



# BLUE WATERS

**SUSTAINED PETASCALE IN ACTION:**  
ENABLING TRANSFORMATIVE RESEARCH

**2019** ANNUAL REPORT



[illegible]

2019 ANNUAL REPORT

**Creative Director**  
Steve Duensing

**Graphic Designer**  
Megan Janeski

**Project Director**  
William Kramer

*The research highlighted in this book is part of the Blue Waters sustained-petascale computing project, which is supported by the National Science Foundation (awards OCI-0725070 and ACI-1238993) and the state of Illinois. Blue Waters is a joint effort of the University of Illinois at Urbana–Champaign and its National Center for Supercomputing Applications.*

Visit <https://bluewaters.ncsa.illinois.edu/science-teams> for the latest on Blue Waters-enabled science and to watch the 2019 Blue Waters Symposium presentations.



CLASSIFICATION KEY

To provide an overview of how science teams are using Blue Waters, researchers were asked if their work fit any of the following classifications (number responding in parentheses):

DI	Data-intensive: Uses large numbers of files, large disk space/ bandwidth, or automated workflows/offsite transfers, etc. (83)
GA	GPU-accelerated: Written to run faster on XK nodes than on XE nodes (55)
TN	Thousand-node: Scales to at least 1,000 nodes for production science (74)
MI	Memory-intensive: Uses at least 50 percent of available memory on 1,000-node run (21)
BW	Only on Blue Waters: Research only possible on Blue Waters (43)
MP	Multiphysics/multiscale: Job spans multiple length/time scales or physical/chemical processes (59)
ML	Machine learning: Employs deep learning or other techniques; includes “big data” (18)
CI	Communication-intensive: Requires high-bandwidth/low-latency interconnect for frequent tightly coupled messaging (33)
IA	Industrial application: Researcher has private sector collaborators or results directly applicable to industry (10)
FS	Frontier science: This science has produced a first-of-its-kind outcome (21)
BI	Big idea: This research relates to the National Science Foundation’s “10 Big Ideas” initiative (83)

TABLE OF CONTENTS

3	NCSA DIRECTOR BILL GROPP – LEADING BY EXAMPLE
4	PROJECT DIRECTOR BILL KRAMER – TOGETHER, WE CHANGED THE WORLD
6	BLUE WATERS THROUGH THE YEARS
10	BLUE WATERS METRIC INFORMATION
12	EXTENDED ABSTRACTS
12	<i>space science</i>
76	<i>geoscience</i>
118	<i>physics &amp; engineering</i>
210	<i>computer science &amp; engineering</i>
236	<i>biology, chemistry &amp; health</i>
318	<i>social science, economics, &amp; humanities</i>
326	<i>mathematical &amp; statistical sciences</i>
330	<i>graduate fellows</i>
372	SCIENCE AND ENGINEERING ADVISORY TEAM COMMITTEE
373	OTHER BLUE WATERS PROJECTS
374	REFERENCES
406	INDEX

KB = kilobytes  
MB = megabytes  
GB = gigabytes  
TB = terabytes  
PB = petabytes  
I/O = input/output  
Knh = thousand node hours  
Mnh = million node hours  
MPI = message-passing interface

Allocations denoted as type/size in extended abstracts.

LEADING BY EXAMPLE



The National Center for Supercomputing Applications (NCSA) was funded in 1986 to enable discoveries not possible anywhere else. We have a long history of doing things that have not been done before. The Blue Waters Project admirably continued that tradition and spawned new avenues of research and discovery for NCSA and our partners.

We are leading the way in multimessenger astrophysics (MMA). In fact, we were recently awarded over \$20 million in grants in MMA areas that would not have been possible without the discoveries Blue Waters helped achieve. MMA combines data from cosmic messengers such as gravitational waves, neutrinos, and electromagnetic waves. We are bringing together large-scale computing, data science, artificial intelligence (AI), machine learning (ML), and software development not only to advance these domains but to develop techniques, technologies and software that may transfer to other domains as well.

We’re helping to marry medicine and technology as a partner with the Carle Illinois College of Medicine, the world’s first engineering-based medical school. And for many years we’ve partnered with Mayo Clinic to advance precision medicine—drugs or treatments designed for small groups rather than large populations—by applying ML and AI.

We are incorporating ML and AI into agriculture as a partner in the University of Illinois’ new Center for Digital Agriculture. With the Institute for Genomic Biology; the College of Agricultural, Consumer and Environmental Sciences; and the Grainger College of Engineering we’ll be facilitating student education in the use of technology in agriculture as well as assisting farmers and agricultural industries to advance digital agriculture to keep pace with the ways technology is transforming how we feed and support a growing global population.

In addition, we recently launched the Center for Artificial Intelligence Innovation at NCSA that will focus on leveraging past and existing research projects to develop software to amplify current AI and ML research areas and explore new ones.

Earlier in 2020 I was privileged to be part of a delegation of University of Illinois leaders that participated in discussions on the next frontier of AI, data science, and design thinking with Illinois alumnus Tom Siebel and his staff at C3.ai. In his remarks, Illinois’ Vice Chancellor for Academic Affairs and Provost Andreas Cangellaris commented that “no matter how powerful the computers are, the real power comes from the people behind them.”

Blue Waters is not just a peak petascale system; it is a true sustained-petascale system, specifically designed to be well balanced and deliver petascale computing power and productivity across a variety of applications. Because of this balance, for almost seven years the Blue Waters Project has enabled remarkable work in biology, chemistry, physics, geosciences, cosmology and astrophysics, atmospheric science, and many other fields such as agriculture, economics, political science, and the social sciences.

You’ll find these remarkable discoveries in the pages that follow. Discoveries that were made by the people behind the computer—the extraordinary NCSA staff and the scientists they support in their quest for frontier science. They are leading by example. And as the leader of NCSA, I would not have it any other way.

Dr. William “Bill” Gropp  
Director  
National Center for  
Supercomputing Applications



# TOGETHER, WE CHANGED THE WORLD



I take great pride in the Blue Waters Project and in the outstanding science and project teams that make Blue Waters an exceptional open resource for research.

Every year I write how Blue Waters enabled computational, data analysis, and now machine learning/artificial intelligence investigations that could not be done otherwise. Once again, as you go through this report you will

see the many accomplishments that carry the badge indicating “Only on Blue Waters” to signify that the research would not have been possible on other open-science resources deployed at the time.

Through the years, increasing numbers of projects have been identified as involving “data intensive” processing or as incorporating machine learning/artificial intelligence. This shift is another indication of the leadership of Blue Waters and the impact it is making. Many projects do not just involve data-intensive processing or machine learning but instead are combining these methods with multiscale and multiphysics simulation to achieve cutting-edge results. As I have said before, Blue Waters is an early catalyst for *convergence* because of its exceptional balance of computing and I/O capabilities, along with thoroughly supporting teams using different methods while also helping science teams integrate different methods within their work to achieve remarkable new results.

As a leadership-class compute system, Blue Waters is intended to help pioneer amazing science and engineering discoveries. We designed it to be a balanced system in which its computational, interconnect, external network, and storage system work together to deliver unprecedented petascale and highly usable computing capabilities to the national research community. Blue Waters was conceived to accelerate the science frontier and to rapidly expand the adoption of new methods into other best-of-breed and “frontier” uses. This supercomputer boot-strapping process is how the world moved from the first 2D atmospheric circulation model to being able to accurately pre-

dict weather 10 to 15 days in advance, from studying basic fluid dynamics to using computing for complete aircraft and ship design, and for moving from studying small molecules to understanding the basic principles of life.

One of the most unexpected and remarkable successes of Blue Waters in 2019 was the role it played in a gerrymandering decision reached by Ohio’s Sixth District Court, where University of Illinois at Urbana–Champaign’s Wendy Cho presented an analysis conducted on Blue Waters and based on more than three million simulations of possible electoral maps for Ohio. Cho used a novel evolutionary Markov chain Monte Carlo algorithm designed to take advantage of Blue Waters’ massively parallel architecture to produce an array of map scenarios that could be compared to the current map. Further, she served as an expert witness and presented results from her simulations on Blue Waters to objectively compare biased and unbiased redistricting.

Making high-performance computing (HPC) available to more of our scientific community is important to the Blue Waters Project team. To expand participation in the use of the Blue Waters resources and services, a new category of Broadening Participation allocations was announced in 2017, a component of the Innovation and Exploration allocations managed by NCSA. Through this, over 3.7-million node-hours of computing resources were awarded to 21 research teams. Included among the principal investigators (PIs) were 10 females and two underrepresented minorities. In addition, there were four female and eight underrepresented minority colleagues listed as co-PIs. Among the lead institutions, five are Minority Serving Institutions and nine are within jurisdictions of the National Science Foundation’s (NSF) Office of Integrative Activities’ Established Program to Stimulate Competitive Research (EPSCoR).

Our Blue Waters training program has served over 12,000 people within the current and future HPC workforce. Our highly competitive Blue Waters Graduate Fellowship program has helped 50 young graduate student researchers, and the Blue Waters Internship Program has enabled more than 150 undergraduates to gain in-depth experience on advanced research methods to take full advantage of leadership-class computing.

We also inspire future generations to become interested in science and technology and educate the public on how HPC facilitates scientific discovery. By offering guided tours of the National Petascale Computing Facility as well as open houses, more than 12,000 visitors have come to see Blue Waters in action, many of them K–12 and college students.

By sharing the many best practices we developed, our staff is distributing the expertise and experiences we have acquired from our accumulated knowledge. More importantly, our Blue Waters science teams have published more than 1,000 articles and papers in their respective domains.

By being conscientious stewards of the grant monies awarded to us, we were able to operate the Blue Waters supercomputer an additional 18 months beyond original plans, providing nearly seven years of access to Blue Waters instead of the five years originally planned by NSF. We were also able to add additional programs not in our original project scope, such as the aforementioned Broadening Participation allocations, as well as the Blue Waters Graduate Fellowships, our Petascale Application Improvement Discovery (PAID) program, and our Virtual School.

Continued service and operations only make sense if a system is reliable and resilient. Blue Waters is the most monitored and studied system in the world, with dozens of publications about its reliability and resiliency. The original contractual requirements for our vendor was that there would be no more than 150 node failures per month, and no more than six systemwide outages per month. In reality, for the past five years we averaged only one or two node outages per day, and the mean time between system interruption ranges from three to six months.

Another way to explain this is that Blue Waters has 49,508 AMD CPU module sockets. We recently went back and reevaluated all the AMD CPU modules that had been taken out of the system, many as a preemptive measure. The analysis found there were 154 modules, or only 0.3 percent, that were defective. When equated to Mean Time Between Failure (MTBF) over the six-and-a-half years, this indicates an MTBF of over 2,000 years for a CPU module.

Talented, dedicated individuals at NCSA and our partner institutions have maintained the system, supported our science and engineering partners in

achieving breakthrough research, improved tools and processes, and trained the next generation of computational researchers. This publication celebrates the accomplishments of our partners and the researchers, educators, and students they support but also celebrates the dedication and innovation of the Blue Waters staff.

The end of the NSF funding for Blue Waters’ “broad access” phase is not the end of the story. Blue Waters will continue providing groundbreaking insight as it transitions to being a high-performance computational and data resource for the National Geospatial-Intelligence Agency (NGA). Blue Waters will focus ninety percent of the system resources on NGA-related projects, with ten percent of Blue Waters reserved for use by researchers at the University of Illinois at Urbana–Champaign. As Illinois’ Vice Chancellor for Academic Affairs and Provost Andreas Cangellaris commented, “no matter how powerful the computers are, the real power comes from the people behind them.”

NGA will be building upon the amazingly impactful ArcticDEM research of Paul Morin of the University of Minnesota, which mapped the Arctic Circle, and the REMA research of Ian Howett of The Ohio State University, which mapped Antarctica. NGA will be funding the operations of Blue Waters to make high-resolution topographic maps of the entire Earth and other related projects associated with its mission to “map and model the world.” Who would have dreamed that was possible when Blue Waters entered production in 2013?

I am so thankful to have been able to work with everyone on the Blue Waters Project these past years. Together, we changed the world. And we’ll keep changing it. I cannot wait to see what discoveries we will uncover!

**Dr. William T.C. Kramer**  
Director and Principal Investigator  
Blue Waters Project



# BLUE WATERS THROUGH THE YEARS

- AUGUST 2007**  
Deployment Award funded for \$208 million for the development, acquisition and deployment of the first sustained-petascale system
- JUNE 2010**  
The National Petascale Computing Facility completed
- DECEMBER 2011**  
Initial Cray equipment arrives at NPCF
- DECEMBER 2012**  
System deployment completed and accepted with NSF concurrence
- APRIL 2, 2013**  
Regular full-service production start date
- JULY 2013**  
Expansion from 276 to 288 cabinets
- OCTOBER 2013**  
Operations and maintenance Award Funded for \$134 million that was planned for 4.5 years but in reality spanned 7 years
- DECEMBER 19, 2019**  
Blue Waters ends its general-access role as NSF Track-1 leadership supercomputer

The Blue Waters supercomputer began full service operations at the National Center for Supercomputing Applications in April 2013. Funded by the National Science Foundation, the University of Illinois and the State of Illinois, the primary goal of Blue Waters is to enable productive frontier research with a well-balanced leadership class computational and data analysis system that can process vast amounts of data in a precise, time-efficient manner. Funds for operation and management included a \$134 million dollar grant awarded by the NSF to cover 7 years of Blue Waters’ service.

Blue Waters ended its role as an NSF Track-1 leadership supercomputer in 2019 but continues to serve as the world’s largest, non-classified geospatial system. The impressive computing capacity of Blue Waters has afforded research, development, and educational opportunities for communities locally and globally.

Throughout its life, teams have used Blue Waters for an impressive number of science discoveries. These discoveries are then shared with the world when teams publish their results. To date, over 1,300 papers have been presented at conferences and published in 244 journals. This number continues to grow as teams further analyze the vast amounts of data generated on Blue Waters.

### THE SYSTEM


Blue Waters is currently the largest system ever built and delivered by supercomputer manufacturer Cray. It comprises 49,500 processor modules in 288 liquid-cooled cabinets, containing both AMD Interlagos CPUs and NVIDIA Kepler (K20x) GPUs.

With all its parts Blue Waters weighs about 576,000 pounds, or more than 288 tons. The size was necessary for the ambitious computing goals of the system.


Blue Waters was the first sustained petascale system, with more than 1.3 quadrillion (1.3 x 10<sup>15</sup>) sustained calculations per second for a very wide range of large scale applications. Operating at peak performance, the supercomputer is capable of reaching 10 times its sustained computing power, reaching up to 13.34 petaflops (13.34 x 10<sup>15</sup>).


Solving scientific problems requires prolonged computation, making the primary focus of Blue Waters sustained petascale performance. Thanks to the work of contributors to the Blue Waters project, researchers can get many science applications to run at a sustained petaflop. Running this many calculations per second opens the door for scientists and engineers to gain new insights into everything from the smallest biological processes to the largest cosmic events.

Storing this data requires a global addressable memory as impressive as its computing ability. Blue Waters has 1.66 petabytes (1,660,000 GB) of memory and has 26 usable petabytes of data storage provided by the Cray Sonexion storage appliances.



**288 TONS**  
IS ABOUT AS MUCH AS  
**22 SCHOOL BUSES**

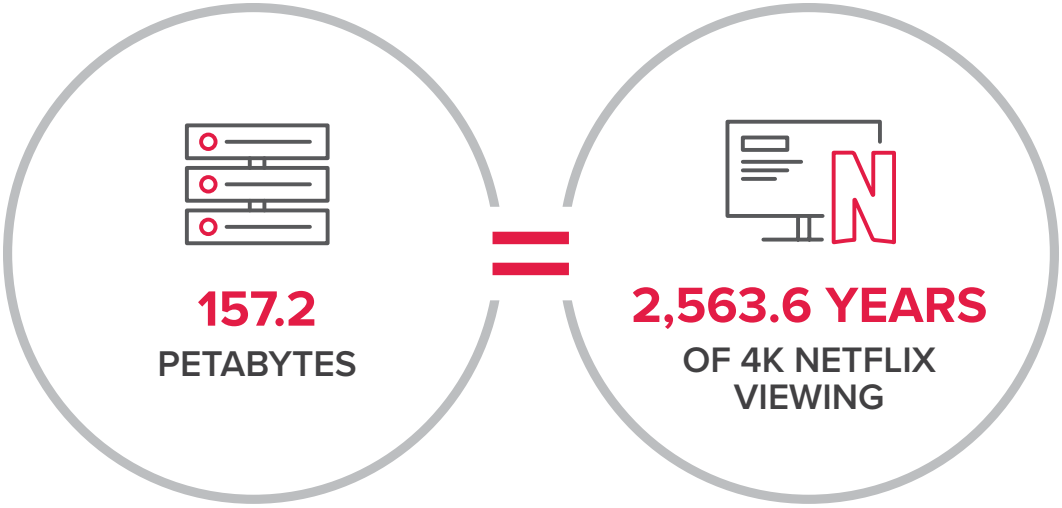




**IT WOULD TAKE A PERSON**  
**423 MILLION YEARS**

**TO DO WHAT BLUE WATERS CAN DO**  
**IN ONE SECOND**





For long periods of storage, a nearline tape environment was built. At its peak, the Blue Waters nearline system stored over 35 petabytes of user data on its 11,600 miles of tape, which would stretch about half-way around the earth's equator.

From July 2012 through March 2020 Blue Waters transferred 157.2 petabytes of data to and from its file systems from external sources.

THE BUILDING

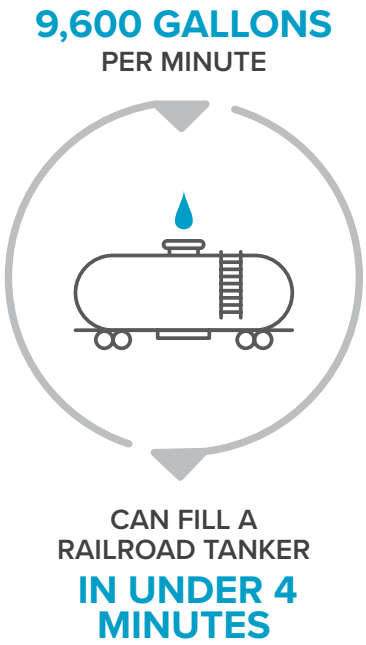
Safely housing and supporting this massive computing system required the construction of the National Petascale Computing Facility. This state of the art facility spans 88,000 square feet, with the machine room housing Blue Waters taking up about 30,000 square feet of six foot raised floor.

NPCF provides 24MW of electricity for computing and data equipment. In order to keep the supercomputer from overheating, NPCF has a water-cooling system capable of cycling at a peak flow rate of 9,600 gallons per minute.

In cool months, about 40-50 percent of the year, the facility uses its own evaporative cooling towers so almost no energy is used for water cooling. This system helped enable the NPCF to become one of the first computing facilities to achieve a Leadership in Energy and Environmental Design (LEED) Gold certification.

OUTREACH

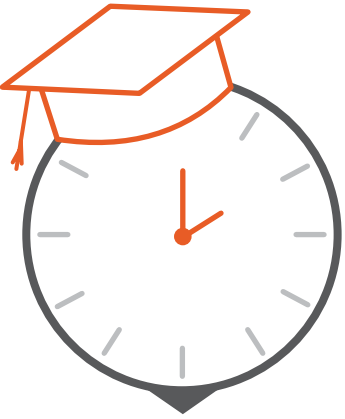
Having exposure to a supercomputer like Blue Waters is not just a boon for research, but also for education. The project has given students first-hand experience with Blue Waters through the creation of 139 undergraduate internship opportunities as well as 50 graduate fellowships. But student involvement does not end there. Almost half of the Blue Waters users are graduate students, and they have been a fundamental part of the science teams. In fact, the average allocation



time used per graduate student was 388,663 node hours. In addition to students being an active part of the research on Blue Waters, over its years of service 12.6 million node hours have been made available for educational projects.

Beyond this direct access to the machine, graduate students and postdocs had access to numerous experiences and training that contributed to their research projects through their Blue Waters experience. These projects serve to help develop a broader community of petascale computing experts. Reaching a community further than just those with accounts on Blue Waters has been important as well. The webinars/seminars created have had over 50,000 viewers and the training sessions and workshops created have had over 10,000 viewers. These are preserved online and will continue to serve the community into the future.

Some of these educational efforts included Virtual School of Computational Science and Engineering summer school events, virtual courses available at 27 institutions—including minority-serving institutions (MSI) and those in the Established Program to Stimulate Competitive Research (EPSCoR) program—and the International High Performance Computing Summer School. These programs gave well over six thousand students the opportunity to learn about petascale computing.



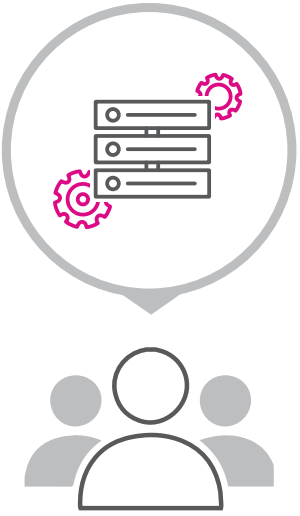
12.6 MILLION NODE-HOURS FOR EDUCATION

Exposure to the supercomputer is not just limited to those wishing to conduct research or work in the field of supercomputing. Over the years of operation over 14,000 people have toured the NPCF to see the Blue Waters supercomputer in person. These tours let people see for themselves how far computing technology has come and learn about the great research advances being made in our community.

WHAT'S NEXT

Although Blue Waters has ended its role as an NSF Track-1 supercomputer it is still serving to help make tremendous gains in our knowledge of the world. The Blue Waters Project, National Geospatial-Intelligence Agency, University of Minnesota and The Ohio State University are collaborating on the production of a digital elevation model (DEM) of the entire Earth.

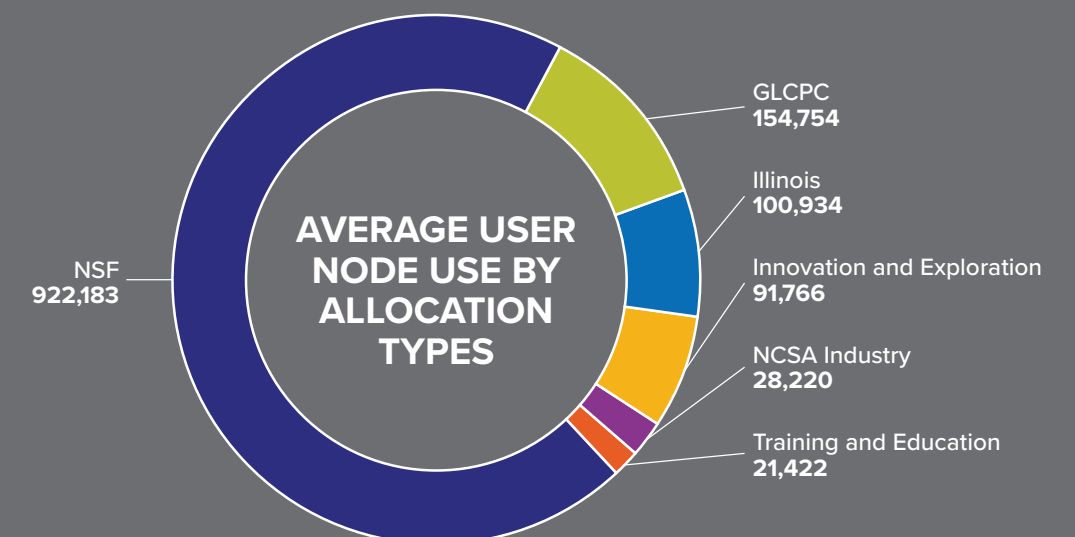
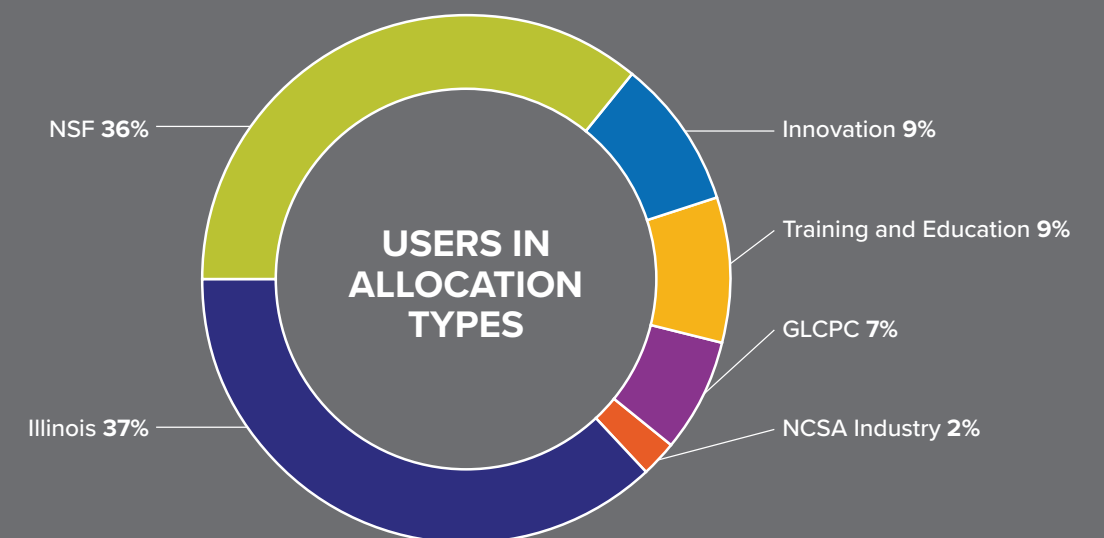
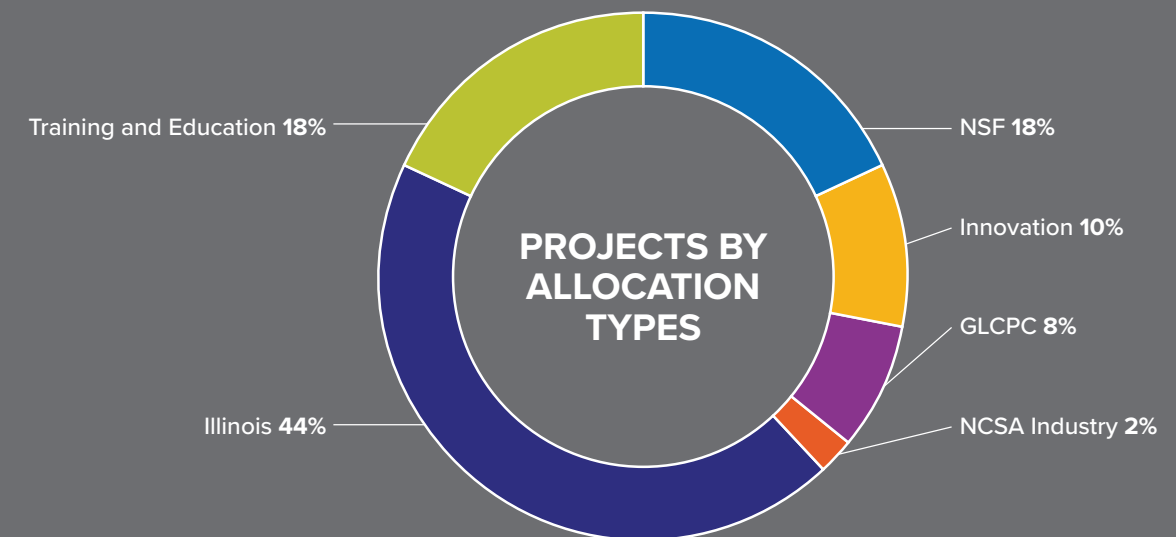
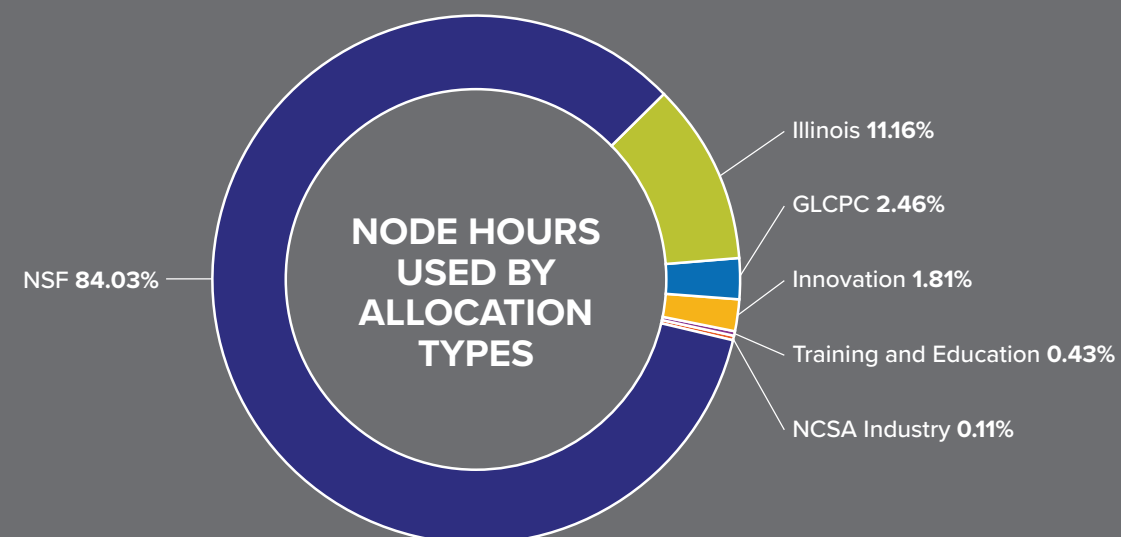
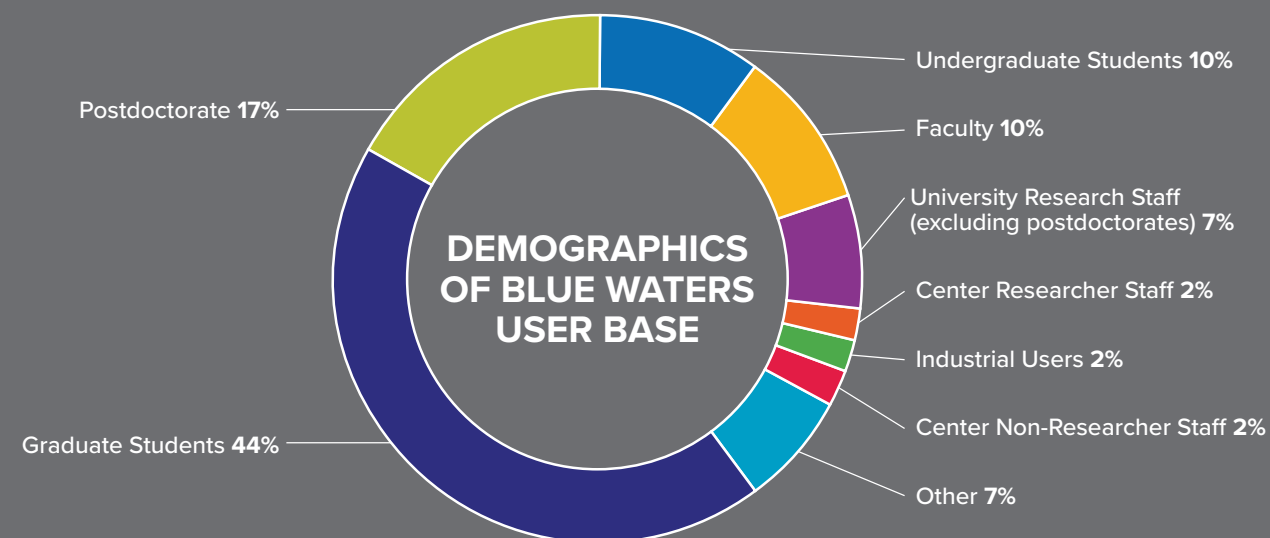
EarthDEM follows the successful ArcticDEM and REMA projects which helped researchers in tracking geological changes, such as ice cap collapse, in the Arctic and Antarctic. Having publicly accessible, high-resolution topography of the entire world gives researchers the ability to assess landscape changes, no matter how remote the area.



14,000 VISITORS



# BLUE WATERS METRIC INFORMATION





# SPACE SCIENCE

## AEROSPACE ENGINEERING

## ASTRONOMY

## ASTROPHYSICS

## COSMOLOGY

## HELIOPHYSICS

## NUCLEAR PHYSICS

14	<i>Simulating Two-Fluid MHD Turbulence and Dynamos in Star-Forming Molecular Clouds and a New Paradigm for Computational Astrophysics for Spherical Systems</i>	42	<i>Data- and Compute-Intensive Challenges for Observational Astronomy in the Great Survey Era</i>
16	<i>The Computational Keys to the Supernova Puzzle: How Multiple 3D Radiation/Hydrodynamic Models Can Unlock the Supernova Mystery</i>	44	<i>Modal Decompositions of Shock Interactions</i>
18	<i>Shedding Light on Supermassive Binary Black Hole Mergers</i>	46	<i>Plume Plasma Spacecraft Interactions</i>
20	<i>Accretion Dynamics of Supermassive Black Hole Binaries</i>	48	<i>The Spreading of Three-Dimensional Magnetic Reconnection in Asymmetric Geometry</i>
22	<i>Achieving Probabilistic Classification of Cosmic Web Particles Using Rapidly Generated Training Data: A Method for Classifying Galaxies into Their Cosmic Web Structural Groups Using Supervised Machine Learning</i>	50	<i>Assembling a Map of the Universe: Shapes and Mass Distribution for the Dark Energy Survey</i>
24	<i>The Epoch of the First Luminous Black Holes: Evolving the BlueTides Simulation into the First Billion Years of Cosmic History</i>	52	<i>Petascale Simulations of Binary Neutron Star Mergers</i>
26	<i>Advancing First-Principle Symmetry-Guided Nuclear Modeling for Studies of Nucleosynthesis and Fundamental Symmetries in Nature</i>	54	<i>Development of a Scalable Gravity Solver for Enzo-E</i>
28	<i>Magnetized Models of Giant Impacts</i>	56	<i>Simulating Galaxy Formation Across Cosmic Time</i>
30	<i>Cosmic Reionization on Computers</i>	58	<i>Processing Dark Energy Camera Data to Make the World's Best Map of the Night Sky</i>
32	<i>Elucidating the Alignment Mechanism for Black Hole Accretion Disks Subjected to Lense-Thirring Torques</i>	60	<i>Coupling the Solar Wind and Local Interstellar Medium in the Era of the New Horizons, Interstellar Boundary Explorer, Parker Solar Probe, Ulysses, and Voyager Spacecraft</i>
34	<i>Understanding the Origins of the Stars and Galaxies in our Universe</i>	62	<i>Interior Dynamics of Young Stars Revealed by 3D Hydrodynamic Simulations</i>
36	<i>Deep Learning at Scale for the Construction of Galaxy Catalogs with the Dark Energy Survey</i>	64	<i>Modeling of Galaxy Populations</i>
38	<i>Characterization of Numerical Relativity Waveforms of Eccentric Binary Black Hole Mergers</i>	66	<i>Effects of Active Galaxy Feedback on the Intracluster Medium</i>
40	<i>Fusing Numerical Relativity and Deep Learning to Detect Eccentric Binary Black Hole Mergers Using Higher-Order Waveform Multipoles</i>	68	<i>Gravitational and Electromagnetic Signatures from Binary Black Hole-Neutron Star Mergers: A Jet Engine for Short Gamma-Ray Bursts</i>
		70	<i>Feeding Black Holes: Tilt with a Twist</i>
		72	<i>Scaling the BATS-R-US MHD Model to Over 100,000 Cores with Efficient Hybrid OpenMP and MPI Parallelization</i>
		74	<i>Numerical Study on the Fragmentation Condition in a Primordial Accretion Disk</i>
		75	<i>Merging Black Holes and Neutron Stars</i>



DI

TN

BW

MP

CI

# SIMULATING TWO-FLUID MHD TURBULENCE AND DYNAMOS IN STAR-FORMING MOLECULAR CLOUDS AND A NEW PARADIGM FOR COMPUTATIONAL ASTROPHYSICS FOR SPHERICAL SYSTEMS

**Allocation:** NSF PRAC/2,700 Knh  
**PI:** Dinshaw S. Balsara<sup>1</sup>  
**Co-PI:** Alexandre Lazarian<sup>2</sup>  
**Collaborator:** Blakesley Burkhart<sup>3</sup>

<sup>1</sup>University of Notre Dame  
<sup>2</sup>University of Wisconsin–Madison  
<sup>3</sup>Flatiron Institute

## EXECUTIVE SUMMARY

We are at the threshold of a new data-rich and simulation-rich era in star-formation studies. The question of how stars form is fascinating in and of itself and has a great impact on several other areas of astrophysics. The consensus is that a predominant amount of star formation in our galaxy takes place in molecular clouds, and more specifically in giant molecular clouds (GMC). The project aims to study magnetic field evolution in partially ionized plasmas as well as developing and applying a new paradigm for simulating magnetohydrodynamics (MHD) on geodesic meshes that cover the sphere with no coordinate singularities and no loss of accuracy at the poles.

## RESEARCH CHALLENGE

Our understanding of the star-formation process has reached the point where advanced observational capabilities are required. Consequently, NASA has made multimillion-dollar investments in the High-resolution Airborne Wideband Camera Plus (HAWC+) instrument aboard the Stratospheric Observatory for Infrared Astronomy airborne observatory with the specific goal of understanding the turbulent nature of star-forming clouds. At the same time, high-resolution simulations that include the appropriate physics of GMCs are also of critical importance. The PIs are theorists who are participating in a multiyear, funded NASA proposal to obtain observational data associated with turbulence in the Perseus GMC.

The PIs have also done the leading simulations of two-fluid MHD simulations on a range of XSEDE and PRAC resources [4,7–10]. At the original resolutions of  $512^3$  and smaller, the team would have been unable to match the observations from HAWC+. With the new generation of simulations, with a  $1,024^3$  zone and upwards resolution, the researchers were able to obtain a well-defined inertial range in two-fluid and dynamo simulations. This ensured that NASA's investment in HAWC+ is being matched by well-resolved simulations.

The most compelling motivation for understanding two-fluid, ambipolar diffusion-mediated turbulence, in fact, comes from recent observations. Differences in the linewidths between neutral and ionized tracers have led to the suggestion that the dissipation

of turbulence from ambipolar diffusion sets in on scales smaller than 0.0018 parsecs in the Messier 17 nebula [2].

A breakthrough realization by Xu and Lazarian [7] claimed that magnetic fields would grow in a partially ionized plasma (known as the dynamo problem) at rates that are very different from the growth in a fully ionized plasma. Their prediction was that while magnetic fields grow exponentially in a fully ionized plasma, they grow only quadratically in a partially ionized plasma. However, confirmation of the theory requires a highly resolved turbulent flow. Fig. 1 (from [10]) shows the evolution of magnetic energy with time in a two-fluid dynamo. The right panel shows magnetic field growth in a partially ionized dynamo. The inset on the left shows the initial growth of magnetic energy, with the red curve representing the research team's highly resolved simulations and the blue curve showing the best fit to a quadratic. The agreement with theory is very good and has been documented. The study of the magnetic energy spectra is also significant and is shown in Fig. 2. The peak in the magnetic energy spectrum is predicted to evolve with time at a set rate, and the results of this project conform to theory. These results have also been published [10].

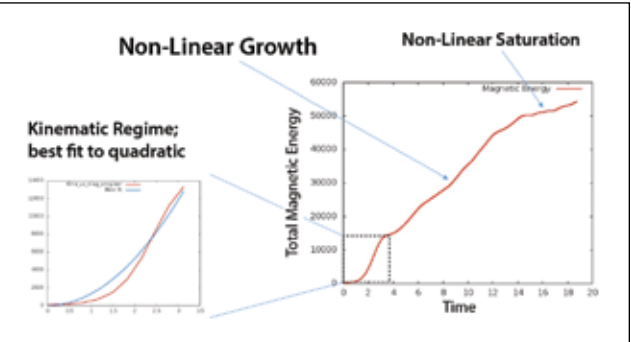


Figure 1: This figure, from [10], shows the evolution of magnetic energy with time in a two-fluid dynamo. The right panel shows magnetic field growth in a partially ionized dynamo. The inset on the left shows the initial growth of magnetic energy with the red curve representing the team's highly resolved simulations and the blue curve showing the best fit to a quadratic.

## METHODS & CODES

The core MHD algorithms in the research group's RIEMANN code are based on higher-order Godunov schemes. The team has been on the forefront of the effort to develop high-accuracy schemes for computational astrophysics in general and computational MHD in particular. Two-fluid methods have been described in [3,5,6]. The team has also published its breakthrough work on divergence-free MHD on geodesic meshes [11]. The method made several advances on higher-order-accurate reconstruction, including the reconstruction of divergence-free magnetic fields on isoparametrically mapped meshes.

## RESULTS & IMPACT

Several large-scale simulations have been completed or are ongoing on Blue Waters. The key points from these simulations have been documented above and are being readied for publication in 2019. The new work represents a substantial improvement in resolution as well as in the details of input physics and accuracy of the simulation code. In addition, the first results on geodesic mesh MHD have been published [11], showing the petascale scalability on Blue Waters for the new code.

## WHY BLUE WATERS

The simulations reported here are extremely CPU-intensive. The goal of this project is to use the petascale computing power of Blue Waters to push the resolution, accuracy, and fidelity of the simulations much higher in order to match theory with the observations coming from NASA-funded instruments.

## PUBLICATIONS & DATA SETS

S. Xu, S. Garain, D. S. Balsara, and A. Lazarian, "Turbulent dynamo in a weakly ionized medium," *Astrophys. J.*, vol. 872, no. 62, pp. 1–12, 2019.  
D. S. Balsara, V. Florinski, S. Garain, S. Subramanyan, and K. F. Gurski, "Efficient, divergence-free high order MHD on 3D spherical meshes with optimal geodesic mapping," *Mon. Notices Royal Astron. Soc.*, vol. 487, p. 1283, 2019.

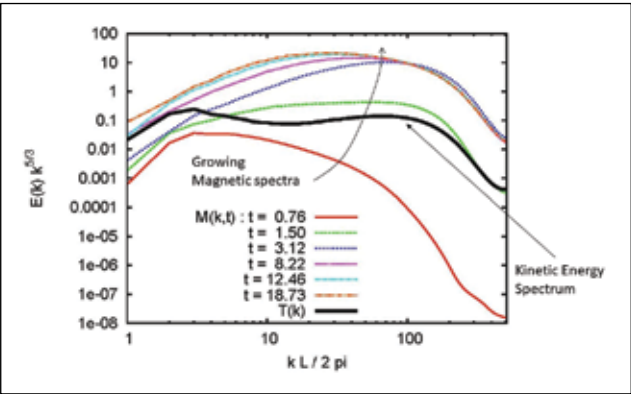


Figure 2: The study of the magnetic energy spectra is also significant. The peak in the magnetic energy spectrum is predicted to evolve with time at a set rate and the results conform to theory.



# THE COMPUTATIONAL KEYS TO THE SUPERNOVA PUZZLE: HOW MULTIPLE 3D RADIATION/HYDRODYNAMIC MODELS CAN UNLOCK THE SUPERNOVA MYSTERY

**Allocation:** NSF PRAC/5,800 and 15,000 Knh  
**PI:** Adam Burrows<sup>1</sup>  
**Collaborators:** David Radice<sup>1,2</sup>, Hiroki Nagakura<sup>1</sup>, David Vartanyan<sup>1</sup>, Aaron Skinner<sup>3</sup>, Joshua Dolence<sup>4</sup>

<sup>1</sup>Princeton University  
<sup>2</sup>Institute for Advanced Study  
<sup>3</sup>Lawrence Livermore National Laboratory  
<sup>4</sup>Los Alamos National Laboratory

## EXECUTIVE SUMMARY

The mechanism of supernova explosions is a long-standing problem in theoretical astrophysics. Its obstinacy is intertwined with the broad range of physics necessary to address the phenomenon and the challenging numerical context, involving the coupling of multidimensional hydrodynamics with neutrino radiation transfer in violently turbulent flow. This project explores solutions to this decades-old conundrum with the state-of-the-art 3D code Fornax. The overarching goal is to determine the mechanism of explosion, explosion energies, residual neutron star masses, nucleosynthesis, and explosion morphologies. Various other observational diagnostics such as radioactive nickel yields, neutrino and gravitational-wave signatures, and newly born pulsar kicks are also of interest. This initiative is enabled by the novel algorithms in Fornax that yield speedups of factors of five to 10 over previous codes. This enables multiple 3D simulations, rather than just one, to be performed each year and, hence, permits a previously unprecedented exploration of parameter space.

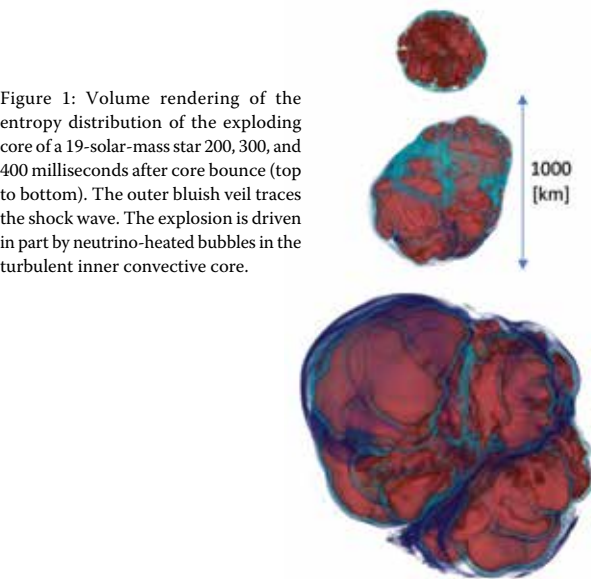


Figure 1: Volume rendering of the entropy distribution of the exploding core of a 19-solar-mass star 200, 300, and 400 milliseconds after core bounce (top to bottom). The outer bluish veil traces the shock wave. The explosion is driven in part by neutrino-heated bubbles in the turbulent inner convective core.

## RESEARCH CHALLENGE

Core-collapse supernovae dramatically announce the death of massive stars and the birth of neutron stars and black holes. During this violent process, a combination of high-density nuclear physics, multidimensional hydrodynamics, radiation transport, and neutrino physics determines whether and how the star explodes. However, the precise mechanism of explosion has not been pinned down, and this 50-year-old puzzle is one of the central remaining unsolved problems in theoretical astrophysics. Nevertheless, we are now at a crossroads. An early phase of modern supernova theory involved routine spherical simulations. This allowed explorations in parameter and progenitor space to fully characterize the phenomenon as a function of all-important quantities. Mistakes could be made quickly and an overarching understanding in 1-D could be achieved.

However, researchers knew that the cores were unstable to hydrodynamic overturn and turbulence that could not be captured in one dimension. The next phase of discovery occurred when technique and hardware advanced sufficiently so that 2D calculations became as routine as 1-D had been. This phase gave scientists a glimpse of the effects of turbulence and convection with state-of-the-art neutrino transport. At the end of this phase, a few 3D simulations appeared, but each of these simulations required approximately one year of simulation time on high-performance computing resources. As a result, researchers obtained only glimpses of the range of outcomes and their dependence on parameters in the full 3D of Nature.

With the help of Blue Waters, the research team has now entered what it believes to be the third phase of modern supernova theory, wherein multiple 3D simulations can now be performed each year, each requiring approximately one month of wall-clock time. This phase ramped up at the beginning of 2019 and is starting to reveal the full systematic dependence of full-physics 3D supernova simulations on progenitor, microphysics, and resolution. Fully characterizing the supernova phenomenon in this way, with multiple 3D simulations each year, has been the goal of supernova theory for decades and is the ultimate astrophysics Grand Challenge.

