my code (check!), parallelizing my preprocessing routines with dask (check!), and getting experience with multi GPU and multi node architecture and programming.

Sharon Tettegah
University of California, Santa Barbara

Research focus on broadening participation in STEM

User Preferences in Engineering & Beyond

I am very interested in helping undergraduate minority students gain access and knowledge in the area of computational/data science.



Sharon Tettegah

University of California, Santa Barbara

Computational challenges

Overall knowledge in the technical areas:

Learning R and also python

Visualization applications

Databases

Opportunities

SDSC: Provide basic knowledge of a variety of tools

Stephanie Labou
University of California San Diego

Data Science Librarian – support all things data/data science related, across all departments, for undergrads, grads, faculty, and staff

Support ranges from one time consultations, to recurring consults, to standing meetings

Also focused on making and maintaining connections across campus for data/compute-intensive education & research

Keeping a foot in the research realm (workflows needing HPC, parallel code, etc.)

Stephanie Labou

University of California San Diego

My computational challenges: any challenges anyone asks me about

My goal: know just enough to be dangerous about a lot of computationally-related things

- Focus on resources to learn more

Helping students/researchers:

- Get started with supercomputing – Linux env, load/retrieve data, parallelize code
- Hot topics: work with containers, ML at scale

Tian Ni

Purdue IE & GM R&D Operations Research

Topic: emergent safety issue detection by text mining external data sources including NHTSA VOQ, Twitter, Chevrolet Forum, Facebook and etc to link to internal data sources including Engineering, Supply Chain, Manufacturing, Quality, Warranty data to detect potential hazardous emergent issues.

Tools: ontology extraction, text mining, machine learning, artificial intelligence

Approach: data streaming vendor tools, in house algorithm development

Tian Ni

Purdue IE & GM R&D Operations Research

Challenges: massive data, slow streaming, computational expensive, low prediction accuracy

SDSC: data flow management, real-time streaming/processing, data partition, parallel computing, performance tuning on external data sources

***Tianyang Chen**, PhD Student in Geography*
University of North Carolina at Charlotte
tchen19@uncc.edu

Role: Research Assistant in Center for Applied GIS

DeepHyd Project: a deep learning-based artificial intelligence approach for the automated classification of hydraulic structures

Data Acquisition Algorithms: Structure from Motion (OpenSfM: https://opensfm.readthedocs.io/en/latest/)

Problem: Speed up this process as per the accessibility of high-resolution photo

Current status: share-nothing parallelism on some steps

Data: Approximate 300 images with 5,184*3,888 resolution per bridge

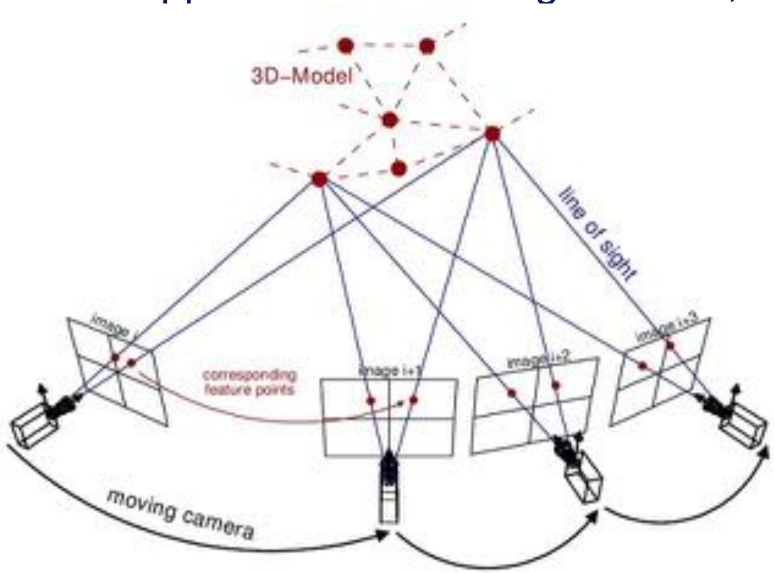
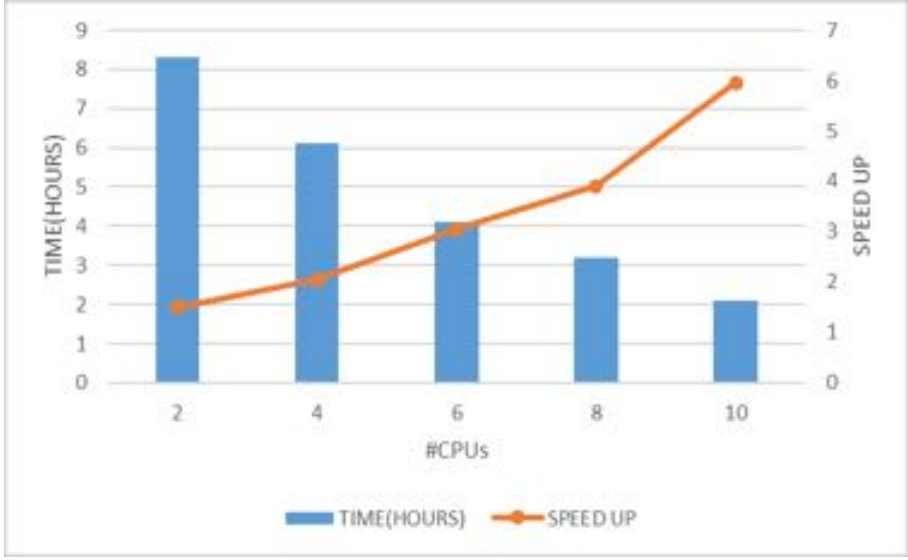