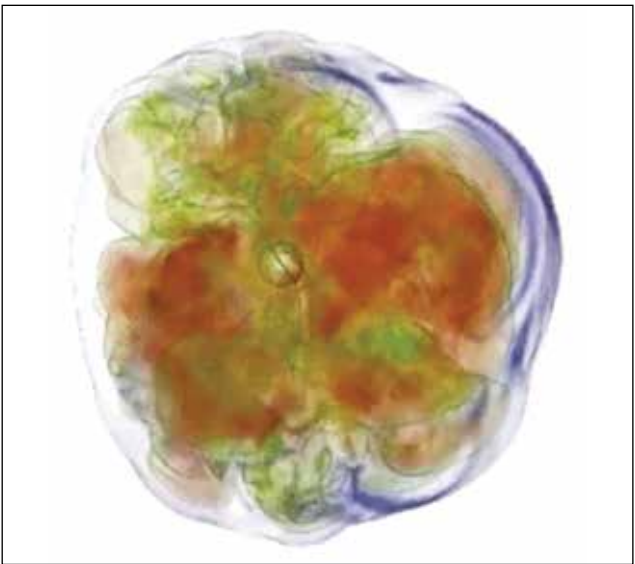


Figure 2: The explosion of the turbulent core of a 60-solar-mass star, volume-rendered in entropy. Red is high entropy, green is intermediate entropy, and blue marks the position of the supernova shock wave. The sphere near the center is the newly born proto-neutron star, depicted as an iso-mass-density contour at  $10^{11} \text{ g cm}^{-3}$ .

A solution to the core-collapse supernova problem would benefit ongoing efforts of observers and instrument designers in the United States and around the world engaged in projects to determine the origin of the elements, to measure gravitational waves (LIGO), study pulsars, and interpret laboratory nuclear reaction rate measurements in light of stellar nucleosynthesis. In addition, such studies support the experimental nuclear physics program of the NSF by exploring nucleosynthesis in astrophysical explosions, the properties of the neutrino, and the equation of state and phases of dense nuclear matter. Moreover, these investigations connect directly with the lower-energy programs of FRIB and FAIR, the high-energy experiments carried out at RHIC and the LHC, and the hyperon–hyperon and hyperon–nucleon programs of JPARC, GSI, JLAB, and NICA.

## METHODS & CODES

The research team developed a new multidimensional, multigroup radiation/hydrodynamic code, Fornax, for the study of core-collapse supernovae. Fornax is a directionally unsplit Godunov-type code that employs spherical coordinates, solves the comoving-frame, multigroup, two-moment, velocity-dependent transport equations to  $O(v/c)$ , uses the M1 tensor closure for the second and third moments of the radiation fields, and employs a dendritic spherical grid. Fluxes at cell faces are computed with an HLLC Riemann solver based on left and right states reconstructed from the underlying volume-averaged states. Three species of neutrino are followed using an explicit Godunov characteristic method applied to the radiation transport operators and an implicit solver for the radiation source terms. In this way, the radiative transport and transfer are handled locally, without the need for a global solution on the entire mesh.

The team thereby significantly reduced the computational complexity and communication overhead of traditional multidimensional radiative transfer solutions by bypassing the need for global iterative solvers. In summary, the main advantages of the Fornax code are its efficiency owing to its explicit nature, its excellent strong scaling to hundreds of thousands of cores, its truly multidimensional transport solution, and interior static mesh derefinement.

## RESULTS & IMPACT

Using Blue Waters, the research team has performed a suite of 3D runs of the collapse, bounce, and explosion (most often) of 9-, 10-, 12-, 13-, 14-, 15-, 16-, 17-, 18-, 19-, 25-, and 60-solar-mass progenitor massive stars. This is the most extensive set of 3D supernova simulations with the necessary realism ever performed. Moreover, the team has been able to conduct the highest-resolution full-physics simulations ever performed, calculate the gravitational wave and neutrino signatures, explore pulsar kick speeds, and establish debris morphologies and compositions.

## WHY BLUE WATERS

For the team’s code, Fornax, the Blue Waters architecture in the MPI/CPU context, with its large per-node memory and rapid interconnect, provides the quickest turnaround for these 3D supernova simulations of any available high-performance computing resource.

## PUBLICATIONS & DATA SETS

- A. Burrows, D. Radice, and D. Vartanyan, “Three-dimensional supernova explosion simulations of 9-, 10-, 11-, 12-, and 13-solar-mass stars,” *Mon. Not. Roy. Astron. Soc.*, vol. 482, p. 3153, 2019, doi: 10.1093/mnras/stz543.
- M. A. Skinner, J. Dolence, A. Burrows, D. Radice, and D. Vartanyan, “Fornax: A flexible code for multiphysics astrophysical simulations,” *Ap. J. Suppl.*, vol. 241, no. 7, 2019, doi: 10.3847/1538-4365/ab007f.
- H. Nagakura, A. Burrows, D. Radice, and D. Vartanyan, “Towards an understanding of the resolution dependence of core-collapse supernova simulations,” *Mon. Not. Roy. Astron. Soc.*, 2019, doi: 10.1093/mnras/stz2730.
- A. Burrows, D. Vartanyan, J. C. Dolence, M. A. Skinner, and D. Radice, “Crucial physical dependencies of the core-collapse supernova mechanism,” *Space Sci. Rev.*, vol. 214, no. 33, 2018, doi: 10.1007/s11214-017-0450-9.
- D. Vartanyan, A. Burrows, D. Radice, M. A. Skinner, and J. C. Dolence, “Revival of the fittest: Exploding core-collapse supernovae,” *Mon. Not. Roy. Astron. Soc.*, vol. 477, no. 3, p. 3091, 2018, doi: 10.1093/mnras/sty809.
- D. Radice, V. Morozova, A. Burrows, D. Vartanyan, and H. Nagakura, “Characterizing the gravitational wave signal from core-collapse supernovae,” *Ap.J. Lett.*, vol. 876, p. L9, 2019, doi: 10.3847/2041-8213/ab191a.



# SHEDDING LIGHT ON SUPERMASSIVE BINARY BLACK HOLE MERGERS

**Allocation:** NSF PRAC/10,200 Knh  
**PI:** Manuela Campanelli<sup>1</sup>  
**Co-PIs:** Scott Noble<sup>2,3</sup>, Julian Krolik<sup>4</sup>  
**Collaborators:** Stephane d'Ascoli, Mark Avara<sup>1</sup>, Dennis Bowen<sup>1</sup>, Vassilios Mewes<sup>1</sup>

<sup>1</sup>Rochester Institute of Technology  
<sup>2</sup>University of Tulsa  
<sup>3</sup>NASA Goddard Space Flight Center  
<sup>4</sup>Johns Hopkins University

## EXECUTIVE SUMMARY

Observing electromagnetic and gravitational waves from supermassive binary black holes and their environments promises to provide important new information about both strong-field gravity and galaxy evolution. Since little theoretical understanding about the details of these accreting binary black hole systems exists, the aim of this project is to continually advance the realism and rigor of simulations of these systems. The problem is complicated because dynamical general relativity, plasma physics, and radiation physics all must be calculated together over vast spatial and temporal scales.

This past year, the research team finished performing the first magnetohydrodynamics (MHD) simulations of black holes with their own minidisks surrounded by a circumbinary disk. The team also created the first detailed electromagnetic predictions consistent with simulations using postproduction radiation transport. These predictions will be critical to the success of electromagnetic searches and source characterization leading up to the launch of the Laser Interferometer Space Antenna (LISA), which is planned for 2034.

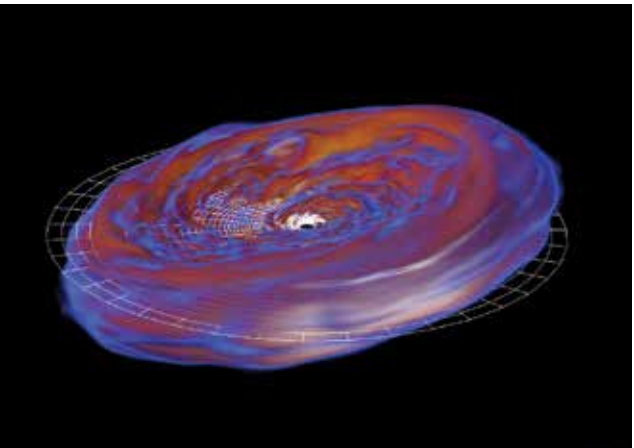


Figure 1: This simulation snapshot shows a partially evolved disk with a 12° tilt with respect to the binary supermassive black hole orbital plane. (Credit: Mark J. Avara)

## RESEARCH CHALLENGE

Realistic accretion disk simulations are particularly challenging because they involve a multitude of physical processes interacting over large dynamic ranges in space and time. In actual systems, gas is collected at scales a million times larger than the black holes themselves, yet many cells per black hole width must be used to capture the relativistic plasma dynamics in their vicinity. Reliable angular momentum transport of gas through the disk requires solving the MHD equations of motion at sufficiently high resolution to adequately resolve the responsible internal magnetic stresses. Consistency between the gas's thermodynamics and radiation model is desirable to produce self-consistent predictions of the light produced by the modeled systems, which is the ultimate goal of this program.

Then, transporting the produced light to a distant observer requires researchers to calculate how light moves in the curved time-dependent spacetime of the black hole binary, and how it scatters and is absorbed by intervening gas; *i.e.*, investigators must solve the general relativistic geodesic and radiative transfer equations. All this sophistication is built so scientists may confidently predict what electromagnetic counterparts may exist to the extremely bright gravitational wave sources LISA will see over cosmological distances. Since LISA is planned to launch in more than a decade, researchers can begin to search for these systems with predictions in hand. The research group's simulations may discover features that are unique to binaries and inform the search for them.

## METHODS & CODES

The research team used the flux conservative GRMHD code called HARM3D. It is written in a covariant way such that arbitrary spacetime metrics and coordinate systems may be used without the need to modify core routines. The team also tested and used the new code PatchworkMHD coupled with HARM3D to enable the most sophisticated and longest simulations yet. This allowed the placement of meshes with different refinement/coordinates and topologies in a way ideal for the system, increasing the efficiency of the simulations and extending them by an order of magnitude in time. It enabled the team to capture the behavior

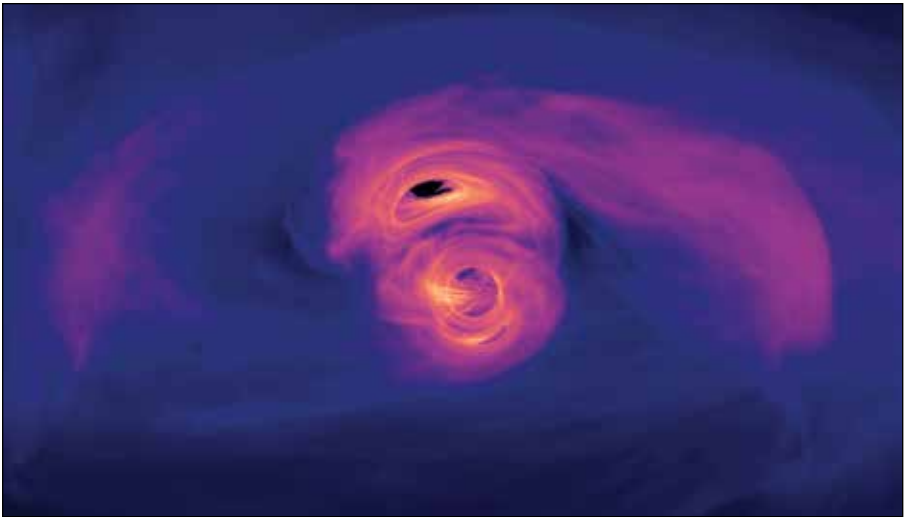


Figure 2: General relativistic ray-traced image of emission from magnetized gas accreting onto a binary system of black holes. The intensity of light is represented by the intensity of the hue. (Credit: S. Noble)

of a region between the black holes that was excluded in prior 3D simulations and to achieve a resolution unprecedented in magnetized 3D simulations of accreting binaries.

## RESULTS & IMPACT

The research team's continued research program focused on 3D MHD simulations of the black hole minidisks and circumbinary disk interactions has advanced the field in a number of ways. During a previous allocation period, the team discovered a new phenomenon in which the irregular circumbinary flow can modulate the rate of accretion onto the minidisks in a quasiperiodic fashion, causing them to be depleted and then refilled as they passed by the over-density feature in the circumbinary disk [1]. During the most recent allocation period, the team continued the simulation of three binary black hole orbits to 12 orbits [3]. This allowed the team to probe the minidisks quasiperiodic behavior deeper into the late stages of the relativistic binary black hole inspiral. The research group showed that the quasiperiodicity in the minidisk evolution is driven by the interaction of the individual black holes with a localized over-density feature in the circumbinary, which orbits the binary at a well-defined beat frequency with respect to the orbiting black holes. Since the mass inflow times onto each black hole during the late stages of inspiral are comparable to the variability of the mass supply owing to the over-density, the minidisk masses and accretion rate onto each black hole are strongly modulated at this beat frequency. This constitutes a distinctive electromagnetic signature, the sort that could distinguish supermassive black hole binary systems from typical accreting single supermassive black holes. The research team subsequently explored how all this changes when the disk is tilted with respect to the binary's orbit, since the gas fed to the system need not always be aligned. The team has completed a first survey and are analyzing the results as of August 2019 (Fig. 1).

The team performed first-of-its-kind radiative transfer calculations in time-dependent general relativity using the simulation's data as an emitting source in order to predict the electromagnetic

emission from the minidisk simulation. A range of viewing angles and observation frequencies were surveyed for all time slices of data to explore the energy, time, and angle dependence of the emission [2]. This calculation resulted in the first electromagnetic spectrum of accreting supermassive black holes in the inspiral regime. These results were shared with the community with the help of several news releases [4–6], which included high-resolution movies and images (Fig. 2).

## WHY BLUE WATERS

The 3D GRMHD mini-disk simulation ran for three orbits and used 1.125 million node hours on Blue Waters. The simulation used a 600 x 160 x 640 grid, or approximately 60 million cells, and evolved through about three million timesteps using 600 nodes. The follow-up simulation, which extended this to 12 orbits, used 148 days of wall-clock time running on 600 nodes for a total of 2.13 million node hours. The tilted circumbinary disk simulations ran on 500 Blue Waters nodes at a time, but required a long duration. NCSA staff members were helpful in arranging reservations for the runs. Without the vast resources available on Blue Waters, the research team would not have been able to perform this work.

## PUBLICATIONS & DATA SETS

D. B. Bowen, V. Mewes, S. C. Noble, M. J. Avara, M. Campanelli, and J. H. Krolik, "Quasi-periodicity of Supermassive Binary Black Hole Accretion Approaching Merger," *Astrophys. J.*, vol. 879, no. 2, p. 76, July 2019, doi: 10.3847/1538-4357/ab2453.  
S. d'Ascoli, S. C. Noble, D. B. Bowen, M. Campanelli, J. H. Krolik, and V. Mewes, "Electromagnetic emission from supermassive binary black holes approaching merger," *Astrophys. J.* vol. 865, no. 2, p. 140, Oct. 2018.  
D. B. Bowen, V. Mewes, M. Campanelli, S. C. Noble, J. H. Krolik, and M. Zilhão, "Quasi-periodic behavior of mini-disks in binary black holes approaching merger," *Astrophys. J. Lett.*, vol. 853, no. 1, p. L17, Jan. 2018.



ACCRETION DYNAMICS OF SUPERMASSIVE BLACK HOLE BINARIES

Allocation: NSF PRAC/16,600 Knh  
PI: Manuela Campanelli<sup>1</sup>  
Collaborators: Scott Noble<sup>2,3</sup>, Julian Krolik<sup>4</sup>, Mark Avara<sup>1</sup>, Dennis Bowen<sup>1</sup>, Vassilios Mewes<sup>1</sup>

<sup>1</sup>Rochester Institute of Technology  
<sup>2</sup>University of Tulsa  
<sup>3</sup>NASA Goddard Space Flight Center  
<sup>4</sup>Johns Hopkins University

EXECUTIVE SUMMARY

The plasma dynamics and electromagnetic emission of a supermassive black hole binary embedded in a gaseous astrophysical environment is integral to the story of galaxy evolution and the growth of supermassive black holes. The future launch of the Laser Interferometer Space Antenna (LISA) and an enhanced view of the sky enabled by upcoming X-ray and time-domain telescopes will provide a multimessenger view of these binary systems. To meaningfully identify the specific electromagnetic signals coming from a supermassive black hole binary approaching merger requires numerical explorations on Blue Waters of their full non-linear behavior. With this allocation, the research team has performed the first 3D GRMHD simulation of an electromagnetically active supermassive black hole binary to over 30 orbits. This became possible with the development of a new multimesh code, PatchworkMHD, which removed significant limitations on meshes, coordinates, and geometry from prior simulations, provided better efficiency, and will ultimately enable the first parameter space survey in 3D–MHD simulations of these binary systems.

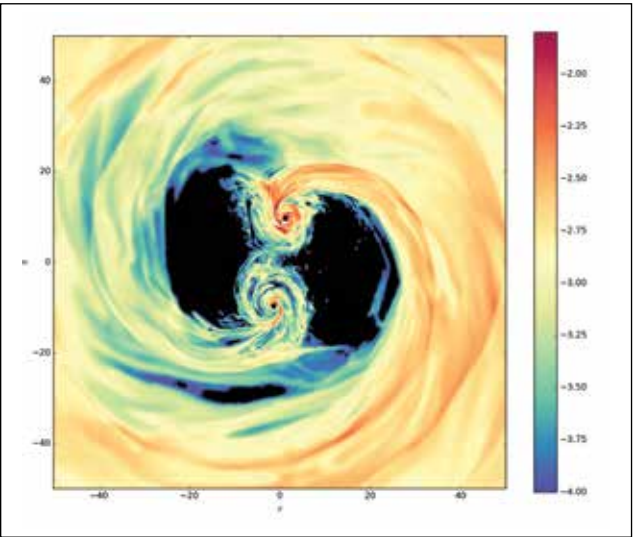


Figure 1: Density is shown in hue on a logarithmic scale in a midorbital-plane slice through an accreting black hole binary simulation [4] in which the binary is aligned with the outer accretion disk. This is the first 3D GRMHD simulation to evolve the entire region close to the binary. (Credit: Mark Avara)

RESEARCH CHALLENGE

Understanding the behavior of supermassive black hole binaries in an empty region of space requires proper evolution of a dynamical general relativistic spacetime. When such a system is embedded in a gaseous environment in which plasma falls toward the black holes forming accretion disks, the system can become incredibly bright, and only then will electromagnetic telescopes detect them. Simulations of these accreting systems are particularly challenging as they involve a multitude of physical processes interacting over large dynamic ranges in space and time. The gas is collected at scales over a million times larger than the black holes themselves, yet a realistic grid-based numerical study requires many grid cells spanning the width of the black hole in order to capture the relativistic plasma dynamics in their vicinity. The demands grow further because realistic angular momentum transport in the gas of the disk requires solving the magnetohydrodynamic (MHD) equations of motion at sufficiently high resolution to adequately resolve the responsible internal magnetic stresses. Furthermore, consistency between the gas’s thermodynamics and a radiation model is desirable to produce reliable predictions of the light produced by the systems, which is the ultimate goal of this program. Transporting the produced light to the equivalent of an Earth-based telescope requires researchers to calculate how light moves in the curved time-dependent spacetime of the black hole binary and how it is scattered and absorbed by intervening gas. This requires numerically solving the general relativistic geodesic and radiative transfer equations.

METHODS & CODES

The research team used the flux conservative GRMHD code called HARM3D. It is written in a covariant way such that arbitrary spacetime metrics and coordinate systems may be used without the need to modify core routines. The team has also used the new code PatchworkMHD coupled with HARM3D to enable the most sophisticated and longest simulations of supermassive black hole binaries yet (Fig. 1). PatchworkMHD allows the placement of meshes with different refinement/coordinates and topologies in a way that is ideal for the system (Fig. 2). This ability has increased the efficiency of the simulations, extending them by an order of magnitude in time. It also allowed the team to cap-

ture the behavior of a region between the black holes that was excluded in prior 3D simulations and to achieve a resolution unprecedented for magnetized 3D simulations of accreting binaries.

RESULTS & IMPACT

This continued research program focused on 3D MHD simulations of minidisk and circumbinary disk interactions has advanced the field in a number of ways. During a previous allocation, the research team discovered a new phenomenon in which the irregular circumbinary flow can modulate the rate of accretion onto the minidisks, leading them to be depleted and then refilled as they pass by the over-density feature in the circumbinary disk, producing quasiperiodic electromagnetic emission [1–3]. An extension of this early simulation to 12 orbits provides enough temporal overlap for quantitative comparison to the team’s newest 30-orbit simulation using the new PatchworkMHD infrastructure.

With better resolution, the first 3D GRMHD evolution of the region where material sloshes from orbit around one black hole in the binary to the other, and with enough orbits to start constraining and identifying light-curve properties unique to these systems, the team’s most recent simulation has provided a number of new insights soon to be published [4]. The research group has discovered that at this separation there is significant interaction among the materials orbiting each black hole. The team has also discovered symmetry breaking important to observations that only 3D simulations capture, and have constrained how the magnetic field threading each black hole evolves through these final stages prior to merger (Fig. 1). Insights in this last domain help the team to understand potential electromagnetic signatures from relativistic jets in supermassive black hole binaries, producing complementary observables to those coming from the disks.

The large number of orbits performed in this simulation is enough to evolve the binary through a gravitational-wave decay to half its starting separation. The team has therefore finally been able to capture the coupled evolution of the circumbinary disk and the accreting black holes through their final stages. This body of work starts to fill the gap between nonrelativistic studies of widely separate binaries and fully relativistic studies that capture the black hole merger. There is much yet to be discovered in how these regimes connect; the tools that the research team has developed and tested on Blue Waters are well suited for taking the next steps.

Analysis and interpretation of the combined postprocessing data products for both the simulation published in [3] as well as the 30-orbit [4] simulations are underway and will provide much more detail, especially with regard to a time-dependent electromagnetic view of supermassive black hole systems.

WHY BLUE WATERS

Using PatchworkMHD, the research team was able to run binary simulations with 30 times the prior efficiency, which—coupled with the scale of resources available only on Blue Waters—has

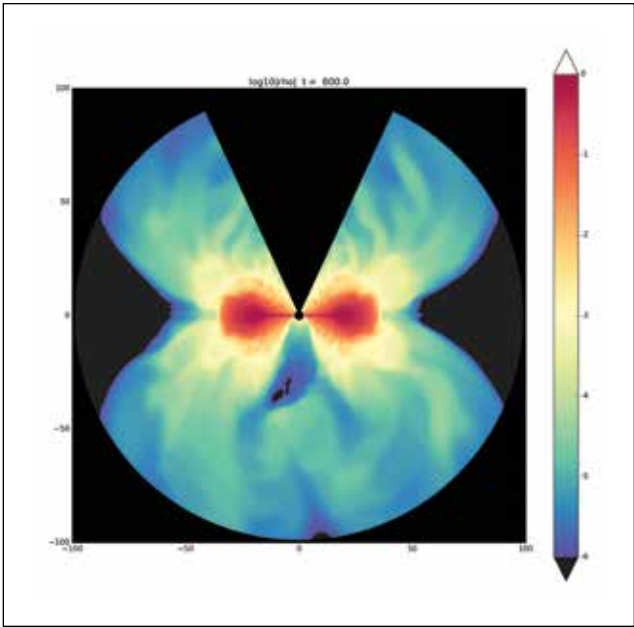


Figure 2: A multipatch configuration for an accretion disk around a single black hole, showing evolution of the “southern” jet region with a patch covering the axial coordinate singularity. (Credit: Mark Avara)

enabled the team to perform the first physical parameter-space study of these systems in 3D GRMHD.

PUBLICATIONS & DATA SETS

- D. B. Bowen, V. Mewes, S. C. Noble, M. J. Avara, M. Campanelli, and J. H. Krolik, “Quasi-periodicity of supermassive binary black hole accretion,” *Astrophys. J.*, vol. 879, no. 2, May 2, 2019, doi: 10.3847/2041-8213/aaa756.
- S. d’Ascoli, S. C. Noble, D. B. Bowen, M. Campanelli, and J. H. Krolik, “Electromagnetic emission from supermassive binary black holes approaching merger,” *Astrophys. J.*, vol. 865, p. 140, 2018.
- D. B. Bowen, V. Mewes, M. Campanelli, S. C. Noble, and J. H. Krolik, “Quasi-periodic behavior of mini-disks in binary black holes approaching merger,” *Astrophys. J. Lett.*, vol. 853, p. L17, 2018, doi: 10.3847/1538-4357/ab2453.
- Mark Avara *et al.*, in preparation, 2020.



ACHIEVING PROBABILISTIC CLASSIFICATION OF COSMIC WEB PARTICLES USING RAPIDLY GENERATED TRAINING DATA: A METHOD FOR CLASSIFYING GALAXIES INTO THEIR COSMIC WEB STRUCTURAL GROUPS USING SUPERVISED MACHINE LEARNING

Allocation: Exploratory/35 Knh  
PI: Matias Carrasco Kind<sup>1</sup>  
Co-PI: Brandon Buncher<sup>1</sup>

<sup>1</sup>University of Illinois at Urbana–Champaign

EXECUTIVE SUMMARY

The cosmic web consists of a network of galaxies and dark matter. Long, strandlike filaments connect between spheroidal galaxy clusters, leaving underdense void regions in between. Knowing whether a galaxy is a member of a halo (the dark matter clump around which clusters form), filament, or void provides substantial information about its surrounding environment, which helps in understanding how it formed and will evolve.

However, current methods are limited: direct classification algorithms are generally inefficient, while deep learning-based methods are inconsistent with one another. Therefore, the research group created a novel classification method using supervised machine learning. They train the algorithm using quickly generated data that visually approximates the cosmic web using predetermined generation algorithms. Then, with the help of the high memory capacity of Blue Waters, the researchers use the trained algorithm to classify galaxies in simulated data. While the training data lack much of the detail seen in observed/simulated data, it takes substantially less computational power to create. A

simulation of cosmic web formation with 16 million particles requires tens of thousands of node-hours on a multinode cluster, whereas the team’s method can generate training data with the same number of particles in less than an hour on a laptop computer. The researchers have demonstrated that this method provides enough information for the machine learning algorithm to “learn” to correctly classify particles in more realistic data sets. Although it is trained using simpler data, the robustness of machine learning helps the algorithm bridge the information gap, providing classification at a substantially cheaper cost.

RESEARCH CHALLENGE

Current methods used to classify galaxies into cosmic web structural groups typically utilize neural networks, as direct algorithms are too computationally intense. However, arbitrary hyperparameters and inconsistent cosmic web class definitions lead to substantial disagreement among methods. In addition, owing to these hyperparameters, it is difficult to establish self-consistency for a given method.

Figure 1: (left) A two Mpc-thick slice of the 3D training data set with 660,000 particles: red particles are members of halos, green of filaments, and blue of voids. (right) A two Mpc-thick slice of the 3D N-body simulation data set with 16.7 million particles; training was performed exclusively using the training data set on the left. A megaparsec (Mpc) is a unit of length used in astronomy.

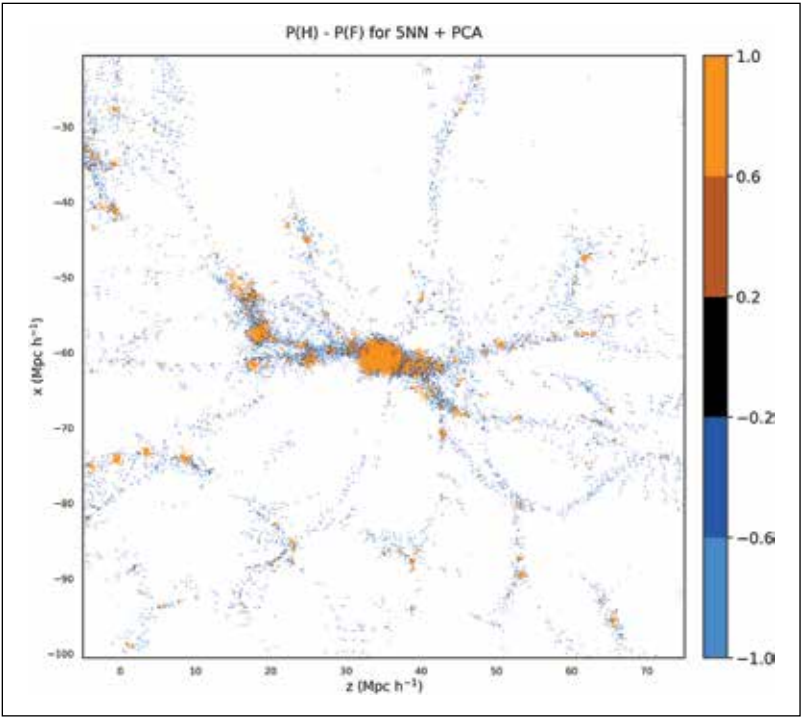
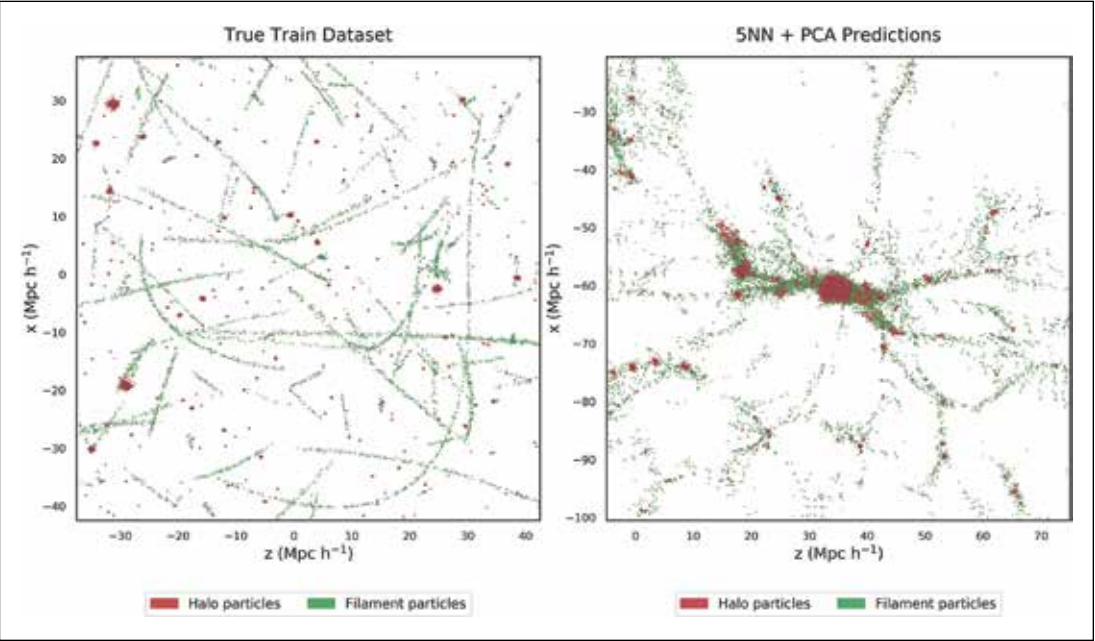


Figure 2: The particle probabilities for N-body simulation halos/filaments. Light-colored particles had a high probability for halo (orange) or filament (blue) classification, while black particles were ambiguous. Note that the majority of ambiguous particles appeared on the border between halos and filaments, where ambiguity naturally exists owing to the low-density contrast.

A particle’s cosmic web class membership provides substantial information about its local environment, which plays a great role in its history and future evolution. The research team aims to improve cosmic web classification through implementing a supervised algorithm, which is easier to verify, lacks arbitrary hyperparameters, and, by simplifying training data generation, is substantially less expensive and time-consuming than direct methods.

METHODS & CODES

Training data is generated by sampling particles from predetermined toy model structures. Halos are produced by sampling particles from a spherical Gaussian distribution and then distributed randomly throughout the region. Filaments consist of particles populated in a uniform cylinder around randomly generated Bezier curves (<https://github.com/dhermes/bezier>), while voids are a uniform background distribution. Measurements of the k-nearest neighbors provide information about the local density magnitude, while the explained variance ratio calculated from a principal component analysis decomposition measures the density field directionality. A random forest algorithm was trained using the results of these measurements on each particle, and the trained algorithm was used to classify particles in an N-body simulation. Though these classifications cannot be directly verified owing to a lack of true class values, the researchers demonstrated the robustness of the method by comparing these predictions to those made on another toy model data set.

RESULTS & IMPACT

Currently, the team is working on a final write-up of the results. Work to date has demonstrated that the methodology does

provide a robust method for classifying particles at a much lower computational cost. In addition, the team’s verification methods have shown that the method achieves probabilistic classification, providing additional information to use when studying the formation and evolution of galaxies. Although the optimal configuration of measurements and toy model parameters remains to be found, the research group has demonstrated the effectiveness of using a supervised method trained with fast generated data, and they expect that future optimization will lead to improvements in particle classification that is naturally accompanied by information on the classification confidence. This will change the field of cosmology by simplifying and improving the efficiency of particle classification, enabling greater understanding of the cosmic web and its effects on other structures.

WHY BLUE WATERS

Blue Waters provided substantial benefit through its high-memory nodes. Although toy model generation is extremely fast and efficient, performing measurements on the N-body simulation requires considerably more memory than is available on a laptop computer or a single Blue Waters node. Segmenting the particle field does not avoid this issue because of the size of the required files, and analysis would take an unreasonable amount of time. Blue Waters’ nodes had enough memory to load the segmented files and analyze them, and by parallelizing the analysis across several nodes, it allowed computations that would normally require several weeks on a high-capacity remote machine to be completed in less than a day.



# THE EPOCH OF THE FIRST LUMINOUS BLACK HOLES: EVOLVING THE BLUETIDES SIMULATION INTO THE FIRST BILLION YEARS OF COSMIC HISTORY

**Allocation:** NSF PRAC/4,400 Knh  
**PI:** Tiziana Di Matteo<sup>1</sup>  
**Co-PIs:** Yu Feng<sup>2</sup>, Rupert Croft<sup>1</sup>  
**Collaborators:** Aklant Bhowmick<sup>1</sup>, Yueying Ni<sup>1</sup>, Ananth Tenneti<sup>3</sup>, Steve Wilkins<sup>4</sup>, Madeleine Marshall<sup>5</sup>, Stuart Wyithe<sup>5</sup>

<sup>1</sup>Carnegie Mellon University  
<sup>2</sup>University of California, Berkeley  
<sup>3</sup>University College, London  
<sup>4</sup>University of Sussex  
<sup>5</sup>University of Melbourne

## EXECUTIVE SUMMARY

Quasars, powered by supermassive black holes, are the most luminous objects known. As a result, they enable unparalleled studies of the Universe at the earliest cosmic epochs. The absolute record holder, a black hole with a mass of 800 million times that of our sun, has recently been discovered from when the Universe was only 690 million years old—just five percent of its current age, and at the very dawn of galaxy formation.

With their large simulation volume, the researchers have made many predictions on what the James Webb Space Telescope (JWST) will see when pointed at such early quasars after its launch in 2021. By comparison, the BlueTides simulation box is ten thousand times larger in angular area than the field of view of JWST’s Near Infrared Camera, allowing the rare regions where quasars form to be translated into mock JWST observations. JWST should reveal an assortment of black hole galaxies, including compact spheroids, and isolated galaxies.

## RESEARCH CHALLENGE

The previous phase of the BlueTides (BT) simulation made direct contact with the observations of the giant record holding black hole, making predictions for its host galaxy. Remarkably, the sample of early Universe quasars has grown tremendously over the last year, with 30 more quasars having been found from the first billion years of the universe. The researchers would like to study these types of objects directly with BlueTides, if they are able to form, to look at the conditions that induce the rapid black hole growth necessary to reach these giant masses and make predictions for upcoming telescopes. In order to make contact with this much larger sample, it was necessary to evolve BlueTides forward in time, reaching into a new regime where hundreds of thousands of galaxies in the simulation volume are forming stars, and the Universe is at a new level of complexity.

## METHODS & CODES

BlueTides is run on the entire set of compute nodes on Blue Waters using the latest version of the research team’s MP-Gadget

code. The general characteristics of the GADGET family of codes are those of a flexible TreePM-SPH solver for cosmological fluids of dark matter, gas, and stars. In addition to the basic physics of gravity and hydrodynamics, the code also contains numerous further physics modules covering aspects of star formation and black hole growth. The MP-Gadget code variant was developed to be lean, efficient, and scalable. As the simulation progresses, the computation starts to become dominated by the hydrodynamics solver as more galaxies form. This part of the code has traditionally been the hardest to scale, with timesteps getting very small in highly clustered regions of very high density. These regions, however, are the most interesting ones, where the observable massive galaxies and quasars form. The MP-Gadget hydro-solver is MPI/thread hybridized, and the team has recently further improved the threading efficiency. More threads mean fewer MPI ranks, and thus easier over-decomposition.

## RESULTS & IMPACT

As the cosmic web of gas and dark matter was evolving in BlueTides at an ever-increasing pace, a large sample of massive black holes had formed in the simulation volume over the course of the current run. With such a set of objects, the researchers can now carry out statistical tests on the whole population, and test cosmological models quantitatively. The host galaxies of these black holes have detailed stellar properties available: They all tend to be highly star-forming but have a wide range of kinematic characteristics (disks, bulges), and lie in a variety of large-scale environments. The assembly of each galaxy happens at the same time that the central black holes grow, and for approximately 30 objects the black hole outshines the galaxy significantly and the object would be observed as a quasar.

With the large simulation volume, the researchers have made many predictions on what the James Webb Space Telescope (JWST) will see when pointed at such early quasars after its launch in 2021. By comparison, the BlueTides simulation box is ten thousand times larger in angular area than the field of view of JWST’s Near Infrared Camera, allowing the rare regions where quasars

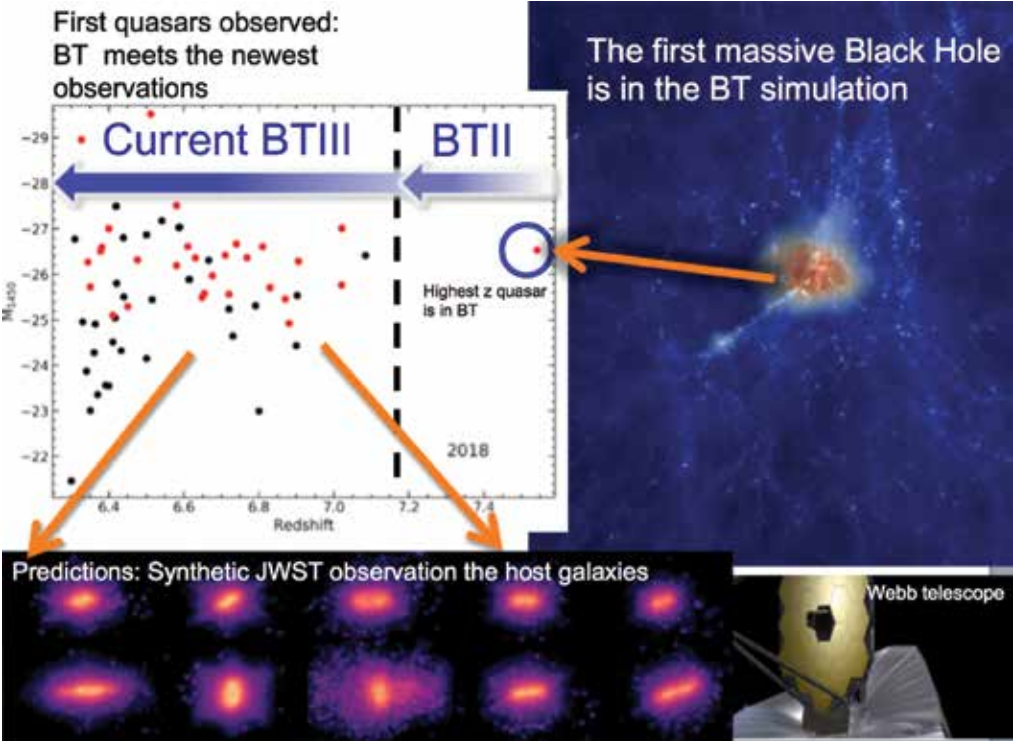


Figure 1: Left—The current (as of 2019) quasar sample in the early Universe. Thirty new quasars at  $z > 6$  (the first billion years) were discovered recently. Right: The large-scale environment surrounding the first massive black hole to form in BlueTides. Bottom: BlueTides galaxies as they would be seen by the Webb telescope.

form to be translated into mock JWST observations. JWST should reveal an assortment of black hole galaxies, including compact spheroids, and isolated galaxies.

## WHY BLUE WATERS

A complete simulation of the Universe at the epochs the researchers are studying requires a small enough particle mass to model the dwarf galaxies, which significantly contribute to the summed ionizing photon output of all sources. It also requires an enormous volume (on the order of 1 cubic gigaparsec or Gpc, where 1 Gpc is  $3 \times 10^{10}$  cubic light years) in order to capture the rarest and brightest objects: the first quasars. The first requirement is therefore equivalent to a high particle density, and the second to a large volume.

Previous calculations on smaller high-performance computing systems have either fulfilled the first requirement in a small volume, or the second with large particle masses, thus only resolved for large galaxies. With Blue Waters, however, the research team has reached the point where the required number of particles (about one trillion) could be contained in memory, and the petaflop computing power was available to evolve them forward in time. Blue Waters, therefore, has made possible this qualitative advance, allowing arguably the first complete simulation (at least in terms of the hydrodynamics and gravitational physics) of the creation of the first galaxies and large-scale structures in the Universe.

The application runs required essentially the full system: The researchers used 20,250 nodes (648,000 core equivalents—the

new version of the code can scale higher, but the team left a safety margin) using 57 GB/node (89%). This application thus uses 1.15 PB of memory—something Blue Waters can provide, as it is 90% of the available memory. Running such large jobs on a regular basis in a very timely fashion obviously requires advanced resource management, and the way the Blue Waters project has been set up made this possible.

## PUBLICATIONS & DATA SETS

A. K. Bhowmick, T. Di Matteo, Y. Feng, and F. Lanusse, “The clustering of  $z > 7$  galaxies: predictions from the BLUETIDES simulation,” *MNRAS*, vol. 474, no. 4, pp. 5393–5405, Mar. 2018.

A. Tenneti, S. M. Wilkins, T. Di Matteo, R. A. C. Croft, and Y. Feng, “A tiny host galaxy for the first giant black hole:  $z = 7.5$  quasar in BlueTides,” *MNRAS*, vol. 483, no. 1, pp. 1388–1399, Feb. 2019.

K.–W. Huang, Y. Feng, and T. Di Matteo, “The early growth of supermassive black holes in cosmological hydrodynamic simulations with constrained Gaussian realizations,” *MNRAS*, submitted (2019), arXiv:1906.00242.

Y. Ni, T. Di Matteo, Y. Feng, R. A. C. Croft, and A. Tenneti, “Gas outflows from the  $z = 7.54$  quasar: predictions from the BlueTides simulation,” *MNRAS*, vol. 481, no. 4, pp. 4877–4884, Dec. 2018.

Y. Ni, M.–Y. Wang, Y. Feng, and T. Di Matteo, “Predictions for the abundance of high-redshift galaxies in a fuzzy dark matter Universe,” *MNRAS*, vol. 488, no. 4, pp. 5551–5565, Oct. 2019.

S. M. Wilkins *et al.*, “Nebular line emission during the epoch of reionization,” *MNRAS*, submitted (2019), arXiv:1904.07504.



ADVANCING FIRST-PRINCIPLE SYMMETRY-GUIDED NUCLEAR MODELING FOR STUDIES OF NUCLEOSYNTHESIS AND FUNDAMENTAL SYMMETRIES IN NATURE

**Allocation:** NSF PRAC/171.5 Knh  
**PI:** Jerry Draayer<sup>1</sup>  
**Co-PIs:** K. D. Launey<sup>1</sup>, T. Dytrych<sup>1,2</sup>  
**Collaborators:** R. Baker<sup>1</sup>, G. Sargsyan<sup>1</sup>, A. Mercenne<sup>1</sup>, D. Langr<sup>3</sup>, M. Kocourek<sup>3</sup>

<sup>1</sup>Louisiana State University  
<sup>2</sup>Czech Academy of Sciences  
<sup>3</sup>Czech Technical University

EXECUTIVE SUMMARY

The Blue Waters (BW) system has enabled modeling of nuclear wave functions with unprecedented accuracy for light- and medium-mass nuclei. This is key to addressing two of the most challenging questions in physics today; namely, the origin of elements and whether the neutrino is its own antiparticle. The work also supports and informs current and projected experimental efforts at state-of-the art radioactive beam facilities, including the upcoming Facility for Rare Isotope Beams at Michigan State University. Breakthrough theoretical advances [1,2] along with the development of novel highly scalable algorithms [3,4] and coupled with the BW cutting-edge computational power have opened a new region—the intermediate-mass nuclei from fluorine to calcium isotopes—for first investigations with *ab initio* (“from first principles”) methods. This targets nuclei far from stability with collective and cluster substructures, while pinpointing key features of astrophysical processes, probing fundamental symmetries in nature as well as supporting current and upcoming experiments at radioactive beam facilities.

RESEARCH CHALLENGE

One of the quintessential open problems in contemporary physics is to design a comprehensive many-body theory for modeling and predicting nuclear structure and reactions starting from in-

ternucleon forces that are consistent with the underlying theory of quantum chromodynamics (QCD). The ultimate goal of *ab initio* theory is to find a solution to this problem, which is a computationally highly intensive endeavor owing to a dual challenge: the nonperturbative nature of QCD in the low-energy regime and the complexity of many-particle nuclei. Because short-lived nuclei are currently difficult or impossible to access by experiment but are key to understanding processes in extreme environments ranging from stellar explosions to the interior of nuclear reactors, first-principle nuclear models that hold predictive capabilities will have tremendous impact on advancing knowledge at the frontiers of multiple branches of physics, including astrophysics, neutrino physics, and applied physics.

METHODS & CODES

The research team has developed an innovative *ab initio* nuclear structure approach, dubbed the symmetry-adapted no-core shell model (SA–NCSM) [1], with concomitant computer code LSU3shell [3–5], that embraces the first-principles concept and capitalizes on a new symmetry of the nucleus. The *ab initio* SA–NCSM solves the time-independent Schrödinger equation as a Hamiltonian matrix eigenvalue problem. The main computational task is to evaluate a large symmetric Hamiltonian matrix and to obtain the lowest-lying eigenvectors that correspond to the experimental regime. Accuracy is based on the degree of convergence, which is linked to the size of the model space that can be achieved. The SA–NCSM utilizes physically relevant model space of significantly reduced dimensionality compared to ultralarge model spaces encountered by standard *ab initio* approaches. These theoretical advances [2,6,7], coupled with the computational power of the Blue Waters system, have allowed the team to reach medium-mass nuclei that are inaccessible experimentally and to other *ab initio* methods.

RESULTS & IMPACT

The nuclei of interest represent a considerable challenge requiring computational power of nearly the entire BW machine and its system memory. Two graduate students have carried forward these studies and have had the unique opportunity to work with

supercomputers and massively parallel programming environments. The following list describes the results and their impact:

- While enhanced deformation and cluster substructures are difficult to describe from first principles, the SA–NCSM and the BW system have allowed the first *ab initio* descriptions of deformed nuclei using chiral interactions [2,6,7]. The team has continued to study emergent deformation and clustering in nuclei, from first principles, for Mg isotopes and their mirror nuclei. In addition to calculations for <sup>21</sup>Mg and <sup>21</sup>F, <sup>22</sup>Mg and <sup>22</sup>Ne, new results are now available for <sup>23</sup>Mg and <sup>23</sup>Na, and the challenging <sup>28</sup>Mg (Fig. 1). For <sup>28</sup>Mg, new observations have revealed an anomaly in a specific type of collective transition, which has been explained by the SA–NCSM model as shape mixing.
- For neutrino experiments, it is important to reduce uncertainties related to the response of the detector nuclei to the neutrino. Response functions are now feasible for intermediate-mass nuclei, such as <sup>40</sup>Ar, the next-generation detector ingredient. As an illustration, BW has allowed the team to carry forward large-scale *ab initio* calculations for the response for <sup>20</sup>Ne (Fig. 2). Such calculations are important for studies of nuclear compressibility, which in turn inform the equation of state for neutron stars. In addition, various peaks in the response function can provide further insight into clustering substructures and collective degrees of freedom.
- *Ab initio* modeling of open-shell intermediate-mass nuclei is now feasible, and these, in turn, are used to study nuclei on the path of the X-ray burst nucleosynthesis, key to further understanding the origin and production of heavy elements. For example, *ab initio* calculations for <sup>20</sup>Ne, and especially the negative parity states (Fig. 2), as well as for <sup>15</sup>O, are used to calculate alpha decay probabilities, which enter as a critical input into reaction rates for (α,p) and proton capture reactions. Indeed, the alpha-induced reaction for <sup>15</sup>O is currently not well understood but has been suggested to have the largest effect on nucleosynthesis simulations. This *ab initio* modeling is important for providing accurate predictions for deformed and, in the future, heavy nuclei of interest to understand the r-process nucleosynthesis, one of the most challenging problems in astrophysics today.

WHY BLUE WATERS

Currently, only the BW system provides resources required for these *ab initio* studies of medium-mass isotopes with cutting-edge accuracy. To illustrate the level of complexity, applications to medium-mass nuclei require more than hundreds of exabytes of memory to store the Hamiltonian matrix. In order to capitalize on advances feasible with the SA–NCSM and Blue Waters’ capabilities, and with the help of the BW staff, the research team managed to improve scalability and performance of its code. As a result, the team’s largest production runs efficiently utilized 715,712 concurrent threads running on 22,366 Cray

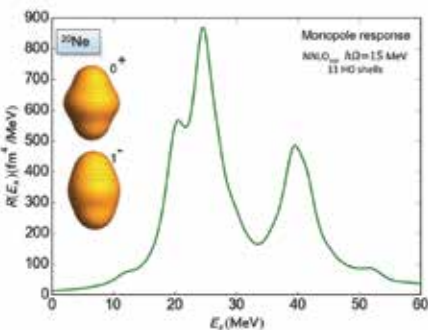


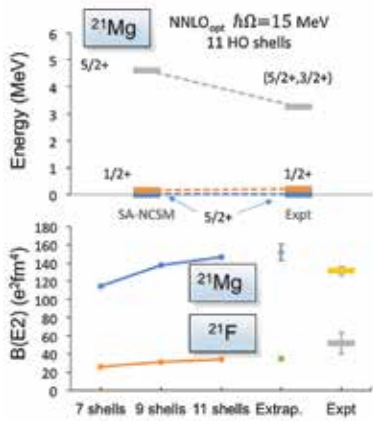
Figure 2: The first *ab initio* response of the deformed <sup>20</sup>Ne to an external probe vs. the energy transfer, revealing a clear evidence (the highest peak) for a giant monopole resonance and informing nuclear compressibility. Other emergent features detected within the SA–NCSM model are also shown: enhanced collectivity (0<sup>+</sup>) and clustering (1<sup>-</sup>).

XE6 compute nodes to solve the nuclear eigenvalue problem with a memory footprint of up to 750 terabytes of data. Clearly, the BW system represents a unique computational platform that already plays a crucial role in advancing *ab initio* nuclear theory toward new domains.

PUBLICATIONS & DATA SETS

- D. Langr, T. Dytrych, K. D. Launey, and J. P. Draayer, “Accelerating many-nucleon basis generation for high performance computing enabled *ab initio* nuclear structure studies,” *Int. J. High Perform. Comp. Appl.*, vol. 33, no. 3, pp. 522–533, 2019, doi: 10.1177/1094342019838314.
- D. Langr, T. Dytrych, J. P. Draayer, K. D. Launey, and P. Tvrdík, “Efficient algorithm for representations of U(3) in U(N),” *Comput. Phys. Commun.*, 2019, doi: 10.1016/j.cpc.2019.05.018.
- F. Knapp, T. Dytrych, D. Langr, and T. Oberhuber, “Importance basis truncation in the symmetry-adapted no-core shell model,” *Act. Phys. Pol. B*, vol. 50, pp. 541–547, 2019, doi: 10.5506/APhysPolB.50.541.
- R.B. Baker *et al.*, “Benchmark calculations of electromagnetic sum rules with symmetry-adapted basis and hyperspherical harmonics,” *Phys. Rev. C*, to be submitted, 2019.
- G. Sargsyan *et al.*, “Emergent collectivity in Mg isotopes from first principles,” in preparation, 2019.
- P. Ruotsalainen *et al.*, “Isospin symmetry in *B*(E2) values: Coulomb excitation study of <sup>21</sup>Mg,” *Phys. Rev. C*, vol. 99, p. 051301 (R), 2019, doi: 10.1103/PhysRevC.99.051301.
- M. Burrows *et al.*, “*Ab initio* folding potentials for nucleon-nucleus scattering based on NCSM one-body densities,” *Phys. Rev. C*, vol. 99, p. 044603, 2019, doi: 10.1103/PhysRevC.99.044603.
- J. Henderson *et al.*, “Testing microscopically derived descriptions of nuclear collectivity: Coulomb excitation of <sup>22</sup>Mg,” *Phys. Lett. B*, vol. 782, p. 468, 2018, doi: 10.1016/j.physletb.2018.05.064.
- M. Burrows *et al.*, “*Ab initio* translationally invariant non-local one-body densities from a no-core shell-model theory,” *Phys. Rev. C*, vol. 97, p. 024325, 2018, doi: 10.1103/physrevc.97.024325.

Figure 1: First-ever *ab initio* descriptions of collective features in magnesium (Mg) isotopes. SA–NCSM calculations are performed for ultralarge model spaces using chiral potentials and compared to experiment: (upper panel) energy of the three lowest excitations of <sup>21</sup>Mg and (lower panel) transition probabilities from the first excited state to the ground state.





# MAGNETIZED MODELS OF GIANT IMPACTS

**Allocation:** Exploratory/30 Knh, 300 Knh  
**PI:** Charles F. Gammie<sup>1</sup>  
**Co-PI:** Patrick Mullen<sup>1</sup>

<sup>1</sup>University of Illinois at Urbana–Champaign

## EXECUTIVE SUMMARY

The giant impact hypothesis suggests that about 4.5 billion years ago, a highly energetic, off-centered collision between a Mars-sized body and the proto-Earth formed a massive debris disk about the early Earth. It is from this orbiting disk of debris that the Moon is thought to have coalesced. Numerical simulations of the giant impact require massively parallel, high-resolution, multiphysics simulations. The research team’s models, applied to the Blue Waters supercomputer, are unique in that they are the first to address the dynamical importance of any magnetic field the impactor and/or proto-Earth may have possessed. These simulations have found that the natural consequence of a magnetized Moon-forming giant impact is a disk of debris hosting a toroidal magnetic field. A series of high-resolution numerical experiments also demonstrated that magnetic field strengths are amplified by as many as three to four orders of magnitude following the impact and early disk evolution.

## RESEARCH CHALLENGE

Shortly after the formation of the solar system, the giant impact hypothesis suggests that a Mars-sized impactor, Theia, struck the proto-Earth in an off-centered collision. This giant impact sent liquid and vaporized silicates into orbit around the proto-Earth, forming a protolunar disk from which the moon is thought to have coalesced. Three-dimensional simulations of this impact have been studied extensively; however, all of them have neglected the potential role of magnetic fields [1,2]. Therefore, the research group employed the Blue Waters supercomputer to study the dynamical importance of magnetic fields in the giant impact and the early evolution of the protolunar disk.

## METHODS & CODES

The team applied the astrophysical magnetohydrodynamics code Athena++ [6]. Athena++, written in C++, operates in a purely Eulerian framework and uses a hybrid parallelization model

based on OpenMP/MPI. The code employs a task-based execution model: each thread works on whichever tasks are presently available during a given timestep. Athena++ is modular; the included physics (such as magnetic fields, self-gravity, equation of state, diffusion, and the like) in addition to the choice of integrator/reconstruction method and coordinate system, are set as configure options at build time.

## RESULTS & IMPACT

The research team’s simulations initialize the impactor and proto-Earth with dipole magnetic fields. The Moon-forming giant impact wraps and winds these magnetic field lines into a toroidal configuration in the debris disk. Shear and turbulence in the disk amplify these field strengths by nearly four orders of magnitude in approximately one month after initial contact [4]. Thus, with even a modest initial field strength for the impactor and proto-Earth, these simulations demonstrate that magnetic fields could become dynamically important in a short timescale compared to the lifetime of the protolunar disk. In particular, magnetic fields may drive a fluid instability in the disk that causes turbulence to mix vaporized rock between the protolunar disk and the Earth. Such magnetic turbulence may help explain the isotopic similarities between the Earth and the Moon [4].

## WHY BLUE WATERS

A large computational domain is required to contain the giant impact. In addition, a high-resolution is necessary to accurately model dynamically important shocks and capture the onset of turbulence in the disk. Further, long integrations are required to evolve the simulation from initial impact to approximately one month after the formation of the protolunar disk. The Blue Waters supercomputer enables these large-scale, multiphysics simulations in which the team has applied as many as 128 Blue Waters XE nodes to a 1,024<sup>3</sup> (~200-km linear resolution) simulation.

## PUBLICATIONS & DATA SETS

P. D. Mullen and C. F. Gammie, “Magnetized models of giant impacts,” in preparation, 2019.

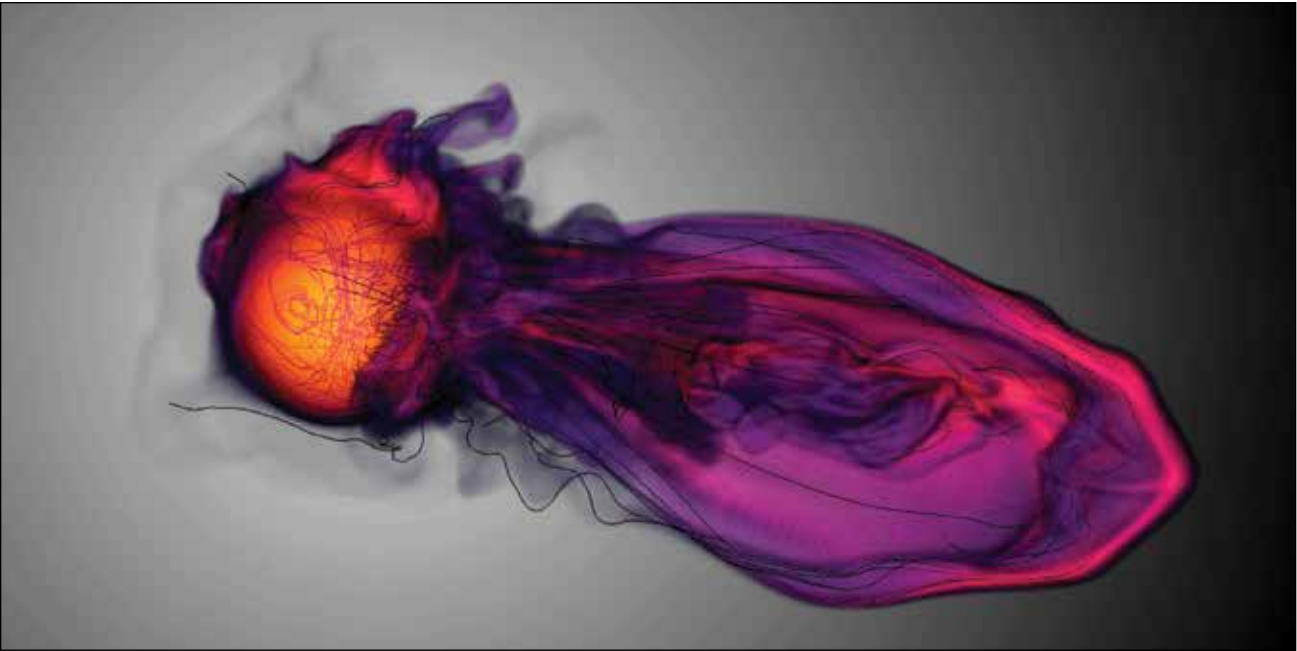


Figure 1: Tidal arm of impact debris ejected approximately one hour after a magnetized Moon-forming giant impact. The log of the composite density is shown in color while black rods show magnetic field lines. Visualization performed with VisIt [3].



COSMIC REIONIZATION ON COMPUTERS

Allocation: NSF PRAC/535 Knh  
PI: Nickolay Gnedin<sup>1</sup>

<sup>1</sup>University of Chicago

EXECUTIVE SUMMARY

Cosmic reionization—the process of ionization of the bulk of cosmic gas by ultraviolet radiation from the first galaxies and quasars—is the last frontier of modern cosmology. The Cosmic Reionization on Computers (CROC) project produced several numerical simulations of reionization that self-consistently modeled all relevant physics from radiative transfer to gas dynamics and star formation in simulation volumes of over 100 comoving megaparsecs (Mpc). This was necessary to model a representative sample of high-mass galaxies, and a spatial resolution approaching 100 parsecs in physical units was necessary to reliably model star formation in galaxies. CROC simulations, therefore, cover the full range of spatial, temporal, and mass scales important for studying reionization. The largest CROC simulations required Blue Waters'-class machines to complete.

RESEARCH CHALLENGE

Study of cosmic reionization has been highlighted by the last Decadal Survey as one of the most promising areas of astrophysical research. Because of the observational constraints on reionization, theoretical modeling, including numerical simulations, plays a relatively larger part in reionization studies than in many other fields of modern astrophysics. While the first simulations of reionization were attempted nearly 20 years ago, major breakthroughs in this field are only possible with modern petascale supercomputing platforms.

Taking advantage of this technological progress, the research team initiated a CROC project. This aims, over the course of several years, to produce numerical simulations of reionization that model self-consistently all relevant physics, ranging from radiative transfer to gas dynamics and star formation, in simulation

volumes of over 100 comoving Mpc, which is necessary to model a representative sample of high-mass galaxies, and with a spatial resolution approaching 100 parsecs in physical units, which is necessary to reliably model star formation in galaxies. These simulations, therefore, cover the full range of spatial, temporal, and mass scales important for studying reionization.

The primary motivation for focusing on reionization now is the expected major advance in observational capabilities: The James Webb Space Telescope (the next flagship NASA mission) is scheduled to launch in 2021, and studying galaxies responsible for cosmic reionization is one of its primary goals. Studies of intergalactic gas will be propelled forward by the deployment of 30-meter telescopes, several of which will become operational in the first half of the next decade. Other novel observational tools will follow in the second half of the next decade.

METHODS & CODES

In order to reach the required dynamic range, the research team relied on the adaptive mesh refinement technique. The simulations are run with the adaptive refinement tree code, a publicly available cosmological simulation code developed and supported by the research group. The code includes all necessary physical modules for simulating cosmic reionization (dynamics of dark matter and gas, atomic processes, interstellar chemistry, star formation and stellar feedback, radiative transfer of ionizing and UV radiation). ART is MPI+OpenMP parallel and scales perfectly on this type of simulation to about 50,000 cores, with parallel scaling remaining acceptable to about 100,000 cores.

RESULTS & IMPACT

CROC simulations are defining the state of the art in this field. By virtue of including all the relevant physics and extending to volumes that are required to properly capture the process of reionization, they are creating a physically plausible model of cosmic reionization that can be matched against any existing observational data. Such comparisons have been made by the research group in the last several years in a series of papers. However, the most constraining observational data set that exists today is the distribution of optical depth in the spectra of distant quasars. Over 100 quasars during the reionization epoch have been discovered so far, and for almost 70 of them, high-resolution and high-quality spectra exist. These data probe the largest spatial scales relevant for reionization of about 70 comoving Mpc.

During the second year of this project, the team focused on two separate scientific questions: (1) how galaxies influence and shape the distribution of ionized gas on large scales; and (2) how the brightest known quasars affect galaxies and gas in their vicinity. Both of these questions are at the frontier of modern research and promise to develop into new, independent areas of study in the next several years. An example of cosmic gas distribution from one of the team's simulations is shown in Fig. 1.

WHY BLUE WATERS

Only two supercomputers in the United States were available for these simulations, which are both memory- and communication-intensive. The largest simulations use 8 billion particles and around 50 billion cells in the adaptively refined mesh. Thus, they require petascale computing capabilities. Blue Waters was much faster than the other available machine and had better I/O support. Hence, it was the platform of choice.

PUBLICATIONS & DATA SETS

H. Chen and N. Y. Gnedin, "Constraints on the duty cycles of quasars at  $z \sim 6$ ," *Astrophys. J.*, vol. 868, no. 2, p. 126, Nov. 2018.

E. Garaldi, N. Y. Gnedin, and P. Madau, "Constraining the tail end of reionization using Ly $\alpha$  Transmission Spikes," *Astrophys. J.*, vol. 876, no. 1, p. 31, April 2019.

H. Zhu, C. Avestruz, and N. Y. Gnedin, "Cosmic reionization on computers: Reionization histories of present-day galaxies," *Astrophys. J.*, vol. 882, no. 2, p. 152, Sept. 2019.

H. Chen, "The role of quasar radiative feedback on galaxy formation during cosmic reionization," in preparation, 2019.

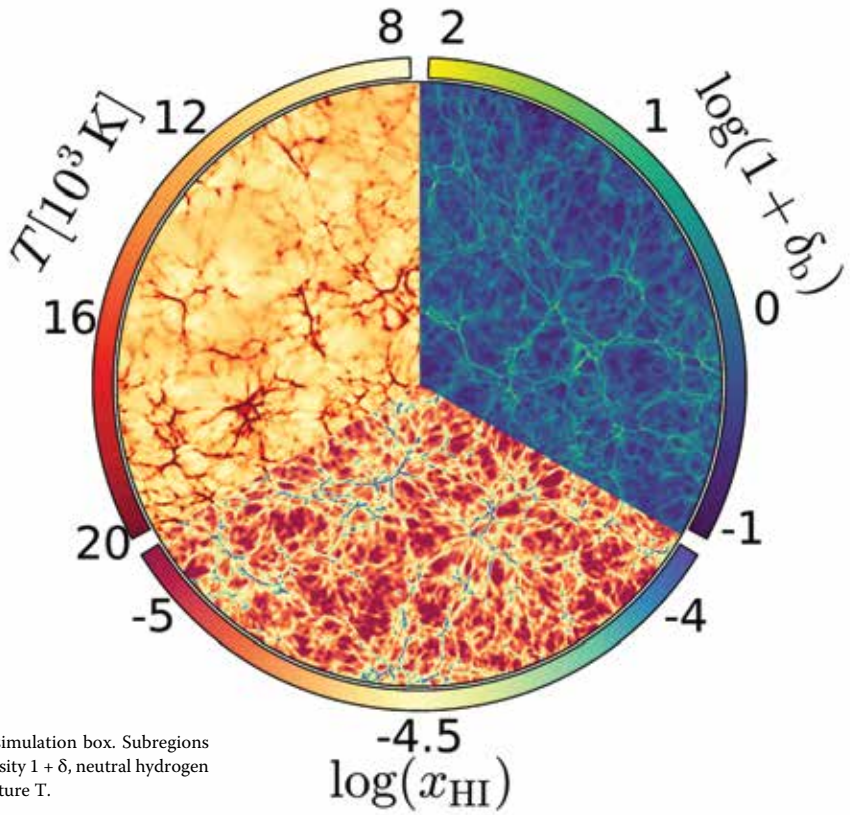


Figure 1: Slice through the simulation box. Subregions show the cosmic gas overdensity  $1 + \delta$ , neutral hydrogen fraction  $x_{\text{HI}}$ , and gas temperature  $T$ .



# ELUCIDATING THE ALIGNMENT MECHANISM FOR BLACK HOLE ACCRETION DISKS SUBJECTED TO LENSE–THIRRING TORQUES

**Allocation:** GLCPC/288 Knh  
**PI:** John Hawley<sup>1</sup>  
**Co-PI:** Julian H. Krolik<sup>2</sup>

<sup>1</sup>University of Virginia  
<sup>2</sup>Johns Hopkins University

## EXECUTIVE SUMMARY

The research team uses Blue Waters to study astrophysical accretion onto a spinning black hole in which there is a misalignment (tilt) between the orbital axis of the incoming gas and the black hole rotation axis. Astrophysicists have long expected that an initially misaligned orbiting accretion disk would align with the black hole’s spin axis at some location near the hole. A detailed understanding of this alignment process has, however, been limited owing to the assumption of a phenomenological viscosity to describe the internal dissipation necessary for alignment. The team’s simulations capture the physical internal stress due to magnetohydrodynamic turbulence with no reliance on phenomenological viscosity; such simulations are only possible with the high grid resolution made feasible by Blue Waters. The investigation probes how a time-steady transition might be achieved between an inner disk region aligned with the equatorial plane of the central mass’s spin and an outer region orbiting in a different plane.

## RESEARCH CHALLENGE

Accretion disks occur in a wide variety of astrophysical systems. Whenever the disk’s angular momentum is oblique to the angular momentum of the central object(s), a torque causes rings within the disk to precess, twisting and warping it. Because the torque weakens rapidly with increasing radius, it has been thought that some unspecified “friction” brings the inner portions of such disks into alignment with the equator of the central object. Despite considerable theoretical effort, researchers are still unable to predict the alignment radius for a disk. Nearly all previous work on this topic has assumed that such a disk’s internal stresses can be described by a parameterized isotropic viscosity. However, there is a well-established physical mechanism for internal stresses in accretion disks—correlated magnetohydrodynamic (MHD) turbulence, driven by the magnetorotational instability [1,2].

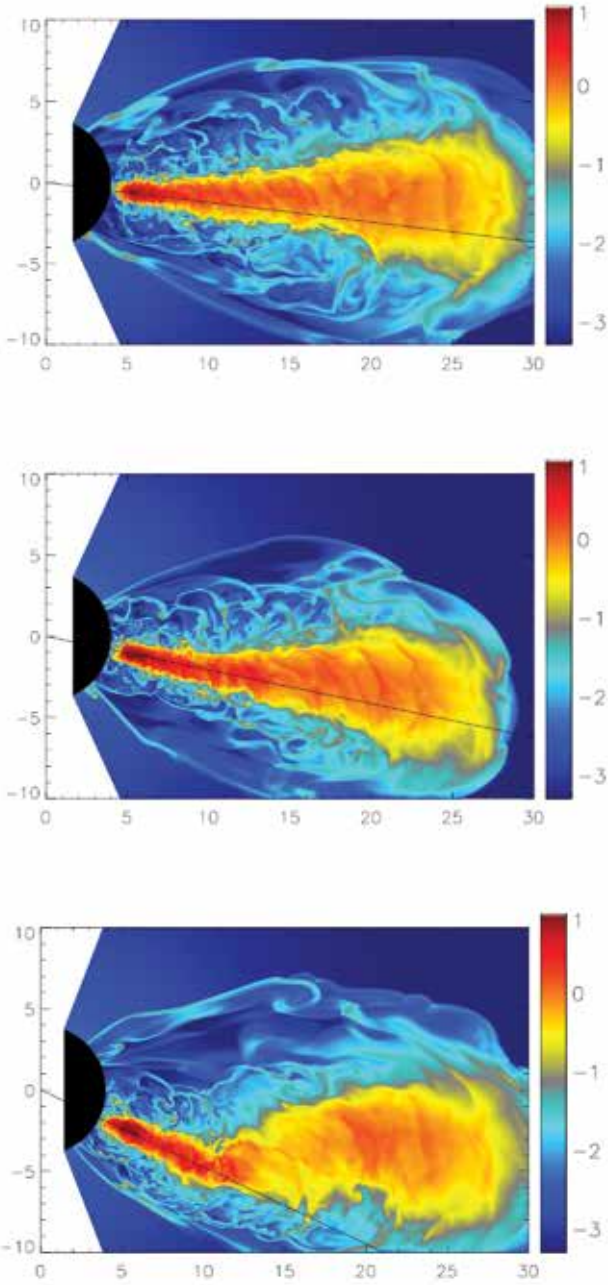


Figure 1: Logarithmic density contour for disk tilt angles of 6° (top), 12° (middle), and 24° (bottom) after 20 orbits of applied torque. The inner disk regions have aligned with the black hole equatorial plane (overlaid line), and the transition from alignment to outer obliquity occurs at approximately the same radius.

Simulating MHD turbulence in a tilted disk is demanding. To evolve MHD turbulence, a simulation must have a timestep that is very short compared to an orbital time, whereas the precession timescale of orientation transition is many orbital periods long. This project’s numerical simulations investigate how twisted disks align when their mechanics are described only in terms of real forces, including MHD turbulence. The aim is to develop a predictive model for the location alignment front in disks subjected to alignment torques.

## METHODS & CODES

The research team used a simplified disk model first studied in [3], consisting of an isothermal disk orbiting a point-mass in Newtonian gravity with a Keplerian angular velocity distribution and including only a lowest-order post-Newtonian term to represent the relativistic Lense–Thirring torque. This idealized model allowed the team to focus on the important physical processes governing alignment; these can be studied in isolation and in detail. They examined the influence of black hole tilt on the alignment process, evolving three models where the tilt angle is changed without altering anything else. This approach stands in contrast to those efforts that simulate model disks in the full general-relativistic context.

The simulations were done with a Fortran-95 version of *Zeus*, an operator-split code that solves the equations of compressible MHD by time-explicit finite-differencing [4] using the constrained transport algorithm [5] to preserve the divergence-free property of the magnetic field. *Zeus* uses domain decomposition and MPI for parallelization.

## RESULTS & IMPACT

The project team investigated the effects of black hole tilt angle on accretion disk alignment, modeling three initial black hole tilts of 6°, 12°, and 24° using both magnetohydrodynamic and (inviscid) hydrodynamic evolutions. By considering a single sound speed and a simple underlying disk model, the researchers isolated the effect of tilt amplitude, which is found to have a limited influence on alignment. The imposed tilt angle simply sets the “unit,” and the resulting dynamics are determined by ratios in terms of that unit. The radial shape of the transition region between the aligned inner disk and unaligned outer disk, as well as the location of the head of the alignment front, are very nearly identical once the black hole tilt (see Fig. 1) is scaled out. Tilt does have some secondary effects. The steady state warp within the alignment front increases at a slightly faster rate with tilt than simple proportionality. Another effect is a small decrease in the initial alignment front velocity with increasing tilt.

Purely hydrodynamic disks behave similarly, but because the amplitude of the disk warp is a function of the tilt angle for a given disk thickness,  $h/r$ , higher tilt angles have larger relative warp values. From this it follows that bending waves produced in the disk are increasingly nonlinear, and hence more dissipative. Bending waves propagating through the disk without hindrance pro-

mote solid-body precession which, in turn, can end alignment at an earlier time compared to an MHD disk where the turbulence inhibits wave propagation [6]. In this work, the hydrodynamic models saw a rapid diminution of precession phase gradient in the outer disk, which brought alignment to a halt and reversed the motion of the alignment front.

Some simulations of tilted disks have either seen evidence for a possible separation of the inner and outer disk [7], or clear disk breaking, whether the central object is a binary or a spinning black hole (e.g., [8]). In this project’s simulations, even when the initial misalignment is 24°, which, for the sound speed studied, is four vertical scale heights at the disk fiducial radius, the surface density remains a smooth function of radius; i.e., the team found no examples in which the disk inner-aligned and outer-misaligned regions separated, or “broke.”

## WHY BLUE WATERS

The research team has used Blue Waters to compute new thin disk simulations subject to Lense–Thirring torque with unprecedented resolution to explore the mechanisms behind, and scaling properties of, disk alignment. The unique high-performance capabilities of Blue Waters enabled key maximum-resolution simulations.

## PUBLICATIONS & DATA SETS

J. F. Hawley and J. H. Krolik, “Sound speed dependence of alignment in accretion disks subjected to Lense–Thirring torques,” *Astrophys. J.*, vol. 866, no. 1, 2018, doi: 10.3847/1548-4357/aadf90.  
J. F. Hawley and J. H. Krolik, “Tilt dependence of alignment in accretion disks subjected to Lense–Thirring torques,” *Astrophys. J.*, vol. 878, 2019, doi: 10.3847/1538-4357/ab1f6e.



UNDERSTANDING THE ORIGINS OF THE STARS AND GALAXIES IN OUR UNIVERSE

**Allocation:** NSF PRAC/4,940 Knh  
**PI:** Philip F. Hopkins<sup>1</sup>  
**Collaborators:** Claude–André Faucher–Giguère<sup>2</sup>, Dušan Kereš<sup>3</sup>, Eliot Quataert<sup>4</sup>, Norman Murray<sup>5</sup>, Mike Boylan–Kolchin<sup>6</sup>, James Bullock<sup>7</sup>, Andrew Wetzel<sup>3</sup>, Robyn Sanderson<sup>8</sup>, Chris Hayward<sup>9</sup>, Robert Feldman<sup>10</sup>, Christine Corbett–Moran<sup>11</sup>

<sup>1</sup> California Institute of Technology	<sup>5</sup> Canadian Institute for Theoretical Astrophysics	<sup>9</sup> Flatiron Institute
<sup>2</sup> Northwestern University	<sup>6</sup> University of Texas at Austin	<sup>10</sup> ETH Zurich
<sup>3</sup> University of California, San Diego	<sup>7</sup> University of California, Irvine	<sup>11</sup> NASA JPL
<sup>4</sup> University of California, Berkeley	<sup>8</sup> University of Pennsylvania	

EXECUTIVE SUMMARY

This research uses Blue Waters to explore the origins of galaxies and stars and the nature of dark matter. At a fundamental level, the study of galaxies and stars seeks to answer the question, “How did we get from the Big Bang to the Milky Way?” This is an immensely challenging question, involving the interplay among gravity, fluid dynamics, radiation, matter, and stars exploding as supernovae, giving rise to explosive outflows of material from galaxies that can reach across the observable Universe. The physics is chaotic and wildly nonlinear, and the range of timescales is tremendous (from one to ten billion years). As such, massive numerical simulations that can follow all of these processes are required. By using numerical simulations, the researchers have gained fundamental insights into why galaxies today look as they do and, in the process, have strongly constrained the allowed nature of the dark matter.

RESEARCH CHALLENGE

The program seeks to understand the origin and nature of galaxies, using massively parallel simulations that follow the birth and evolution of galaxies and stars from the very early Universe to the present day. The simulations model the origins, evolution, internal structure, and observable properties of galaxies ranging in size from the smallest observed “dwarf” galaxies (with just a few thousand stars) to the Milky Way and Andromeda (the “Local Group”). Deep and fundamental questions remain unsolved in this area, including, simply, “How did we get from the Big Bang to the Milky Way?” as well as, “Why did the Universe form so few stars (compared to what it could have done)?” Further questions include, “Why did stars form where and when they did?” and “How can we use galaxies to probe the fundamental nature of dark matter?” At the heart of these issues lies the fact that stars, once they form, are not passive actors within a galaxy: they shine and emit tremendous amounts of energy in the form of light (radiation), stellar winds, and supernova explosions. This energy can blow material out of the galaxy entirely and completely alter the evolutionary history of galaxies.

But these stellar and galactic processes remain poorly understood, in large part because they: (1) couple very small and very large scales in the Universe and require simulations with enormous dynamic range to model them, and (2) involve a diverse range

of physics including (but not limited to) gravity, fluid dynamics, magnetic fields, conduction and viscosity, radiation–matter interactions, interstellar chemistry, and stellar evolution. The simulations run on Blue Waters incorporated all of these processes into the highest-resolution simulations yet completed, allowing the research team to address these questions for the first time at the level of detail needed to make observable predictions. Billions of dollars are being invested in new telescopes and instruments to explore these questions experimentally; these simulations are critical tools to make detailed predictions and leverage these transformative observations.

METHODS & CODES

The researchers have run a large suite of cosmological, high-resolution simulations including detailed treatments of the physics of the interstellar medium, star formation, feedback in radiation and supernovae, magnetic fields, and cosmic rays. The simulations use the feedback in realistic environments (FIRE) physics methods in the GIZMO code, a new massively parallel multimethod, hybrid Lagrangian–Eulerian finite-element, high-order, radiation-hydrodynamics code (unique in numerical methods employed and physics supported).

RESULTS & IMPACT

These cosmological simulations target galaxies from the faintest dwarfs through to the Milky Way and run at the ultrahigh resolution and realism required to interpret the next generation of observations. The petascale resources of Blue Waters allow the researchers to resolve each galaxy with approximately one billion particles and follow them self-consistently over their entire history in realistic cosmological settings. When the interstellar medium is resolved into dense molecular clouds, massive stars naturally form and then inject large quantities of energy and momentum into the surrounding medium via “stellar feedback”; this feedback is critical to produce realistic galaxies and generate the powerful galactic winds observed, radically altering the baryon cycle between galaxies and the circumgalactic medium. The simulations model the physics of galaxy formation with unprecedented realism, uniquely incorporating not only all of the important stellar feedback mechanisms (radiation pressure, photo-heating, stellar winds, supernovae, cosmic rays) but also magnetic fields,

physical (anisotropic) Braginskii conduction and viscosity, passive scalar (metal) diffusion, and explicit, multiwavelength radiation hydrodynamics.

This work represents the culmination of several years of research supported by the National Science Foundation, and has been critical in enabling the science of the FIRE project: a collaboration of theorists across 13 different major institutions. The program has revealed fundamental new insights into how stars alter their galactic environments and has changed observational inferences about the nature of dark matter in those galaxies. The simulations are also being used to support an outreach component involving high school students and teachers, and undergraduate students, as well as a large science team using these simulations. The simulations have already been utilized to make predictions specifically for next-generation telescopes including (but not limited to) JWST, LSST, Gaia, and HST, in order to constrain the origin of the heavy elements in the Universe, and test theories of galaxy and star formation, the reionization history of the early Universe, the effects of fundamental plasma physics in the circum- and intergalactic medium, and the nature of cold dark matter.

WHY BLUE WATERS

Blue Waters was critical for this research because the enormous computational challenges detailed in this report required more than 100 million CPU-hours on tens of thousands of processors with tens of terabytes of active memory to store and evolve the immensely complex physical systems, which produced petabytes of data products. No other facility could have enabled this research.

PUBLICATIONS & DATA SETS

E.–R. Moseley, J. Squire, and P. F. Hopkins, “Non-linear evolution of instabilities between dust and sound waves,” *Mon. Notices Royal Astron. Soc.*, vol. 489, p. 325, 2019.  
T. K. Chan, D. Kereš, P. F. Hopkins, E. Quataert, K.-Y. Su, C. C. Hayward, C.–A. Faucher-Giguère, “Cosmic ray feedback in the FIRE simulations: constraining cosmic ray propagation with GeV  $\gamma$ -ray emission,” *Mon. Notices Royal Astron. Soc.*, vol. 488, p. 3716, 2019.  
M. Y. Grudić and P. F. Hopkins, “The elephant in the room: the importance of the details of massive star formation in molecular clouds,” *Mon. Notices Royal Astron. Soc.*, vol. 488, p. 2970, 2019.  
M. Y. Grudić, P. F. Hopkins, E. J. Lee, N. Murray, C.–A. Faucher-Giguère, and L. C. Johnson, “On the nature of variations in

the measured star formation efficiency of molecular clouds,” *Mon. Notices Royal Astron. Soc.*, vol. 488, p. 1501, 2019.  
Z. Hafen *et al.*, “The origins of the circumgalactic medium in the FIRE simulations,” *Mon. Notices Royal Astron. Soc.*, vol. 488, p. 1248, 2019.  
K.–Y. Su *et al.*, “The failure of stellar feedback, magnetic fields, conduction, and morphological quenching in maintaining red galaxies,” *Mon. Notices Royal Astron. Soc.*, vol. 487, p. 4393, 2019.  
S. Garrison–Kimmel *et al.*, “The Local Group on FIRE: Dwarf galaxy populations across a suite of hydrodynamic simulations,” *Mon. Notices Royal Astron. Soc.*, vol. 487, p. 1380, 2019.  
V. P. Debattista *et al.*, “Formation, vertex deviation, and age of the Milky Way’s bulge: Input from a cosmological simulation with a late-forming bar,” *Mon. Notices Royal Astron. Soc.*, vol. 485, p. 5073, 2019.  
D. Guszejnov, P. F. Hopkins, and A. S. Graus, “Is it possible to reconcile extragalactic IMF variations with a universal Milky Way IMF?” *Mon. Notices Royal Astron. Soc.*, vol. 485, p. 4852, 2019.  
D. Seligman, P. F. Hopkins, and J. Squire, “Non-linear evolution of the resonant drag instability in magnetized gas,” *Mon. Notices Royal Astron. Soc.*, vol. 485, p. 3991, 2019.  
C.–L. Hung *et al.*, “What drives the evolution of gas kinematics in star-forming galaxies?” *Mon. Notices Royal Astron. Soc.*, vol. 482, p. 5125, 2019.  
P. F. Hopkins and M. Y. Grudić, “Numerical problems in coupling photon momentum (radiation pressure) to gas,” *Mon. Notices Royal Astron. Soc.*, vol. 483, no. 3, pp. 4187–4196, 2019.  
A. Lamberts *et al.*, “Predicting the binary black hole population of the Milky Way with cosmological simulations,” *Mon. Notices Royal Astron. Soc.*, vol. 480, no. 2, pp. 2704–2718, 2018.  
C. Corbett–Moran, M. Y. Grudić, and P. F. Hopkins, “The effects of metallicity and cooling physics on fragmentation: Implications on direct-collapse black hole formation,” *Mon. Notices Royal Astron. Soc.*, submitted for publication, 2018, arXiv:1803.06430.  
R. Sanderson *et al.*, “Reconciling observed and simulated stellar halo masses,” *The Astrophysical Journal*, vol. 869, no. 1, p. 12, 2018.  
S. Garrison–Kimmel *et al.*, “The origin of the diverse morphologies and kinematics of Milky Way-mass galaxies in the FIRE-2 simulations,” *Mon. Notices Royal Astron. Soc.*, vol. 481, no. 3, pp. 4133–4157, 2018.  
P. F. Hopkins, “A new public release of the GIZMO code,” arXiv preprint, 2017, arXiv:1712.01294.

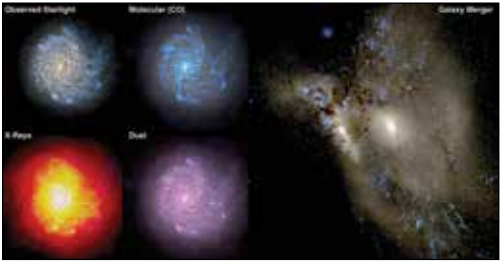


Figure 1: Simulations of a Milky Way-like galaxy. Observed starlight (mock images) is shown with overlaid intensity maps showing mock carbon monoxide (molecular gas), X-rays, and dust emission. The “Galaxy Merger” portion shows a mock Hubble image during a galaxy collision, where violent “bursts” of star formation are triggered.



Figure 2: Mock Hubble map of a simulated galaxy, as seen from a Sun-like star. Filamentary molecular cloud complexes and young star clusters are visible within the “Milky Way” (galactic disk). The research team’s combination of physics and resolution allows them to model galactic structure with unprecedented realism.



# DEEP LEARNING AT SCALE FOR THE CONSTRUCTION OF GALAXY CATALOGS WITH THE DARK ENERGY SURVEY

**Allocation:** Illinois/600 Knh  
**PI:** Eliu Huerta<sup>1</sup>  
**Co-PIs:** Gabrielle Allen<sup>1</sup>, Roland Haas<sup>1</sup>, Ed Seidel<sup>1</sup>, Zhizhen Zhao<sup>1</sup>  
**Collaborators:** Asad Khan<sup>1</sup>, Sibow Wang<sup>1</sup>, Robert Gruendl<sup>1</sup>, Elise Jennings<sup>2</sup>, Huihuo Zheng<sup>2</sup>

<sup>1</sup>University of Illinois at Urbana–Champaign  
<sup>2</sup>Argonne Leadership Computing Facility

## EXECUTIVE SUMMARY

The scale of ongoing and future electromagnetic surveys poses formidable challenges to classifying astronomical objects. Pioneering efforts on this front include citizen science campaigns adopted by the Sloan Digital Sky Survey (SDSS). SDSS data sets have recently been used to train neural network models to classify galaxies in the Dark Energy Survey (DES) that overlap the footprint of both surveys. The research team has demonstrated that knowledge from deep learning algorithms, pretrained with real-object images, can be transferred to classify galaxies that overlap both SDSS and DES surveys, achieving a state-of-the-art accuracy of 99.6%. In addition, the team has demonstrated that this process can be completed within just eight minutes using distributed training. The researchers also used their neural network classifier to label 10,000 DES galaxies that do not overlap previous surveys. Further, the team has shown that these new data sets can be combined with recursive training to create DES galaxy catalogs in preparation for the Large Synoptic Survey Telescope era.

## RESEARCH CHALLENGE

The classification of astrophysical objects has been pursued in the past using a diverse set of tools. For instance, galaxies

have been classified using their photometric properties, achieving classification accuracies of around 85% [1]. Other methods of classifying galaxies according to their morphology have taken into account their physical properties across multiple wavelengths. For instance, the method introduced in [2] considered a sample of galaxies from the Sloan Digital Sky Survey (SDSS) [3] using the five SDSS filters (u, g, r, i, z) and then used a combination of shapelet decomposition and principal components analysis (PCA). Other methods for galaxy classification include Concentration–Asymmetry–Smoothness [4] and machine learning, including artificial neural networks and PCAs [5].

In recent years, citizen science campaigns have played a key role in classifying thousands of celestial objects in astronomical surveys. SDSS is an archetypical example of a successful approach to classifying hundreds of thousands of galaxies. As electromagnetic surveys continue to increase their depth and coverage, campaigns of this nature may lack scalability. For example, within six years of operation, the Dark Energy Survey (DES) [6] observed over three hundred million galaxies, a number that will be surpassed by the observing capabilities of the Large Synoptic Survey Telescope (LSST) [7]. In brief, there is a pressing need to explore new approaches to maximize the science throughput of

Figure 1: Top panels: Labeled images of the SDSS training data set. Bottom panels: Sample of galaxies from SDSS DR7 and the corresponding cross-matched galaxies from DES DR1.

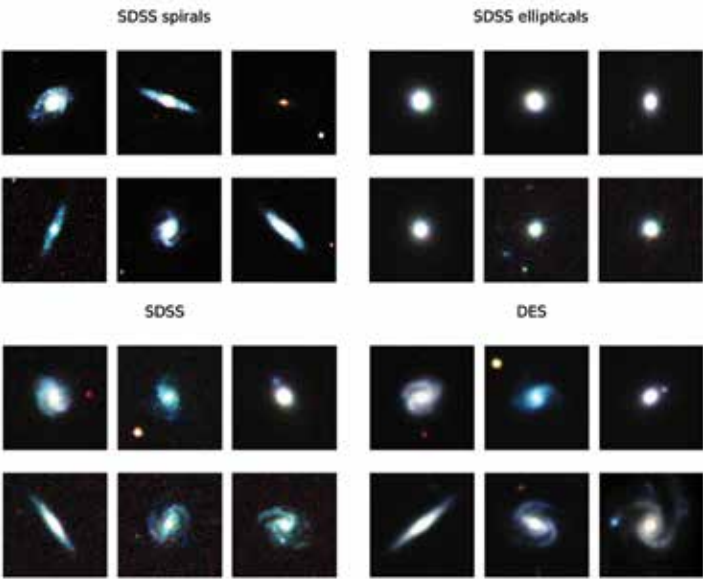


Figure 2: t-SNE visualization of the clustering of HP SDSS and DES test sets, and unlabeled DES test set.

next-generation electromagnetic surveys. A promising paradigm is the convergence of deep learning and large-scale computing to address the imminent increase in data volume, complexity, and latency of observations of LSST-type surveys.

## METHODS & CODES

The research team used a subset of SDSS Data Release (DR) 7 images for which it had high confidence classifications through the Galaxy Zoo project; *i.e.*, the team only chose galaxies with a debiased probability greater than 0.985 for combined spirals and 0.926 for ellipticals, respectively. Samples of these images are shown in Fig. 1. The team chose these cutoff thresholds to ensure that: (1) the galaxies used for training the neural network had robust and accurate classifications; and (2) the representation of both classes in the training and test data sets were balanced. The team then divided these images into three separate data sets for training, validation, and testing. The validation set was used to monitor the accuracy and loss when training and fine-tuning the deep neural network, and hence served to optimize hyperparameters such as learning rate and number of epochs for training.

Two test sets were carefully constructed so that the images in each set lay in both the SDSS and DES footprints. The first test set consisted of images with a Galaxy Zoo classification confidence similar to that of the training set, *i.e.*, a high probability cut-off was introduced. This test set was hence labeled High Probability (HP) Test Set, and there were two versions, one for each survey: HP SDSS and HP DES. Just as in the training set, the images for SDSS were obtained from SDSS DR7 and the corresponding images for DES were obtained from the DES DR1 data release. Furthermore, a second test set was created without introducing any probability thresholds on the Galaxy Zoo classification confidence. This set consisted of almost all galaxies lying in both the SDSS and DES footprints, and hence was labeled Full Overlap (FO) Test Set. Again, there were two versions: FO SDSS and FO DES. The motivation behind creating this second test set was that the galaxy profiles in the unlabeled DES data set would more closely match those in the FO test sets. Hence, the FO test set served as a good evaluation metric of the performance of the neural net on the ultimate task of classifying all unlabeled galaxies in the DES catalogue.

The team used open-source software stacks for its studies. The deep learning APIs used were Keras [8] and TensorFlow [9]. For the classification problem, the team did transfer learning starting with the Xception model [10], which had been pretrained with the ImageNet data set [11]. The researchers chose this neural network model because it outperforms many other state-of-the-art neural network models, including Inception-v3, ResNet-152, and VGG16, on the ImageNet validation data set, and it has been suggested that better ImageNet architectures are capable of learning better transferable representations [12]. More importantly, the research team carried out several experiments and found that Xception performed as well as or nominally better on the validation and testing galaxy data sets compared to many other state-of-the-art architectures.

## RESULTS & IMPACT

This is the first application of deep transfer learning combined with distributed training for the classification of DES galaxies that overlap the footprint of the SDSS survey, achieving state-of-the-art accuracies of 99.6%. The research team has also used its neural network classifier to label over 10,000 DES galaxies that had not been observed in previous surveys. Using t-SNE visualizations (see [13] and Fig. 2), the research group found that deep transfer learning was effective to abstract morphological information from the galaxy images to clearly identify two distinct classes of galaxies in the unlabeled DES data set.