Image source: http://theia-sfm.org/sfm.html

Source: https://cybergis.uncc.edu/deephyd/

***Tianyang Chen**, PhD Student in Geography*
University of North Carolina at Charlotte
tchen19@uncc.edu

Sequential time:16 hours per bridge



CPU: Intel® Xeon® Processor E5 v3 10cores 3.1GHz

- **Copperhead**: **Linux**-based HPC Cluster
  - 96 compute nodes
  - 2,060 computing cores (CPUs)
  - Total Memory: 18,004 GB
  - Located at University Research Computing at the University of North Carolina at Charlotte (https://urc.uncc.edu/)

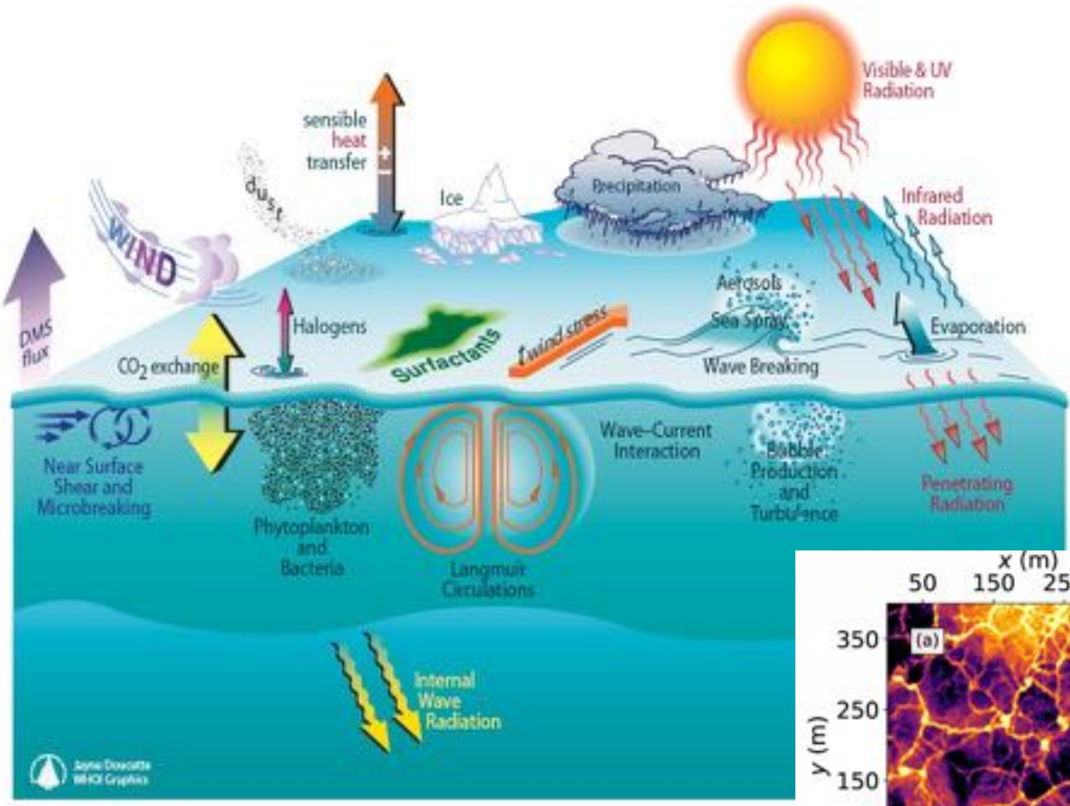
Estimated time for all bridges in NC:
18,000 bridges*2 hours ≈ **4 years**
To accelerate: MPI,GPU, and optimization