## WHY BLUE WATERS

Blue Waters was essential to extract and curate the data sets used to train, validate, and test the team’s neural network models at scale. Staff provided support to deploy the software stacks, to use them at scale, and to carry out all the analyses reported in this study.

## PUBLICATIONS & DATA SETS

A. Khan, E. A. Huerta, S. Wang, R. Gruendl, E. Jennings, and H. Zheng, “Deep learning at scale for the construction of galaxy catalogs with the Dark Energy Survey,” *Phys. Lett. B*, vol. 795, pp. 248–258, 2019, doi: 10.1016/j.physletb.2019.06.009.



CHARACTERIZATION OF NUMERICAL RELATIVITY WAVEFORMS OF ECCENTRIC BINARY BLACK HOLE MERGERS

Allocation: Illinois/325 Knh  
PI: Eliu Huerta<sup>1</sup>  
Co-PIs: Roland Haas<sup>1</sup>, Gabrielle Allen<sup>1</sup>, Ed Seidel<sup>1</sup>  
Collaborator: Sarah Habib<sup>1</sup>

<sup>1</sup>University of Illinois at Urbana–Champaign

EXECUTIVE SUMMARY

An accurate description of the physics of eccentric binary black holes (BBHs) throughout the late-inspiral, merger, and ringdown requires numerical relativity (NR). Once NR simulations are post-processed and extracted [1], it is necessary to quantify the eccentricity and other orbital parameters that uniquely identify them. Existing methods to measure eccentricity based on the BH trajectories in NR simulations do not provide a sound approach, given that these trajectories are gauge-dependent. The research team circumvents this limitation by introducing a gauge-invariant method that characterizes an NR waveform by comparing it to a large array of semianalytical waveforms, dubbed ENIGMA [2] waveforms, that are written in the gauge used to detect gravitational waves with the Laser Interferometer Gravitational-Wave Observatory detectors. The researchers quantify the circularization of eccentric BBHs near merger and quantify the impact of higher-order waveform modes in the morphology of eccentric NR waveforms. This study is timely and relevant to characterizing future observations of eccentric BBHs.

RESEARCH CHALLENGE

This research aims to quantify the eccentricity and mean anomaly of NR waveforms that describe eccentric BBH mergers using a gauge-invariant approach and to develop open source algorithms to conduct this analysis on Blue Waters to process catalogs of NR waveforms at scale. The work is of interest to the NR and gravitational wave (GW) astrophysics communities.

METHODS & CODES

The research team explored a variety of gauge-invariant objects to directly compare NR and ENIGMA (eccentric, nonspinning, inspiral, Gaussian-process merger approximant) waveforms. The researchers found that the dimensionless object  $M\omega$ , where  $\omega$  is the mean orbital frequency and  $M$  stands for the total mass of the BBH, provides a robust approach to capture the signatures of eccentricity. To compute  $M\omega$  for NR waveforms, the team uses the relation  $M\omega = \frac{1}{2} \frac{a-b-a/b}{a^2+b^2}$ , where  $a$  and  $b$  represent the plus and cross polarizations of an NR waveform  $h=a-1b$ . The ENIGMA waveform model produces this gauge-invariant quantity by providing, as input parameters, the mass-ratio of the BBH system and the

Figure 1: Inner product, indicated by the color bar, between eccentric binary black hole systems and their quasicircular counterparts. As the eccentric systems reach merger, marked at  $t=0M$ , eccentricity is gradually radiated away until the systems become effectively circular, which corresponds to an overlap value of 1 with quasicircular waveforms.

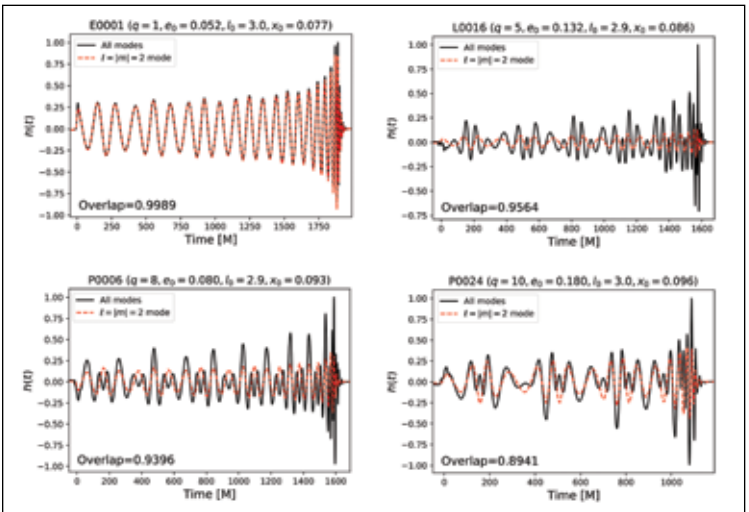
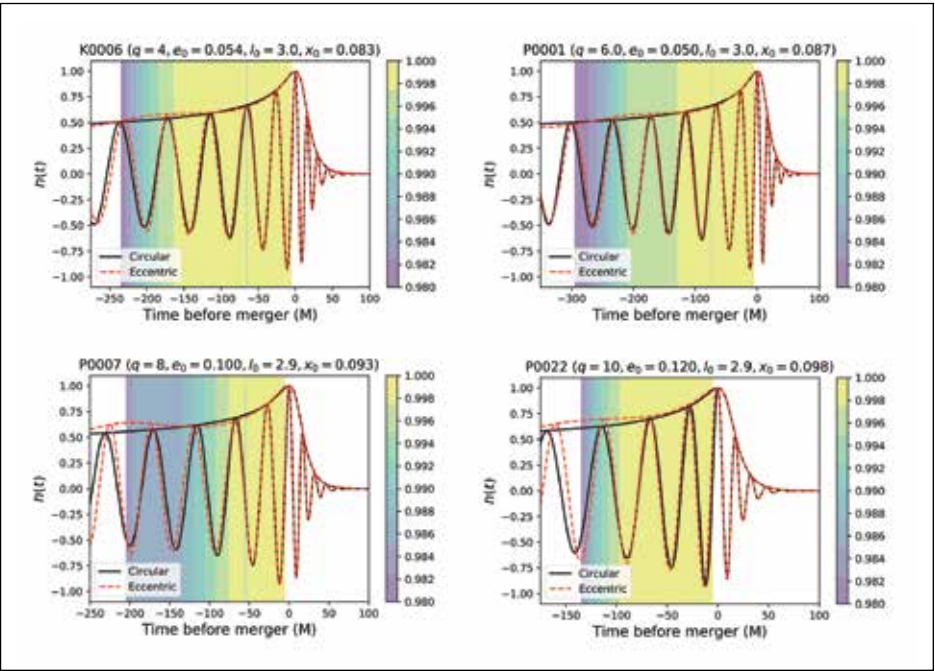


Figure 2: Overlap between numerical relativity waveforms that include either the modes  $(l, |m|) = \{(2, 2), (2, 1), (3, 3), (3, 2), (3, 1), (4, 4), (4, 3), (4, 2), (4, 1)\}$ , or just the  $l=|m|=2$  mode. Higher-order modes become more significant for asymmetric mass-ratio binary black holes.

initial eccentricity,  $e_0$ , and mean anomaly,  $l_0$ , of the system at a fiducial GW frequency  $f_0$  from which the waveform is produced.

The team removes junk radiation from the NR waveform by applying a Savitsky–Golay filter. Since input parameters are required to produce ENIGMA waveforms, seed values are initialized for  $f_0$  and  $e_0$ . The researchers provide an informed guess of the GW frequency using the relation  $f_0 \sim \omega_0/(\pi M)$ , where  $\omega(t=t_0)=\omega_0$ , and  $t_0$  is the time at which the NR waveform is free from junk radiation. Mean anomaly is initialized to  $\pi$ , a value manually determined to be optimal through verification of a few NR waveforms. Orbital eccentricity does not require a seed value since the range of possible values is consistent for all catalogued waveforms.

The algorithm starts with a grid search in the 2D parameter space of  $(f_0, e_0)$ , and iteratively refines it. The researchers densely sample the frequency range  $f \in [f_0 - 5\text{Hz}, f_0 + 5\text{Hz}]$  and the eccentricity range  $e \in [0.1, 0.3]$ . For each coordinate pair, an ENIGMA  $M\omega$  is produced using the specified  $(f_0, e_0)$  values and the seeded  $l_0$ . The resulting  $M\omega$  time evolution is then compared to that of the original NR waveform. Parameters are chosen that minimize the difference between two properties of the ENIGMA and NR evolutions: time duration of the first orbital cycle and the maximum change in  $M\omega$  during the first cycle. Throughout the entire search,  $l_0$  is held constant.

After completing the grid search, the chosen  $(f_0, e_0)$  parameters are further refined iteratively. In this stage, initial GW frequency and orbital eccentricity are independently varied stepwise to increase precision. The team used this method to characterize the 89 NR waveforms presented in [3].

RESULTS & IMPACT

This research addresses two topics: (1) the rate of circularization of eccentric BBH mergers, and (2) the importance of including higher-order waveform modes for an accurate modeling of these astrophysical systems.

Regarding the first analysis, Fig. 1 indicates that for the more eccentric systems, circularization only happens about 50M be-

fore merger, and that the increase in overlap, as  $t_0 \rightarrow 0M$ , is not monotonic. Rather, it has an oscillatory behavior that tracks the eccentric trajectory of the BBH system, which is clearly captured by the waveform amplitude. As soon as the waveform amplitude of the eccentric signals becomes increasingly monotonic, so does the overlap.

For the second topic, the researchers have constructed NR waveforms that include either the modes  $(l, |m|) = \{(2, 2), (2, 1), (3, 3), (3, 2), (3, 1), (4, 4), (4, 3), (4, 2), (4, 1)\}$ , or just the  $l=|m|=2$  mode. For the NR waveforms that include higher-order modes, the team has quantified the regions of parameter space that maximize the contribution of these modes for GW detection. Upon constructing these NR waveforms, the scientists compute the overlap between these NR waveforms so that those only include the leading-order quadrupole term. Fig. 2 presents results for these calculations for a variety of astrophysically motivated scenarios. These indicate that the inclusion of higher-order modes does not quantitatively modify the morphology of  $l=|m|=2$  NR waveforms that describe equal-mass eccentric BBH mergers. However, NR waveforms that describe asymmetric mass-ratio and eccentric BBH mergers have a much richer topology that requires the inclusion of higher-order waveform modes.

WHY BLUE WATERS

Blue Waters was critical in producing a catalog of NR waveforms and enabling the research team to postprocess and characterize these waveforms *in situ* and at scale. While this analysis would be prohibitive in a campus cluster-sized resource, the analysis can be completed within a few minutes on Blue Waters for a catalog of over 100 NR waveforms.

PUBLICATIONS & DATA SETS

S. Habib and E. A. Huerta, “Characterization of numerical relativity waveforms of eccentric binary black hole mergers,” *Phys. Rev. D*, vol. 100, no. 4, pp. 044016–044026, Aug. 2019.



FUSING NUMERICAL RELATIVITY AND DEEP LEARNING TO DETECT ECCENTRIC BINARY BLACK HOLE MERGERS USING HIGHER-ORDER WAVEFORM MULTIPOLES

Allocation: Illinois/845 Knh  
PI: Eliu Huerta<sup>1</sup>  
Collaborators: Adam Rebei<sup>1</sup>, Sibó Wang<sup>1</sup>, Daniel Johnson<sup>1</sup>, Sarah M. Habib<sup>1</sup>, Roland Haas<sup>1</sup>, Daniel George<sup>1</sup>

<sup>1</sup>University of Illinois at Urbana–Champaign

EXECUTIVE SUMMARY

An ever-increasing number of gravitational wave detections with the LIGO (Laser Interferometer Gravitational-Wave Observatory) and Virgo observatories has firmly established the existence of binary black hole mergers. Using a catalog of numerical relativity simulations that describes eccentric black holes mergers with mass-ratios  $1 \leq q \leq 10$ , and eccentricities  $e_0 \lesssim 0.18$  ten gravitational wave cycles before the merger event, the research team determined the mass-ratio, eccentricity, and binary inclination angle combinations that maximize the contribution of the higher-order waveform multipoles  $(l, |m|)$  for gravitational wave detection. The researchers then explored the implications of these results in the context of stellar-mass black holes that are detectable by LIGO detectors and showed that compared to models that only include the  $(l, |m|) = (2, 2)$  mode, the inclusion of higher-order waveform multipoles can increase the signal-to-noise ratio of eccentric binary black hole mergers by up to approximately 45% for mass-ratio binaries  $q \leq 10$ .

RESEARCH CHALLENGE

Motivated by recent electromagnetic observations that suggest the existence of compact binary populations in galactic cluster M22 [1] and in the galactic center [2], and considering that eccentricity provides one of the cleanest signatures to identify these compact binary populations, the research team studied the detectability of these signals in LIGO. LIGO relies on waveform models to detect gravitational waves in its data stream. While numerical relativity simulations are free from approximation errors and represent the actual gravitational waves produced by colliding black holes, these simulations are too expensive to use directly. Approximate models, on the other hand, will often ignore all subdominant modes in the gravitational waves, potentially resulting in a reduced detection accuracy in LIGO. Assessing the impact of these approximations is only possible using high-fidelity numerical simulations covering a sizable region of the black hole parameter space accessible to LIGO. Computing this many numerical relativity waveforms with sufficient accuracy is a formidable

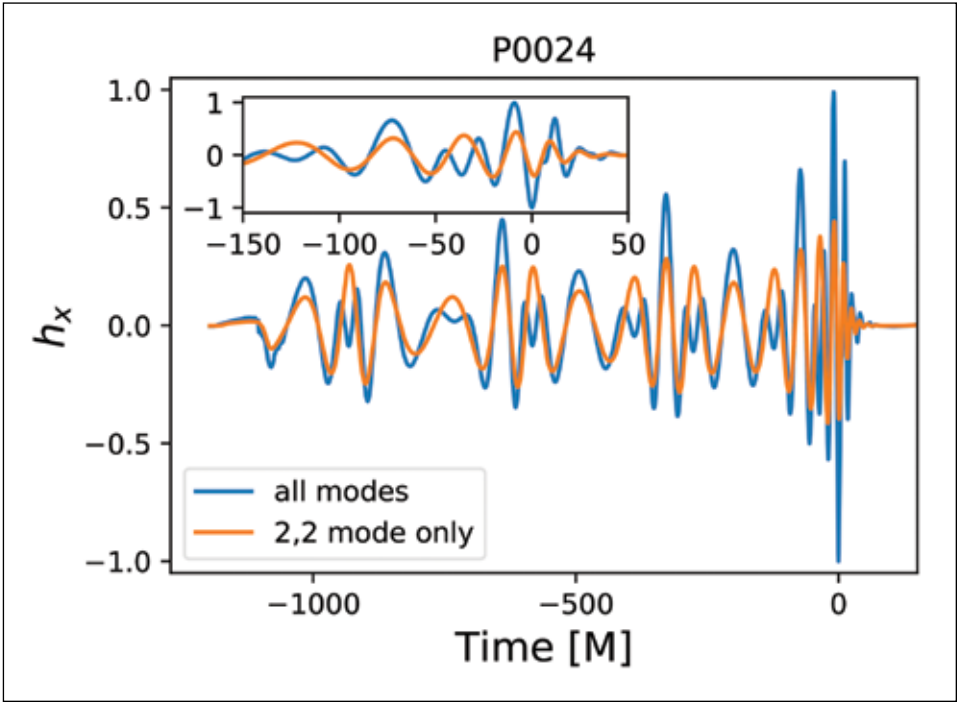


Figure 1

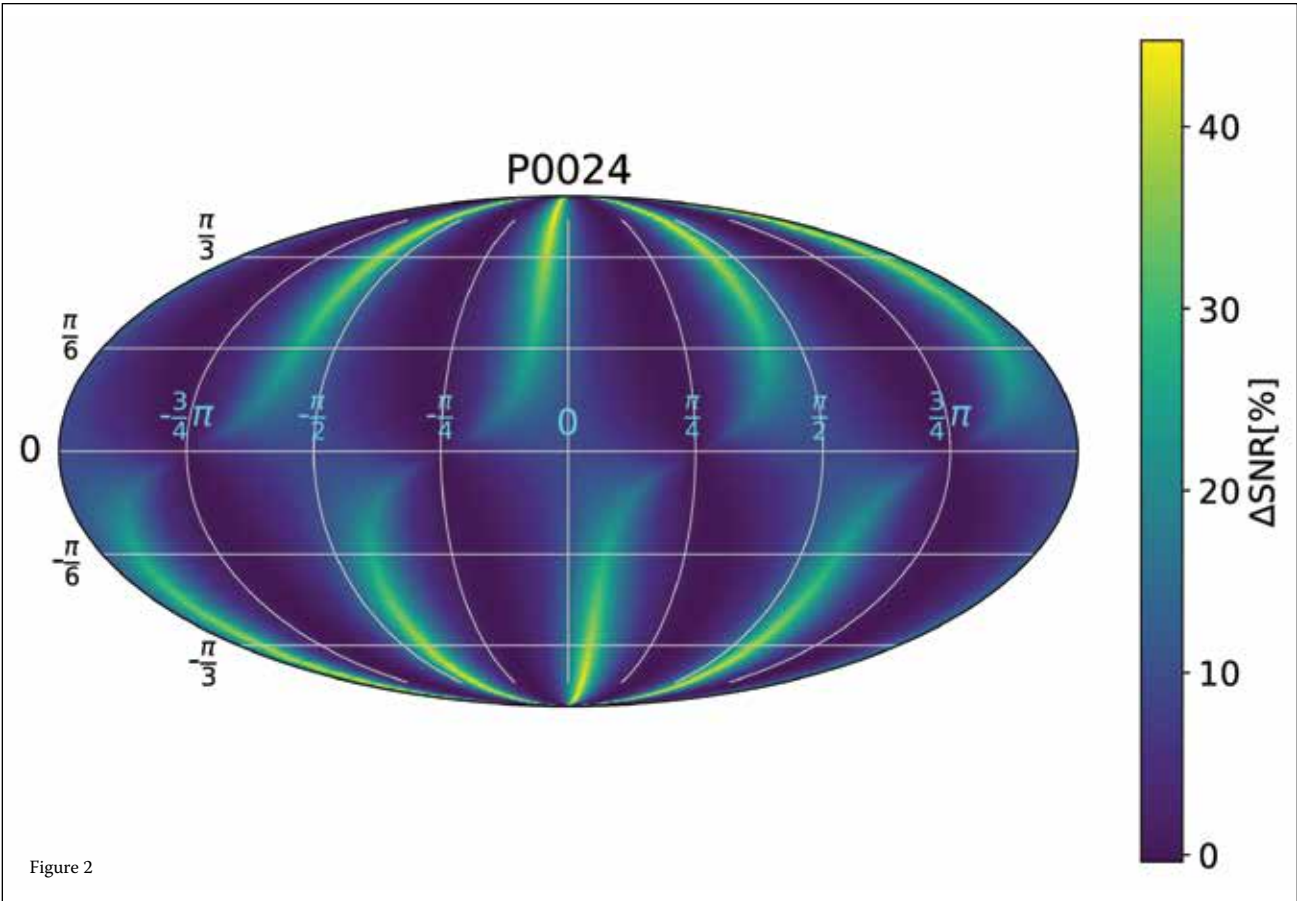


Figure 2

challenge since Einstein's equations of general relativity are among the most complex partial differential equations encountered in modern physics.

METHODS & CODES

The researchers used the Einstein Toolkit, a community-driven, open source astrophysics framework. The toolkit is OpenMP+MPI hybrid parallelized and uses automatic code generation to create compute kernels for general relativity. It employs eighth-order finite-differencing stencils in space and fourth-order accurate time integration. Adaptive mesh refinement including subcycling in time is used to resolve both the centers of the black holes as well as the gravitational waves far away from the black holes where they are detectable by LIGO.

RESULTS & IMPACT

The research team showed that the inclusion of higher-order waveform multipoles can increase the signal-to-noise ratio of eccentric binary black hole mergers by up to approximately 45% for mass-ratio binaries  $q \leq 10$ . Fig. 1 shows a representative waveform prediction with or without including higher-order modes. Generally speaking, higher-order modes add extra structure to the waveform, with more complex structure being present in eccentric

black hole collisions than in purely circular inspirals and collisions. Fig. 2 shows the relative increase in signal-to-noise ratio depending on the sky-location of the source. The improvement obtained by including higher-order modes is most prominent in regions of the sky where the  $(l, |m|) = (2, 2)$  mode is strongly suppressed. Including these modes will thus be critical to search for and find astrophysically motivated eccentric black hole collisions. Furthermore, the team showed that machine learning can accurately reconstruct higher-order waveform multipole signals embedded in real LIGO data.

WHY BLUE WATERS

Only Blue Waters provides the compute capabilities to simulate the hundreds of binary black hole collisions required to construct a waveform catalog suitable to explore the effect of higher-order multipoles on detectability by LIGO.

PUBLICATIONS & DATA SETS

A. Rebei *et al.*, "Fusing numerical relativity and deep learning to detect higher-order multipole waveforms from eccentric binary black hole mergers," *Phys. Rev. D*, vol. 100, no. 4, pp. 044025–044040, Aug. 2019.



DATA- AND COMPUTE-INTENSIVE CHALLENGES FOR OBSERVATIONAL ASTRONOMY IN THE GREAT SURVEY ERA

Allocation: Blue Waters Professor/ 300 Knh  
PI: Athol J. Kemball<sup>1</sup>  
Collaborators: Di Wen<sup>1</sup>, Michael Katolik<sup>1</sup>, Samantha Thrush<sup>1</sup>, Jamila Serena Taaki<sup>1</sup>

<sup>1</sup>University of Illinois at Urbana–Champaign

EXECUTIVE SUMMARY

This Blue Waters project is motivated by the data- and compute-intensive challenges posed by the large-survey telescopes of the coming decade, including the Large Synoptic Survey Telescope (LSST) and the Square Kilometer Array (SKA), among others. The science goals of these telescopes require high data acquisition rates and concomitant innovations in data analysis methods, some very compute-intensive and requiring resources on the scale of Blue Waters. The specific subprojects included in this parent project report include: (1) exoplanet detectability in large-scale surveys and optimal new approaches to exoplanet detection and parameter estimation in general, and (2) development of new statistical estimators for dark matter distribution and clustering in large-scale cosmological simulations and surveys in anticipation of LSST and related surveys.

RESEARCH CHALLENGE

This project is focused on data- and compute-intensive problems posed by the next generation of telescopes in observational astronomy, specifically those forming part of the imminent Great Survey Era, such as the LSST [1] and the Square Kilometer Array (SKA) [2]. Their science goals require high data acquisition rates with concomitant requirements for novel compute-intensive analysis approaches given their considerably more complex instrumental data models. Specific areas of focus include: (1) development and characterization of new algorithmic approaches for transiting exoplanet detection and parameter estimation, particularly for future large surveys; (2) development of new statistical estimators for the distribution and properties of dark matter for use as discriminants of cosmological models and in preparation for future large-scale surveys such as LSST; and (3) algorithmic challenges that are foundational to radio-interferometric imaging for highly data-intensive future telescopes such as the SKA.

The relevance of this research with Blue Waters centers on the fact that the telescopes of the Great Survey Era such as LSST and SKA represent multibillion-dollar investments in the observational astronomy research infrastructure for the next decade. This research is critical to this key transition in observational astronomy.

METHODS & CODES

For radio-interferometry projects, the research team has used the generic software framework eM [6], which encapsulates community codes to allow efficient prototyping in the initial stages of

algorithm exploration. Prototyping for other projects has been performed in parallelized Python. However, the team has developed optimized MPI codes when needed, such as in the cosmological analysis problem described below. When using community-based codes, the team carefully considers I/O and memory-access patterns in order to match the Blue Waters architecture and ensure resiliency.

RESULTS & IMPACT

The research subproject into exoplanet detection and estimation (with Jamila Serena Taaki) is strongly informed by the contemporary state-of-practice in spectral estimation and detection. A code framework has been developed and deployed on Blue Waters to evaluate the statistical performance of new transiting exoplanet estimators against data from the Kepler satellite [3], including algorithms based on marginalization, Monte Carlo methods, and new Bayesian approaches.

In a research subproject (with Di Wen), the team has explored the cosmological analysis of large-survey data with application for future telescopes such as LSST, specifically the development and characterization of estimators for the statistical properties of the distribution of dark matter in large-scale N-body cosmological simulations. This specific research project focuses on the use of the estimator for the counts-in-cells probability distribution function (PDF)  $f(N,V)$ , which defines the probability of finding  $N$  galaxies in a cosmological volume  $V$ . Although computationally intensive to estimate, this PDF contains significant embedded information on the properties of dark matter. The team's specific research interest is on the use of this distribution function as a discriminator of cosmological models and also as an assessment of its application to future survey data from LSST.

In the past year, the research team has applied the counts-in-cells PDF estimator for dark matter halos to the Dark Energy Universe Simulations (DEUS) [4] on Blue Waters. They have obtained measurements between redshifts  $z = 0$  to  $z = 4$  at both linear and nonlinear scales and have compared the best fits of four analytical models to the measured counts-in-cells distributions. The resulting counts-in-cells distributions  $f(N,V)$  as a function of redshift  $z$ , cell-size, and cosmological model are shown in Fig. 1. The team has also used Blue Waters to compute particularly expensive fits to PDF models where variable precision arithmetic is required owing to very large numerical coefficients. The percentage difference between  $f(N,V)$  for different cosmological dark

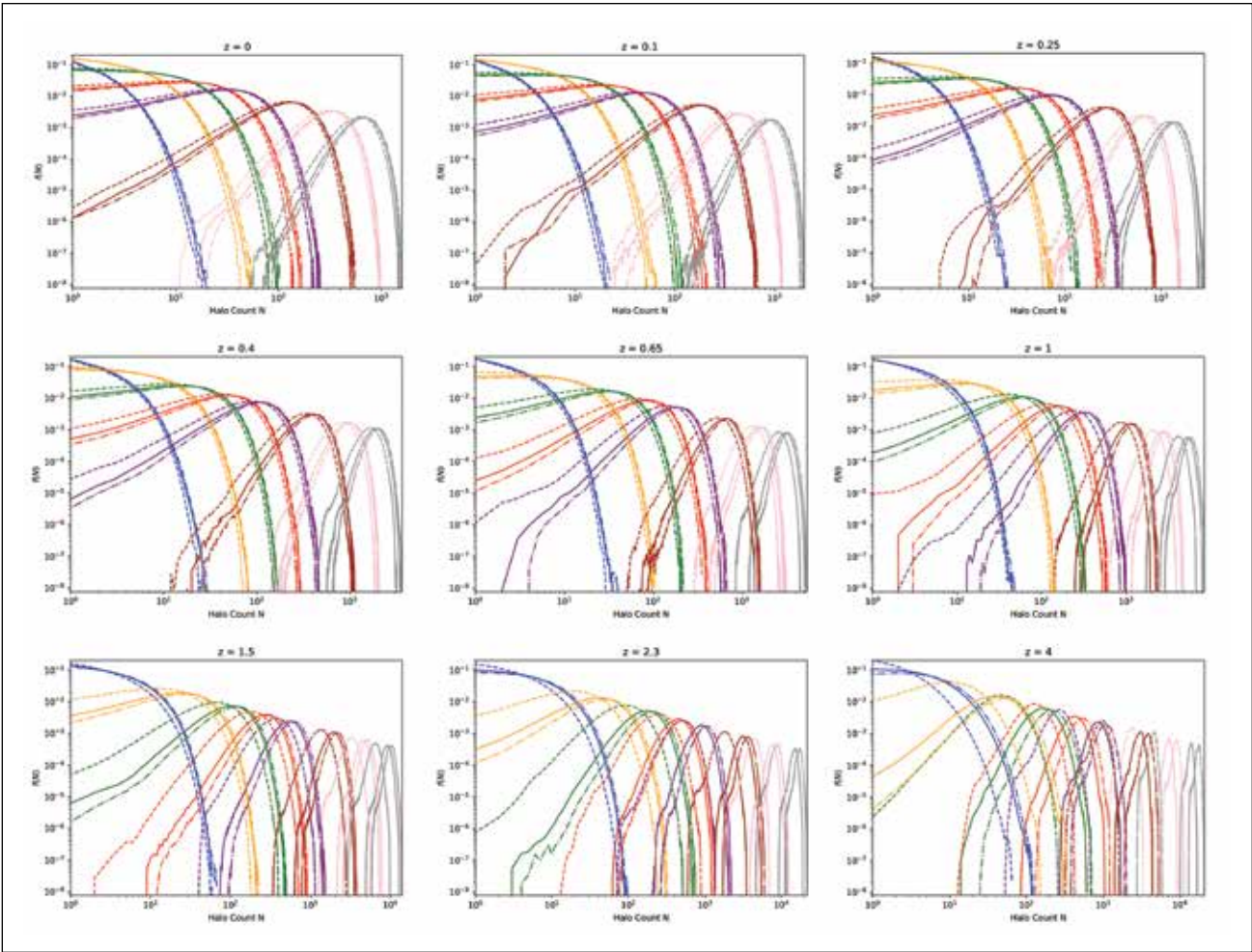


Figure 1: Counts-in-cells distributions  $f(N)$  for the  $\Lambda$ CDM (solid line), RPCDM (Ratra–Peebles; dash line) and wCDM (phantom dark energy; dash dot line) models at various redshifts. Line series colors indicate spherical cell radii  $R = 2h^{-1}$  (blue),  $4h^{-1}$  (orange),  $6h^{-1}$  (green),  $8h^{-1}$  (red),  $10h^{-1}$  (purple),  $15h^{-1}$  (brown),  $20h^{-1}$  (pink) and  $25h^{-1}$  (gray) [all Mpc].

energy models is a few percent at small  $z$  but can be significantly higher at high  $z$ , and therefore provide a useful manner to separate different dark energy models; here, the researchers consider quintessence and phantom dark energy against standard cold dark matter cosmology.

The research team's work in radio-interferometric calibration and imaging complications concerns the most computationally challenging problems for future data-intensive telescopes in this field.

WHY BLUE WATERS

The specific research conducted under this allocation explores new computationally intensive approaches to data analysis in this discipline. Several key projects are not practical without petascale resources. In addition, the code bases used in a number of these research subprojects are based on serial community codes in a parallel framework and therefore benefit particularly from the balanced architecture of Blue Waters.

PUBLICATIONS & DATA SETS

D. Wen, A. J. Kemball, and W. C. Saslaw, “Halo counts-in-cells for cosmological models with different dark energy,” *Astrophys. J.*, submitted, 2019.

D. Wen, “Counts-in-cells distribution of dark matter halos,” 12th Great Lakes Cosmology Workshop, Rochester, NY, U.S.A., Aug 6–8, 2019.



DI

TN

MODAL DECOMPOSITIONS OF SHOCK INTERACTIONS

Allocation: Illinois/625 Knh  
PI: Deborah A. Levin<sup>1</sup>

<sup>1</sup>University of Illinois at Urbana–Champaign

EXECUTIVE SUMMARY

Hypersonic flows involving Edney type IV/V shock wave boundary layer interaction (SWBLI) systems are characterized by complex multilength-scale features that include the triple point, separation bubble, shear layer, and shocklets. These features lead to a complex flow dynamics that is characterized by thermal nonequilibrium downstream of the bow shock as a result of the small residence time of the flow, a low-frequency unsteadiness associated with the laminar separation bubble (caused by a strong pressure rise along the surface), the potential growth of the Kelvin–Helmholtz instability in the shear layer at the contact surface downstream of the triple point, periodic oscillations of the triple point, and Goertler vortices (secondary flows that appear in a boundary layer flow along a concave wall) owing to curved streamlines on the outer side of the separation bubble. The PI seeks to understand these phenomena using the particle-based direct simulation Monte Carlo (DSMC) method, which is valid for a greater range of local Knudsen numbers than Navier–Stokes equations. This work is important because the accurate modeling of SWBLIs is crucial to design hypersonic vehicles with the best possible performance.

RESEARCH CHALLENGE

Given the complexities of the 3D finite span simulation, it is pertinent first to understand the behavior of the 2D base flow under self-excited, spanwise homogeneous perturbations by simulating flow over the double wedge of a chosen span length with periodic flow domain boundaries in the spanwise direction. Such perturbations could be generated inside the laminar separation bubble and grow, or the shock system may amplify the disturbances naturally present in the free stream. Growing perturbations may have a range of different wavelengths; however, this work involves the wavelength for which the growth is maximum since the spanwise periodic flow structures will exhibit this particular wavelength. The spanwise periodic case can shed light on the stability of the base flow under self-excited perturbations, the wavelength of spanwise periodic structures, their origin, and their time-accurate temporal behavior. In addition, the perturbations present in the separation bubble may have a range of dominant frequencies. One of the broader objectives of this research is to find out if there is a link between the frequency of shock oscillation and the characteristic frequency associated with the growth and shrinkage of the separation bubble. The second challenge of this work is the development and verification of a linear stability analysis framework that takes into account thermal nonequilibrium, thereby extending its application to hypersonic SWBLI systems.

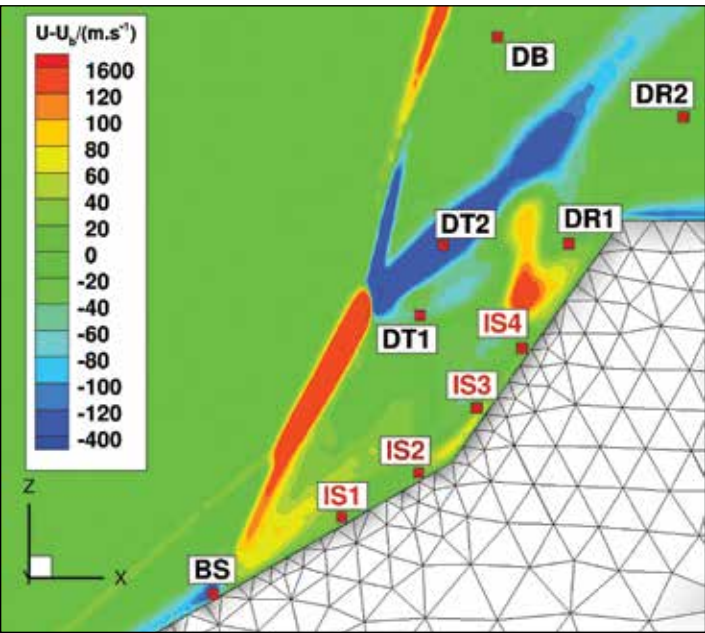


Figure 1: Location of numerical probes in the hypersonic flow over the double wedge showing the difference between the instantaneous and steady state streamwise velocity fields.

METHODS & CODES

The scalable unstructured gas-dynamics adaptive mesh refinement (SUGAR) DSMC code was developed and optimized on Blue Waters by making several improvements in the data structure of particles and how they are accessed from computational cells [1]. Two key findings are summarized here. First, with profiling and comparison with an older Cartesian grid code with two-level mesh adaption, SMILE, the PI found that the significant computational bottleneck was not the collision step (the usual problem) but rather the particle-mapping step to an unstructured grid. This led to the conclusion that since particle selection and mapping are performed extensively in DSMC, the previous use of the recursive pointer approach was unacceptably high. The SUGAR computational tool instead utilizes linearized data structures based on Morton Z space-filling curves. In this approach, only the leaves of an octree are stored with each leaf cell assigned a unique Morton index with which it can be accessed. During the mapping phase, based on a particle’s position, maximum depth of the root, and its minimum node position, a unique Morton index is computed that is the particle’s leaf cell index.

RESULTS & IMPACT

A number of key results from this work were presented in a poster at the recent International Union of Theoretical and Applied Mechanics (IUTAM) Transition meeting [2]. Toward the goal of studying the stability of base flows under self-excited perturbations, a preliminary spanwise periodic case was simulated with a span length of 72 mm, which is twice the size of the separation length defined; *i.e.*, the distance between the separation and reattachment points. The case started from a precomputed 2D base flow field extruded in the spanwise direction and was run for 0.25 ms, such that one cycle of the low-frequency wave through the oblique bow shock was captured. To compute the frequency of this wave, a numerical probe was placed downstream from the bow shock at the point labeled “DB” in Fig. 1. Fig. 2

(left) shows the  $L_2$  norm of instantaneous residual kinetic energy at probe location DB. Planes 1 to 5 listed in the legends of Fig. 2 (left) correspond to data sets obtained at the same X and Z coordinates of the probe DB but at different spanwise (Y) coordinates of 1, 15.4, 29.8, 44.2, 58.6 mm, respectively. As the changes in the bow shock are felt downstream in the residual kinetic energy at probe DB, two peaks can be identified after the initial transient period of 0.05 ms. Based on the difference between the time at which those two peaks occur, the corresponding shock frequency,  $f_s$ , is calculated as 8 kHz. At probe locations IS1, IS2, IS3, and IS4, which lie inside the separation bubble, the residual kinetic energy grows linearly, whereas at all other probe locations it decays linearly (not shown). Fig. 2 (center) shows the unsteady dynamics in the residual kinetic energy at probe location IS1, where the frequency of fluctuations inside box one is approximately 21 times larger than the shock frequency. The zoom of the data in box two, shown in Fig. 2 (right), reveals differences in the residual history among spanwise shifted locations of five copies of probe IS1 indicating spanwise flow perturbations. A detailed analysis of the frequencies associated with the separation bubble and an estimate of the spanwise wavelength can be found in the PI’s recently presented work [3].

WHY BLUE WATERS

These simulations show for the first time that self-excited perturbations are present even at two orders of magnitude lower unit Reynolds number in hypersonic flow. They also show that the estimate of the spanwise wavelength of flow structures of 7.2 mm [3] is reasonable. With this estimate, the size of the simulation can be reduced by decreasing the span length to 28.8 mm from 72 mm. Such numerical investigations can only be performed with petascale computational assets such as Blue Waters, where this research has used about 60 billion particles and 4.5 billion computational cells. Such runs require about 86,000 Blue Waters node-hours to simulate an unsteady flow for 0.1 ms.

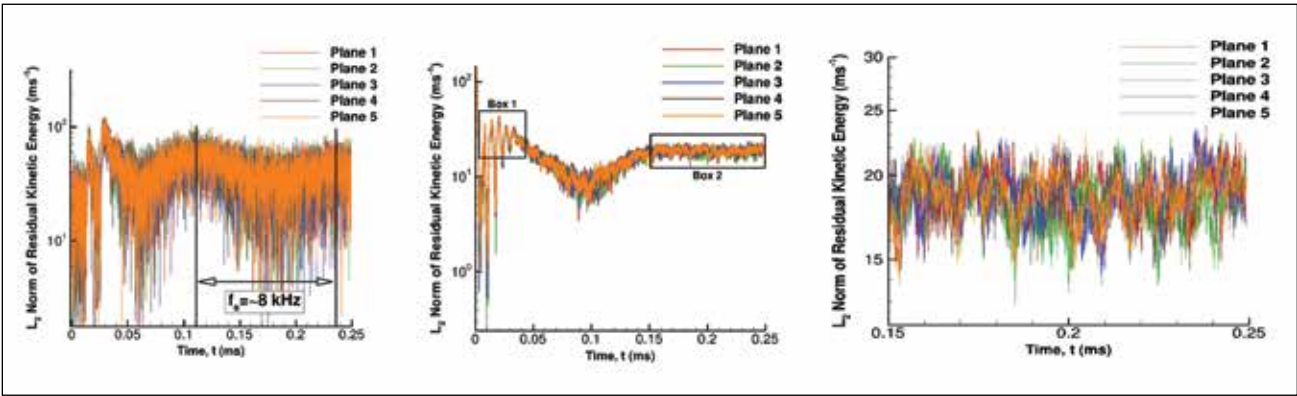


Figure 2: The time history of  $L_2$  norm of residual kinetic energy at probe locations DB (left) and IS1 (center), shown in Fig. 1 and zoomed (right) at the rectangular box marked in the center figure. DB and IS indicate portions of the flow field that are downstream of the strong bow shock and in the separation bubble, respectively.



PLUME PLASMA SPACECRAFT INTERACTIONS

Allocation: Illinois/630 Knh  
PI: Deborah A. Levin<sup>1</sup>

<sup>1</sup>University of Illinois at Urbana–Champaign

EXECUTIVE SUMMARY

High-fidelity numerical plasma modeling has been a key aspect for predicting physics in plasma. Using Blue Waters, the researcher conducted a series of numerical studies on a variety of plasma-based flows such as space plasma–surface interactions including modeling spacecraft contamination from electric propulsion thrusters and anomalous spacecraft charging in low Earth orbit. These simulations are unique in that both electrons and ions are treated as computational particles. In order to do this, a GPU-based code, called CHAOS, was developed during the researcher’s previous Blue Waters projects, and has demonstrated an 85% strong scaling. Given the immense range of length and timescales in plasma-based flows, a parallel computing architecture such as Blue Waters is necessary for their self-consistent numerical modeling.

RESEARCH CHALLENGE

Solar cell arrays provide onboard power for most spacecraft in orbits around Earth. These arrays have about a 1,000 V potential drop along their span, using wires to interconnect multiple dielectric photocells (Fig. 1). The positive potential of the interconnects attracts ambient electrons, creating an additional current loop that drains power generated by the solar panels. When the voltage across the interconnect is increased beyond a certain limit, the current dramatically shoots up, causing a large parasitic current. This phenomenon is known as “snap over.” The major contribution of the collection current to the interconnect comes from the secondary electrons emitted by nearby dielectric surfaces. The objective of this work is to model plasma physics near solar panel interconnects using a kinetic particle-in-cell approach and to establish how plasma surface interactions affect parasitic current.

METHODS & CODES

The modeling was performed with the direct simulation Monte Carlo (DSMC) particle-in-cell (PIC) code, CHAOS (CUDA-based hybrid approach for octree simulations) that was developed to study neutral flows through porous media under a previous Blue Waters effort and was reported earlier. The approach utilizes the Morton Z-curve octree structure and a volume-of-fluids approach used to compute volume of cut-leaf nodes using ray-tracing, which is very efficient on GPUs [1]. The code has been used previously to model ion thruster plume contamination of spacecraft in the backflow region in a fully kinetic manner [2]. To model the snap-over phenomenon, the secondary electron emission process has

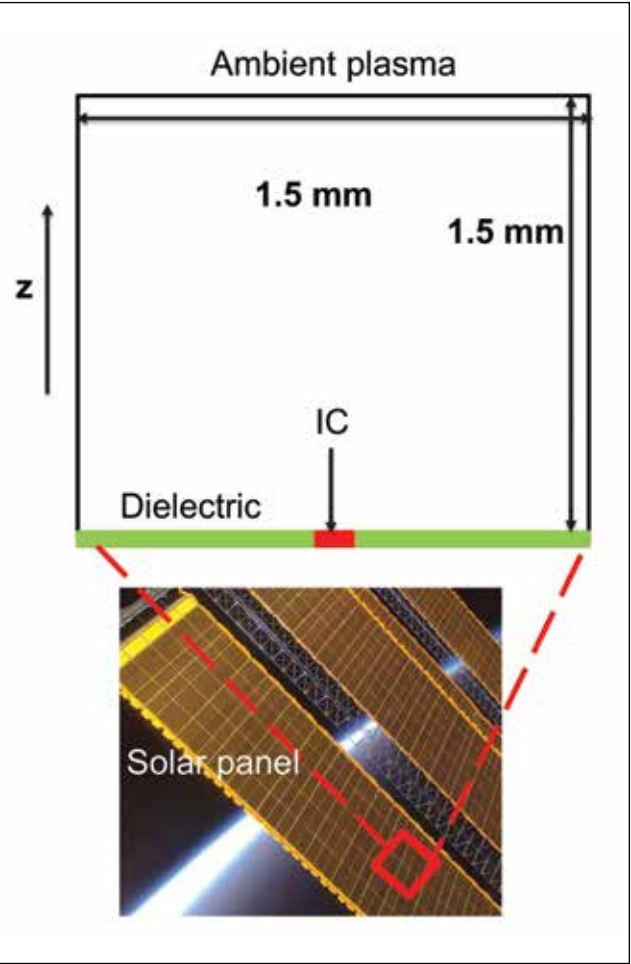


Figure 1: Computational domain of ambient plasma in front of solar cell array. Two dielectric elements are shown with the interconnect (IC) lying at 0.75 mm on the surface and their relationship to a typical solar cell array on a spacecraft.

been added to the code to model electron emission yields from magnesium fluoride, a typical dielectric, as a function of the incident ambient electron energy at the surface. It is assumed that every secondary electron that is emitted by the dielectric surface also leaves a positive charge on the surface. Emitted electrons are assumed to have an energy of 1 eV with velocities taken from a cosine distribution.

Fig. 2 shows a key result from the research that was recently presented [3]. The left-hand side (LHS) chart shows the time variation of surface potential at three numerical probes that are

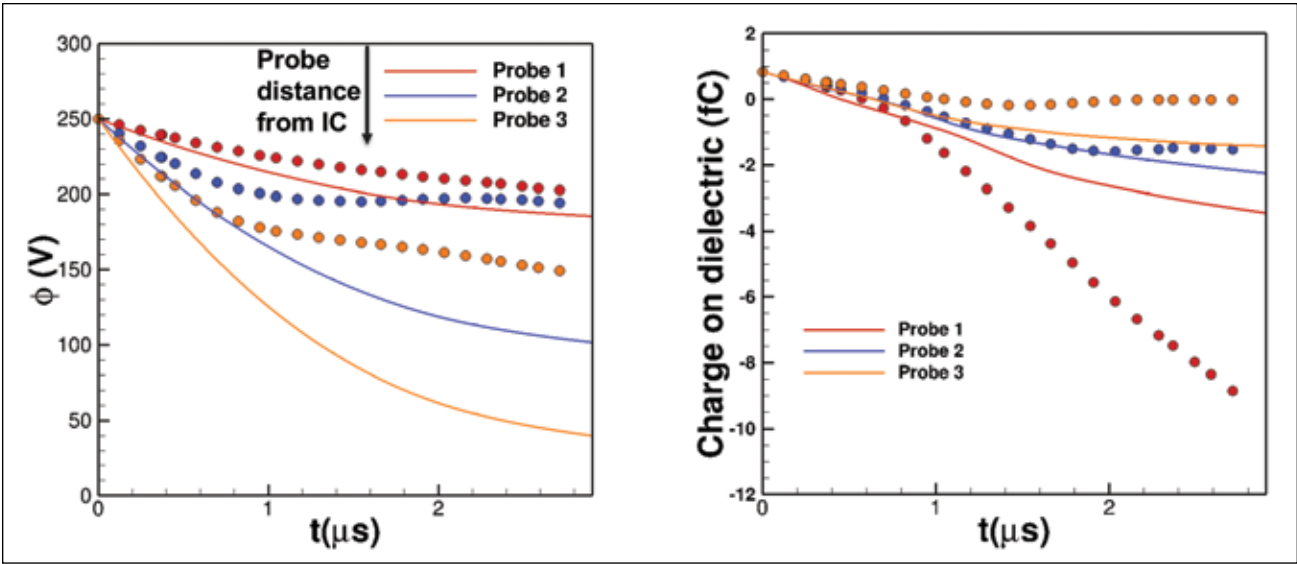


Figure 2: Electric potential and dielectric charging as a function of distance from the interconnect for simulations with and without secondary electron emission (SEE). Solid lines and closed circles indicate no SEE and with SEE, respectively.

approximately 0.075, 0.15, and 0.31 mm from the interconnect (IC), and the right-hand side (RHS) chart shows the accumulating surface charge on the dielectric photosensitive material as a function of time. When the secondary electron emission (SEE) is included, the dielectric discharges more slowly and to a higher potential, attracting more primary electrons from the ambient plasma. The surface charge on probes two and three for an emissive dielectric surface reaches a steady value at a higher charge than the charge for a nonemissive dielectric surface, as expected. However, at probe one, which lies very close to the IC, the charge collected on the surface keeps falling (RHS), even when the electric potential has reached a steady value (LHS). This shows that the region near probe 1 behaves as a conductor and collects an electron current similar to that of the IC.