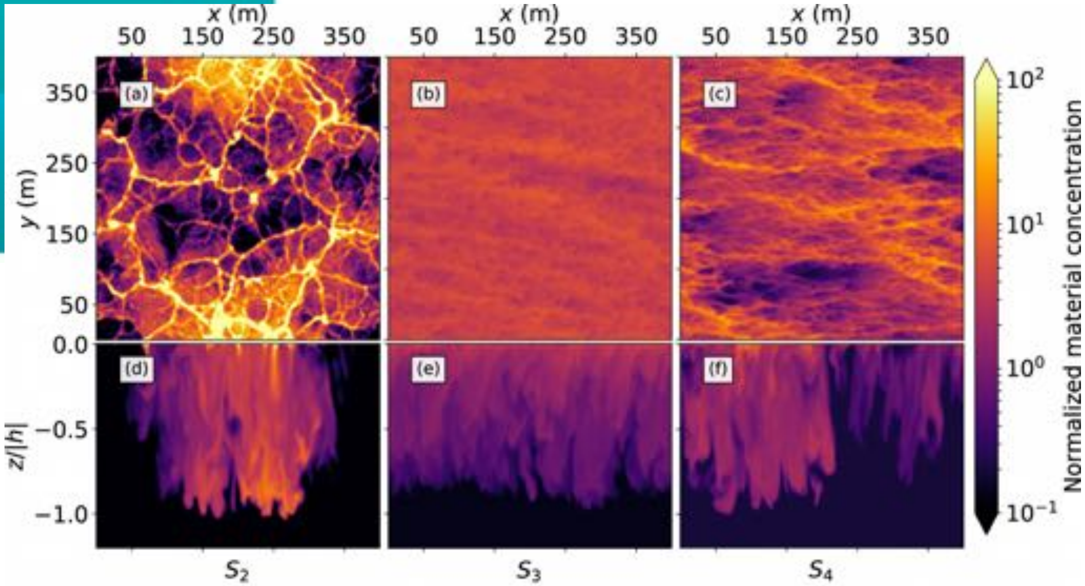
Bridge info. source: https://www.ncdot.gov/initiatives-policies/Transportation/bridges/Pages/default.aspx

Tomas Chor
UCLA

- Simulations of ocean and atmosphere
- Fluid dynamics simulations
  - How do scalars (oil, plankton) get transported?
  - How can we parameterize it on global models?

- Fortran code to generate data
- Python code to post-process it

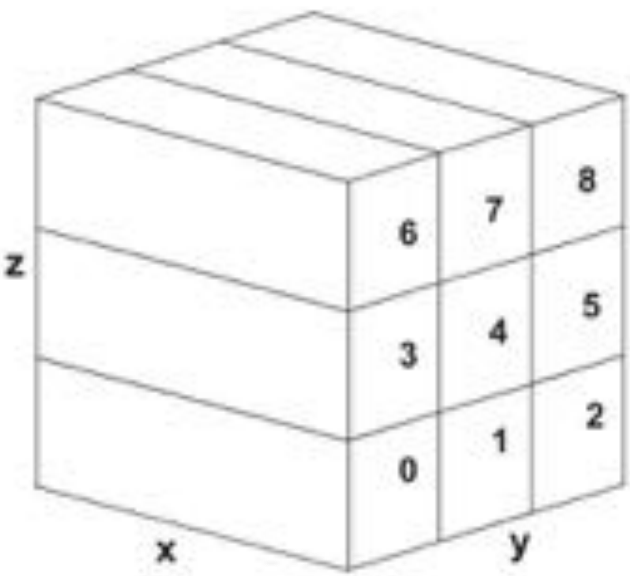




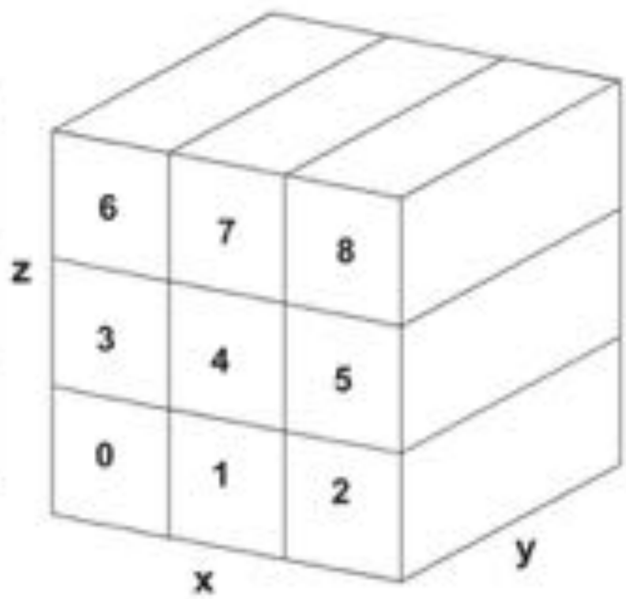
Tomas Chor
UCLA

- Improve post-processing of large datasets (Python)
- Effectively parallelize the fluid dynamics simulation code (Fortran)

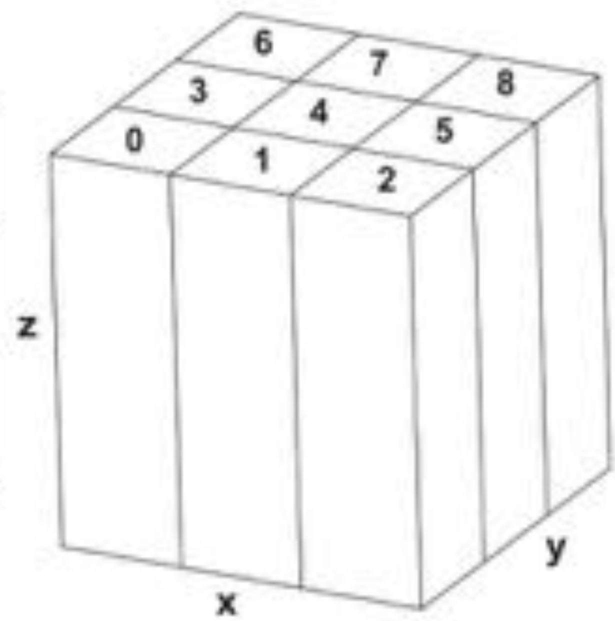
X-derivative



Y-derivative

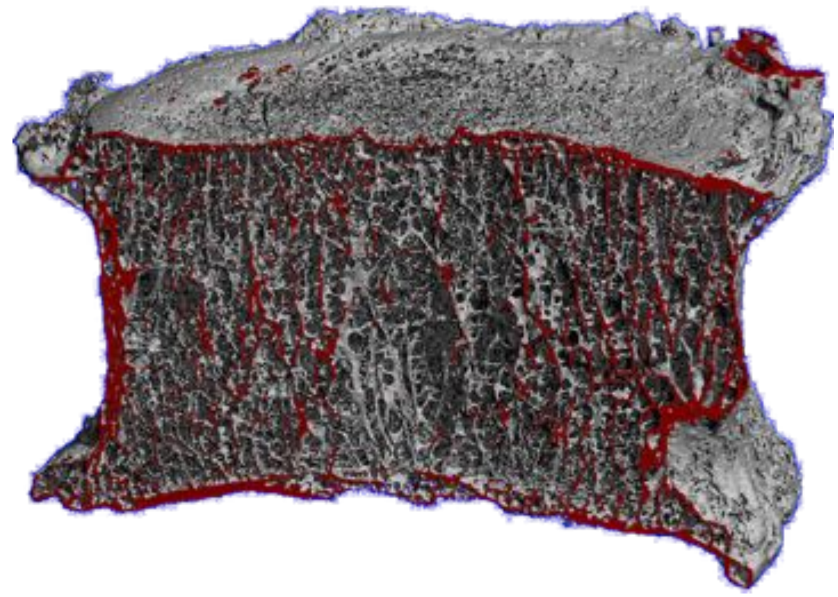


Z-derivative



Tongge Wu
UC Berkeley

- Bone Fracture
- Finite Element Method
- Degree of Freedom (10M~1B)
- TACC – Stampede2
- Customized Customized Solver
- Parallel post-processing



Micro-CT reconstruction of human vertebral body[1]

$$\int_R \xi \cdot \rho \mathbf{a} \mathrm{d}v + \int_R \frac{\partial \xi}{\partial \mathbf{x}} \cdot \mathbf{T} \mathrm{d}v = \int_R \xi \cdot \rho \mathbf{b} \mathrm{d}v + \int_\Gamma \xi \cdot \mathbf{t} \mathrm{d}a$$

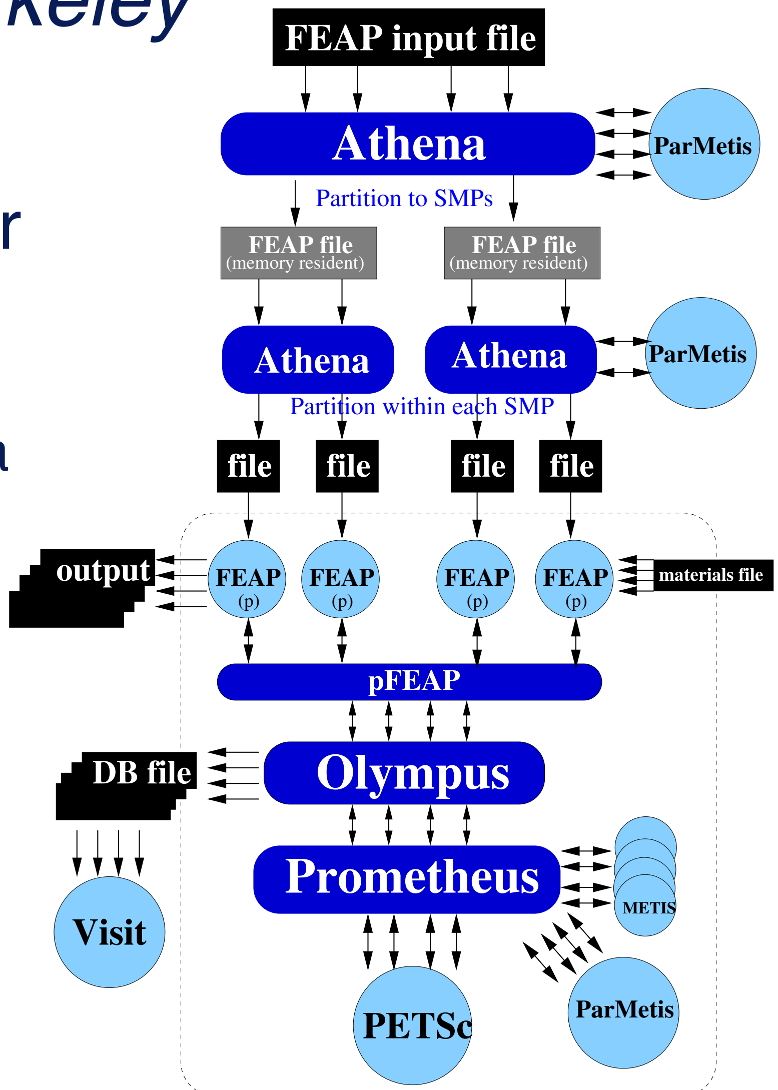
$$A\mathbf{u} = \mathbf{b} \quad \text{or} \quad R(\mathbf{u}) = \mathbf{0}$$