These results have been obtained with a potential drop of 250 V across the IC. The effect of SEE is to cause electrons to gain kinetic energy of up to 250 eV initially when the surface is charged to a potential of 250 V at  $t = 0$ . This incident electron energy gives a yield of about 3.0 secondary electrons that charges the surface locally positively and emits two additional electrons near the dielectric. The more positive dielectric surface allows electrons in front of the dielectric with SEE to gain a higher  $z$ -velocity than without SEE, thereby increasing the total electron flux on the solar panel surface.

To understand the sensitivity of the results to the energy at which electrons are emitted from the dielectric material, additional simulations were performed wherein the emitted energy was reduced from 1 to 0.1 eV. The PI found that electrons have a higher density near the dielectric surface for the higher emitted energy case because the electrons emitted from the dielectric surface are attracted toward the same positively charged dielec-

tric surface, and the ones that are emitted with the lower kinetic energy are more likely to be reincident on the dielectric surface, effectively negating the positive charge on the surface caused by the secondary electron emission. This was also evident in the time discharging of the surface electric potential where the time variation of surface potential on probes two and three for the lower emitted energy case was closer to the case without SEE than the higher energy emitted case. Since a major fraction of emitted secondary electrons is reincident on the dielectric surface for the lower energy emitted case, they do not reach the interconnect, causing a 25% decrease in the IC current for the lower energy emitted case.

WHY BLUE WATERS

As efficient as the algorithms are in the CHAOS code, the kinetic modeling of electrons in the presence of heavy ion masses such as positive atomic oxygen or xenon is rare because such simulations are computationally challenging. The ability to perform heterogeneous computing on a large number of combined CPU/GPU nodes is unique to the Blue Waters architecture. This hardware has enabled the PI to demonstrate the first ever such simulations of plasma-based flows.



DI

TN

MI

MP

BI

# THE SPREADING OF THREE-DIMENSIONAL MAGNETIC RECONNECTION IN ASYMMETRIC GEOMETRY

**Allocation:** NSF PRAC/90 Knh  
**PI:** Yi-Hsin Liu<sup>1</sup>  
**Co-PIs:** Michael Hesse<sup>2</sup>, William Daughton<sup>3</sup>  
**Collaborators:** Tak Chu Li<sup>1</sup>

<sup>1</sup>Dartmouth College  
<sup>2</sup>University of Bergen  
<sup>3</sup>Los Alamos National Laboratory

## EXECUTIVE SUMMARY

Earth’s magnetosphere—a region formed from its magnetic fields—shields the planet from constant bombardment by supersonic solar winds. However, this magnetic shield, called the magnetopause, can be eroded by various plasma mechanisms. Among them, magnetic reconnection is arguably the most active process. Reconnection not only allows the transport of solar wind plasmas into Earth’s magnetosphere but also releases the magnetic energy and changes the magnetic topology. At Earth’s magnetopause, magnetic reconnection proceeds between the shocked solar wind plasmas and the magnetosphere plasmas. Many three-dimensional properties of magnetic reconnection in such asymmetric geometry remain unclear. The research team used first-principle simulations to explore the 3D kinetic physics that control this critical energy conversion process.

## RESEARCH CHALLENGE

Massive solar eruptions drive magnetic storms that impact Earth’s magnetosphere and space weather. The consequential electromagnetic waves, electric currents, and energetic particles can harm satellites, astronauts, GPS systems, radio communication, and power grids on the ground. Magnetic reconnection is the

critical player in solar wind–magnetosphere coupling, and space weather in general. Fundamental questions in reconnection research include: Is there a simple principle that determines the orientation of the reconnection x-line (the null line where magnetic reconnection occurs) in such an asymmetric current sheet? How fast does reconnection spread out from a point source? The answers to these questions remain unclear given our current understanding of magnetic reconnection, and the research team aims to study its 3D nature. This work is a crucial step in the quest for predicting the location and rate of flux transfer at Earth’s magnetopause, and will thus improve the forecast of space weather.

## METHODS & CODES

This project employed the particle-in-cell code VPIC, which solves the relativistic Vlasov–Maxwell system of equations using an explicit charge-conserving approach. Charged particles were advanced using Leapfrog with sixth-order Boris rotation, then the current and charge density were accumulated on grid points to update electromagnetic fields. Marder divergence cleaning was employed to ensure the divergence was free of the magnetic field. The level of error was bounded by the numerical round-off effect.

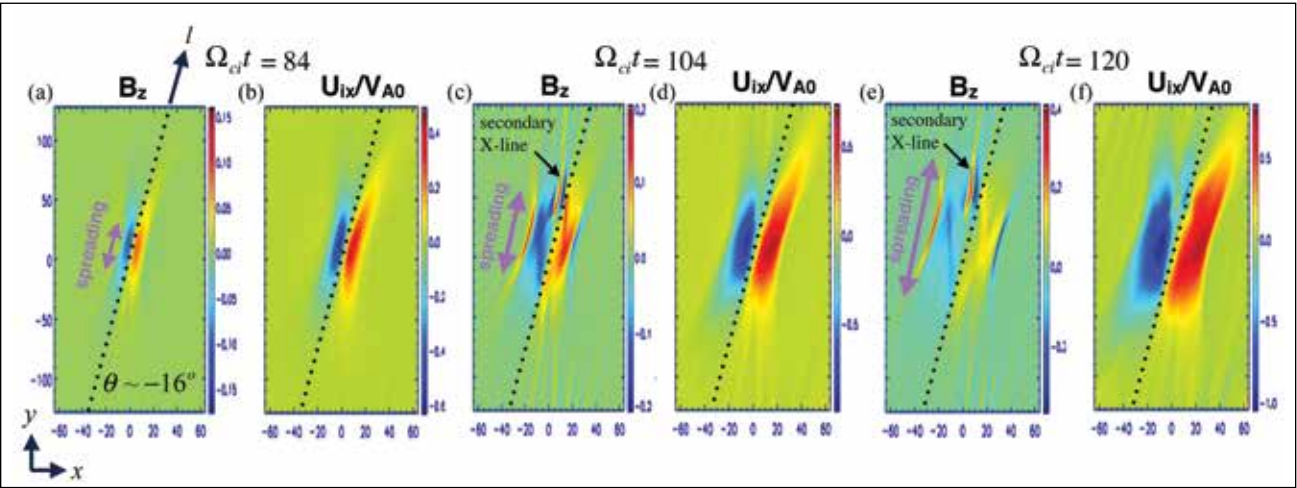


Figure 1: X-line spreading observed in the (normalized) reconnected magnetic field  $B_z/B_0$  and ion outflow  $U_{ix}/V_{A0}$  at three different times. The x-line spreads along the orientation labeled as the  $I$  axis that is indicated by dotted black lines.

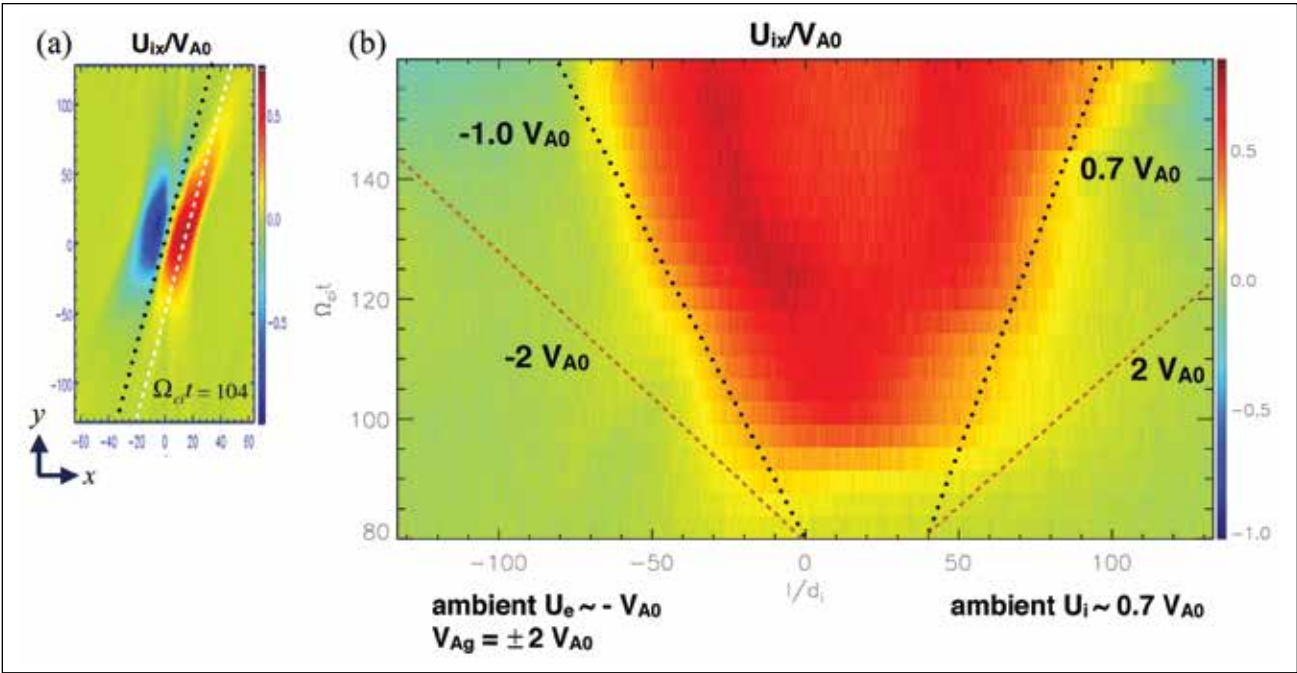


Figure 2: X-line spreading measured from the ion outflow speed  $U_{ix}$ : (a) A cut of  $U_{ix}$  across the initial x-line on the  $x$ – $y$  plane at time  $t = 104/\Omega_{ci}$ . The spreading of  $U_{ix}$  is sampled at the dashed white line. (b) The time stack plot of the sampled  $U_{ix}$ .

## RESULTS & IMPACT

The spreading of magnetic reconnection in asymmetric current sheets was studied in the third year of this project. The result has an application to the reconnection that occurs at Earth’s magnetopause. The research team initiated reconnection at the center of a large simulation domain to minimize the boundary effect. The resulting x-line had sufficient freedom to develop along an optimal orientation, and the signal of fast reconnection spread out from two ends of the x-line. The team measured the spreading speed and found that it exhibited a strong dependence on the thickness of the initial current sheet. They mapped out a criterion for the Alfvénic spreading of fast reconnection and explored its physics origin. While the criterion appears to be stringent, this study suggests that under typical conditions at Earth’s magnetopause, the reconnection x-line is unlikely to demonstrate Alfvénic spreading at the local Alfvén speed regardless of the guide field strength. An Alfvénic spreading is expected only if the typically thick magnetopause current sheet is substantially compressed (perhaps by the solar wind) to a state strongly unstable to the collisionless tearing instability.

These results compared favorably with recent magnetopause observations using SuperDARN radars and THEMIS satellites. In addition, this new result is important to NASA’s ongoing Magnetospheric Multiscale Mission, which was designed to study the kinetic physics of the reconnection x-line. The results could also be relevant to the ongoing European Space Agency (ESA)–Japan Aerospace Exploration Agency Mercury mission, BepiColombo, and the upcoming ESA–Chinese Academy of Sciences joint

mission, Solar Wind Magnetosphere–Ionosphere Link Explorer, which will study the development of reconnection lines at Earth’s magnetopause using X-ray and UV imagers.

## WHY BLUE WATERS

Because the x-line has a dimension down to the electron scale, a fully kinetic description is necessary. Given the available computational capability, it has become possible to use a first-principle kinetic simulation to investigate the dynamics of the x-line in a reasonably large 3D system that spans from the electron kinetic scale to the magnetohydrodynamics scale. A representative 3D run in this project traces the motion of two trillion charged particles under the interaction of self-generated electromagnetic fields, which are evaluated on six billion grids. The output data easily have a size of hundreds of terabytes for each run. Blue Waters provides not only the computational resource for the calculation but also the online storage for the output and restart files.

## PUBLICATIONS & DATA SETS

Y.–H. Liu, M. Hesse, T. C. Li, M. Kuznetsova, and A. Le, “Orientation and stability of asymmetric magnetic reconnection x-line,” *J. Geophys. Res.*, vol. 123, pp. 4908–4920, May 2018.  
T. C. Li, Y.–H. Liu, M. Hesse, and Y. Zhu, “Three-dimensional x-line spreading in asymmetric magnetic reconnection,” *J. Geophys. Res.*, submitted, 2019, arXiv: <https://arxiv.org/abs/1907.02025v1>.



# ASSEMBLING A MAP OF THE UNIVERSE: SHAPES AND MASS DISTRIBUTION FOR THE DARK ENERGY SURVEY

**Allocation:** Illinois/460 Knh  
**PI:** Felipe Menanteau<sup>1,2</sup>  
**Collaborators:** Robert Gruendl<sup>1,2</sup>, Donald Petravick<sup>1</sup>, Erin Sheldon<sup>3</sup>

<sup>1</sup>National Center for Supercomputing Applications (NCSA)  
<sup>2</sup>University of Illinois at Urbana–Champaign  
<sup>3</sup>Brookhaven National Laboratory

## EXECUTIVE SUMMARY

The Dark Energy Survey (DES) has performed a 5,000 square-degree wide field survey in five optical bands of the southern sky and a 30 square-degree deep supernova survey with the aim of understanding the nature of dark energy and the accelerating Universe. DES used the 3 square-degree CCD (Charge-Coupled Device) camera (DECam) installed at the prime focus on the Blanco 4-m telescope with the goal of recording the positions and shapes of 500 million galaxies up to redshift 1.4. Over the course of the survey, DES produced roughly 500 TB of raw data, which has been processed with the DES Data Management system (DESDM). The Blue Waters allocation has facilitated the use of new processing algorithms that make detailed measurements of the galaxy properties (size, shape, brightness) that enable cosmological investigation into the history of structure formation in the Universe.

## RESEARCH CHALLENGE

DES aims to understand the origin of cosmic acceleration and the nature of dark energy using four complementary methods: weak gravitational lensing, galaxy cluster counts, large-scale galaxy clustering (including baryon acoustic oscillations), and Type Ia supernovae. DES comprises two multiband imaging surveys: a 5,000 square-degree  $g, r, i, z, Y$  wide survey of the southern sky to approximately the 24th mag and a deeper time-domain 30 square-degree  $g, r, i, z$  deep DES Supernova Survey with a cadence of approximately five days.

## METHODS & CODES

DES uses the state-of-the-art 3 square-degree DECam, a 570 megapixel camera installed at the prime focus on the Blanco 4-m telescope at the Cerro Tololo Interamerican Observatory (CTIO) in Northern Chile. DECam consists of 62 fully depleted, 250-mi-

cron-thick  $2,048 \times 4,096$  CCDs combined with four  $2,048 \times 2,048$  guider and eight  $2,048 \times 2,048$  autofocus CCDs.

For 575 nights from 2013–2019, DES scanned the sky to perform a 5,000 square-degree wide field survey. Over five observing seasons, DES measured shapes, positions, fluxes, and colors for approximately 300 million galaxies and discovered and measured light curves for 3,500 supernovae, using these measurements to deliver powerful new constraints on cosmic acceleration and dark energy. Each image arrives from CTIO in Chile at the National Center for Supercomputing Applications (NCSA; Urbana, Illinois) within minutes of being observed and is usually processed by the nightly processing pipeline within the next 24 hours. These nightly pipelines are critical for the near-real-time supernovae analyses, and also provide rapid feedback about overall data quality as the survey progresses. The other cosmological probes (weak lensing and galaxy clustering) rely on the combination of all survey observations to form a data release with unprecedented depth but at the cost of averaging over instrumental changes and varying atmospheric conditions.

The DESDM software and workflows (using HTCondor) have already been implemented and exercised as part of previous allocations on Blue Waters. In addition, the atomic pieces of the DES Y6A1 release comprised 131,602 reduced, calibrated exposures and 10,169 COADD tiles and catalogs that cover the 5,000 square degree survey volume. These form the inputs for the joint analyses.

## RESULTS & IMPACT

During the early years of the DES survey, new algorithms for measuring the galaxy properties were developed, but these require a joint analysis of combined images (COADDs) with the individual nightly observations that make up the survey on a per-object (galaxy) basis. The Blue Waters allocation was used to perform these analyses.

New measurement algorithms have been developed to cope with the increased astronomical survey depth, as the density of sources (stars and galaxies) projected along any line of sight in the survey volume is high enough that the projected object images overlap. Further, modern survey methodologies to obtain those depths rely on the combination of many images taken over many years in varied conditions (*e.g.*, atmospheric turbulence). Three advanced measurement algorithms have been developed by the DES Collaboration to make precise measurements based on the initial data release products. These include:

- MOF (Multi-Object Multiband Fitting), which simultaneously fits basic galaxy shapes and fluxes to each object and its neighbors;
- SOF (Single-Object Multi-Band Fitting), which performs a similar fit but masks neighboring object as a cross-check to the MOF analysis; and
- Shear, which makes a detailed shape analysis of an object and its neighbors.

These new catalogs are more precise and minimize the systematic uncertainties imprinted on the survey data. As an example of how these new catalogs will be used by the DES Collaboration, Fig. 1 shows the mass map from the first year of DES observations inferred from weak lensing measurements [1].

## WHY BLUE WATERS

DESDM is led by NCSA, where all images are archived and served to the community. Data processing on Blue Waters is more robust and performant than distributing workloads to remote sites. Moreover, the proximity of DESDM scientists and Blue Waters staff enables rapid feedback and clear communication, which are important for the success of complex implementations. Blue Waters’ primary utilization as a massively parallel resource also affords a complementary high-throughput resource by providing an elastic pool that can accommodate the intermediate demands of DES’s campaign processing to obtain results within a month.

## PUBLICATIONS & DATA SETS

Data set: The DES Collaboration Year 3 Annual release (Y3A1)

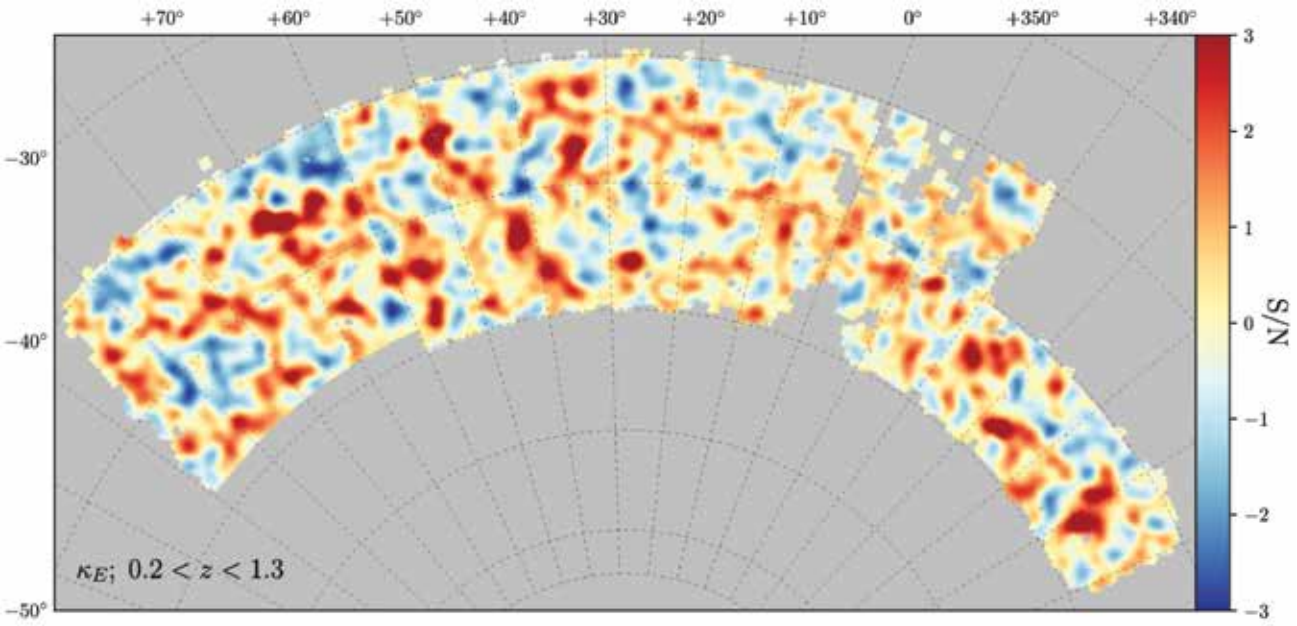


Figure 1: The mass map from the Year 1 DES observations inferred from weak lensing measurements of galaxies, where red represents higher mass density [1].



PETASCALE SIMULATIONS OF BINARY NEUTRON STAR MERGERS

**Allocation:** NSF PRAC/200 Knh  
**PI:** Philipp Mösta<sup>1</sup>  
**Co-PI:** David Radice<sup>2</sup>  
**Collaborators:** Roland Haas<sup>3</sup>, Erik Schnetter<sup>4</sup>, Sebastiano Bernuzzi<sup>5</sup>

<sup>1</sup>University of California, Berkeley  
<sup>2</sup>Princeton University  
<sup>3</sup>National Center for Supercomputing Applications  
<sup>4</sup>Perimeter Institute  
<sup>5</sup>University of Jena

EXECUTIVE SUMMARY

The gravitational wave and electromagnetic data from the collision between two neutron stars, GW170817 [1], is revolutionizing our understanding of the origin of the heavy elements, of the nature of short gamma-ray bursts, and of the properties of matter at extreme densities. To fully exploit the potential of multimessenger observations, the research team performed *ab initio* simulations of binary neutron star mergers and computed their gravitational wave and electromagnetic signatures. These simulations were enabled by both Blue Waters’ capacity for massively parallel high-resolution simulations and high throughput for large sets of simulations at smaller scales.

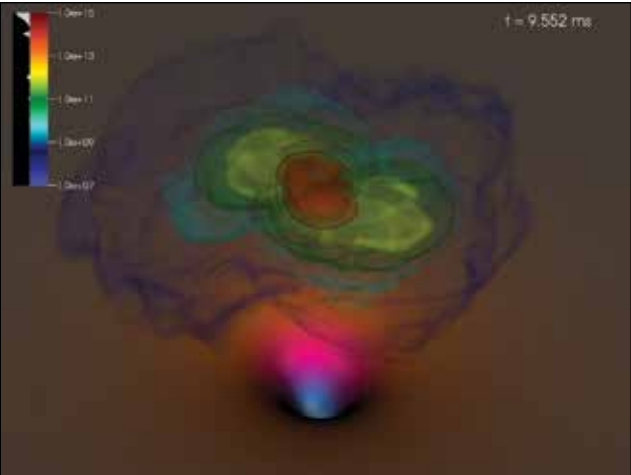


Figure 1: Volume rendering of the fluid rest-mass density (in g/cm³) in a simulation performed on Blue Waters of the collision between two neutron stars. The figure also shows the value of the lapse function, a proxy for the curvature of spacetime in the orbital plane of the binary. Tidal streams and shocks generated during these mergers generate neutron-rich outflows that synthesize heavy elements and power bright optical transients.

RESEARCH CHALLENGE

The gravitational wave and electromagnetic data from the collision between two neutron stars, GW170817 [1], is revolutionizing our understanding of the origin of the heavy elements, of the nature of short gamma-ray bursts, and of the properties of matter at extreme densities. There are analytical and semianalytical models with various degrees of complexity and sophistication that can be used to separately interpret gravitational wave and electromagnetic data. However, to fully harness this and future multimessenger observations, it is necessary to construct models that can combine the different “messengers.” Toward this aim, the research group performed *ab initio* simulations of binary neutron star mergers and computed their gravitational wave and electromagnetic signatures.

METHODS & CODES

The research team used two simulation codes: WhiskyTHC [2], and GRHydro [3]. Both codes are based on the Einstein Toolkit [4] but implement different treatments for matter and use different numerical schemes. WhiskyTHC is optimized for the study of the merger phase. It models neutron stars using finite-temperature tabulated nuclear equations of state and has a simplified treatment for neutrino radiation. GRHydro has a similar treatment of the neutron star matter but was originally designed to study core-collapse supernovae, so its neutrino treatment is specialized to the case where there is a centrally condensed source. GRHydro can model magnetic effects directly by solving the equations of general-relativistic magnetohydrodynamics, while WhiskyTHC treats angular momentum transport owing to magnetic stresses using a subgrid model. This project leverages the strengths of both codes; the team performed merger simulations using WhiskyTHC, and then mapped the results into GRHydro to study their long-term evolution.

RESULTS & IMPACT

The simulations revealed a new mechanism that could power mildly relativistic outflows during mergers and could be revealed by radio observations months to years after merger. The team also found that the electromagnetic (EM) signal is very sensitive to the merger outcome; *i.e.*, whether a black hole was formed promptly

or with some delay. They used this observation to develop a novel way to constrain neutron star radii using joint electromagnetic and gravitational wave observations of neutron star mergers. The simulations showed that the outflows powered by mergers could inject positrons in the interstellar medium, and the researchers suggested that neutron star mergers could provide a solution to the decades-old puzzle concerning the origin of cold positrons in the galactic center and some of the dwarf galaxies. Gravitational waveforms, ejecta data, and nucleosynthesis yields from the simulations are publicly available at [www.computational-relativity.org](http://www.computational-relativity.org) and [zenodo.org](https://zenodo.org).

The team has performed the first general-relativistic magnetohydrodynamics (MHD) turbulence simulations of neutron star merger remnants including nuclear and neutrino physics. These simulations allowed them to quantify the role magnetic fields play in angular momentum transport in the remnant, and they showed that the lifetime of the remnant can change significantly when including magnetic field effects. The results from these simulations will also be used to inform initial conditions for long-term follow-up simulations of merger remnants.

WHY BLUE WATERS

The MHD turbulence simulations of neutron-star merger remnants that resolve the magneto-rotational instability would have been impossible without Blue Waters’ capability. Blue Waters’ capacity is also crucial for a quick turnaround in the team’s numerical relativity neutron-star merger simulations. No other machine allows for the throughput of many of these simulations concurrently to generate the theoretical predictions needed for Laser Interferometer Gravitational-Wave Observatory and EM follow-up.

PUBLICATIONS & DATA SETS

A. Perego, S. Bernuzzi, and D. Radice, “Thermodynamics conditions of matter in neutron star mergers,” *Eur. Phys. J. A*, vol. 55, no. 8, Art. no. 124, Aug. 2019, doi: 10.1140/epja/i2019-12810-7.

G. M. Fuller, A. Kusenko, D. Radice, and V. Tikhonov, “Positrons and 511 keV radiation as tracers of recent binary neutron star mergers,” *Phys. Rev. Lett.*, vol. 122, no. 12, p. 121101, Mar. 2019, doi: 10.1103/PhysRevLett.122.121101.

D. Radice and L. Dai, “Multimessenger parameter estimation of GW170817,” *Eur. Phys. J. A*, vol. 55, no. 4, art. 50, April 2019, doi: 10.1140/epja/i2019-12716-4.

D. Radice, A. Perego, K. Hotokezaka, S. Bernuzzi, S. A. Fromm, and L. F. Roberts, “Viscous-dynamical ejecta from binary neutron star merger,” *Astrophys. J. Lett.*, vol. 869, no. 2, p. L35, Dec. 2018, doi: 10.3847/2041-8213/aaf053.

D. Radice, A. Perego, K. Hotokezaka, S. A. Fromm, S. Bernuzzi, and L. F. Roberts, “Binary neutron star mergers: mass ejection, electromagnetic counterparts and nucleosynthesis,” *Astrophys. J.*, vol. 869, no. 2, p. 130, Dec. 2018, doi: 10.3847/1538-4357/aaf054.

P. Mösta and D. Radice, “3D general-relativistic MHD simulations of binary neutron-star merger remnants,” in preparation, 2019.



DEVELOPMENT OF A SCALABLE GRAVITY SOLVER FOR ENZO-E

Allocation: NSF PRAC/100 Knh  
PI: Michael L. Norman<sup>1</sup>  
Co-PI: James Bordner<sup>1</sup>

<sup>1</sup>University of California, San Diego

EXECUTIVE SUMMARY

Blue Waters was used to develop and profile a highly scalable version of the Enzo cosmological adaptive mesh refinement (AMR) code called Enzo-E (E for extreme scale) [4]. In this project, the research team illustrated the weak scaling results of the domain-decomposed AMR Poisson solver using a suite of dark matter clustering simulations.

RESEARCH CHALLENGE

Improved computational models of the intergalactic medium (IGM) are needed to extract information encoded in the high-resolution optical spectra of distant quasars. That information includes the physical state of the mostly primordial gas pervading the Universe but also the dark matter that shapes the gas into discrete intergalactic absorption line systems (the Lyman- $\alpha$  forest). Standard computational models are discrepant with respect to certain aspects of the observational data [1] suggesting there is some key ingredient missing in the models. Previously, the research team explored whether modeling the population of quasars that provide a photoionizing bath of ultraviolet radiation as discrete point sources rather than a homogeneous background could improve agreement with observations. The team found it does not, using the Enzo hydrodynamic cosmology code [2] enhanced with multigroup flux-limited diffusion radiative transfer. The team is, therefore, investigating its next hypothesis: that dense gas bound to galaxies that is unresolved in the Enzo simulations supplies significant absorption of the quasar light and modifies the key observables in such a way as to improve agreement with observations.

METHODS & CODES

Including galaxies in simulations of the IGM poses severe resolution requirements that can be addressed using adaptive mesh refinement (AMR) around galaxies. However, Enzo's AMR capability is not sufficiently scalable to permit a full frontal assault on this problem. For this reason, the research team has been developing a successor to the Enzo code called Enzo-E built on an entirely new highly scalable AMR framework called Cello [3]. The combined code—Enzo-E/Cello—uses the Charm++ parallel object framework for parallelization. The team has implemented the proven-scalable Forest-of-Octrees AMR algorithm [4] on top of Charm++ and has obtained excellent parallel scaling results on Blue Waters as a prelude to the target application.

Solving the elliptic Poisson equation on the Forest-of-Octrees adaptive mesh is a prerequisite for cosmological applications. In the past year, the team has developed such a solver and profiled it on Blue Waters. Fig. 1 shows its application to a pure dark matter N-body simulation. The left side shows the projection of the dark matter density onto the adaptive mesh at redshift 7.5 in a cubic volume 6 comoving megaparsecs on a side. The right side shows the projection of the adaptive mesh, with levels of refinement encoded in color: root grid = dark blue; levels 1, 2, 3, 4 = light blue, green, yellow, and red, respectively. Each square is the projection of a block of 16 x 16 x 16 computational cells, on which the gravitational potential was solved using the following method.

First, the dark matter particle masses were assigned to the grid block that contains them using Cloud-In-Cell interpolation. Second, the matter densities were projected to the root grid (lev-

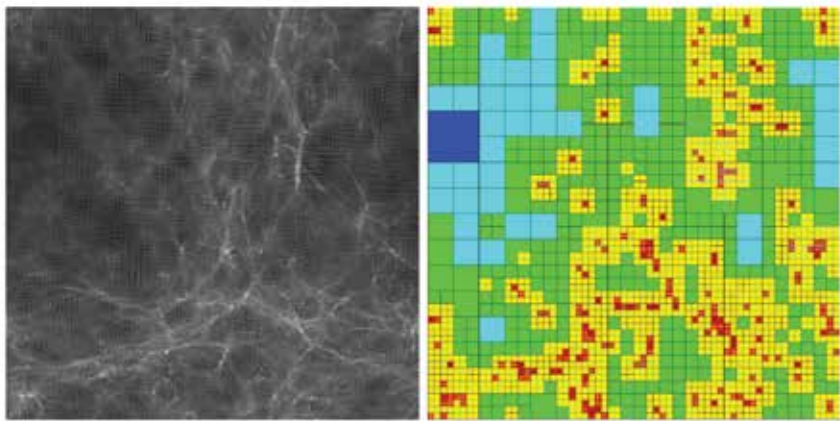


Figure 1: Left—The gravitational clustering of dark matter in a six megaparsec volume of the Universe. Right—The array-of-octree computational mesh. Each square is the projection of a cubic region of space resolved by 16 x 16 x 16 cells. This computation uses five levels of mesh refinement, color coded by level.

el 0). Third, the global gravitational potential was computed on the root grid using a V-cycle multigrid solver. Fourth, the gravitational potential was interpolated from the root grid to the faces of each octree. Fifth, the gravitational potential was computed for each octree using a multilevel iterative solver (BiCGStab). Finally, the potential was smoothed across all the leaf nodes of all octrees using Jacobi smoothing.

RESULTS & IMPACT

The DD (domain-decomposed) Poisson solver described above was implemented and tested on Blue Waters in the context of dark matter-only cosmological structure formation simulations. Weak scaling results are shown in Fig. 2. The research team plotted wall time per timestep versus timestep for four simulations in box sizes 3, 6, 12, and 24 comoving megaparsecs on a side, with root grid resolutions of 64<sup>3</sup>, 128<sup>3</sup>, 256<sup>3</sup>, and 512<sup>3</sup> cells. These correspond to 4<sup>3</sup>, 8<sup>3</sup>, 16<sup>3</sup>, and 32<sup>3</sup> arrays of blocks of 16<sup>3</sup> cells each. Each root grid block is the base of a separate octree, which refines as structure forms. With 8 root grid blocks per core, the simulations were run on 8, 64, 512, and 4,096 cores of Blue Waters. At early times, before cycle 100, mesh refinement had not begun, and the execution time was dominated by the global multigrid solver. At later times, up to a factor of 100 more refined blocks at all levels were created to resolve structure formation. Execution time was dominated by the BiCGStab tree solver. The fact that the curves from different runs bunch together at late times indicated favorable weak scaling of the DD solver.

WHY BLUE WATERS

It is essential to have a petascale resource like Blue Waters to develop an application for exascale machines. The sheer size, the favorable queue structure, and the excellent throughput at large core counts were instrumental during the code development and test phase of this research. In the process of developing the DD solver, certain synchronization issues only showed up at large scale (P > 1,000). It was essential to be able to turn around large test runs quickly as the research team diagnosed the problem.

PUBLICATIONS & DATA SETS

J. O. Bordner and M. L. Norman, "Computational cosmology and astrophysics in adaptive meshes using Charm++," presented at the SC'18 PAW-ATM Workshop. [Online]. Available: <https://arxiv.org/abs/1810.01319>

Enzo -E GitHub repository. [Online]. Available: <https://github.com/enzo-project/enzo-e>

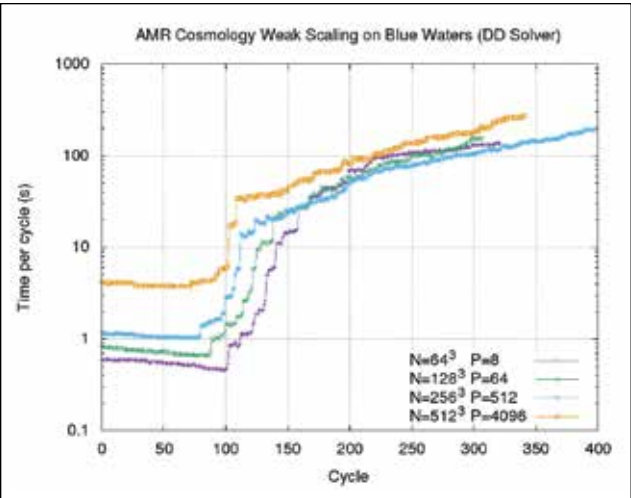


Figure 2: Wall time per timestep vs. timestep for four cosmological AMR simulations performed with Enzo-E comprising a weak scaling study. Growth in wall time reflects the creation of additional blocks as structure forms. The four simulations have 4<sup>3</sup>, 8<sup>3</sup>, 16<sup>3</sup>, and 32<sup>3</sup> octrees, with eight octrees assigned to each core.



DI

SIMULATING GALAXY FORMATION ACROSS COSMIC TIME

**Allocation:** NSF PRAC/1,500 Knh  
**PI:** Brian O'Shea<sup>1</sup>  
**Co-PIs:** David C. Collins<sup>2</sup>, John H. Wise<sup>3</sup>  
**Collaborators:** Cameron J. Hummels<sup>4</sup>, Britton D. Smith<sup>5</sup>, Molly S. Peeples<sup>6</sup>, John D. Regan<sup>7</sup>, Jason Tumlinson<sup>6</sup>, Lauren Corlies<sup>8,9</sup>

<sup>1</sup>Michigan State University  
<sup>2</sup>Florida State University  
<sup>3</sup>Georgia Institute of Technology  
<sup>4</sup>California Institute of Technology  
<sup>5</sup>University of Edinburgh  
<sup>6</sup>Space Telescope Science Institute  
<sup>7</sup>Dublin City University  
<sup>8</sup>LSST Corporation  
<sup>9</sup>University of Arizona

BW

MP

BI

EXECUTIVE SUMMARY

This project addresses two fundamental questions in galaxy formation: How do the first stars and galaxies influence the rest of cosmological structure formation over the following 13 billion years? And how does the cycling of gas into and out of galaxies through cosmic inflows and supernova-driven winds regulate the behavior of these galaxies?

The research team addresses these questions via a suite of physics-rich, high-dynamic-range adaptive mesh refinement simulations done with the Enzo code (enzo-project.org), which has been modified to scale to large core counts on Blue Waters. Through these simulations, the researchers have gained key insights into the formation of the first supermassive black holes and into the ways in which the circumgalactic gas around galaxies acts as a “thermostat” that regulates star formation.

RESEARCH CHALLENGE

The primary goals of this research are to understand two fundamental issues in galaxy formation: the birth and growth of the first generations of galaxies and their connections to present-day galaxies like the Milky Way, and the “baryon cycle” in galaxies like the Milky Way; in other words, the movement of gas into and out of galaxies, and how this regulates the behavior of star formation in galaxies. Both of these questions are critical to interpreting observations of galaxies over the age of the Universe through both current observatories (such as the Hubble Space Telescope and the 10-meter Keck telescope on Mauna Kea) and future observatories (such as the James Webb Space Telescope and the Large Synoptic Survey Telescope). All of the calculations needed to study these problems require simulations with extremely high dynamic range in space and time, complex physics (often including radiation transport and nonequilibrium gas chemistry), and large simulation volumes.

METHODS & CODES

The researchers’ simulation tool of choice is the Enzo code [1,2; <http://enzo-project.org>], an open source and community-devel-

oped software platform for studying cosmological structure formation. Enzo allows the inclusion of all the critical physical components needed to study galaxy formation—gravity, dark matter dynamics, fluid dynamics, the microphysics of plasmas, and prescriptions for star formation and feedback—and can scale to large numbers of CPUs. All analysis was done with the yt code [3; <http://yt-project.org>].

RESULTS & IMPACT

The main results involve the growth of supermassive black holes in the early Universe and the cycling of gas into and out of galaxies. The former result [4,5] demonstrates a novel “direct collapse” mechanism for gas clouds to turn into intermediate-mass black holes (with masses of thousands or tens of thousands of solar masses) rather than first turning into stars. This is important because it provides the first plausible formation mechanism for the extremely massive (hundreds of millions to billions of solar masses) black holes seen in the early Universe. The project has also demonstrated that the frequency with which these black holes form is consistent with the frequency that early-Universe supermassive black holes are observed.

The second set of important results involves the exploration of gas cycling into and out of galaxies [6–8]. Simulations demonstrate that massively increased physical resolution in the “circumgalactic medium”—the gas outside of the stellar disk of a galaxy but is bound to the galaxy by gravity, and composes almost half of the mass of the baryons in the galaxy—is incredibly important. The work suggests that increasing the resolution by more than an order of magnitude beyond previous state-of-the-art calculations results in the appearance of both spatial and chemical features that are seen in observations but not in previous simulations. This research is revolutionizing the understanding of the interface between the stellar component of galaxies and the diffuse corona of gas that surrounds them and provides predictions of observational quantities relating to quasar absorption line spectra and to the direct emission of radiation from the circumgalactic medium, as observed by Keck’s KCWI instrument.

WHY BLUE WATERS

The simulations used to properly model galaxies in both the early Universe and the present day require extremely high spatial and temporal dynamic ranges and also require complex physics—most importantly, radiation transport and nonequilibrium gas chemistry. Furthermore, large simulation volumes (and thus many resolution elements) are needed to model the many early galaxies that will merge together to create a Milky Way-like galaxy at the present day, and in the research team’s present-day galaxy simulations, huge numbers of cells are required to accurately resolve the circumgalactic gas. Taken together, this project requires the use of a supercomputer with large memory and disk space to accommodate the tremendous data set sizes; large computational resources; and an extremely high bandwidth, low-latency communication network to enable significant scaling of the radiation transport code. Blue Waters is the only machine available to the academic community that fits all of these requirements.

PUBLICATIONS & DATA SETS

J. Wise *et al.*, “Formation of massive black holes in rapidly growing pre-galactic gas clouds,” *Nature*, vol. 566, pp. 85–88, 2019.  
[5] J. Regan *et al.*, “The emergence of the first star-free atomic cooling haloes in the Universe,” *MNRAS*, submitted, 2019, arXiv:1908.02823.  
[6] M. Peeples *et al.*, “Figuring out gas and galaxies in Enzo (FOGGIE), I. Resolving simulated circumgalactic absorption at  $2 \leq z \leq 2.5$ ,” *Astrophys. J.*, vol. 873, p. 129, 2019.  
[7] L. Corlies *et al.*, “Figuring out gas & galaxies in Enzo (FOGGIE), II. Emission from the  $z=3$  circumgalactic medium,” *Astrophys. J.*, submitted, 2019, arXiv:1811.05060.  
[8] C. Hummels *et al.*, “The impact of enhanced halo resolution on the simulated circumgalactic medium,” *Astrophys. J.*, vol. 882, p. 156, 2019.



PROCESSING DARK ENERGY CAMERA DATA TO MAKE THE WORLD’S BEST MAP OF THE NIGHT SKY

**Allocation:** Director Discretionary/80.1 Knh  
**PI:** Donald Petravick<sup>1</sup>  
**Collaborators:** Eric Morganson<sup>1</sup>, Monika Adamow<sup>1</sup>, Doug Friedel<sup>1</sup>, Robert Gruendl<sup>1</sup>, Michael Johnson<sup>1</sup>

<sup>1</sup>University of Illinois at Urbana–Champaign

EXECUTIVE SUMMARY

The Dark Energy Camera (DECam) on the Blanco 4-meter telescope has been a premier instrument for making astronomical surveys during its seven years of operation. The largest of these surveys is the Dark Energy Survey (DES), whose data was processed by the National Center for Supercomputing Applications (NCSA). In addition to DES, DECam has produced 200,000 exposures from various smaller surveys. These smaller surveys were processed with multiple pipelines, many of which have known deficiencies. In addition, their processed data are not publically available for search. The research team is processing these 200,000 exposures along with the original DES data. This meta-survey will immediately be much deeper than any survey of similar size. It will be used to study Milky Way structure, galaxy clusters, and

solar system objects. It will also provide the “before” image for astronomical transients in multimessenger astronomy.

RESEARCH CHALLENGE

The DECam has a three square degree field of view, and each image it produces contains 50 million pixels [1]. This has made it a major instrument for making astronomical surveys that cover large areas of the sky. The largest of these surveys, the Dark Energy Survey (DES) [2], was made with nearly half of the DECam observing time, and its data were processed at NCSA with a highly developed image processing pipeline. Smaller DECam surveys like the Dark Energy Camera Legacy Surveys [3] and the Dark Energy Camera Plane Survey [4] comprise months of DECam data and cover thousands of square degrees. However, they

were processed with multiple pipelines that are known to be quantitatively inferior to the DES pipeline. In addition, while the raw data for these surveys are publicly available, searchable catalogs are not. The research team is processing all DECam data with the DES pipeline and releasing it publicly. For many applications, this will increase the effective DES area from 5,500 square degrees to 25,000 square degrees, or roughly 60% of the sky.

METHOD & CODES

The research team uses the DES image processing pipeline [5]. In addition to correcting or masking camera artifacts with the most recent proven algorithms, the pipeline uses human-inspected calibration images and a superior background subtraction method based on principle component analysis of the background [6]. The net result of these improvements is that DES reduced data have a photometric (brightness) precision of 0.5%. DECam data reduced with other pipelines have a photometric precision of between 2% and 6%, depending on the pipeline.

RESULTS & IMPACT

Using Blue Waters, the researchers have processed 66,740 DECam exposures. This comprises 33 TB of raw data turned into 100 TB of processed data. These data cover roughly 15,000 square degrees of the sky, with an average coverage of 12 exposures in any given area. These images are the deepest available over most of this new area.

The research team has given early access to these data to DES scientists. Preliminary results (with publications forthcoming) include the discovery of new outer Milky Way structures, improved observations of galaxy clusters, and the discovery of solar system objects (asteroids). The Milky Way structures will be used to trace the ancient collisions the Milky Way has had with its neighbors that have led to its current form. Images of galaxy clusters will be used to determine their masses, and through these measurements, to understand the growth of cosmological structure in the Universe. The discovery of new solar system objects helps complete a census of nearby objects, including those that may someday impact Earth. All of these discoveries required the new depth and wide area that these data provide and could not have been made had the data not been processed with Blue Waters and made available through this project.

In addition to this ongoing research, the team’s survey will provide the template, or “before” image, needed to detect astronomical transients, including those initially detected with the Laser Interferometer Gravitational-Wave Observatory and other multimessenger detectors. Having archival templates will allow these objects to be studied as they are dynamic instead of hours or days later when the data have been fully analyzed.

WHY BLUE WATERS

The challenges of processing 33 TB of raw data into 100 TB of usable images and catalogs extend beyond the hundreds of thousands of core hours needed to perform the processing. This work

involved the importation of a large, specialized software stack; constant transfer of large files across limited network space; and millions of calls to a central database to organize the processing and data. The Blue Waters staff showed the system’s versatility by working with the research team to solve each of these problems so that data could be processed quickly with minimal human interaction and transferred back to the home system reliably. This allowed the team not only to perform the necessary computations but to present the astronomical community with a usable and exciting new data set before other groups even thought to begin to reduce the data.

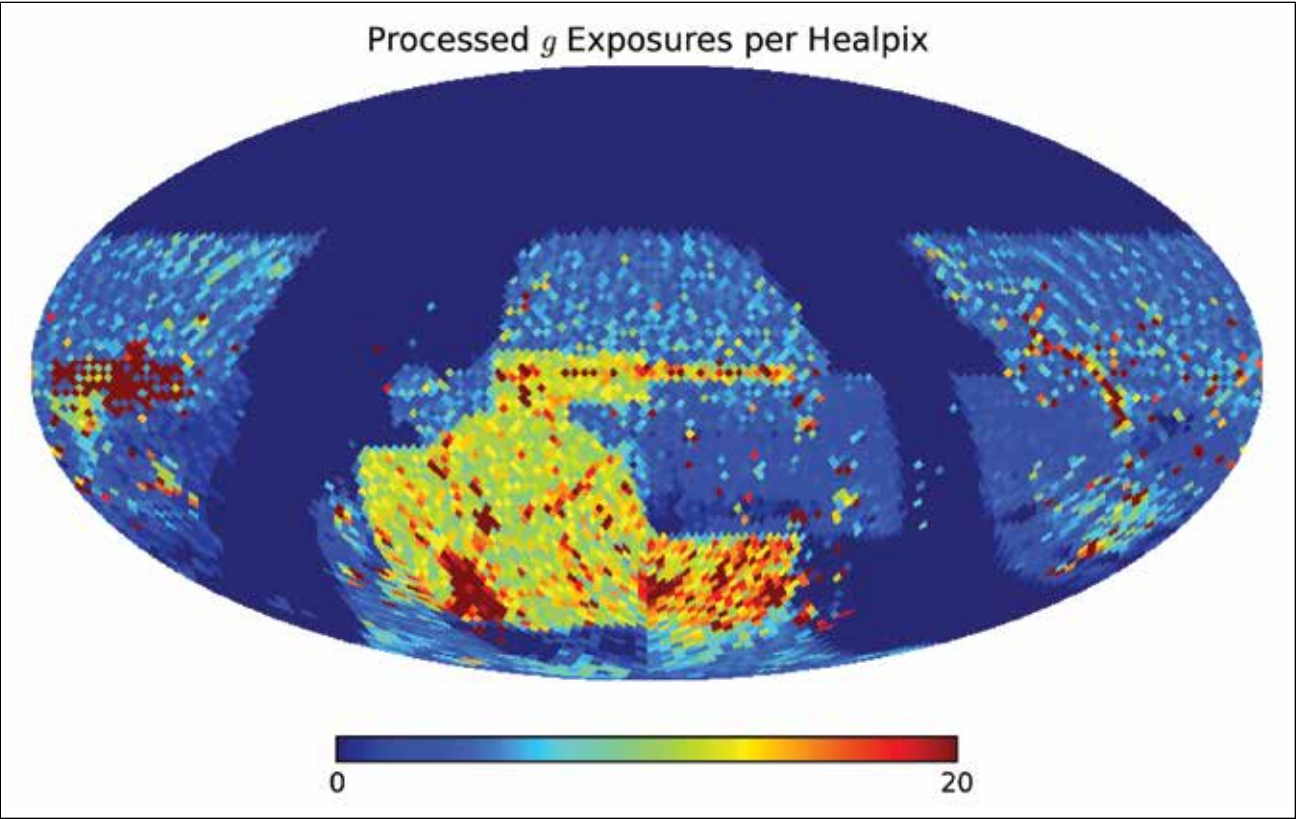


Figure 1: The spatial distribution of g band DECam data processed by the DECADE (DECam All Data Everywhere) team. These data cover half the night sky to unprecedented depth.



# COUPLING THE SOLAR WIND AND LOCAL INTERSTELLAR MEDIUM IN THE ERA OF THE NEW HORIZONS, INTERSTELLAR BOUNDARY EXPLORER, PARKER SOLAR PROBE, ULYSSES, AND VOYAGER SPACECRAFT

**Allocation:** NSF PRAC/2,900 Knh  
**PI:** Nikolai Pogorelov<sup>1</sup>  
**Co-PI:** Jacob Heerikhuisen<sup>1</sup>  
**Collaborators:** Sergey Borovikov<sup>1</sup>, Tae Kim<sup>1</sup>, Mehmet Sarp Yalim<sup>1</sup>

<sup>1</sup>University of Alabama, Huntsville

## EXECUTIVE SUMMARY

The research team has investigated physical phenomena occurring when the solar wind (SW) interacts with the local interstellar medium (LISM). These include: (1) the effect of non-Maxwellian plasma distribution on the SW–LISM interaction; (2) the propagation of coronal mass ejections, constrained by multi-viewpoint images, through the solar wind flow governed by photospheric magnetograms; (3) transient phenomena affecting space weather at Earth and other planets; (4) magnetohydrodynamic (MHD) instabilities and magnetic reconnection in the presence of turbulence; (5) the effect of nonthermal pickup ions (PUIs) on spacecraft measurements; (6) energetic neutral atom observations from the Interstellar Boundary Explorer (IBEX) through direct 3D, MHD–kinetic simulations; (6) the heliospheric effect on the observed anisotropy of galactic cosmic rays with energies on the order of TeV and the origin of this anisotropy; and (7) the global structure of the solar wind flow along the Parker Solar Probe (PSP) trajectory. Our simulations help interpret IBEX, New Horizons, PSP, Ulysses, and Voyager measurements, as well as air shower observations.

## RESEARCH CHALLENGE

The Grand Challenge of this research is to investigate fundamental physical phenomena that start on the solar surface and result in solar wind acceleration and propagation through interplanetary space toward the boundary of the heliosphere, where the solar wind interacts with the local interstellar medium. The research team studied plasma instabilities, magnetic reconnection, cosmic ray transport, and kinetic effects of partial ionization in plasma. These included the birth/death of secondary neutral atoms and nonthermal PUIs, and phenomena driven by MHD turbulence. Most of the research team’s simulations were data-driven and also validated by observations from such space missions as IBEX [1], New Horizons (NH) [2], PSP [3], Ulysses [4], Voyager [5], and the fleet of near-Earth spacecraft.

To drive the coronal model, the researchers used the wealth of magnetogram data accompanied by the satellite missions STEREO and SOHO observations. This allowed the team to develop a new approach, preserving the shape and speed of a coronal

mass ejection as well as the plasma mass and poloidal magnetic field fluxes carried by it. Simulations were especially focused on the interstellar mission of Voyager 1 and 2 (V1 and V2) spacecraft, which crossed the entire heliosphere and are now traversing the LISM. For the first time in the history of mankind, we are acquiring *in situ* information on the properties of LISM plasma, energetic particles, and magnetic fields at the heliospheric boundary. Voyager data are complemented by the IBEX observations of energetic neutral atoms (ENAs) in different energy bands. Since ENAs are born owing to charge exchange of nonthermal PUIs with other neutrals, the team was able to investigate the effect of non-Maxwellian proton distribution function on the heliospheric structure and support the interstellar origin of the IBEX ribbon.

## METHODS & CODES

The researchers solved the equations of ideal MHD coupled with the kinetic Boltzmann equation describing the transport of neutral atoms. In a less strict approach, the flow of atoms was

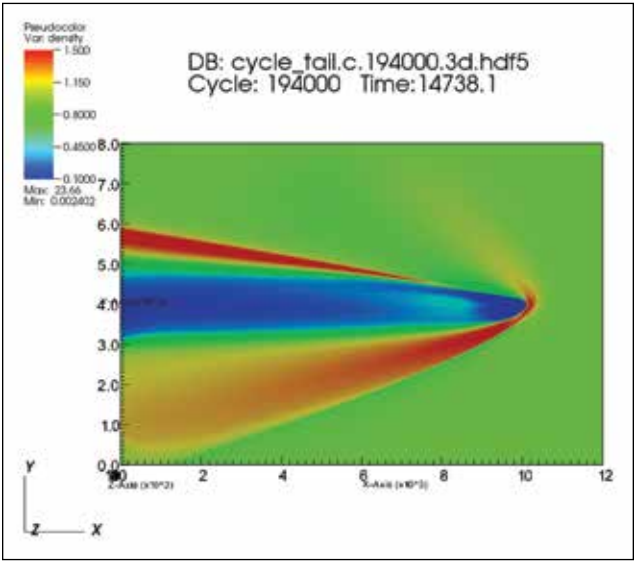


Figure 1: Solar wind interaction with the local interstellar medium. Plasma density distribution in the cross-section of maximum flaring of the heliopause exhibits a very long heliotail and additional, unexpected discontinuities.

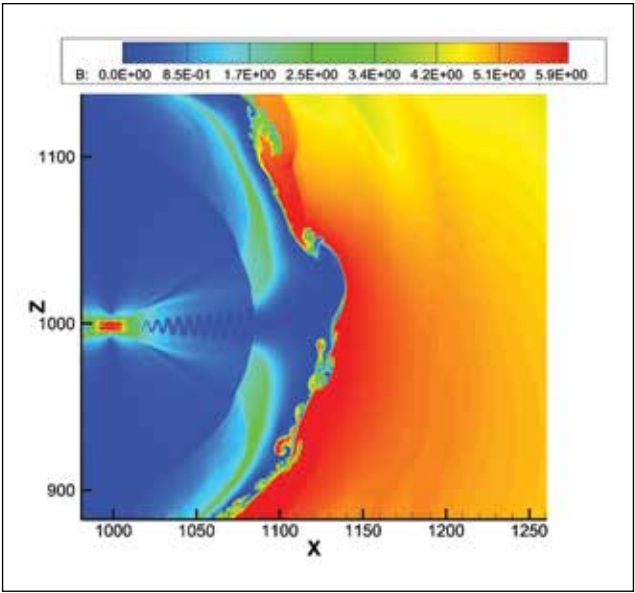


Figure 2: Solar wind interaction with the local interstellar medium. A snapshot of the time-dependent distribution of magnetic field magnitude at the heliospheric boundary. The heliopause instability is stronger in the northern hemisphere, but there are signatures of magnetic reconnection in the southern hemisphere.

modeled with a few systems of the Euler gas dynamic equations describing the atom populations that differ by the domain of their origin. The team has developed both fluid dynamics and kinetic models for PUIs and turbulence generated by kinetic instabilities of their distribution function. All these are components of a Multi-Scale Fluid-Kinetic Simulation Suite—an adaptive mesh refinement code built on the Chombo framework.

## RESULTS & IMPACT

- The team has developed a powerful data-driven solar model that allowed for the simulation of the solar wind flow and interplanetary magnetic field as a function of time along the Earth; New Horizons spacecraft; and Pluto, Neptune, and Uranus trajectories.
- This work represents a substantial breakthrough in modeling flows of partially ionized plasma in the presence of PUIs by designing special boundary conditions for the latter at the heliospheric termination shock. This is of utmost importance because such boundary conditions are intrinsically kinetic.
- The team has developed a unique solar wind model that is based on synchronous vector magnetograms from the Solar Dynamics Observatory. This made it possible, for the first time in the history of solar wind simulations, to create a mathematically consistent model of solar corona.
- This study has analyzed quantitatively the distribution of quantities in the heliospheric boundary layer—a region of interstellar plasma in front of the heliopause that is characterized by depressed plasma density and enhanced interstellar magnetic field. The MHD instabilities and magnetic reconnection have

been analyzed with high resolution in space and time, owing to the high computing power of Blue Waters.

- The research group’s numerical simulations have shown features of magnetic reconnection near the heliopause in the southern hemisphere and, in particular, near the point where Voyager 2 crossed the heliopause.
- The researchers have reproduced the 5 TeV cosmic ray anisotropy observed in the Tibet air shower experiment. It has been demonstrated that the heliosphere provides most of the higher-than-dipole contributions to this anisotropy. The cosmic ray anisotropy in the pristine LISM has been derived and the likely source of this anisotropy has been identified.
- The team’s new, data-driven coronal mass ejection simulations have been extended to involve the poloidal magnetic flux observations in the photosphere, which improves the quality of simulations tremendously.
- This project’s numerical simulation results have found their way to the broader space science community and received publicity. A few internal and national press releases were focused on this research, powered by Blue Waters.

## WHY BLUE WATERS

Blue Waters is not just a supercomputer with higher-than-usual allocation opportunities. It comes with an efficient and highly professional support staff, who responded to all our concerns and were very helpful in the development of job scheduling and visualization strategies. The PAID and Student Fellowship opportunities were also extremely valuable.

## PUBLICATIONS & DATA SETS

- E. Fraternali, N. V. Pogorelov, J. D. Richardson, and D. Tordella, “Magnetic turbulence spectra and intermittency in the heliosheath and in the local interstellar medium,” *Astrophys. J.*, vol. 874, p. 40, 2019.
- J. Heerikhuisen, E. J. Zirnstein, N. V. Pogorelov, G. P. Zank, and M. Desai, “The effect of suprathermal protons in the heliosheath on the global structure of the heliosphere and heliotail,” *Astrophys. J.*, vol. 874, p. 76, 2019.
- T. Singh, M. S. Yalim, and N. V. Pogorelov, “A data-constrained model for coronal mass ejections using the graduated cylindrical shell method,” *Astrophys. J.*, vol. 864, p. 18, 2018.
- T. Singh, M. S. Yalim, N. V. Pogorelov, and N. Gopalswamy, “Simulating solar coronal mass ejections constrained by observations of their speed and poloidal flux,” *Astrophys. J. Letters*, vol. 875, p. L17, 2019.
- E. J. Zirnstein, J. Heerikhuisen, D. J. McComas, N. V. Pogorelov, D. B. Reisenfeld, and J. R. Szalay, “Simulation of the solar wind dynamic pressure increase in 2014 and its effect on energetic neutral atom fluxes from the heliosphere,” *Astrophys. J.*, vol. 859, p. 104, 2018.



# INTERIOR DYNAMICS OF YOUNG STARS REVEALED BY 3D HYDRODYNAMIC SIMULATIONS

**Allocation:** Innovation and Exploration/400 Knh  
**PI:** Jane Pratt<sup>1</sup>  
**Collaborators:** Isabelle Baraffe<sup>2</sup>, Tom Goffrey<sup>3</sup>, MULTidimensional Stellar Implicit Code developers' team

<sup>1</sup>Georgia State University  
<sup>2</sup>University of Exeter  
<sup>3</sup>University of Warwick

## EXECUTIVE SUMMARY

Our current understanding of the evolution of stars is drawn from one-dimensional calculations based on simple phenomenological approaches and fitted observational data. To interpret the data produced from the recent space missions CoRoT, Kepler, and GAIA, as well as to understand new high-quality data from the upcoming missions TESS and PLATO, it is necessary to study realistic stellar conditions using three-dimensional hydrodynamic simulations.



Toward the center of a star, plasma becomes progressively hotter and more dense; convective plumes characterize some layers, while others contain a variety of waves. Mixing of chemically different material and heat between these layers, and the mixing due to the different dynamics within these layers, are basic ingredients needed to predict the course of stellar evolution. This work involves large and physically complex 3D simulations that address the multiscale dynamics of the entirety of the stellar interior, a problem for which Blue Waters is ideally suited.

## RESEARCH CHALLENGE

Through simulations of a young, low-mass star, the research team studied convection in a large convective envelope, as well as convective penetration, rotation, and internal gravity waves. The main challenge for the simulations was to design them to be realistic to the stellar interior. The team set up the simulations using a realistic stratification produced with the MESA stellar evolution code and used a realistic tabulated equation of state and opacity in the calculations. To produce realistic simulation results, the group also needed sufficient spatial resolution to accurately reproduce the multiscale flows in the large convection zone and lower radiative zone. They then needed to run these simulations for a long period of simulation time to capture a full range of the dynamics. This is, therefore, a project of large computational scale that requires Blue Waters' unique capabilities.

## METHODS & CODES

This project used the MULTidimensional Stellar Implicit Code (MUSIC), a code that has been designed and developed under a European Research Council Advanced Grant over the last six years. MUSIC solves fully compressible fluid equations in a spherical geometry using a second-order finite volume method and MPI-FORTRAN. MUSIC also uses fully implicit time integration in order to cover long windows of stellar dynamics. The time integration algorithm is second-order and centers on a Jacobian-Free Newton-Krylov method using physics-based preconditioning. Each aspect of the design of MUSIC has been chosen to seamlessly extend one-dimensional stellar evolution calculations into three dimensions.

Figure 1: Visualization of radial velocity in 3D simulation of a young sun. Red indicates outward flows while grey indicates inward flows. Simulation performed with the MUSIC code.

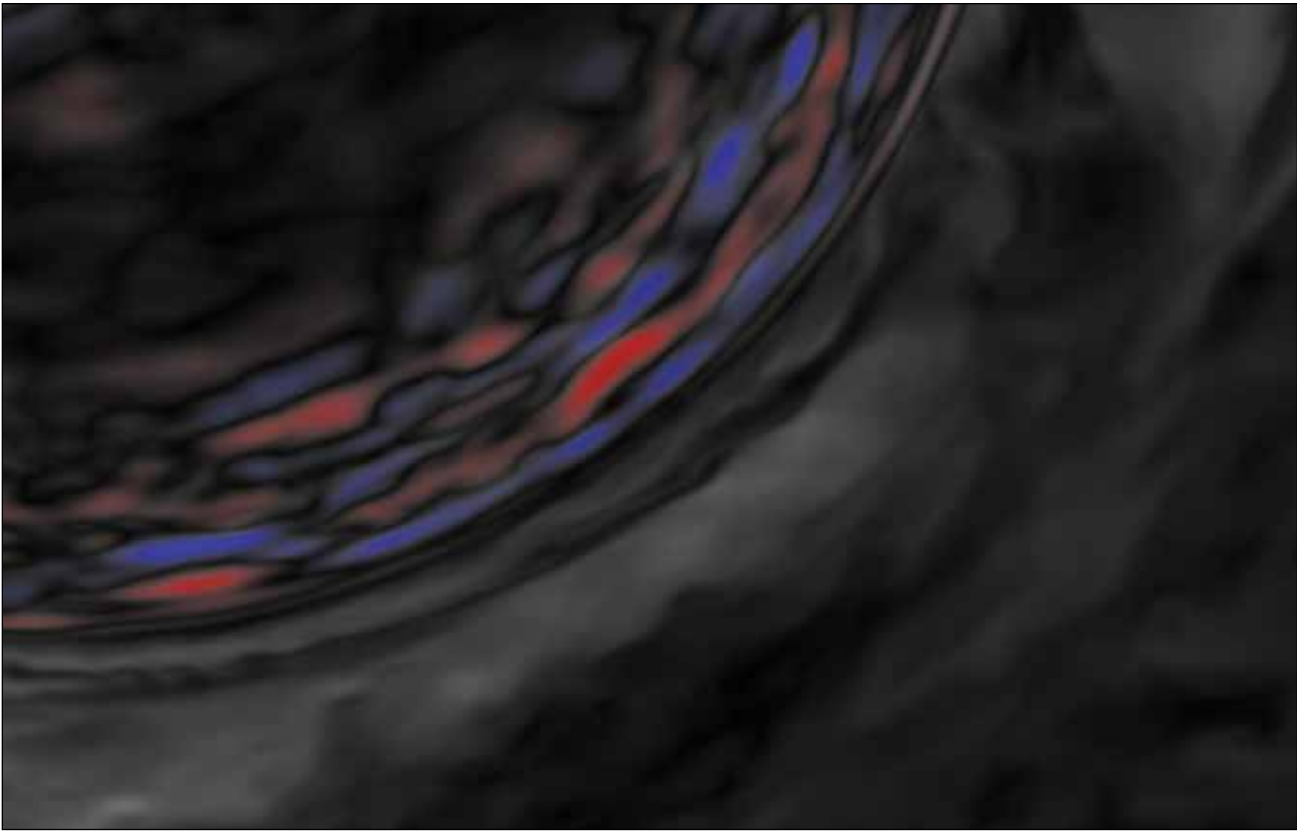


Figure 2: Visualization of large thermal fluctuations, zoomed in around the bottom of the convection zone and the top of the radiative zone of the young sun. Blue indicates large negative thermal fluctuations while red indicates large positive thermal fluctuations. Greyscale shows the velocity magnitude in this region.

## RESULTS & IMPACT

Early results from a simulation performed on Blue Waters are shown in Figs. 1 and 2. In Fig. 1, radial velocities are shown in a spherical wedge that fills 80% of the star's radius. Smaller-scale convective flows are visible near the surface, while the extent of these radial flows is larger, deep in the stellar interior.

Although these flows are of a length scale similar to the stellar radius, they also clearly have small-scale features that are important for diffusion and chemical mixing. The research team is currently producing a long time-sequence of data in order to perform a complete statistical analysis of these convective flows and the consequences for mixing when they overshoot the convection zone.

One of the consequences of convective plumes overshooting into the radiative zone is the mixing of colder fluid into a hotter layer, potentially changing the temperature gradients within the stellar interior. Fig. 2 shows that heat becomes trapped in waves that are excited in the overshooting layer. A full study of this phenomenon is underway and will be completed using the simulations currently running on Blue Waters.

This work is expected to result in improved models for stellar evolution, which the researchers expect to be included in open-source code such as MESA. As a result of this project, stellar physicists will be better able to understand how a young star such as that in this project's simulations can evolve into a sun.

## WHY BLUE WATERS

The research team has carried out its simulations on Blue Waters because of the large grid size and large number of timesteps necessary for simulations of a star. Simulations of this size would be expected to run continuously for several years on a university-size system and can be completed on Blue Waters in less than a year. Blue Waters' staff played an essential role in the success of this project by installing the Trilinos library, used by MUSIC.



MODELING OF GALAXY POPULATIONS

**Allocation:** NSF PRAC/9,000 Knh  
**PI:** Thomas Quinn<sup>1</sup>  
**Co-Pi:** Fabio Governato<sup>1</sup>  
**Collaborators:** Michael Tremmel<sup>2</sup>, Arif Babul<sup>3</sup>, Iryna Butsky<sup>1</sup>

<sup>1</sup>University of Washington  
<sup>2</sup>Yale University  
<sup>3</sup>University of Victoria

EXECUTIVE SUMMARY

Clusters of galaxies are both a useful probe of cosmology and a laboratory for understanding galactic feedback processes. However, modeling galactic-scale feedback processes in the context of a cluster presents a computational challenge because of the large dynamic range involved. Through the use of a highly scalable N-body/smooth-particle hydrodynamics code running on Blue Waters, the research team is tackling this challenging problem. The results show that models that have successfully reproduced the morphology and number densities of field galaxies can also produce realistic models of cluster galaxies. Large computational resources with high-performance networks are necessary for these calculations, however.

RESEARCH CHALLENGE

Groups and clusters of galaxies are the largest bound objects in the Universe, containing more than a third of its warm-hot diffuse gas and a significant fraction of all galaxies. Consequently, understanding the physical processes that occur in group and cluster environments, including the interactions among the dark matter, hot diffuse gas, stars, and active galactic nuclei (AGN), is key to gaining insights into the evolution of baryons and galaxies across the age of the Universe. Furthermore, galaxy clusters are one of the few places where the majority of the baryons are visible via X-ray and microwave. In contrast to field galaxies, where feedback from supernovae and AGN puts gas into a mostly invisible circumgalactic medium, feedback from cluster galaxies will impact the state of the intracluster medium (ICM). Hence, clusters will provide very tight constraints on our understanding of galactic feedback processes.

Clusters of galaxies are also key probes of cosmology and large-scale structure. Their size makes them visible across a wide range of redshifts and their population statistics are sensitive to cosmological parameters such as the amplitude of the initial power spectrum and the evolution of the cosmic expansion rate. However, using clusters as cosmological probes requires understanding of the relationship between observables and the total mass of the cluster, which in turn requires the detailed modeling of the gravitational/hydrodynamic processes using large simulations.

METHODS & CODES

The research team used the highly scalable N-body/hydrodynamics code ChaNGa to model the formation and evolution of a population of galaxies in a Coma-sized galaxy cluster, including their contribution to and interaction with the ICM. This code is built on the Charm++ [1] parallel programming infrastructure. It leverages the object-based virtualization and data-driven style of computation inherent in Charm++ to adaptively overlap communication and computation and achieve high levels of resource utilization on large systems. The code has been shown to scale well to 500,000 cores on Blue Waters [2].

The code includes a well-constrained model for star formation and feedback, with improved implementation of supermassive black hole formation, growth, mergers, and feedback [3,4]. In a previous Blue Waters allocation, the team demonstrated that these models can reproduce populations of field galaxies at intermediate-to-high redshift [5] and can reproduce the observed stellar mass-halo mass relationship of galaxies from dwarfs up to galaxy groups [4].

These simulations are being compared to observations of cluster galaxies to understand the physical and temporal origin of their morphologies. The model ICM will be compared to X-ray and microwave data via the Sunyaev-Zeldovich effect to understand the relationship among these observables and the underlying gas properties. Finally, the overall mass distribution will be used to better understand how these clusters gravitationally amplify the light from background galaxies.

RESULTS & IMPACT

The team's simulations are advancing the state of the art in simulations of galaxy clusters, particularly in the relationship among galactic processes and the cluster environment. The research has shown that the supermassive black holes (SMBHs) at the center of galaxies can effectively regulate the star formation rate in cluster galaxies just as they do in field galaxies. Furthermore, the process that establishes the observed relationship between the SMBH mass and the stellar mass of the galaxy is independent of the galactic environment and is determined by the correlation between SMBH accretion rate and the star formation rate. On the other hand, the state of the gas in the core of the cluster is not determined by the SMBH feedback but, rather, is set by the mergers of larger substructures. The team's high-resolution simulations

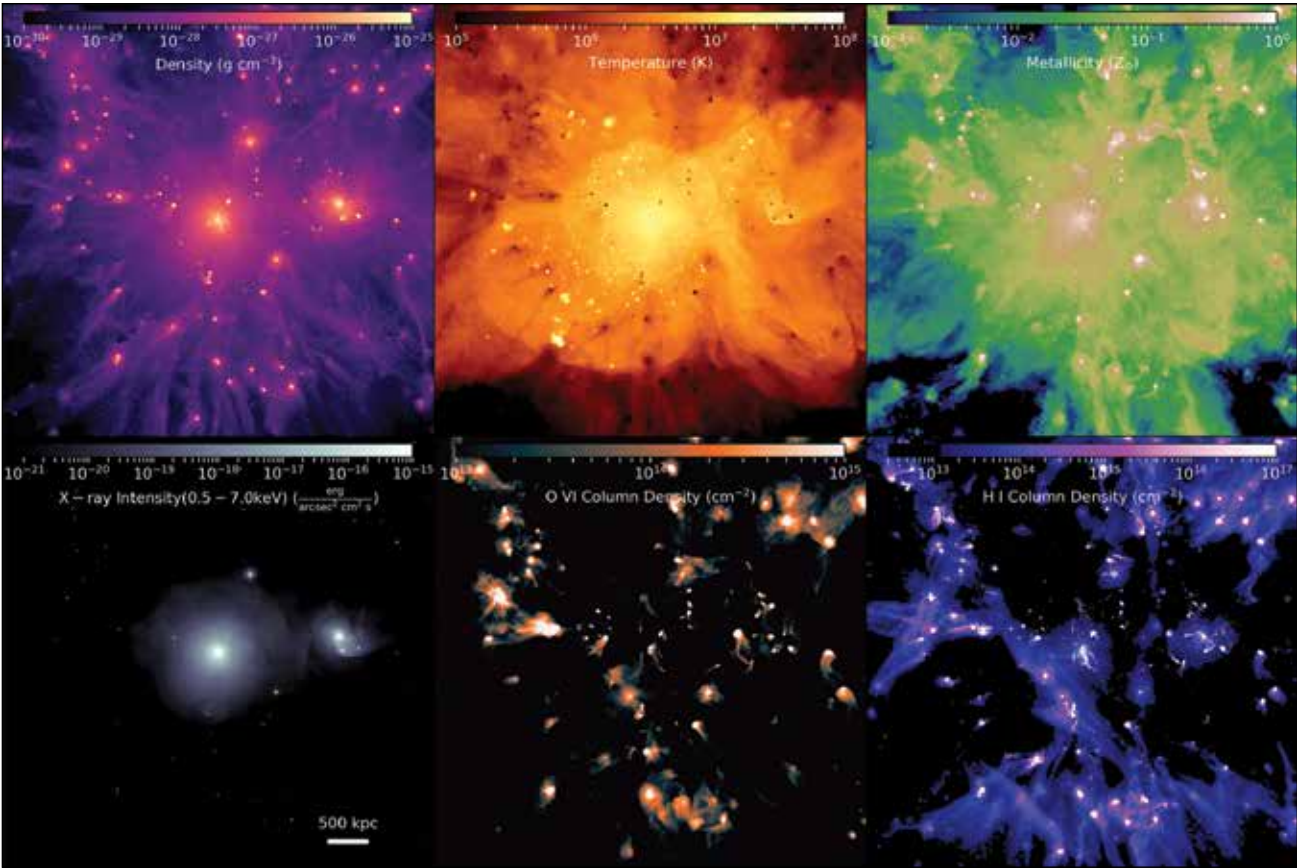


Figure 1: The state of the intracluster gas in the cluster simulation. The top row shows the intrinsic state of the gas: density temperature and metal content. The bottom row shows observable properties: X-ray intensity, column density of oxygen VI, and column density of neutral hydrogen.

allow them to make predictions about the multiphase structure of the cluster gas and how it can be observed with the Hubble Space Telescope and future UV-capable space telescopes (Fig. 1).

WHY BLUE WATERS

Our scientific goals require modeling over a large dynamic range in mass and space. We have demonstrated that mass resolutions on the order of 10<sup>5</sup> solar masses are needed to accurately follow star formation and galaxy morphology. Likewise, we need to model a galaxy cluster on the order of 10<sup>15</sup> solar masses that is comparable to those observed over a range of redshifts. Hence, simulations require approximately 10 billion particles. Such simulations can only be run on the largest computers available. Furthermore, the long-range nature of gravity requires a high-performance, low-latency network to perform the calculations.

PUBLICATIONS & DATA SETS

M. Tremmel *et al.*, “Dancing to ChaNGa: A self-consistent prediction for close SMBH pair formation time-scales following galaxy mergers,” *Mon. Notices Royal Astron. Soc.*, vol. 475, no. 4, pp. 4967–4977, 2018, doi: 10.1093/mnras/sty139.

M. Tremmel *et al.*, “Wandering supermassive black holes in Milky-Way-mass halos,” *Astrophys. J. Letters*, vol. 857, no. 2, pp. L22–27, 2018, doi: 10.3847/2041-8213/aabc0a.

E. Lentz *et al.*, “A new signal model for axion cavity searches from N-body simulations,” *Astrophys. J.*, vol. 845, no. 2, pp. 121–125, 2017, doi: 10.3847/1538-4357/aa80dd.

M. Tremmel *et al.*, “Dancing to ChaNGa: The formation of close pairs of supermassive black holes in cosmological simulations and implications for LISA,” presented at the 233rd Meeting of the American Astronomical Society, Seattle, WA, U.S.A., Jan. 7, 2019, Paper id.141.10.

M. Tremmel *et al.*, “Nature versus nurture: unraveling the mystery of ultra diffuse galaxy formation with the RomulusC galaxy cluster simulation,” presented at the 233rd Meeting of the American Astronomical Society, Seattle, WA, U.S.A., Jan. 10, 2019, Paper id.416.04.

M. Tremmel *et al.*, “Introducing RomulusC: A cosmological simulation of a galaxy cluster with unprecedented resolution,” *Mon. Notices Royal Astron. Soc.*, vol. 483, no. 3, pp. 3336–3362, 2019, doi: 10.1093/mnras/sty3336.



# EFFECTS OF ACTIVE GALAXY FEEDBACK ON THE INTRACLUSTER MEDIUM

**Allocation:** Illinois/935 Knh  
**PI:** Paul Ricker<sup>1</sup>  
**Co-PI:** Yinghe Lü<sup>1</sup>

<sup>1</sup>University of Illinois at Urbana–Champaign

## EXECUTIVE SUMMARY

The research team has used Blue Waters to apply a novel approach to a key question regarding the intracluster medium (ICM), the hot plasma trapped in the gravitational potential wells of galaxy clusters: namely, how relativistic jets produced by accreting black holes regulate the radiative cooling of the ICM. The team uses hydrodynamical simulations of cluster cores with a subgrid model for black hole accretion to study the ability of jets to efficiently and uniformly deposit heat into the ICM. The model improves on existing techniques by directly measuring the accretion rate onto the black hole, linking it to a model of the accretion disk, and using the result to determine the feedback efficiency. In 2019, the researchers have used parts of this framework to study the role of jet precession and ICM turbulence in making feedback more isotropic.

## RESEARCH CHALLENGE

Loss by the ICM should cause the gas to lose pressure support against gravity, condensing to form stars at prodigious rates. Generally, this star formation is not observed; instead, cluster central galaxies are overwhelmingly “red and dead,” with little or no star formation and very old stellar populations. Some heating process must offset the radiative cooling. The most likely candidate is energy input by the supermassive (approximately 109 solar mass) black holes (SMBHs) found at the centers of clusters. When actively accreting matter and producing relativistic jets, these are the central engines of what are called active galactic nuclei (AGN). High-resolution X-ray and radio observations clearly show AGN disturbing and heating the ICM gas.

However, tuning the amount of feedback to match the cooling rate and distributing it so that the gas is evenly heated are serious theoretical problems. This requires connecting processes occurring in accretion disks around black holes smaller than the solar system with plasma physics as much as 30,000 light years away. The modeling of AGN feedback usually involves a subgrid model and considerable simplification of the complex physics involved in the region surrounding the AGN’s central black hole. Such approaches still have major shortcomings. Often, the multiphase structure of the gas is not incorporated, and accretion rates are estimated using spherically symmetric models applied to data on scales much larger than the accretion region. The efficiency of feedback is taken to be a constant, tuned to roughly reproduce the central entropies of clusters.

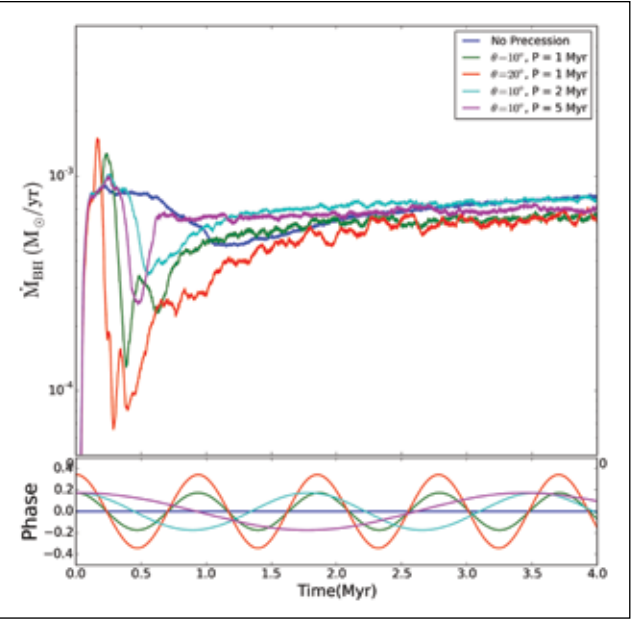


Figure 1: (Top panel) Accretion rate versus time for precessing jets with different periods and precession angles. (Bottom panel) Precession phase versus time. Time is measured in millions of years (Myr).

The research team’s approach, which borrows from techniques used in star formation and stellar evolution, applies a sink particle model to directly measure the accretion rate onto the black hole plus accretion disk system. The core of the cluster is resolved at scales of a few light years, allowing the complex gas structure in the vicinity of the AGN to be directly simulated. The efficiency and mode of feedback, as well as the growth rate of the black hole, are determined by matching the sink particle formalism to a model of the accretion disk that is informed by accretion disk simulations in the literature.

## METHODS & CODES

This work uses the adaptive mesh refinement (AMR) hydrodynamics plus N-body code FLASH 4. FLASH 4’s AMR is an Eulerian flow simulation method that allows high-resolution meshes to be placed only in regions of interest. It also uses a hydrodynamics solver based on the piecewise-parabolic method. In local idealized simulations, turbulent stirring is imposed in a manner that follows a Kolmogorov-like spectrum by adding a divergence-free,

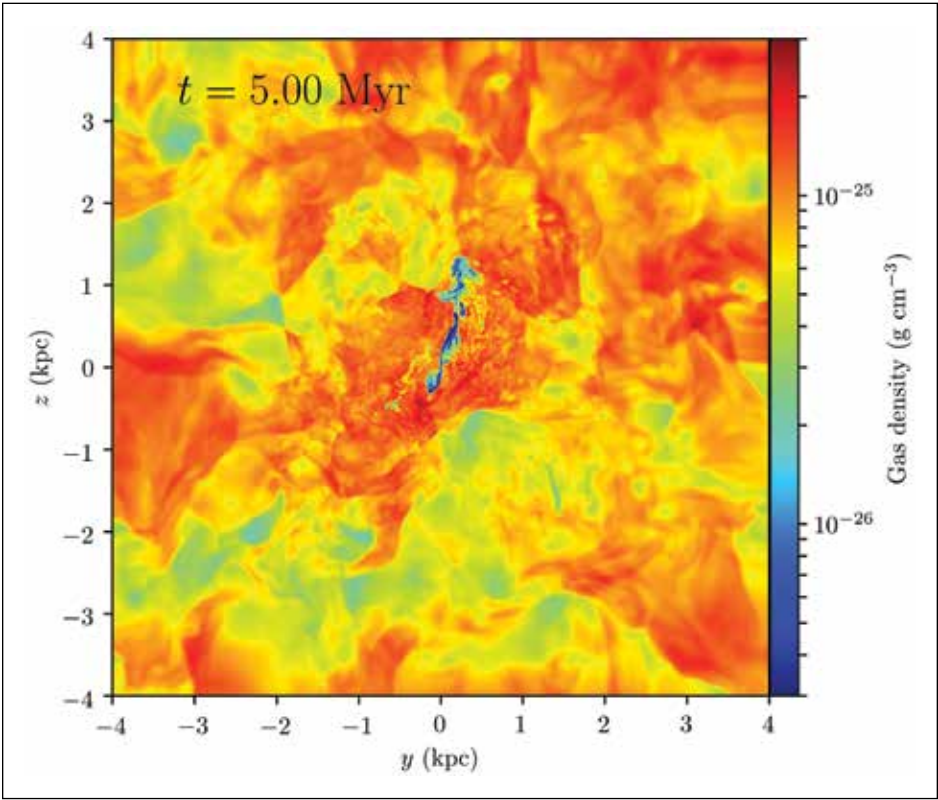


Figure 2: Gas density slice from a FLASH simulation including jet precession and turbulent driving. The region shown is 8 kiloparsecs (about 26,000 light years) across.

time-correlated velocity evolved by an Ornstein–Uhlenbeck random process. Radiative cooling using a metallicity-dependent cooling function assures that the gas reaches the densities needed to trigger feedback events.

## RESULTS & IMPACT

The researchers have used parts of the subgrid framework outlined above to study the ability of jet precession and turbulence to isotropize heating of the ICM. Accretion rates are measured using fluxes on a control surface surrounding the AGN, and feedback efficiency is held constant. They allow the AGN jet to precess, constraining the precession angle and period with observations of jet morphology in radio and X-ray maps. In some simulations, they introduce turbulent stirring to model the effects of galaxy wakes and cluster mergers. The team used recent Hitomi observations of the Perseus cluster to set the stirring energy.

They have conducted isolated cluster-core simulations with stationary and precessing jets as well as without turbulence and with weak or strong turbulence corresponding to different turbulent velocity dispersion values. Fig. 1 shows an example of the evolution of accretion rate in the team’s simulations. The researchers have found that precession helps jets deposit energy in a more distributed manner but suppresses accretion onto the SMBH by sweeping away a larger volume of gas. Larger precession angles seem to contribute to a time lag between accretion and feedback.

Fig. 2 shows an example of one of the runs with both precessing jets and turbulent driving. The researchers have discovered

that while turbulent driving itself enhances the kinetic energy of the ICM and triggers accretion, with precessing jets and weaker turbulent driving, the gas primarily passes through strong shocks produced by the jet, and cavity-like structures are formed. However, the situation changes with stronger turbulence, where the jet material gets blown away and the accretion process is enhanced by inflow of hot gas, allowing more energy to be deposited in the ICM. This coupling between jet precession and turbulent driving thus helps to regulate AGN feedback.

## WHY BLUE WATERS

Numerous complex physical processes are involved in the ICM. The required dynamic ranges in space and time are very large. These features make the ICM a natural setting for simulation studies that exploit the unique characteristics of Blue Waters. The researchers’ improved approach to measuring accretion rates and introducing feedback increases the complexity of the simulations in the innermost region close to the SMBH, which inevitably increases the cost. The computing ability of Blue Waters provides resources for such calculations to be done.

## PUBLICATIONS & DATA SETS

Y. Lu and P. M. Ricker, “AGN feedback from precessing jets on the turbulent intracluster medium,” presented at the 234th Meeting of the American Astronomical Society, St. Louis, MO, U.S.A., June 11, 2019, Paper id. 228.04.



GRAVITATIONAL AND ELECTROMAGNETIC SIGNATURES FROM BINARY BLACK HOLE–NEUTRON STAR MERGERS: A JET ENGINE FOR SHORT GAMMA-RAY BURSTS

Allocation: Illinois/2,774 Knh  
PI: Stuart Shapiro<sup>1</sup>  
Co-PIs: Milton Ruiz<sup>1</sup>, Antonios Tsokaros<sup>1</sup>  
Collaborator: Lunan Sun<sup>1</sup>

<sup>1</sup>University of Illinois at Urbana–Champaign

EXECUTIVE SUMMARY

Recently, the LIGO/Virgo scientific collaboration reported the detection of gravitational waves likely produced by a black hole–neutron star (BHNS) system (source S190426c). No electromagnetic counterparts were linked to this event. Using general relativistic magnetohydrodynamic simulations of a BHNS undergoing merger, the research team surveyed different configurations that differ in the spin of the BH ( $a/M_{\text{BH}} = -0.5, 0, 0.5, 0.75$ ); in the mass ratio ( $q = 3:1, q = 5:1$ ); and orientation of the magnetic field (aligned and tilted by  $90^\circ$  with respect to the orbital angular momentum). Only for configurations with  $a/M_{\text{BH}} \geq 0.5$  and aligned magnetic fields did the team find collimated, magnetically confined jets whose luminosity was consistent with typical short gamma-ray bursts and significant mass outflows that can induce detectable kilonova. By contrast, in case  $q = 5:1$  the remnant disk and magnetic field were too small to drive a jet and generate significant mass outflows or counterpart electromagnetic luminosity. High mass ratio BHNSs may therefore be the progenitors of S190426c.

RESEARCH CHALLENGE

Inspiraling and merging black hole–neutron star binaries are not only important sources of gravitational waves (GWs) but are also promising candidates for coincident electromagnetic (EM) counterparts. In particular, these systems are thought to be progenitors of short gamma-ray bursts (sGRBs) [1–4].

Coincident detection of GWs with EM signals from compact binary mergers containing neutron stars (NSs) could give new insight into their sources: GWs are sensitive to the density profile of NSs and their measurement enforces tight constraints on the equation of state of NSs [5]. Postmerger EM signatures, on the other hand, can help to explain, for example, the phenomenology of sGRBs and the role of BHNS mergers in triggering the nucleosynthesis processes in their ejecta.

BHNS scenarios have attracted a great deal of attention recently because of the first-ever candidate detection of GWs from a BHNS system (90% confidence) reported by the LIGO/VIRGO scientific collaboration [6]. As a crucial step to solidifying the role of BHNSs as multimessenger systems, the research team reported results from general relativity simulations of BHNS configura-

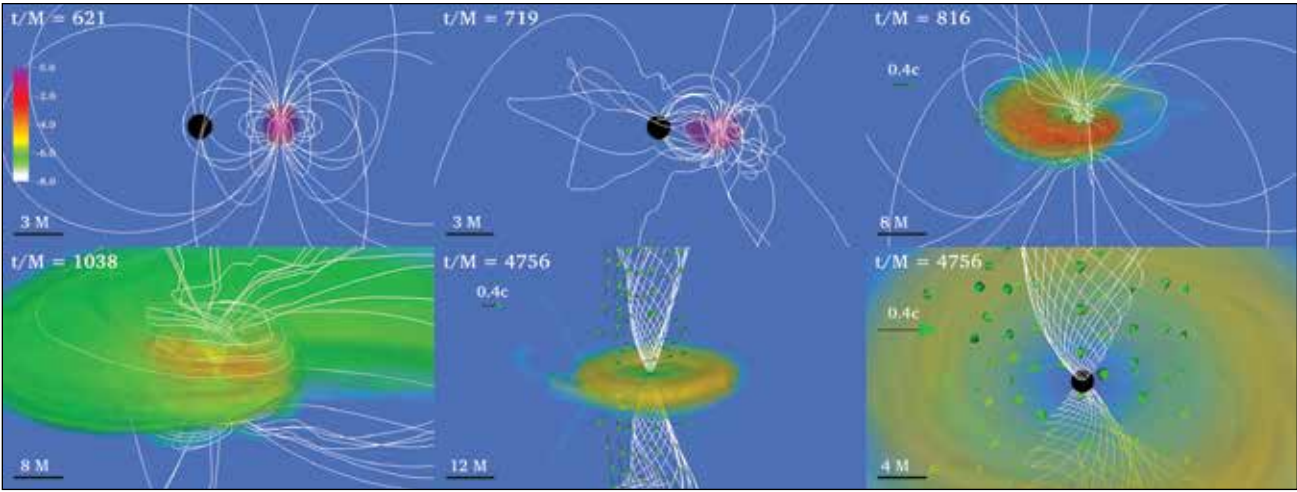
tions undergoing merger that differ in the spin of the BH ( $a/M_{\text{BH}} = -0.5, 0, 0.5, 0.75$ ); in the mass ratio ( $q = 3:1, q = 5:1$ ); and in the orientation of the magnetic field (aligned and tilted by  $90^\circ$  with respect to the orbital angular momentum), to determine their impact in the EM luminosity, ejecta, and other EM counterparts [7,8].

METHODS & CODES

Magnetohydrodynamic (MHD) numerical simulations in full general relativity require the solution of the Einstein field equations to determine the gravitational field as well as the relativistic MHD equations to determine the flow of matter and the electromagnetic fields. Together, the equations constitute a large system of highly nonlinear, multidimensional, partial differential equations in space and time. The researchers solved the above equations through their completely independent Illinois GRMHD code, which has been built over many years on the Cactus infrastructure and uses the Carpet code for adaptive mesh refinement but employs the team’s own algorithms and coding [9]. This code utilizes state-of-the-art high-resolution shock-capturing methods to evolve scenarios involving either vacuum or matter spacetimes, with or without magnetic fields. It utilizes the Baumgarte–Shapiro–Shibata–Nakamura formulation of the Einstein field equations with puncture gauge conditions. It solves the magnetic induction equation by introducing a vector potential and employs a generalized Lorentz gauge condition to reduce the spurious appearance of strong magnetic fields on refinement level boundaries [9].