[1] Adopted from: Scanco Medical.

Tongge Wu

UC Berkeley

- Maintain the "fragile" solver
- Post-processing
  - Matlab / Python / R / Julia
- Visualization
  - Paraview / VisIt



Parallel finite element application[1]

[1] M. Adams+ 2004, SC2004.

Vania Wang
UC Santa Barbara

RESEARCH INTERESTS

- Slum and urban health
- Spatial ontologies
- Social network analysis
- Exponential random graph models
- Dynamic modeling of HIV transmission

CURRENT PROJECTS

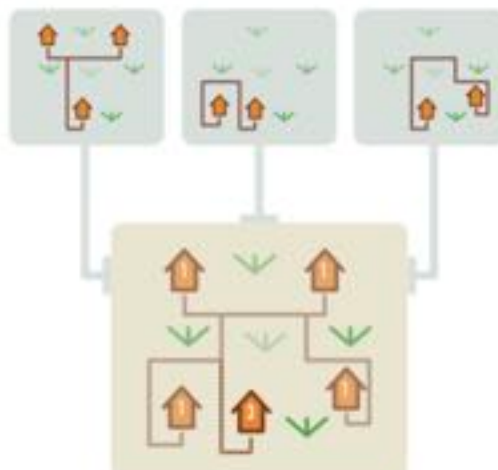
- Developing mapping methods that utilize volunteered geographic information and agent-based stochastic simulation to map dynamic urban spaces in least developed countries
- Exploring bridging as a potential for sustaining HIV epidemics by modeling sexual ties between men who have sex with men (MSM) and women in Seattle, Washington State
- Population size estimation of men who have sex with men in Kunming City, Yunnan Province, China, by using online respondent driven sampling methods

Vania Wang

UC Santa Barbara

Unplanned settlements in least developed countries experience frequent structural change caused by migration and adverse weather events affecting travel paths.

Merged ***travel trajectories*** from volunteered geographic information (**VGI**) can be leveraged to create maps that show frequented locations and paths in unmapped spaces.



Using interdisciplinary methods from geography, public health, and mathematical modeling, ***WeMap*** aims to address these challenges by mapping pathways within unplanned villages in Malawi using VGI.

NETWORK SCIENCE IGERT supported by the NSF

NIH Fogarty

Wenbin Luo

St. Mary's University of San Antonio

Research topic and areas of interest:

- Data structures and algorithms; digital image processing; data mining; computer security;

- Interested in exploring deep learning;

My role related to computation at my institution:

- I am the Linux system administrator for the HPC at St. Mary's University of San Antonio;

Wenbin Luo

St. Mary's University of San Antonio

- **Linux cluster of 156 CPU Cores** running **Red Hat Enterprise Linux Server** with **Platform Cluster Manager**;

- **Lambda Quad from Lambda Labs**:
  - **Operating System**: Ubuntu 18.04 + Lambda Stack
  - **Software**: TensorFlow, PyTorch, Caffe, Caffe 2, Keras, CUDA, cuDNN
  - **Processor**: Intel Core i9-9980XE (18 Cores, 3.00 GHz)
  - **GPUs**: 4x NVIDIA Titan V
  - **Memory**: 256 GB

Computational Challenges:

- Support students and faculty learning and research in HPC and deep learning;
- Be aware of the state-of-the-art in the field;

Susie Zhao, Ph.D. *Email: Xiang.zhao@aamu.edu*

Alabama A&M University

Research areas/projects:

Numerical modeling and simulation, data mining, scientific visualization

Roles:

Professor in Computer Science,

Director of Laboratory of Advanced Modeling & Simulations,

Principle Investigator

Problems & Algorithms:

Multi-phase flows, fluid-structure interaction, CNT-enhanced materials

Grid adaptation

Status & Experience:

- Already doing supercomputing;my codes are serial and parallel together
- HPC and data analysis are very critical to the goals of my projects
- Developed my own software and also use community codes
- Use databases for project management
- Data generated from simulation are large and well formatted

Susie Zhao, Ph.D. *Email: xiang.zhao@aamu.edu*

Alabama A&M University

Computational Challenges:

- Communication overhead
- Dynamic load balance
- Data visualization and analysis

My Plan & Expectation:

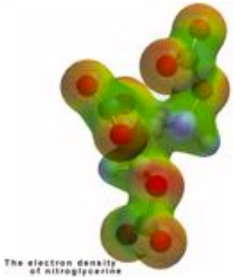
- Understand the latest HPC advancement and available tools
- Gain hands-on HPC skills using the resources at SDSC
- Network with the professionals in HPC community
- Identify new ideas for grant proposals and/or curriculum enhancement

Yanqing Su, Postdoctoral Scholar

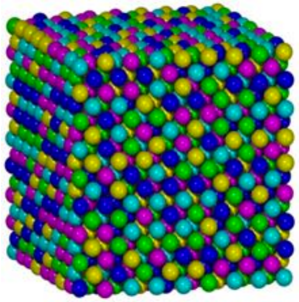
Department of Mechanical Engineering

University of California, Santa Barbara

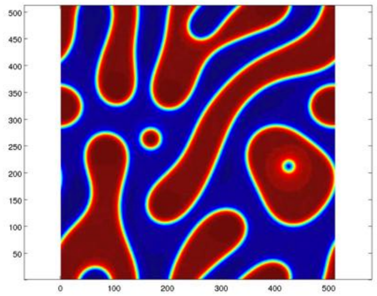
My research centers around dynamics of defects in metallic materials at the nano/micro scales



- Density functional theory (subatomic scale)
  - Theory: A series of approximated Schrödinger equations
  - An MPI-equipped academic code:Vienna Ab initio Simulation Package (VASP)



- Atomistic simulations (atomic/submicro scales)
  - Theory: Newton's second law with interatomic interactions
  - An MPI-equipped open source code: Large-scale Atomic/ Molecular Massively Parallel Simulator (LAMMPS)



- Phase-field method (submicro/micro scales)
  - Theory: Diffusion equation with energy functional of "phases"
  - An in-house phase-field dislocation modeling code, written in C and parallelized with MPI

Yanqing Su, Postdoctoral Scholar

Department of Mechanical Engineering

University of California, Santa Barbara

Challenges in my research

- Density functional theory
  - Large memory usage and low parallel efficiency with large # of cores
- Phase-field method
  - Currently, the simulation cell is divided into slices along one direction
  - Full 3D parallelization will be more efficient

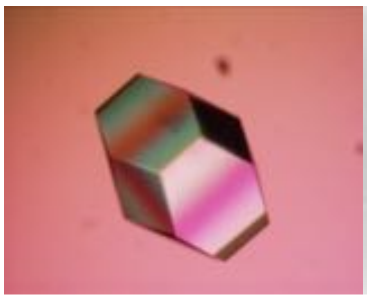
Learning goals at the Summer Institute

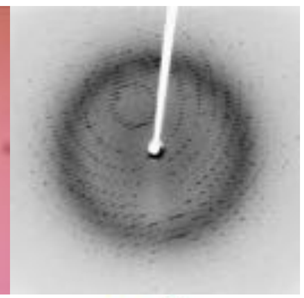
- Learn the computational tools, especially VisIt, GPU, and OpenMP
- Be a better programmer and improve the efficiency of the in-house phase-field modeling code