RESULTS & IMPACT

In agreement with the researchers’ earlier calculations, where the star is seeded with a dipole magnetic field that extends from the interior to the stellar exterior [4,10], they found that the BHNS mergers listed above led to a disk + BH remnant with a rest-mass ranging from  $\sim 10^{-6} M_{\text{sun}}$  to  $\sim 10^{-3} M_{\text{sun}}$ , and dimensionless spin ranging from  $a/M_{\text{BH}} \sim 0.33$  to  $\sim 0.85$ . The early evolution, tidal disruption, and the merger phases are unaltered by the dynamically weak initial magnetic field. In the postmerger phase, the research team found that by around  $\Delta t \sim 3500 M \approx 88 (M_{\text{NS}}/1.4 M_{\text{sun}}) \text{ms}$  after the GW peak emission a magnetically driven jet is launched in the case where the initial spin of the BH companion is equal or larger than  $a/M_{\text{BH}} = 0.5$ . The lifetimes of the jet [ $\Delta t \sim 0.7 (M_{\text{NS}}/1.4 M)$



Volume rendering of rest-mass density  $\rho_0$  normalized to the initial NS maximum value (log scale) at selected times for a black hole (BH) with spin 0.5 and an irrotational neutron star. White lines denote the magnetic field while the arrows denote fluid velocities. The BH horizon is shown as a black sphere.

s] and outgoing Poynting luminosity [ $L_{\text{jet}} \sim 10^{52} \text{ erg/s}$ ] are consistent with observations of typical sGRBs [11], as well as with the Blandford–Znajek [12] mechanism for launching jets and their associated Poynting luminosities. In contrast, by the time the team terminated its simulations, they did not find any indication of an outflow in the other cases; in the nonspinning case ( $a/M_{\text{BH}} = 0$ ) and mass-ratio  $q = 3:1$ , a persistent fallback debris toward the BH was observed until the end of the simulation, the magnetic field above the BH poles was wound into a helical configuration, but the magnetic pressure gradients were still too weak to overcome the fallback ram pressure, and thus it is expected that a longer simulation will be required if a jet were to emerge. However, if the fallback debris timescale is longer than the disk accretion timescale [ $\Delta t \sim 0.36 (M_{\text{NS}}/1.4 M_{\text{sun}}) \text{s}$ ], the launching of a jet in this case may be suppressed. In the counter-rotating  $q = 3:1$  BHNS configuration ( $a/M_{\text{BH}} = -0.5$ ) the star plunges quickly into the BH, leaving an “orphan” BH with a negligibly small accretion disk containing less than 1% of the rest-mass of the NS. Similar behavior was observed in the BHNS configuration with mass ratio  $q = 5:1$ . Finally, in the tilted magnetic field case, the team did not find a coherent poloidal magnetic field component remaining after the BHNS merger; hence, the key ingredient for jet launching was absent. The dynamical ejecta produced in the nonspinning  $q = 5:1$  and the counter-rotating BHNS configurations were too small to be resolved ( $\sim 10^{-5} M_{\text{sun}}$ ) and it may indicate that high mass-ratio BHNS or counter-rotating BHNS configurations may not be accompanied by either kilonovae or short gamma-ray bursts because they unbind a negligible amount of mass and form negligibly small accretion disks onto the remnant BH. The researchers have concluded that these two configurations may be the progenitors of the BHNS candidate S190426c.

WHY BLUE WATERS

Blue Waters provides the required computational power to simulate these cosmic sources in a timely manner. By adding OpenMP support to our message-passing interface (MPI)-based code, scalability on multicore machines has improved greatly. With the Blue Waters high-performance interconnect and processors, the team’s hybrid OpenMP/MPI code exhibits greater scalability and performance than on any other supercomputer they have used. Recently, the researchers were able to build their code with the Blue Waters Intel compilers. This resulted in a significant boost of the code’s performance by about 30%, making Blue Waters unique for tackling the astrophysical problems the team wants to address.

PUBLICATIONS & DATA SETS

M. Ruiz, A. Tsokaros, V. Paschalidis, and S. L. Shapiro, “Effects of spin on magnetized binary neutron star mergers and jet launching,” *Phys. Rev. D*, vol. 99, no. 8, p. 084032, Apr. 2019.

L. Sun, M. Ruiz, and S. L. Shapiro, “Magnetic braking and damping of differential rotation in massive stars,” *Phys. Rev. D*, vol. 99, no. 6, p. 064057, Mar. 2019.

M. Ruiz, S. L. Shapiro, and A. Tsokaros, “Jet launching from binary black hole–neutron star mergers: Dependence on black hole spin, binary mass ratio, and magnetic field orientation,” *Phys. Rev. D*, vol. 98, no. 12, p. 123017, Dec. 2018.

L. Sun, M. Ruiz, and S. L. Shapiro, “Simulating the magnetorotational collapse of supermassive stars: Incorporating gas pressure perturbations and different rotation profiles,” *Phys. Rev. D*, vol. 98, no. 10, p. 103008, Nov. 2018.



FEEDING BLACK HOLES: TILT WITH A TWIST

**Allocation:** NSF PRAC/5,000 Knh  
**PI:** Alexander Tchekhovskoy<sup>1</sup>  
**Collaborators:** Matthew Liska<sup>3,5</sup>, Casper Hesp<sup>3</sup>, Adam Ingram<sup>7</sup>, Michiel van der Klis<sup>3</sup>, Sera Markoff<sup>3</sup>, Koushik Chatterjee<sup>3</sup>, Zack Andalman<sup>1,4,6</sup>, Nicholas Stone<sup>2</sup>, Eric Coughlin<sup>2</sup>

<sup>1</sup>Northwestern University  
<sup>2</sup>Columbia University  
<sup>3</sup>University of Amsterdam  
<sup>4</sup>Evanston Township High School  
<sup>5</sup>Harvard University  
<sup>6</sup>Yale University  
<sup>7</sup>Oxford University

EXECUTIVE SUMMARY

Accretion disks are typically tilted relative to the black hole rotational equator. Lack of symmetries makes numerical studies of such disks particularly challenging, especially on the curved spacetime of a rapidly spinning black hole. The GPU partition of Blue Waters enabled the research team for the first time to simulate such disks in full general relativity. The team found that, contrary to standard expectations, tilted disks tear up into individually precessing subdisks. If torn disks are prevalent around black holes, this calls for a reconsideration of the black hole accretion theory.

RESEARCH CHALLENGE

Tilted accretion is common in astrophysical systems. In fact, researchers expect that nearly all black hole accretion disks are tilted at some level relative to the black hole rotational equator. This is because the gas that approaches the black hole from large distances has no idea which way the black hole is spinning. However, studies of such tilted accretion are extremely challenging, especially in the crucial regime of luminous, radiatively efficient accretion that powers bright quasars. Such accretion disks are razor-thin and difficult to resolve numerically, requiring high resolutions and adaptive grids to follow the body of the disk as it moves through the computational grid.

Figure 1: The inner part of a disk of half-thickness  $h/r = 0.03$  tilted by  $60^\circ$  tears off from the outer misaligned part of the disk and precesses independently. This is the first demonstration of the disk tearing in a GRMHD (general relativistic magnetohydrodynamic) numerical simulation.

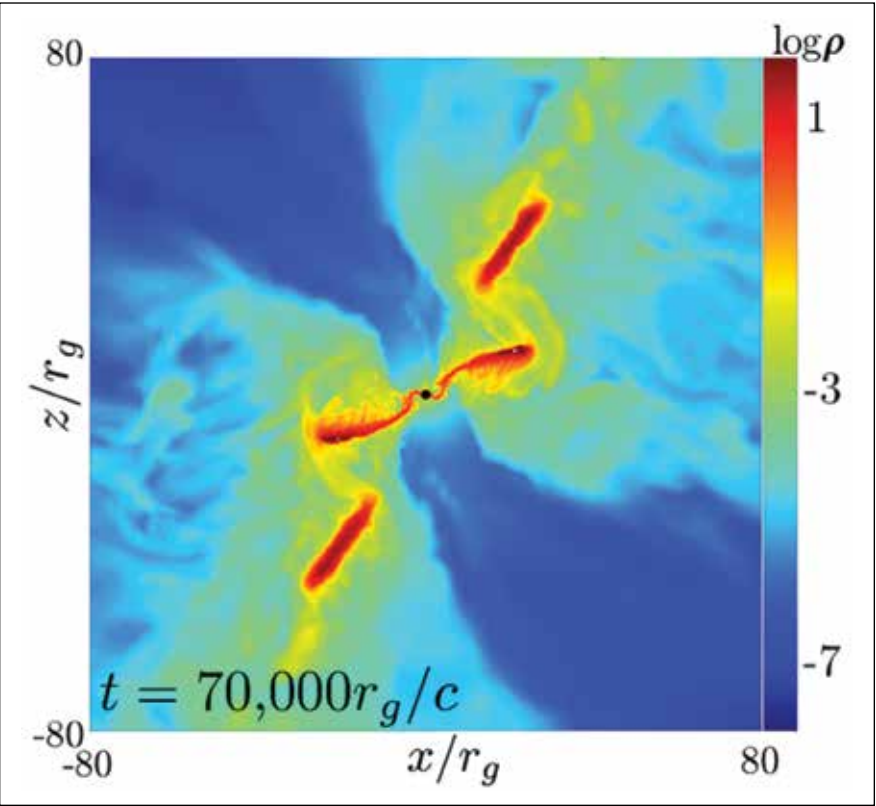


Figure 2: A tilted accretion disk around a spinning black hole tears up into several individual, independently precessing subdisks (blue). The jets (orange), produced by magnetized rotation of the central black hole, escape along the path of least resistance.

METHODS & CODES

Using the research team’s new code, H-AMR [2] (pronounced “hammer”), which includes adaptive mesh refinement, local adaptive timestepping, and runs efficiently on GPUs, the research team was able to overcome the above challenges. H-AMR performs 10 times faster on a GPU than on a similar vintage 16-core CPU. H-AMR is parallelized via MPI with domain decomposition and scales well to thousands of GPUs, achieving weak scaling efficiency of 85% on 4,096 GPUs on the Blue Waters supercomputer. The performance of the code allowed the team to study tilted discs at higher resolutions and over longer durations than was previously possible.

RESULTS & IMPACT

The simulations carried out on Blue Waters’ GPU partition revealed that the frame dragging of a spinning black hole can tear up tilted accretion disks into several individually precessing subdisks [1]. This can lead to complex, variable disk emission as the orientation of subdisks changes in time, and high-energy emission, as a result of jets running into subdisks.

WHY BLUE WATERS

Access to Blue Waters has been instrumental in obtaining these groundbreaking results, which require not only enormous amounts of computing power but also fast interconnect speeds to make use of hundreds of XK nodes. As in the past, Mark Van Moer helped enormously with 3D visualization.

PUBLICATIONS & DATA SETS

M. Liska *et al.*, “Disc tearing and bardeen–petterson alignment in grmhd simulations of highly tilted thin accretion discs, submitted to *Mon. Notices Royal Astron. Soc.*, 2019, arXiv:1904.08428.  
M. Liska, C. Hesp, A. Tchekhovskoy, A. Ingram, M. van der Klis, and S. Markoff, “Formation of precessing jets by tilted black hole discs in 3D general relativistic MHD simulations,” *Mon. Notices Royal Astron. Soc.*, vol. 474, no. 1, pp. L81–L85, Feb. 2018.



SCALING THE BATS-R-US MHD MODEL TO OVER 100,000 CORES WITH EFFICIENT HYBRID OPENMP AND MPI PARALLELIZATION

Allocation: GLCPC/360 Knh  
PI: Gabor Toth<sup>1</sup>  
Collaborator: Hongyang Zhou<sup>1</sup>

<sup>1</sup>University of Michigan

EXECUTIVE SUMMARY

This project aims to optimize and improve multilevel parallelization of the computationally most expensive components of the space weather modeling framework. One of the most important and computationally expensive models in the framework is the BATS-R-US magnetohydrodynamic (MHD) code. With pure MPI parallelization it is limited to about 32,000 cores owing to memory constraints. The research team has designed and imple-

mented an efficient hybrid MPI + OpenMP parallelization. The main idea is to assign grid blocks consisting of hundreds of grid cells to each OpenMP thread. This is much easier to implement than a cell-by-cell multithreading approach, and it is also more efficient. The new version of the code can scale to over 100,000 cores on Blue Waters while maintaining high efficiency.

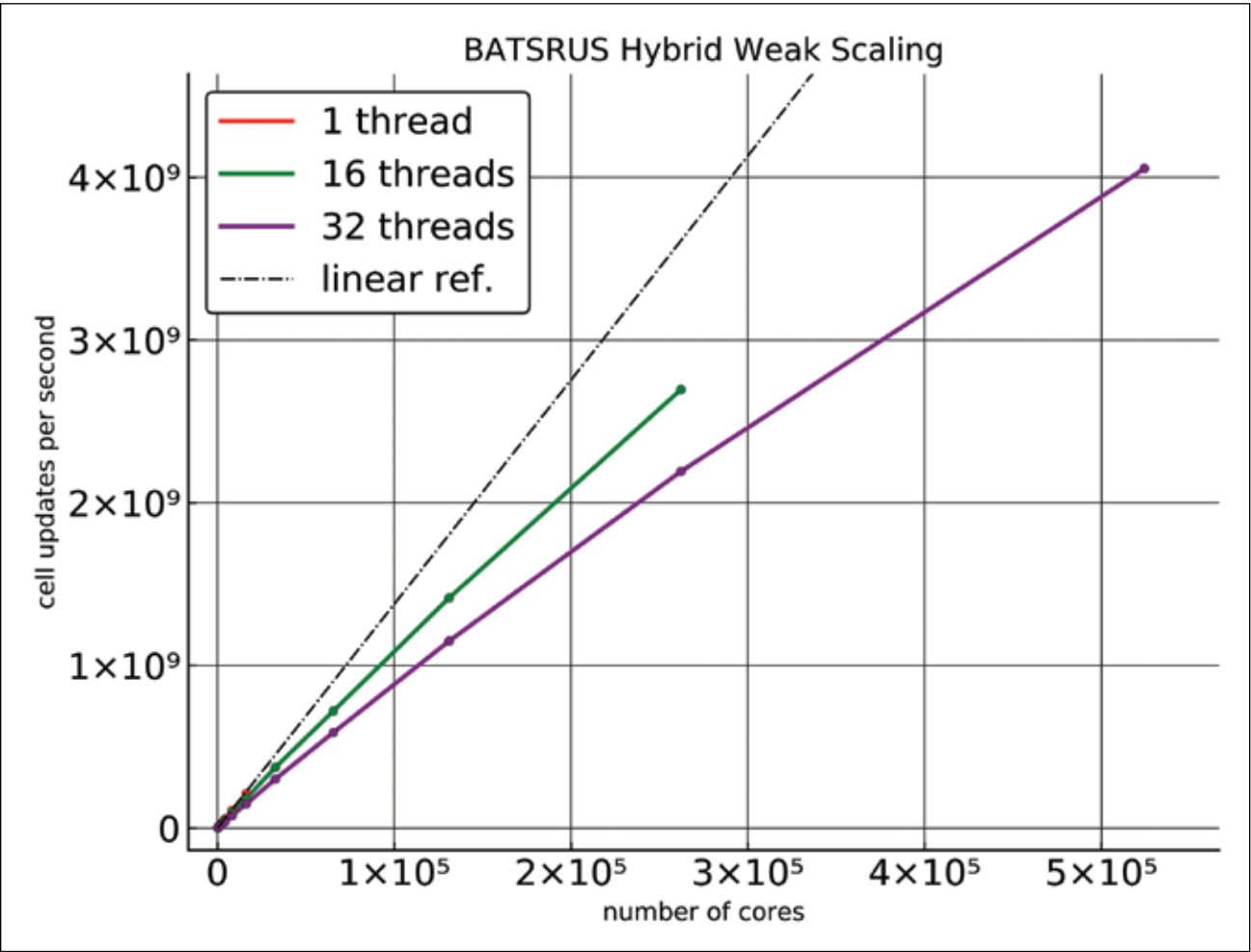


Figure 1: Weak scaling of BATS-R-US on Blue Waters. A 3D MHD problem is solved on a grid consisting of 131,072 grid cells per core using 8 x 8 x 8-cell blocks. There are 256 to 8,192 grid blocks per MPI process depending on the number of OpenMP threads. The dashed line indicates ideal linear scaling.

RESEARCH CHALLENGE

While pure MPI parallelization can be near optimal in terms of speed, the total number of cores used can be limited by the memory usage. In our BATS-R-US MHD code [1], for example, most but not all of the data are distributed over the processors. In particular, the description of the adaptive tree of the grid blocks is stored on all the MPI processes so that finding neighbor blocks or finding the grid block covering some spatial location can be done efficiently. The size of this array grows with the problem size and eventually can become a bottleneck. This issue can be mitigated by using OpenMP parallelization because the amount of memory occupied by the repeated arrays on a given node will be reduced proportionally with the number of MPI processes per node. Correspondingly, one can scale the model to a larger number of cores. With dozens of cores per node on current architectures this difference is significant, by a factor of 15 to 30 or even more.

Adding OpenMP into a complex MPI parallel code is not simple. BATS-R-US consists of about 250,000 lines of Fortran 90+ code without counting comments and empty lines. The workload is distributed over a large fraction of the source code. The goal is to add OpenMP parallelization that obtains good performance with a reasonable development time investment.

The team's primary goal is to allow scaling BATS-R-US to hundreds of thousands of cores on new computers to solve large problems efficiently. Further, the researchers' work and experience should help in making similar projects run smoothly and efficiently.

METHODS & CODES

The researchers' multiphysics code BATS-R-US can solve various partial differential equations with a large variety of numerical schemes on a block adaptive grid. The grid blocks consist of a fixed number of grid cells, typically 4 x 4 x 4 to 8 x 8 x 8 cells, although their physical size may vary. The code loops over grid blocks and performs various computations such as calculating fluxes and sources for each grid cell and then updates the grid cell value, etc. There are two possible implementation strategies for OpenMP: (1) parallelize the loops over grid cells (fine-grained) or (2) parallelize loops over grid blocks (coarse-grained).

The team chose the coarse-grained option for several reasons. In the overall algorithm, the loops over the cells contain less work than the loops over the blocks. There are many more loops over cells than loops over blocks, so a fine-grained approach would require adding more OpenMP parallel sections. For many loops over the grid cells the work per cell may be insufficient to make the OpenMP parallelization efficient given the overhead of starting and closing the multithreaded section. A significant difficulty with the coarse-grained approach is finding and resolving race conditions. It took significant effort for the team to find the appropriate tool: Intel's Inspector turned out to be invaluable.

RESULTS & IMPACT

The team's strategy required relatively modest code changes: only 609 OpenMP directive lines were added to the 250,000-line source code. Most of the changes were declaring variables as thread private or moving module variables into subroutines when convenient. The team also learned to look out for variable initializations, which make local subroutine variables behave as shared by default.

Fig. 1 shows that by using the OpenMP + MPI parallelization, BATS-R-US can run on more than 500,000 cores. Running 32 threads per node requires communication between the two sockets; still, the performance is about 50% of the ideal scaling, which is quite reasonable. With 16 threads per node BATS-R-US can scale up to approximately 250,000 cores and obtain around 80% of the ideal performance. In comparison, the pure MPI parallelization can run only up to about 16,000 cores before running out of memory. The researchers obtained similar results when running the code with an implicit solver.

The team also learned that for some compilers, switching on the OpenMP library can severely impact code performance (a slowdown of up to a factor of three) even if only one thread is used per MPI process. It is important to make sure that the code is portable and performs well for a compiler that runs efficiently with the OpenMP library. The researchers also found that the pinning of the OpenMP threads (assigning them to the proper CPU cores with respect to the MPI processes) can be quite complicated and the proper settings vary from platform to platform and even from compiler to compiler. A simple C code reporting of which core is used by a certain thread and MPI process is invaluable to verify that the pinning works as expected.

WHY BLUE WATERS

Blue Waters provided a platform with the appropriate hardware, software, and computing environment to make good progress with this project. On most systems it is practically impossible to run on more than about 10,000 cores with reasonable turnaround times. On Blue Waters, the team could scale up to about 500,000 cores. The variety of compilers allowed for testing the efficiency of the hybrid parallelization comprehensively and identifying some of the not-so-well-known problems.

PUBLICATIONS & DATA SETS

H. Zhou and G. Toth, "Efficient OpenMP parallelization to a complex MPI parallel magnetohydrodynamics code," *J. Parallel Distributed Comput.*, vol. 139, pp. 65–74, May 2020.



# NUMERICAL STUDY ON THE FRAGMENTATION CONDITION IN A PRIMORDIAL ACCRETION DISK

**Allocation:** Illinois/1,000 Knh  
**PI:** Matthew Turk<sup>1</sup>  
**Co-PIs:** Wei-Ting Liao<sup>1</sup>, His-Yu Schive<sup>1</sup>

<sup>1</sup>University of Illinois at Urbana-Champaign

## EXECUTIVE SUMMARY

This project uses Blue Waters to study the accretion flow that surrounds the first luminous objects known as Population III stars. These early-stage protostars are believed to be surrounded by massive accretion disks where competition between gravity and turbulence leads to the formation of multiple stellar objects. The goals of the project are to understand the multiplicity of the first stellar systems as well as to find the limits on the final mass of individual Population III stars. The research team also aims to uncover the unique roles of both turbulence and gravity on the resulting variability of the accreted mass that governs the final mass of Population III stars.

## RESEARCH CHALLENGE

The simulation will provide a numerical model that furthers our understanding of the formation of Population III stars and their possible observable quantities.

## METHODS & CODES

This work uses GAMER-2 [1] in conjunction with GRACKLE [2]. The numerical scheme allows the team to explore the accretion flow instability while self-consistently evolving the primordial chemistry.

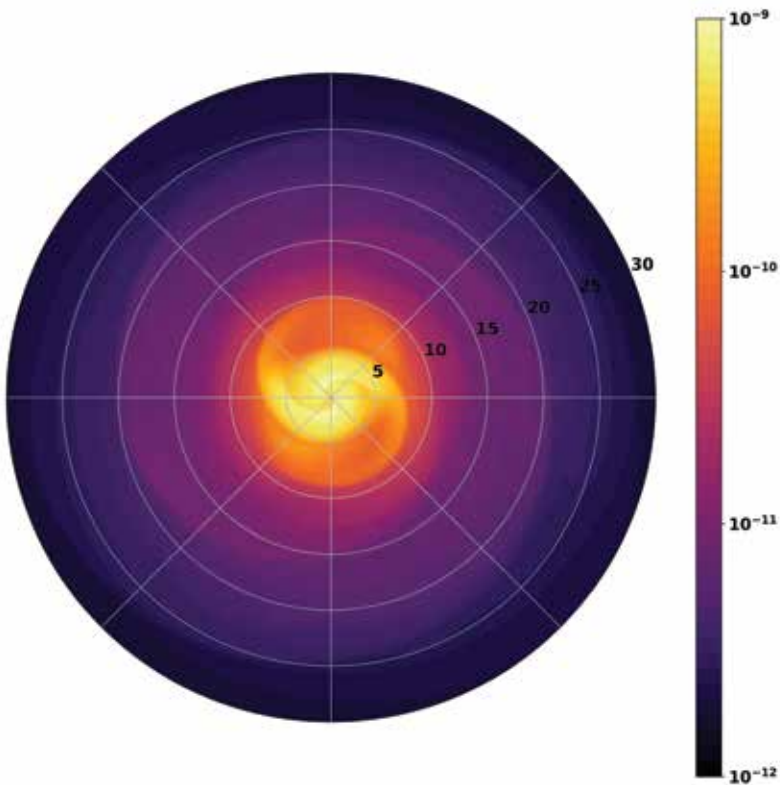
## RESULTS & IMPACT

The simulation suggests that the primordial accretion disks are highly turbulent and, thus, the formation of companion stars could be delayed. Using Blue Waters, the team aims to deepen the understanding of the conditions under which companion protostars may form and how it may impact the initial mass function of primordial stars.

## WHY BLUE WATERS

Blue Waters provides the environment that ensures high performance and scalability. GAMER-2 has been optimized and tested to run efficiently on Blue Waters.

Figure 1: A map of density within a primordial accretion disk at an early evolutionary stage. Two overdense spiral arms have built up, are gravitationally unstable, and are the most likely region for emergence of a companion protostellar object.



# MERGING BLACK HOLES AND NEUTRON STARS

**Allocation:** NSF PRAC/200 Knh  
**PI:** Saul A. Teukolsky<sup>1</sup>

<sup>1</sup>Cornell University

## EXECUTIVE SUMMARY

The primary purpose of the project is the numerical solution of Einstein's equations of general relativity. The goal is to track the coalescence and merger of binary black hole systems and to calculate the emitted gravitational waves (GWs). Another goal is to carry out a similar project for binary systems containing a black hole and one or two neutron stars. The work is aimed at providing theoretical predictions that can be compared with the signals measured by the National Science Foundation's LIGO (Laser Interferometer Gravitational-Wave Observatory) GW detector.

## RESEARCH CHALLENGE

The primary scientific objective of this project is to theoretically underpin and improve the ability of LIGO to extract the rich information that the observed GWs carry. Gravitational waves provide a new window on the universe that will enable scientists to test their understanding of fundamental physics as well as learn about the most extreme events in the cosmos.

## METHODS & CODES

Most of the computations are done with the SpEC code (spectral Einstein code) developed by the collaboration. The numerical methods that are used make it the fastest and most accurate code for treating black holes. The research team is also developing a new code, SpECTRE, that will include innovative methods to treat neutron star systems.

## RESULTS & IMPACT

The researchers have released a new version of their public catalog of gravitational waveforms for use by all scientists, largely through simulations on Blue Waters. The new version increased the size of the catalog from 174 waveforms to 2,018. These waveforms have already been employed to produce a very accurate waveform model that LIGO can use in its data analysis, and will meet most of LIGO's needs for the next two to three years.

## WHY BLUE WATERS

The research team's numerical code runs most efficiently on 50 to 70 processors for each waveform. Blue Waters' nodes are perfectly sized for using one or two nodes per waveform and exploring hundreds of different parameter values to develop the team's catalog.

## PUBLICATIONS & DATA SETS

M. Boyle *et al.*, "The SXS Collaboration catalog of binary black hole simulations," *Class. Quant. Grav.*, vol. 36, p. 195006, 2019.



# GEOSCIENCE

## CLIMATE

## ENVIRONMENT

## GEOLOGY

## GEOMAGNETISM

## SEISMOLOGY

## WEATHER

- 78

Image Processing to Build a Multitemporal Vegetation Elevation Ecosystem Model of the Great Lakes Basin
- 80

Petascale Processing of Satellite Earth Observations
- 82

Large-Scale Remote Monitoring of Invasive Species Dynamics Through a Petascale High-Performance Computing System
- 84

Deforestation of the Amazon Forest: Understanding Hydroclimate Impacts by Tracing the Water that Evaporates from the Forest
- 86

Forecasting Volcanic Unrest and Eruption Potential Using Statistical Data Assimilation
- 88

Monitoring Field-Scale Crop Water Use Using a Satellite Data-Driven Mechanistic Modeling Approach
- 90

Building an Objective Seasonal Forecasting System for U.S. Corn and Soybean Yields
- 92

Inflow and Outflow from Thunderstorms: Tracking Their Influence on Precipitation and Further Growth
- 94

Petascale Polar Topography Production
- 96

High-Resolution Numerical Simulations of Convection Initiation over the Sierras de Córdoba Mountains in Argentina
- 98

Simulations of Violently Tornadoic Supercells and Damaging Thunderstorms
- 100

Prediction of Geomagnetic Secular Variation with Large-Ensemble Geomagnetic Data Assimilation
- 102

Machine Learning for Error Quantification in Simulating the Climate Impacts of Atmospheric Aerosols
- 104

Simulating Hydroclimate Change in Southwest North America at 21,000 Years Ago
- 106

Implementation and Use of a Global Nonhydrostatic Model for Extended-Range Weather Prediction during the RELAMPAGO Field Campaign
- 108

Simulating Large California Earthquakes Before They Occur
- 110

Materials Simulations in Geophysics
- 112

Simulating Aerosol Impacts on Climate, One Particle at a Time: A Regional-Scale, Particle-Resolved Aerosol Model to Quantify and Reduce Uncertainties in Aerosol–Atmosphere Interactions
- 114

Evolving Air Quality Under the Changing Climate
- 116

Sensitivity of Arctic Sea Ice Thickness Distribution to Sea Ice Internal Dynamics in a Changing Climate



IMAGE PROCESSING TO BUILD A MULTITEMPORAL VEGETATION ELEVATION ECOSYSTEM MODEL OF THE GREAT LAKES BASIN

**Allocation:** GLCPC/495 Knh  
**PI:** Jennifer Corcoran<sup>1</sup>  
**Co-PIs:** Brian Huberty<sup>2</sup>, James Klassen<sup>3</sup>, Keith Pelletier<sup>4</sup>  
**Collaborators:** Paul Morin<sup>4</sup>, Joe Knight<sup>4</sup>, Laura Bourgeau–Chavez<sup>5</sup>

<sup>1</sup>University of Minnesota, Twin Cities  
<sup>2</sup>U.S. Fish & Wildlife Service  
<sup>3</sup>SharedGeo  
<sup>4</sup>University of Minnesota  
<sup>5</sup>Michigan Tech Research Institute

EXECUTIVE SUMMARY

Over a period of two-and-a-half years, the research team has acquired, processed, and created high-resolution, multitemporal vegetation elevation ecosystem models (MTVEEM) across the entire Great Lakes Basin (GLB) using stereo, submeter, optical satellite imagery obtained over the last 15 years. The team continues to process new imagery and has added processing steps to monitor changes in ecosystems and water extent over time. By providing this critical data set to stakeholders, both the understanding of the complex processes at work in the GLB and the ability to address adverse changes to the world’s largest freshwater ecosystem will vastly improve. The amount of data processed and analyzed in the GLB is well beyond the capacity of most academic, private, and government systems; the team could not have done this work without a leading-edge petascale resource such as Blue Waters.

RESEARCH CHALLENGE

Ecosystem management requires knowing the type, size, structure, and density of vegetation over time. These important features need to be mapped repeatedly. Stereo submeter, optical satellite imagery, and derived surface vegetation models can be used to better characterize these features and their changes over time, with the added dimension of height.

High-resolution vegetation surface canopy mapping over large geographic regions such as the Great Lakes Basin (GLB) has never been obtained from either aerial or satellite surveys. Additionally, the binational management by Canada and the United States of the GLB limits consistent, repeatable coverage by either country working independently. While a few scattered vegetation surface models exist from expensive airborne active laser sensors within the GLB, these data sets represent single time points and were not planned as continuous, basinwide acquisitions. Having high-resolution multitemporal information in three dimensions enables the kind of science that can address a multitude of critical questions that surround the ecosystems of the GLB.

These research questions remain: How are the ecosystems of the GLB changing and what can we as a society do about it? Con-

tinuous monitoring of surface elevation will detect both natural changes (such as flooding, forest blowdown, fire, insects, and disease outbreaks) and anthropogenic changes (such as harvest and land cover change). Further, MTVEEM will improve habitat and biological modeling. Finally, MTVEEM will be used binationally to better visualize canopy change in forested habitats and freshwater wetland resources within the Great Lakes Basin.

METHODS & CODES

Stereo-mode image acquisition through Digital Globe over the entire GLB began in 2016 and continues through 2019 as clouds and other higher-tasking priorities permit. The research team has processed over 150,000 stereo pairs, where each job consists of converting the input imagery into a standard format (GeoTIFF) and then calls the elevation extraction software (SETSM) [1]. The team expects 50,000 additional satellite image stereo pairs in the GLB each year going forward. Each stereo pair task is run on a single node, submitted in batches of two to 100 tasks per job on the low queue to maximize the scheduler throughput. Complete processing of one stereo pair to 2 meters takes an average of 12 node hours (charged as six node hours owing to using the low queue), totaling 300,000 node hours. Additionally, the researchers estimate it will take 150,000 to 200,000 node hours to process ortho-images and explore producing classifications based on the total number of image pairs they will have in hand.

RESULTS & IMPACT

As the data are processed, the resulting surface canopy models will be openly available in 2019 initially through the University of Minnesota. Other partners’ online distribution systems, such as the National Oceanic and Atmospheric Administration’s Digital Coast and the Great Lakes Observing System, will also be used. The final product, a seamless and registered surface elevation ecosystem model (MTVEEM) of the GLB will enable a large range of science activities at substantially higher resolution than currently available (current status shown in Fig. 1). These canopy maps and change detection products will provide positional accuracies of less than a couple meters with the added ground control points. The team is assessing semidecadal changes in prior-



Figure 1: Great Lakes Basin digital surface model (DSM) production status as of May 2019. The source of the stereo imagery used to produce these DSMs spans the archive, starting in about 2008 to early 2019 (2019 DigitalGlobe NextView License).

ity GLB areas where light detection and ranging-derived digital surface models from six to nine years ago are available. They are also assessing intraseasonal differences by processing and differencing surface models from satellite stereo image pairs within a single growing season. The preliminary results continue to show great promise for providing valuable data to myriad coastal and terrestrial ecosystem science researchers and decision-makers across the entire GLB [2,3].

The research team’s primary concentration will continue to be the GLB with extended temporal and geographical footprints as efficiencies and capacities improve. Though they will continue to make significant progress in the GLB, the team will also explore habitat types and terrain characteristics in depth that are not found within the pilot study areas. This will allow for process adjustments and quality control checks, enabling the team to scale the project to much larger geographic regions. In addition, the team has begun to process pilot areas at a one-half meter spatial resolution.

WHY BLUE WATERS

Stereo satellite imagery allows for the generation of highly accurate surface elevation models; the researchers have already tasked stereo-mode acquisition through Digital Globe over the entire GLB. Each stereo pair is about 1.25 GB; the total number of pairs processed to date is about 120,000 and soon will exceed 150,000. The amount of stereo imagery in a study area the size of the GLB and the computational burden to process each of these image pairs is well beyond those resources available from standard academic, private, and government systems. This is precisely why this project requires a leading-edge petascale resource such as Blue Waters.



PETASCALE PROCESSING OF SATELLITE EARTH OBSERVATIONS

**Allocation:** Blue Waters Professor/187 Knh  
**PI:** Larry Di Girolamo<sup>1</sup>  
**Collaborators:** Matias Carrasco Kind<sup>2</sup>, Gregory Daues<sup>2</sup>, Yulan Hong<sup>1</sup>, Ralph Kahn<sup>3</sup>, Lusheng Liang<sup>4</sup>, Donald Petravick<sup>2</sup>, John Towns<sup>2</sup>, Kent Yang<sup>5</sup>, Ping Yang<sup>6</sup>, Yizhe Zhan<sup>1</sup>, Guangyu Zhao<sup>1</sup>

<sup>1</sup>University of Illinois at Urbana–Champaign  
<sup>2</sup>National Center for Supercomputing Applications  
<sup>3</sup>NASA Goddard Space Flight Center  
<sup>4</sup>NASA Langley Research Center/Science Systems and Applications, Inc.  
<sup>5</sup>The HDF Group  
<sup>6</sup>Texas A&M University

EXECUTIVE SUMMARY

Through collaborative efforts among NASA, HDF Group, and NCSA, the research team fully validated and enhanced the open source tool that resamples Terra satellite data into a common grid with a python interface to make it applicable beyond Terra to any Earth observation data. The Terra satellite, launched in 1999, continues to collect earth science data using five instruments: the Moderate-resolution Imaging Spectroradiometer (MODIS), the Multi-angle Imaging SpectroRadiometer (MISR), the Advanced Spaceborne Thermal Emission and Reflection Radiometer, the Clouds and the Earth's Radiant Energy System (CERES), and the Measurements of Pollution in the Troposphere. Terra data not only serve the scientific community but also the governmental, commercial, and educational communities.

The researchers further used the Terra data set to: (1) characterize the ice crystal roughness of cirrus clouds for a better understanding of ice cloud optical properties; (2) correct biases in the MODIS cloud effective radius retrievals through MISR and MODIS fusion; (3) evaluate MISR and CERES Arctic cloud albedo retrievals, indicating excellent consistency for certain solar zenith angles; and, (4) examine decadal changes in the Earth's radiance fields, revealing little cloud change in the global mean beyond calibration (Fig. 1) but large changes in certain regions. The research group extended the study by processing observations from spaceborne active sensors, showing a strong correlation between cloud occurrence frequency and climate regimes (Fig. 2).

RESEARCH CHALLENGE

The need for Terra data fusion and for scientists to perform large-scale analytics with long records has never been greater [1], given the growing data volume (greater than 1 petabyte), the storage of different instrument data at different NASA centers, the different data file formats and projections, and inadequate cyber-infrastructure [2]. The team initiated the Terra Data Fusion Project to tackle two long-standing problems: (1) the need for efficient generation and delivery of Terra data fusion products; and (2) the need to facilitate the use of Terra data fusion products by the community by generating new products and knowledge through national computing facilities, and to disseminate them

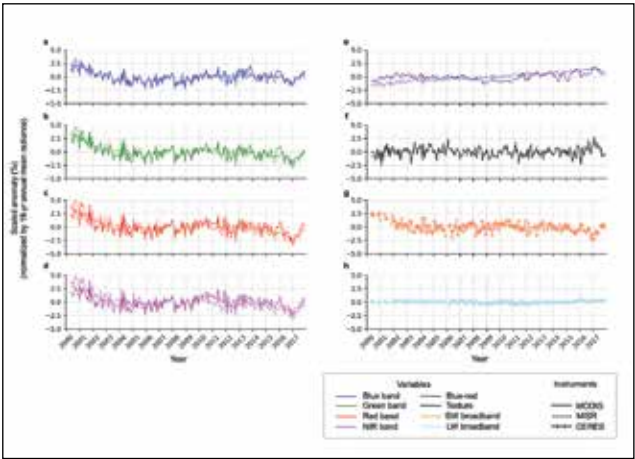


Figure 1: Time series of the deseasonalized and scaled anomalies of the monthly global mean spectral and broadband radiance and texture for MISR, MODIS, and CERES between 2000 and 2017.

through national data-sharing services. Solutions to these problems will significantly facilitate discovery and accelerate progress in earth science research.

METHODS & CODES

Key steps in the Terra Data Fusion Project include: (1) transferring the entire Terra record to Blue Waters from NASA centers (Level 1B radiance; greater than 1 petabyte); (2) building software optimized for whole-mission processing on Blue Waters to create basic fusion products; (3) optimizing data granularity and HDF user-defined settings that best support parallel I/O on Blue Waters; (4) archiving and distributing Terra fusion products through existing NASA services and commercial cloud services; and (5) driving development through processing and analysis of scientifically important use cases. Past Blue Waters reports have highlighted results on many of these key points. The research group has since extended their studies beyond Terra to include other satellite records and has developed an open source tool to resample or reproject the radiance fields into a common grid.

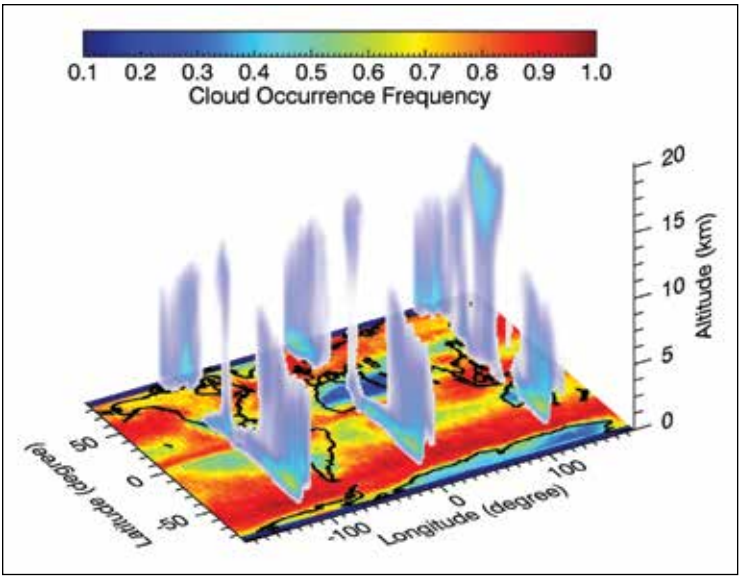


Figure 2: A 3D view of cloud occurrence frequency derived from the CloudSat/CALIPSO 2B-CLDCLASS-LIDAR product from 2007 to 2010.

RESULTS & IMPACT

The research team explored archiving and distributing Terra fusion products through existing NASA and commercial cloud services. They tested accessing and processing the basic fusion data on the Amazon cloud and found its processing time was slightly more than Blue Waters for the same settings but that costs to investigators' grants by using commercial clouds for a data-intensive project can be prohibitive. The team also fully validated and enhanced the open source tool that resamples or reprojects the radiance fields into a common grid with a Python interface to make the tool applicable beyond Terra to any Earth observation data.

Scientific investigation using the fusion data set was carried out primarily in four study areas. (1) The team characterized biases in the MODIS standard product of cloud drop effective radius (Re) and further examined the underlying causes. The results paint a radically different picture of the distributions of Re from what the original MODIS product provided. The bias-corrected cloud drop sizes are now in line with spot measurements from field campaigns and outputs from climate models. (2) The team retrieved ice cloud microphysical properties through MISR–MODIS fusion, showing regional dependence on ice crystal structure. Results have been published in an open access journal, *Remote Sensing*. (3) The group evaluated the MISR and CERES Arctic cloud albedo retrievals, showing excellent consistency between the instruments for solar zenith angles of less than 70°. These results were also published in *Remote Sensing*. (4) The team extended a global and regional radiance and texture trend study previously performed on Blue Waters [3] by using additional instruments: MODIS and CERES. The time series of the deseasonalized scaled anomalies of the monthly global mean spectral radiance and texture are shown in Fig. 1. The global blueing shown in the researchers' previous study [3] mainly results from a calibration drift in the MISR red band. There is no significant glob-

al trend in the Earth's reflected radiation field, but large regional changes were found.

The research team also analyzed cloud data from two spaceborne active sensors, CloudSat and CALIPSO, and found that large cloud occurrence frequencies are closely related to climate regimes at both horizontal and vertical scales (Fig. 2). A large frequency of high-level clouds at altitude ranges of 9 to 17 km in the tropics indicates the dominance of cirrus clouds, whereas shallow clouds with heights below 3 km are mostly observed in the subtropics. In the midlatitudes, clouds mostly occur below 5 km, although some clouds can reach to the tropopause (approximately 8 to 10 km).

WHY BLUE WATERS

The key advantages of using Blue Waters for access, usage, and distribution of Terra fusion products are that the Terra data and processing are local, whereas access and sharing are global. The research team demonstrated that having the Terra data local, with processing tuned to a massively parallel system with excellent sharing services on Blue Waters, provides an optimum framework for large-scale processing, analytics, and mining of the entire Terra record. In addition, the project staff provide expertise critically needed to optimize workflows.

PUBLICATIONS & DATA SETS

Y. Wang, S. Hioki, P. Yang, M. King, L. Di Girolamo, D. Fu, and B. A. Baum, "Inference of an optimal ice particle model through latitudinal analysis of MISR and MODIS data," *Remote Sens.*, vol. 10, no. 12, p. 1981, 2018, doi: 10.3390/rs10121981.  
Y. Zhan, L. Di Girolamo, R. Davies, and C. Moroney, "Instantaneous top-of-atmosphere albedo comparison between CERES and MISR over the Arctic," *Remote Sens.*, vol. 10, no. 12, p. 1882, 2018, doi: 10.3390/rs10121882.



DI

TN

BW

FS

BI

# LARGE-SCALE REMOTE MONITORING OF INVASIVE SPECIES DYNAMICS THROUGH A PETASCALE HIGH-PERFORMANCE COMPUTING SYSTEM

**Allocation:** Innovation and Exploration/110 Knh  
**PI:** Chunyuan Diao<sup>1</sup>

<sup>1</sup>University of Illinois at Urbana–Champaign

## EXECUTIVE SUMMARY

The rapid expansion of exotic saltcedar along riparian corridors has drastically altered landscape structures and ecosystem functions throughout the United States. Conducting the large-scale monitoring of spatio-temporal dynamics of this invasive species over the past 40 years is essentially critical to understanding its invasion mechanism. Previous studies indicated that the leaf senescence stage is the optimal time window to remotely monitor saltcedar distributions. However, the computational complexity in predicting the leaf senescence timing, along with the massive volume of satellite data in both spatial and temporal dimensions, makes large-scale invasive species monitoring prohibitive. Instead, the research team developed a parallel computational network model on Blue Waters that can accommodate the spatio-temporal variation in plant phenology to facilitate the large-scale monitoring of invasive species dynamics. Blue Waters provides unprecedented opportunities to achieve regionwide monitoring of saltcedar distribution to revolutionize scientific understanding of saltcedar invasion processes.

## RESEARCH CHALLENGE

Remote monitoring of invasive saltcedar dynamics is essential for conservation agencies to develop cost-effective control strategies [1,2]. Acquiring satellite imagery during the leaf senescence stage of saltcedar is necessarily crucial to facilitate the large-scale repetitive mapping of this invasive species [3–5]. However, owing to climate variability and anthropogenic forcing, the timing of saltcedar leaf senescence varies over space and time [6]. Given that the leaf senescence stage of saltcedar lasts only for three to four weeks, it is challenging to pinpoint the appropriate satellite imagery that can accommodate this spatial and interannual variation over wide geographical regions [7]. The dearth of high-performance computational systems that can tackle this phenological issue makes the large-scale saltcedar mapping prohibitive and hinders the understanding of its invasion mechanism.

## METHODS & CODES

To accommodate the spatio-temporal variation in plant phenology, the research team developed a complex network-based phenological model using satellite time series. The network-based phenological model constructs an undirected network for each pixel based on its spectral reflectances along the phenological

trajectory. Specifically, the spectral reflectance of the pixel obtained on each day is represented as a node. Those nodes that share similar spectral similarity are connected by edges. The network model groups the spectral reflectances along the trajectory into three clusters: a “pretransition cluster,” a “transition cluster,” and a “posttransition cluster” (Fig. 1). Further, the team developed several network measures to accommodate the spatio-temporal variation in invasive species phenology.

To overcome the computational limits, the research group proposed an innovative pixel-based remote sensing parallel computational system for large-scale saltcedar monitoring. The parallel system adopts hybrid computation models, including a core-level computation model and a node-level data distribution model. In the core-level computation model, a computing node stores the information of a number of pixels and processes them using OpenMP. The node-level data distribution model includes a two-level data decomposition strategy. The first level uses a

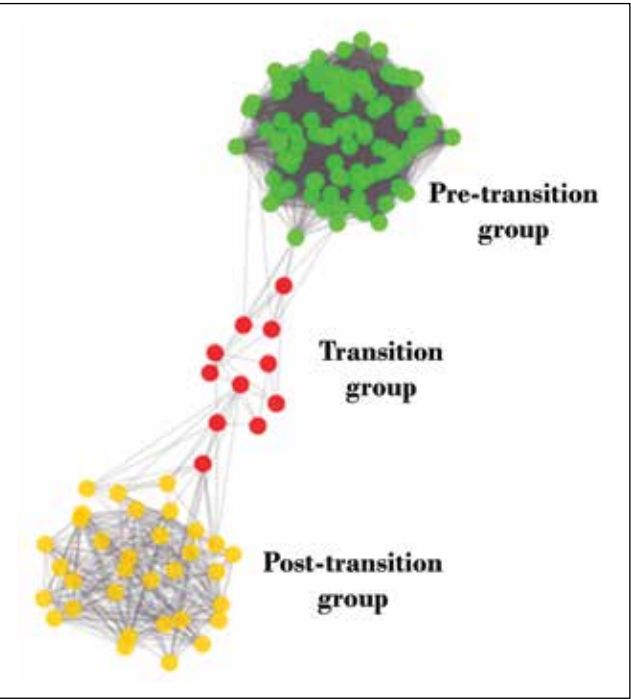


Figure 1: Complex network of phenological progress of invasive species with three clusters.

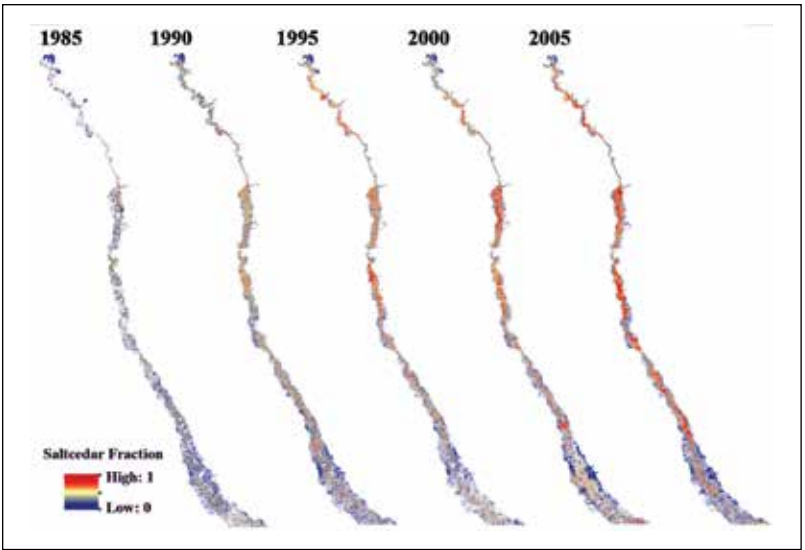


Figure 2: The spatial distributions of invasive saltcedar from 1985 to 2005 along the Forgotten River Reach.

space decomposition. All computing nodes are assigned to different node groups such that each group is responsible for all imagery tiles covering the same spatial extent. At the second-level decomposition, both space and time decomposition are applied within each group to improve the I/O performance. Using this strategy, the huge number of data and I/O operations are evenly distributed among all computing nodes. This greatly reduces the memory requirement and time for reading data files. Furthermore, this data distribution model is highly scalable and can be used efficiently on Blue Waters.