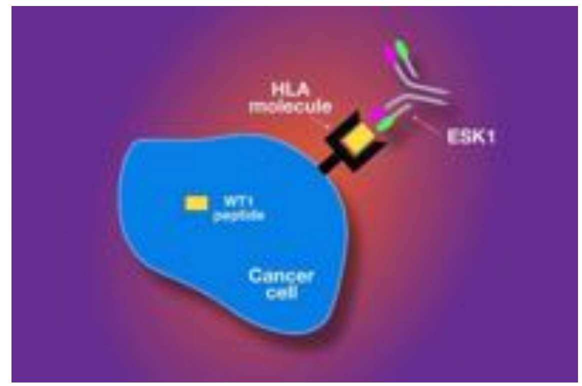
Ye Zou

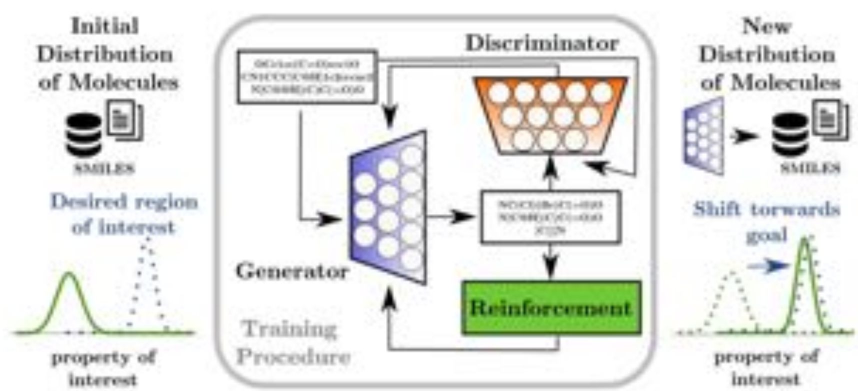
Kansas State University

- Structure based pharmacology for cancer and immunology





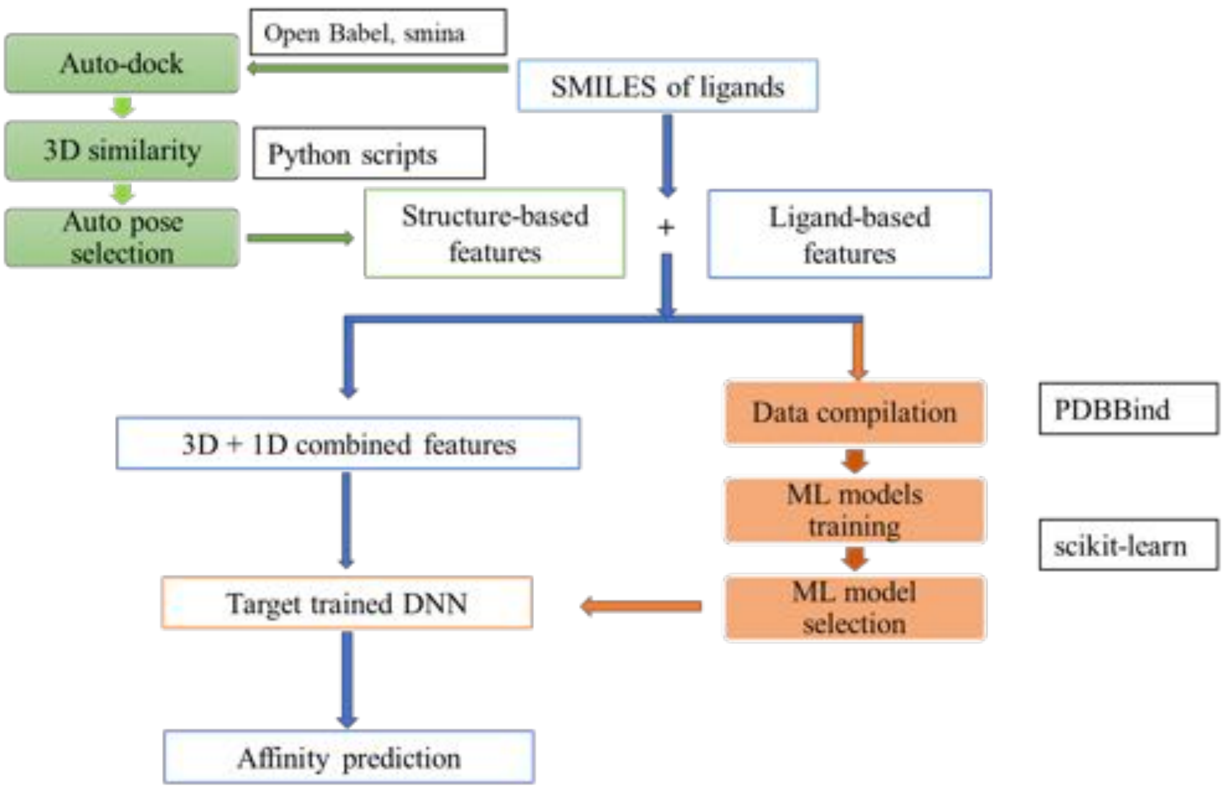




Ye Zou

Kansas State University

- Applications of Machine learning to chemistry and drug design and increase the accuracy.

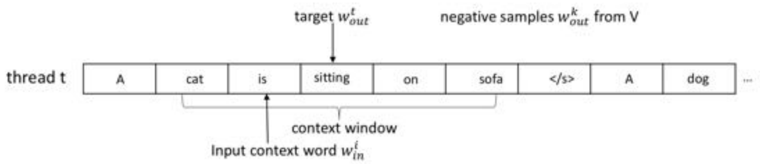


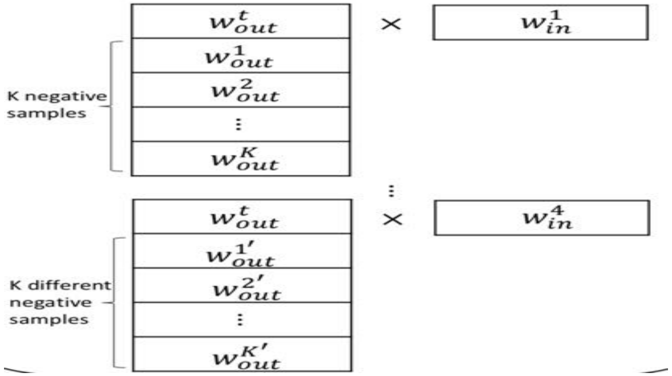
Yohn Jairo Parra, PhD
Florida A&M University

Research Area

- ❖ **Unstructured Behavior Data: An Examination of Lexicon-Based and Deep Learning Approaches**
    - **Algorithms: Word2vec, Glove, Fast text**
    - **Problem: Using deep learning techniques to construct prediction models with sentiment and context to address ambiguity concerns in financial complaint data.**

- ❖ **Libraries: TensorFlow, keras**





Word2vec(dot product)

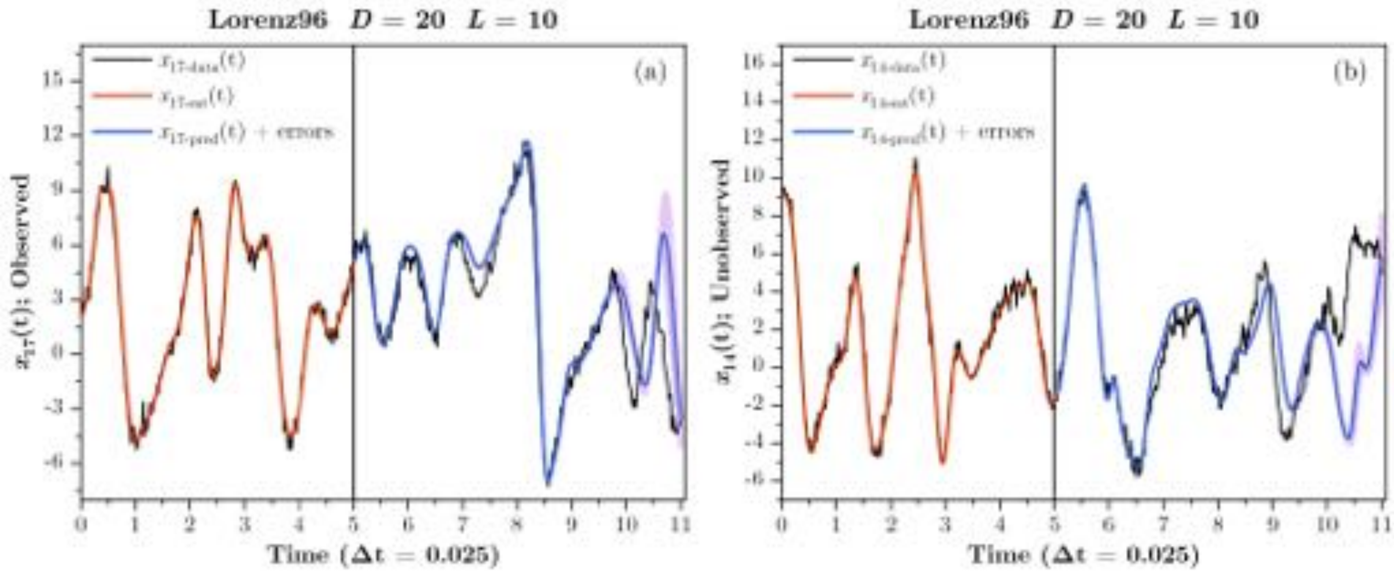
Yohn Jairo Parra, PhD
Florida A&M University

Computational Challenges:

1. Scaling Word2Vec on GPUs
2. More Understanding of different parallelization methods (Amazon SageMaker Blazing Text)
3. Make Reproducible the experiments using docker images, containers, workflows

Zheng Fang

- Research: theory and algorithms for nonlinear dynamics and deep learning using ideas from statistical physics.
- Computational tasks: non-convex optimization; (Hamiltonian) Monte Carlo.
- Problem size: from 10^4 to 10^7 (degrees of freedom).



Zheng Fang

UC San Diego

- Computational challenge: to perform high-dimensional integrals of the form

$$\langle G(\mathbf{X})\rangle = E\left[G(\mathbf{X})\,|\,\mathbf{Y}\right] = \frac{\int d\mathbf{X}\, G(\mathbf{X})\, e^{-A(\mathbf{X})}}{\int d\mathbf{X}\, e^{-A(\mathbf{X})}},$$

- Parallelization is needed for Hamiltonian Monte Carlo calculations.
- Would like to learn more about general HPC and GPU programming at the Summer Institute.