## RESULTS & IMPACT

The network-based phenological model presents a new representation of the complex phenological process of vegetation. This representation characterizes the vegetation’s phenological status through tracking the continuing changes of spectral signatures of vegetation along the temporal trajectory. The network results revealed a gradual phenological shift from north to south along the latitudinal gradient, following the Hopkins bioclimatic law. It suggested that temperature decreased with increasing latitude and would likely be an important force driving the phenological variations in vegetation over wide geographic regions. Along the latitudinal gradient, the interannual variation in the estimated average timing of leaf senescence was generally less than 10 days.

The team has successfully implemented this parallel computational system on Blue Waters and analyzed the massive amount of time series of remotely sensed data to monitor the phenological dynamics, as well as mapping the large-scale invasive species distributions. The team found that invasive saltcedar has expanded dramatically over the riparian corridors of the Southwestern United States (Fig. 2). It has amalgamated many smaller patches into fewer, larger patches; formed a greater number of high-density patches; and outcompeted native vegetation, especially along the river. The landscapes along the riparian corridors, as a consequence, have been transformed enormously. The

monitoring results show great promise in helping conservation agencies to conduct systemic restoration of the affected riparian ecosystems. The parallel system is transferable to many other types of competing vegetation species aside from saltcedar, enhancing the system’s utility and value across a wide range of agencies concerned with invasions.

## WHY BLUE WATERS

The computational challenges of the project lie in two aspects: (1) the high computational costs of processing a single pixel, and (2) the large number of pixels to process at the large-scale. The network algorithm devised for detecting saltcedar leaf senescence outperforms most existing algorithms but has a high computational cost. A huge number of pixels and associated time series data are required for the large-scale monitoring. This number adds to the total amount of computation time. Further, it requires a large amount of storage space. Blue Waters is the most ideal resource to tackle these computational challenges. In addition, Blue Waters’ file system provides high-speed data access. This could dramatically reduce the time of I/O operations of the massive imagery involved in this project, and hence reduce the total computation time.

## PUBLICATIONS & DATA SETS

- C. Diao, “Complex network-based time series remote sensing model in monitoring the fall foliage transition date for peak coloration,” *Remote Sensing Environ.*, vol. 229, pp. 179–192, 2019.
- C. Diao and L. Wang, “Landsat time series-based multiyear spectral angle clustering (MSAC) model to monitor the inter-annual leaf senescence of exotic saltcedar,” *Remote Sensing Environ.*, vol. 209, pp. 581–593, 2018.
- C. Diao, “Innovative pheno-network model in estimating crop phenological stages with satellite time series,” *ISPRS J. Photogrammetry Remote Sensing*, vol. 153, pp. 96–109, 2019.



DEFORESTATION OF THE AMAZON FOREST: UNDERSTANDING HYDROCLIMATE IMPACTS BY TRACING THE WATER THAT EVAPORATES FROM THE FOREST

**Allocation:** Innovation and Exploration/136.8 Knh  
**PI:** Francina Dominguez<sup>1</sup>  
**Collaborators:** Zhao Yang<sup>1</sup>, Jorge Eiras-Barca<sup>1</sup>

<sup>1</sup>University of Illinois at Urbana–Champaign

EXECUTIVE SUMMARY

The Amazon Forest has undergone significant deforestation in the past decades with natural forests being replaced by agriculture and pasturelands. In this work, the research team evaluated the impacts of continued deforestation on the hydroclimate of the South American continent. They initiated 10-year climatological simulations created with the Weather Research and Forecast (WRF) model with added water vapor tracers (WRF–WVT). The water vapor tracers track the water that originates from the Amazonian forest and follow it in space and time as the moisture is advected and contributes to precipitation. In the water-limited southern Amazon, the researchers found that the effects of deforestation are locally strong with distinct changes in the deforested areas. Although area-averaged precipitation decreases, the team also found regions with increased precipitation owing to changes in the atmospheric circulation from changes in land cover. This reveals both positive and negative feedback from deforestation and complex land–atmosphere interactions in the hydroclimate of South America.

RESEARCH CHALLENGE

Deforestation of the Amazon Forest in the past decades has seen natural forests being replaced by agriculture and pasturelands (see mapbiomas.org for the evolution of land cover since 1985). A very high rate of deforestation in the year 2005 (19,000 km<sup>2</sup>) was followed by a sharp decline in 2012; unfortunately, the rate has since increased again to almost 8,000 km<sup>2</sup> in 2018 [2]. The Amazon is the largest tropical rainforest on Earth. Up to 50% of Amazonian precipitation originates as evapotranspiration from the forest [7,9]. Furthermore, downwind regions in the La Plata and Orinoco Basins are dependent on Amazonian moisture for their precipitation [4,8]. The critical question is how continued deforestation of the Amazon Forest will affect the hydroclimate of the South American continent.

METHODS & CODES

The researchers incorporated Water Vapor Tracers (WVT) into the Weather Research and Forecasting model (WRF) [3,5]. This allows users to trace moisture that originates as evapotranspiration from a tagged Amazon region. (Evapotranspiration that originates from the Amazon is numerically “tagged” as it under-

goes the same physical processes as total moisture such as advection, convection, phase change, and the like.) In addition, WRF–WVT saves tracer moisture-related variables. In particular, tracer moisture advection and horizontal diffusion follow the exact same transport equations in WRF for scalar variables, including water vapor and all micrometeors. Tracer moisture changes phase and is converted to precipitation in the same proportion as full moisture in all cases. All tracer moisture species are generated or converted from one to another and to precipitation, mimicking their full-moisture counterparts.

The research team performed two 10-year continuous simulations for the period 2004–2013 over the domain shown in Fig.



Figure 1: Domain for WRF–WVT simulations, including South America and surrounding bodies of water. The green outline is the Amazon Basin and the red outline is the La Plata River Basin. The regions shown in dark red are deforested to cropland in the year 2050 based on the projections in [6].

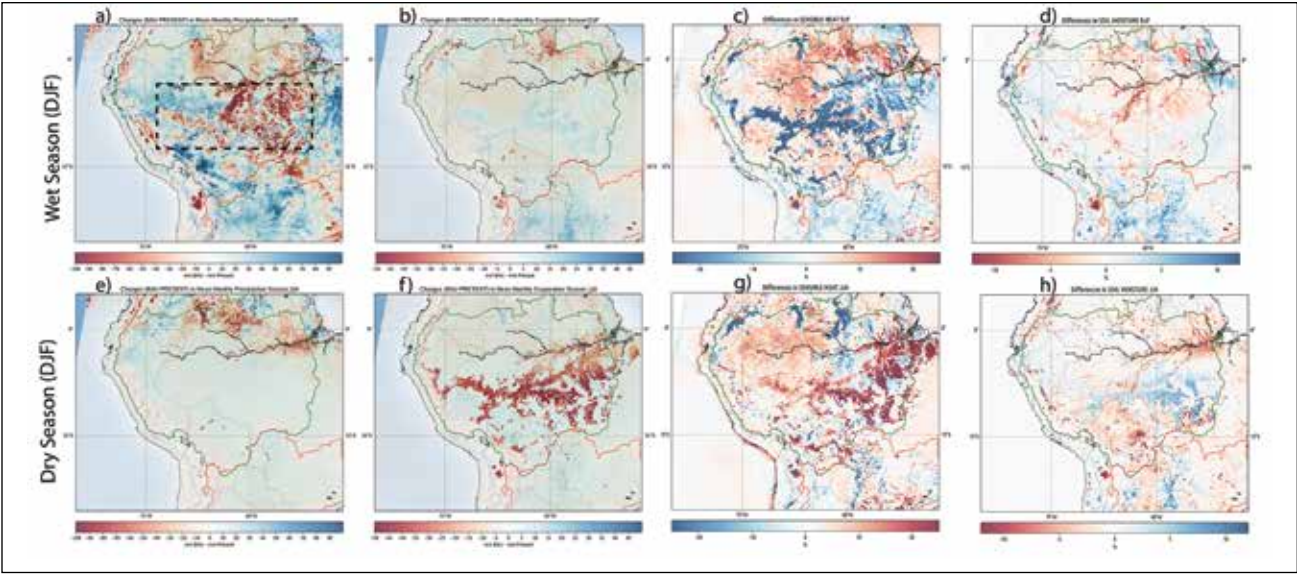


Figure 2: Ten-year average wet season–DJF (top) and dry season–JJA (bottom): (a, e) changes in precipitation; (b, f) evaporation; (c, g) sensible heat; and (d, h) soil moisture. The changes are between the business-as-usual (BAU) deforestation scenario and the control simulation.

1 using lateral boundary conditions from the ERA–Interim re-analysis data set [1]. The simulation has a horizontal resolution of 20 km and 40 levels in the vertical direction (see Yang and Dominguez, listed under “Publications & Data Sets” for details). The first simulation used observed land use for the period 2004–2013, while the second set of simulations has the projected land cover change as simulated by SimAmazonia–1 (Fig. 1, red areas), which provides land-use projections up to the year 2050 that incorporate road development in the Amazon basin [6].

RESULTS & IMPACT

The research team has focused its results on the Southern Amazon Forest (dashed black box in Fig. 2a). This region is characterized by transitional forests where precipitation is less than in the northern tropical forests, and there is a marked dry season [10]. The team has analyzed the results as changes between the deforested scenario and the control simulation for different hydroclimate variables including precipitation, evapotranspiration (ET), sensible heat, and soil moisture. In the discussion that follows, ET and latent heat are used interchangeably, and the results are for both the wet and dry seasons as they show contrasting behavior.

*December–January–February (DJF).* During the wet season (DJF), the deforested regions show very distinct decreases in precipitation (Fig. 2a, red areas). These changes in precipitation result in a decrease in cloudiness and increase in net radiation (short-wave radiation) that increases ET (Fig. 2b, blue areas). Soil moisture is depleted owing to increased ET/latent heat (Fig. 2d, red areas). As more energy goes into latent heat, this leads to a significant decrease in sensible heat (Fig. 2c, blue areas) and surface temperature (not shown). These results indicate that during this season of water abundance, the deforested region behaves as a radiation-limited ecosystem. Interestingly, a marked increase in

moisture flux (not shown) results, rather counterintuitively, in more precipitation downwind of the deforestation area.

*June–July–August (JJA).* During the dry season (JJA), the deforested region shows strong decreases in ET (Fig. 2f), indicating that the vegetation is not able to access the soil moisture. This leads to increased temperature throughout the deforested region and increased sensible heat flux (Fig. 2g). Because there is less ET, the soil moisture increases (Fig. 2h). Changes in precipitation are negligible (Fig. 2e).

In conclusion, when focusing on the transitional forests of the southern Amazon, the research team has found contrasting behavior during the wet and dry seasons. During the wet season, the main control on ET is through radiation as there is still an ample supply of moisture. During the dry season, the region is water-limited and ET decreases significantly. Precipitation during the wet season is clearly decreased over the deforested areas, but precipitation downwind of the deforested areas is actually increased owing to increased moisture transport. During the dry season, there is very little change in precipitation. These results show the complex behavior of land–atmosphere interactions when changes in land cover occur.

WHY BLUE WATERS

Blue Waters was critical to performing the simulations owing to their very high computational expense. The research team would not have been able to perform them on their local cluster.

PUBLICATIONS & DATA SETS

Z. Yang and F. Dominguez, “Investigating land surface effect on the moisture transport over South America with a moisture tagging model,” *J. Clim.*, vol. 32, no. 19, pp. 6627–6644, Oct. 2019.



DI

FORECASTING VOLCANIC UNREST AND ERUPTION POTENTIAL USING STATISTICAL DATA ASSIMILATION

Allocation: Illinois/85 Knh  
PI: Patricia M. Gregg<sup>1</sup>  
Collaborator: Seid Koric<sup>1,2</sup>  
Team Members: Yan Zhan<sup>1</sup>, Jack Albright<sup>1</sup>, Haley Cabaniss<sup>1</sup>

<sup>1</sup>University of Illinois at Urbana–Champaign  
<sup>2</sup>National Center for Supercomputing Applications

EXECUTIVE SUMMARY

A primary motivation for investigating volcanic systems is developing the ability to predict eruptions and mitigate disaster for vulnerable populations. Over the past three years, the Gregg Lab has been developing approaches for forecasting the evolution of volcanic systems in collaboration with the National Center for Supercomputing Applications (NCSA). The research team has implemented a high-performance computing (HPC) workflow using COMSOL Multiphysics finite-element software that links multiphysics model outputs with geophysical monitoring data for volcano forecasting. This project focuses on conducting large system-scale numerical experiments to investigate eruption potential and triggering mechanisms for three volcano targets utilizing the unique computational configuration of Blue Waters. In addition to the scientific outcomes of this effort, the experiments mark the largest distributed implementations of COMSOL Multiphysics. This achievement is of great practical importance for finite-element applications and provides benchmarking for future efforts in other fields, such as engineering, in addition to earth sciences.

RESEARCH CHALLENGE

Currently, 500 million people worldwide live on or near active volcanoes. The team’s current efforts on Blue Waters are focused on developing strategies for rapid assimilation of volcano monitoring data sets into evolving geodynamic models to provide near real-time forecasts and assessment of volcanic unrest. To that end, the group is adapting data assimilation strategies developed in other fields to combine observations of volcanoes experiencing unrest with thermomechanical finite-element models to calculate volcano evolution. By combining multiphysics finite-element models with volcano monitoring data the team is able to track the stress evolution of a magmatic system and provide probability forecasts of volcanic stability during periods of unrest. Utilizing ensemble-based methods, hundreds to thousands of models are run simultaneously to track the evolution of volcanic systems. This method allows the team to evaluate stress accumulation and failure in the lead-up to volcanic eruption and to test for potential eruption-triggering mechanisms to provide a framework for early warning probability forecasts for monitoring agencies. The ultimate goal of this work is to provide a trans-

ferrable data assimilation approach that can be utilized by volcano monitors worldwide.

Three volcano targets were chosen for this study owing to their excellent, real-time geophysical monitoring data sets and past eruption records: (1) Sierra Negra Volcano, Galápagos, Ecuador; (2) Laguna del Maule Volcano, Chile; and (3) Axial Volcano, Juan de Fuca Ridge—a submarine volcano located off the coast of Oregon, U.S.A. Each volcano application provides unique computational and data challenges to allow the team to evaluate potential roadblocks in transferability of the data assimilation approach.

METHODS & CODES

The research team has developed an HPC workflow using Python to efficiently distribute COMSOL Multiphysics models across Blue Waters’ compute nodes and compile model outputs for Ensemble Kalman Filter (EnKF) data assimilation at each timestep. The main computational task is evaluating hundreds of large multiphysics, mechanical finite-element models at each timestep and compiling the model data to provide a probabilistic forecast of volcanic unrest.

RESULTS & IMPACT

The team has applied its Blue Waters allocation to investigate three active volcanic systems. In 2018, the researchers had the opportunity to track the unrest of Sierra Negra Volcano, Galápa-

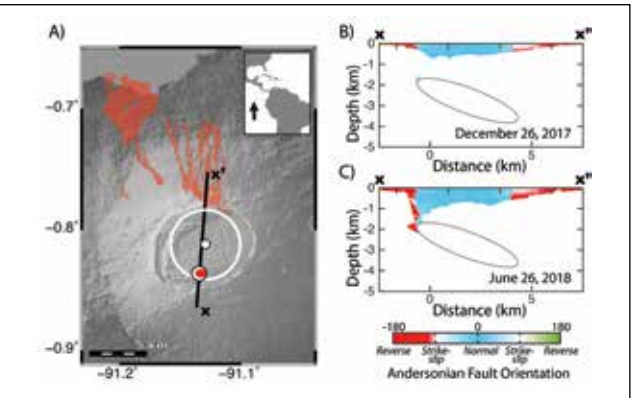


Figure 1: Sierra Negra’s June 26, 2018 Mw 5.4 earthquake (beachball) and eruption (red shaded region). White circle indicates the center of the hindcast source, with its extent outlined by the white line. B) Mean model calculated Mohr-Coulomb failure for December 26, 2017. C) Mean model failure for June 26, 2018.

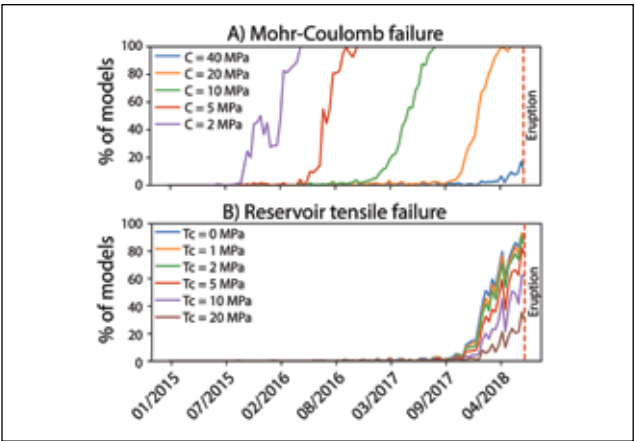


Figure 2: EnKF hindcast of the 2018 eruption of Sierra Negra. A) The percentage of finite element models in the EnKF ensemble experiencing Mohr-Coulomb failure as a function of cohesion, C. B) The percentage of models exhibiting Mode-I, tensile failure along the magma reservoir boundary as a function of tensile strength, Tc.

gos, Ecuador, during an extensive period of unrest (Fig. 1). The EnKF approach provided a forecast that indicated an eruption window that coincided with the timing and location of the 2018 eruption almost six months in advance of the event. Subsequent hindcasting on Blue Waters of the eruption further indicates that it was likely triggered by an earthquake that struck the southern portion of the Sierra Negra caldera 10 hours preceding the eruption (Fig. 2). The successful forecast of the 2018 eruption of Sierra Negra is a critical proof of concept for the volcano data assimilation HPC approach.

In addition to Sierra Negra, Ph.D. candidate and NASA Fellow Yan Zhan has been investigating the unrest and eruption potential of the Laguna del Maule volcano in Chile. Laguna del Maule has experienced significant inflation over the past decade indicating the potential for a large-scale eruptive event. However, current HPC forecasts reveal that the magmatic system appears to be on a stable trajectory currently. Future efforts will continue to monitor real-time data from Laguna del Maule to provide updated model forecasts as the system evolves.

The volcano EnKF approach was applied to a submarine volcanic system. Axial Volcano is an active underwater caldera volcano located off the Pacific Northwest coast of Oregon and Washington. Three eruptions have been recorded by bottom pressure recorders (1998, 2011, 2015) and ocean bottom seismometers (2011, 2015). Axial Caldera is currently monitored in real time via the Ocean Observatory Initiative (OOI). New hindcasts, conducted on Blue Waters as part of NSF-funded Ph.D. candidate Haley Cabaniss’s dissertation, indicate that while the current OOI data stream provides unique insights into this underwater system, the sparse spacing of the bottom pressure recorders does not provide adequate data coverage for the EnKF approach. Cabaniss is utilizing these outcomes to develop new data targets for the OOI and to investigate the incorporation of new data streams.

Finally, the large-scale distributed implementation of COMSOL Multiphysics was tested on Blue Waters. The research group

determined a key bottleneck when distributing COMSOL across hundreds of nodes. COMSOL begins to slow down exponentially owing to the inherent overhead of distributing and then compiling the model data. Unfortunately, this issue limits the scalability of the current version of COMSOL Multiphysics 5.4. Ph.D. candidate and NSF Fellow John Albright has been working with undergraduate NCSA SPIN (Students Pushing Innovation) Fellow Keon Park to test open source finite-element modeling approaches to replace COMSOL to provide better scalability. This effort will continue in the coming years.

WHY BLUE WATERS

EnKF is an ensemble-based sequential data assimilation method that requires calculating hundreds to thousands of finite-element models at each timestep. As such, the EnKF implementation utilizes high-throughput computing with hundreds to thousands of independent concurrent runs that require many nodes. While the EnKF analysis step has been optimized to run very swiftly, the computational expense of running and storing hundred to thousands of finite-element models for each timestep in the EnKF analysis is cost prohibitive for even the large XSEDE clusters. Blue Waters is uniquely positioned to handle the computational needs of this project and has allowed the research group to make rapid progress and to think ambitiously without being hampered by computational limitations.

PUBLICATIONS & DATA SETS

P. M. Gregg *et al.*, “Forecasting the June 26, 2018, eruption of Sierra Negra volcano, Galápagos, Ecuador,” in preparation, 2019.

Y. Zhan, P. M. Gregg, H. Le Mével, C. A. Miller, and C. Cardona, “Integrating reservoir dynamics, crustal stress, and geophysical observations of the Laguna del Maule magmatic system by FEM models and data assimilation,” in preparation, 2019.

H. E. Cabaniss, P. M. Gregg, S. L. Nooner, and W. W. Chadwick, “Triggering the eruption of Axial Seamount, Juan de Fuca Ridge,” in preparation, 2019.

P. M. Gregg, Y. Zhan, H. Le Mével, J. A. Albright, and H. E. Cabaniss, “Linking thermomechanical models with geodetic observations to assess magma reservoir evolution and stability,” presented at the 27th IUGG General Assembly, Montréal, Québec, Canada, Jul. 8–18, 2019.

P. M. Gregg *et al.*, “Forecasting the June 26, 2018, eruption of Sierra Negra, Galápagos,” presented at the 27th IUGG General Assembly, Montréal, Québec, Canada, Jul. 8–18, 2019.

P. M. Gregg, Y. Zhan, F. Amelung, D. Geist, and P. Mothes, “Model forecasts of the 26 June 2018 Eruption of Sierra Negra Volcano, Galápagos,” presented at the AGU Fall 2018 Meeting; Washington, D.C., U.S.A., Dec. 10–14, 2018.

P. M. Gregg, Y. Zhan, J. A. Albright, Z. Lu, J. Freymueller, and F. Amelung, “Imaging volcano deformation sources through geodetic data assimilation,” presented at the 2018 UNAVCO Science Workshop, Broomfield, CO, U.S.A., Mar. 27–29, 2018.



MONITORING FIELD-SCALE CROP WATER USE USING A SATELLITE DATA-DRIVEN MECHANISTIC MODELING APPROACH

Allocation: Illinois/573.143 Knh  
PI: Kaiyu Guan<sup>1</sup>  
Co-PIs: Jian Peng<sup>1</sup>, Chongya Jiang<sup>1</sup>, Bin Peng<sup>1</sup>, Sib0 Wang<sup>1</sup>

<sup>1</sup>University of Illinois at Urbana–Champaign

EXECUTIVE SUMMARY

High-spatiotemporal-resolution evapotranspiration (ET) products with reliable accuracy have many potential applications in agriculture. The research team developed BESS–STAIR, a new framework to estimate high-spatiotemporal-resolution ET that can be used for field-level precision water resources management. BESS–STAIR couples a satellite-driven water–energy–carbon-coupled biophysical model (BESS) with a generic and fully automated fusion algorithm (STAIR) to generate gap-free 30-m-resolution daily ET estimations. Comprehensive evaluation of BESS–STAIR ET estimations revealed: (1) reliable performance over 12 flux tower sites across the U.S. Corn Belt; and (2) reasonable spatial patterns, seasonal cycles, and interannual dynamics. The proposed BESS–STAIR framework has demonstrated its ability to provide significant advancements with regard to daily field-level estimations of ET at regional and decadal scales. When scaled up, which is in process, BESS–STAIR ET products could be very useful for precision water resources management and other precision agriculture applications for the U.S. Corn Belt and elsewhere.

RESEARCH CHALLENGE

With increasing crop water demands and drought threats, mapping and monitoring of cropland evapotranspiration (ET) at high spatial and temporal resolutions becomes increasingly critical for water management and sustainability. However, estimating ET from satellite data for precise water resources management is still challenging owing to limitations in both existing ET models and satellite input data. Specifically, the process of ET is complex and difficult to model, and existing satellite remote sensing data cannot fulfill high resolutions in both space and time.

METHODS & CODES

To address the above two issues, the research team developed a new high-spatiotemporal-resolution ET mapping framework, BESS–STAIR, that integrates the satellite-driven water–carbon–energy-coupled biophysical model BESS (Breathing Earth System Simulator) [1,2] with a generic and fully automated fusion algorithm STAIR (SaTellite dAta IntegRation) [3]. In this framework, STAIR provides daily 30-m multispectral surface reflectance by fusing Landsat and MODIS satellite data to derive fine-resolution leaf area index and visible/near-infrared albedo, all of which,

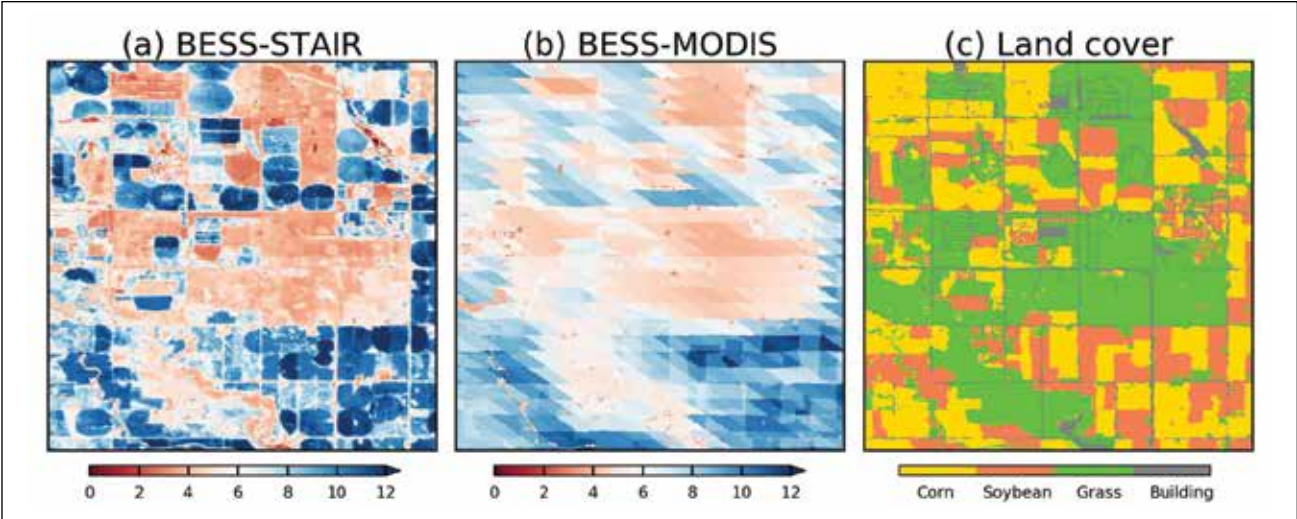


Figure 2: Example daily evapotranspiration (MJ m-2 d-1) derived from: (a) BESS–STAIR (30 m) and (b) BESS–MODIS (500 m) at Mead, Nebraska.

along with coarse-resolution meteorological and CO<sub>2</sub> data, are used to drive BESS to estimate gap-free 30-m resolution daily ET. The team applied BESS–STAIR from 2000 through 2017 in six areas across the U.S. Corn Belt and validated BESS–STAIR ET estimations using flux tower measurements over 12 sites (85 site-years). Results showed that BESS–STAIR daily ET achieved an overall R<sup>2</sup> = 0.74, with RMSE = 0.96 mm d-1 and relative error = 28.6%. In addition, BESS–STAIR ET estimations captured the spatial patterns, seasonal cycles, and interannual dynamics of ET well when benchmarked with the flux measurements. The high performance of the BESS–STAIR framework primarily resulted from: (1) the implementation of coupled constraints on water, carbon, and energy in BESS; (2) high-quality daily 30-m data from STAIR fusion algorithm; and (3) BESS’s applicability under all-sky conditions. BESS–STAIR has great potential to be a reliable tool for water resources management and precision agriculture applications for the U.S. Corn Belt, and even for other agricultural regions worldwide, given the global coverage of its input data.

RESULTS & IMPACT

This project represents the first attempt to couple a satellite-driven, physical process-based model with data fusion techniques to provide daily 30-m-resolution ET estimations at regional and decadal scales. To the research team’s knowledge, there is no explicit bottom-up, biophysically-based high-spatiotemporal resolution, long time series, and regional level methodological framework to estimate daily ET at a 30-m resolution. Other existing methodologies such as ALEXI–STARFM and SEBS–ESTARFM are LST (land surface temperature)-based models, and thus suffer from gap-filling and spatial-sharpening issues. Moreover, validation results from the project show that the BESS–STAIR framework can actually outperform those existing LST-based ET estimations.

The research team currently is upscaling the BESS–STAIR framework to the entire state of Illinois and other parts of the U.S. Midwest. They can foresee a number of potential applications of the high-resolution (30m) ET estimation either for small-scale precision agriculture or large-scale water resource planning and management.

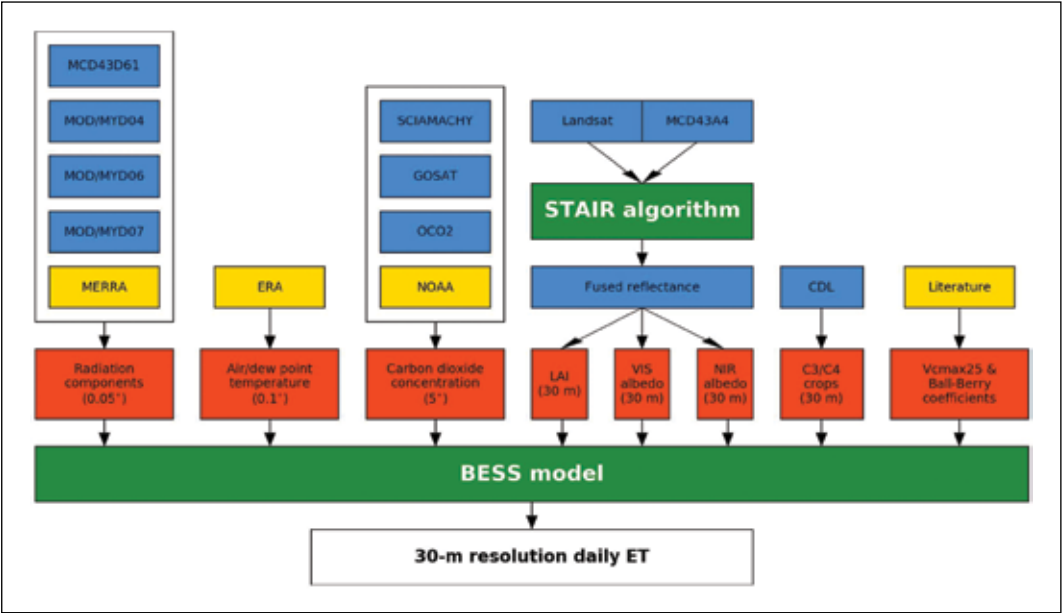
WHY BLUE WATERS

Blue Waters is essential for this research since other resources, such as those available from XSEDE, are not suitable for the project, considering the petabyte-level storage demand, data availability, intensive I/O, and computation demands.

PUBLICATIONS & DATA SETS

C. Jiang, K. Guan, M. Pan, Y. Ryu, B. Peng, and S. Wang, “BESS-STAIR: A new framework to estimate daily, 30-meter, and all-weather evapotranspiration using multi-source satellite data for the U.S. Corn Belt,” *Hydrol. Earth Syst. Sci. Discuss.*, in press, 2019, doi: 10.5194/hess-2019-376.  
Y. Luo, K. Guan, and J. Peng, “STAIR: A generic and fully-automated method to fuse multiple sources of optical satellite data to generate a high-resolution, daily and cloud-/gap-free surface reflectance product,” *Rem. Sensing Environ.*, vol. 214, pp. 87–99, 2018.

Figure 1: The BESS–STAIR framework for field-scale (30-m) evapotranspiration mapping.





BUILDING AN OBJECTIVE SEASONAL FORECASTING SYSTEM FOR U.S. CORN AND SOYBEAN YIELDS

Allocation: Blue Waters Professor/250 Knh  
PI: Kaiyu Guan<sup>1</sup>  
Co-PIs: Jian Peng<sup>1</sup>, Chongya Jiang<sup>1</sup>, Bin Peng<sup>1</sup>, Sib0 Wang<sup>1</sup>

<sup>1</sup>University of Illinois at Urbana–Champaign

EXECUTIVE SUMMARY

Despite significant advances in both seasonal climate prediction and satellite remote sensing, the produced data have not been fully used in crop yield forecasting at the regional scale, compared to survey-based approaches. In this project, the research team built a seasonal forecasting system for U.S. corn (maize) and soybean yield by bridging the most advanced seasonal climate prediction products from the National Centers for Environmental Prediction (NCEP) and satellite remote sensing within a statistical crop modeling framework. The researchers then evaluated the benefits of using seasonal climate prediction and satellite remote sensing data in forecasting U.S. corn and soybean yield at both national and county levels. They found they could not achieve a better forecasting performance than the official survey-based forecast from the United States Department of Agriculture (USDA) until they used both climate and remote sensing observations in their model. Compared with using historical climate information for the unknown future in each growing season, using climate predictions from NCEP gave better forecasting performance once the team corrected the bias in the seasonal climate prediction products. Using the climate–remote sensing combined model and bias-corrected climate prediction from NCEP, the researchers achieved a better forecasting performance than the USDA forecast. The team’s system will be useful for stakeholders in the agriculture industry and commodity markets.

RESEARCH CHALLENGE

More frequent extreme events and ongoing climate change puts food production at a higher risk [1–4]. Seasonal forecast of agricultural production thus becomes increasingly more important for early warning of food security issues, supply chain planning for the agriculture industry, and market prediction [5–7]. Although many countries and regions around the world have their own operational crop yield forecasting systems with varied modeling and data configurations, using combined seasonal climate prediction and remote-sensing data for crop yield prediction at large scales is still rare.

METHODS & CODES

The research team built a seasonal forecasting system for U.S. corn and soybean yield by bridging the North American multi-model ensemble (NMME) seasonal climate prediction products [8] and satellite remote sensing within a statistical crop modeling framework. The seasonal climate prediction products were first bias-corrected and spatially downscaled to 4 km using the percentile mapping algorithm and then aggregated to the county level. Multiple remote-sensing products were used in this system, including the MODIS NDVI, EVI, NIRv, LST, and OCO-2/TROPOMI SIF products. Both traditional statistical regression and machine learning algorithms were used to build the yield prediction models. The model performances were evaluated using

the out-of-sample validation method. The crop yield forecasting performances were benchmarked with that from USDA World Agricultural Supply and Demand Estimates (WASDE) reports.

RESULTS & IMPACT

The research group demonstrated that incorporating satellite information significantly improved the yield forecasting performance, compared with other climate-only models using monthly air temperature, precipitation, and vapor pressure deficit. The bias-corrected climate prediction from NMME showed better yield forecasting performance than the historical climate ensemble. Among the remote-sensing features, using NIRv and EVI can generally have better yield prediction performances. The multi-model ensemble approach can lead to the best yield prediction performance. Finally, the team’s yield forecast outperformed the WASDE reports released by the USDA.

WHY BLUE WATERS

Blue Waters was essential for this research because of the need to run ensemble models (more than 200 ensembles for prediction at each time) at a regular updating frequency. The computational and storage requirements can only be fulfilled by using Blue Waters.

PUBLICATIONS & DATA SETS

B. Peng, K. Guan, M. Pan, and Y. Li, “Benefits of seasonal climate prediction and satellite data for forecasting U.S. maize yield,” *Geophys. Res. Lett.*, vol. 45, pp. 9662–9671, 2018.  
Y. Li *et al.*, “Toward building a transparent statistical model for improving crop yield prediction: Modeling rainfed corn in the U.S.,” *Field Crops Res.*, vol. 234, pp. 55–65, 2019.

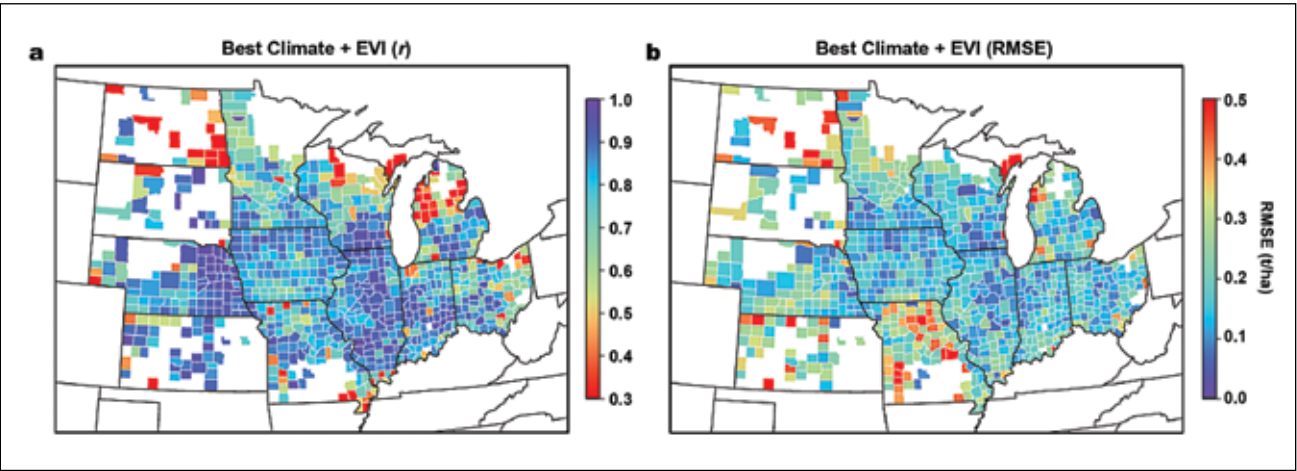


Figure 1: (a) Temporal correlation between actual yield and yield predicted by the “Best Climate + EVI” model and (b) its median prediction root mean square error (RMSE) from 2003 to 2016 at each county. Counties with yield observations of fewer than five years over the evaluation period are not shown.



# INFLOW AND OUTFLOW FROM THUNDERSTORMS: TRACKING THEIR INFLUENCE ON PRECIPITATION AND FURTHER GROWTH

**Allocation:** Blue Waters Professor/240 Knh  
**PI:** Sonia Lasher-Trapp<sup>1</sup>

<sup>1</sup>University of Illinois at Urbana-Champaign

## EXECUTIVE SUMMARY

“Entrainment” describes how clouds and storms bring dry air from outside the cloud inward by their own turbulent motions. Its effects can limit storm development, longevity, and precipitation. The understanding of entrainment has been limited by inadequate model resolution in past studies. Therefore, the research team is running high-resolution 3D storm simulations on Blue Waters to quantify the interactions among entrainment, precipitation, and the generation of new storms.

The latest results suggest that: (1) The entrainment occurring in developing storms shows a dependency on grid resolution, at least down to scales of 100 meters (m); (2) mature, rotating storms do in fact continue to entrain dry air but intermittently, by features that ascend within the cores of the storms; and (3) the maintenance of thunderstorm outflows is most dependent upon the amount of large ice particles called graupel, but their speed, depth, and strength are more correlated with the amount of evaporating rain beneath the storm.

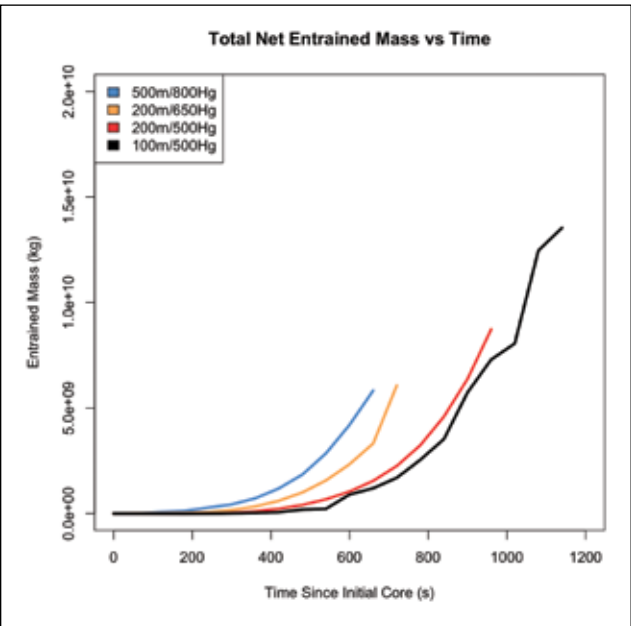


Figure 1: Time series of net mass of air entrained into the core of a developing supercell thunderstorm, for simulations conducted with increasingly smaller grid spacing (resolution) or forcing strengths (labeled). Results have essentially converged between 200-m and 100-m grid spacing, indicating the resolution of the most important turbulent scales.

## RESEARCH CHALLENGE

Deep convective clouds produce the majority of the earth’s precipitation, and yet it is difficult to predict if developing cumulus clouds will attain the depth and longevity required to create heavy rainfall that can produce flooding in some cases. In addition, strong winds (outflows) from some storms are capable of forcing new storms, leading to thunderstorm outbreaks that can affect society over large geographical areas. “Entrainment” is the term for the process by which the turbulent motions within clouds bring dry air from outside the cloud inward. In time, entrainment can limit storm development and its precipitation, but is not always effective in doing so. Long-standing problems in meteorological models have been to understand why they tend to predict rain formation too early, and/or in excessive amounts, and why they often miss predicting outbreaks of storms. The team’s approach here is to create a new understanding of how entrainment operates in thunderstorms, to investigate if deficiencies in past models are related to poor representation of entrainment or simply insufficient resolution, and to investigate the types of precipitation that produce the strongest storm outflows that may generate new storms. This problem affects a broad range of atmospheric science problems ranging from short-term weather forecasts from numerical weather prediction models to climate forecasts from regional and global climate models.

## METHODS & CODES

This work uses the National Center for Atmospheric Research’s CM1 model [1] to simulate convective clouds and storms at high resolution by employing its Message Passing Interface capabilities on the many nodes available on Blue Waters. The method makes use of the NSSL microphysical scheme [2] within CM1 to model the details of precipitation formation. Simulations are conducted in both idealized and realistic environments. The team evaluates entrainment with its own code [3] that calculates mass fluxes into the core of the storm as it evolves and relates it to the amount of precipitation the storm produces. The team also uses its own code that calculates the most important melting/evaporating precipitation types that can strengthen storm outflows and potentially produce outbreaks of thunderstorms.

## RESULTS & IMPACT

The latest results are providing new quantitative information that can assist atmospheric scientists researching ways to represent thunderstorm entrainment, precipitation, and outflows in

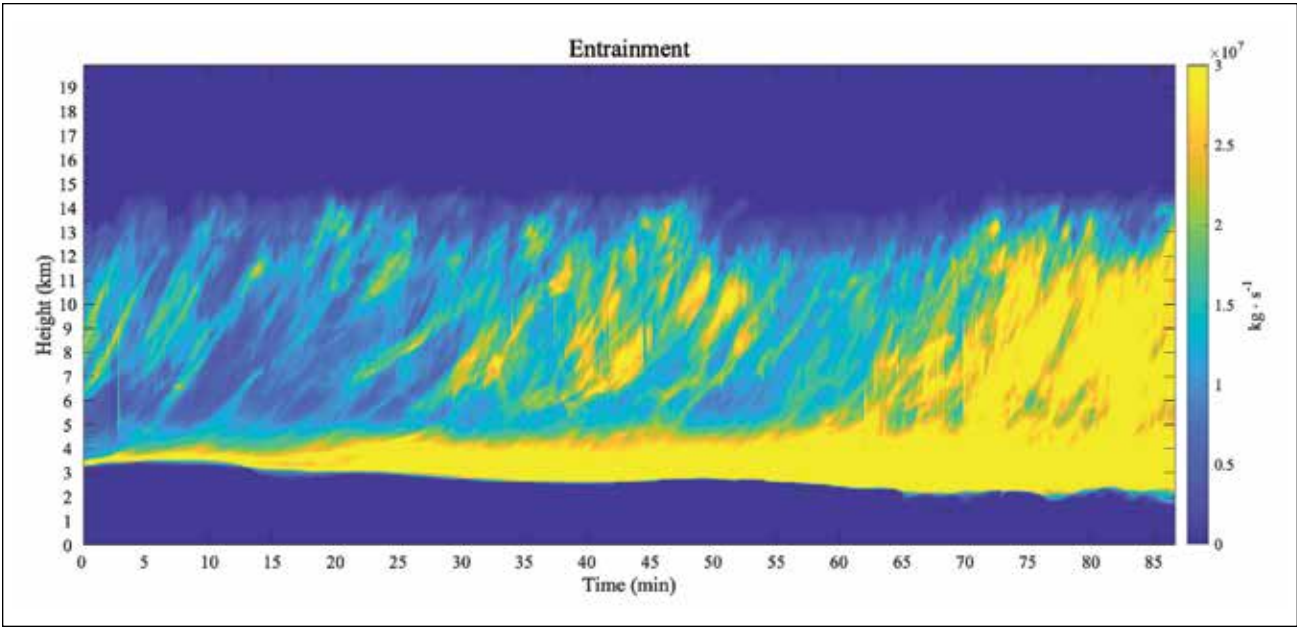


Figure 2: Time series of the rate air is entrained into a mature, rotating thunderstorm as a function of height above the ground. Maxima ascend from the lower part of the storm upward in time, which is an aspect that may contribute to pulslike behavior in its precipitation.

larger-scale weather and climate prediction models. They will also be of use to weather forecasters. The results suggest:

- The entrainment quantified in developing thunderstorms growing in an environment where the winds increase strongly with height is indeed resolution-dependent (Fig. 1), and the team’s calculations suggest that grid spacing of at least 100 m is required for accuracy.
- Although the common thinking is that mature, rotating thunderstorms do not entrain much dry air, the results are showing that they do indeed continue to entrain air (*i.e.* past the developing stage), but that the main entrainment events are transient in time and space (Fig. 2). This is a unique finding, and the team continues to explore the origin of these signatures and to quantify their overall importance to the total amount of air entrained into the storm as well as their possible effect on precipitation.
- Calculations of the contribution of different kinds of precipitation to the storm outflows have revealed that the amount of large ice particles called graupel that fall from the storm appears to be most important in maintaining them, but that characteristics of the outflow (speed, depth, strength) are more influenced by the evaporation of rain beneath the storm bases. This suggests that certain aspects of precipitation production in storms may be critical to forcing thunderstorm outbreaks.

## WHY BLUE WATERS

This Blue Waters allocation has been essential for achieving the high resolution required within a given simulation to properly represent the smaller cloud motions that are important for entrainment and precipitation development over the larger spatial and temporal domains required for thunderstorms and their outflows. As a result of its huge number of nodes, its high speed, large memory, and its large storage capability for high-resolution model output and analysis, Blue Waters enables detailed calculations to be conducted over millions of grid points. The hardware needed to run these kinds of simulations quickly supersedes the limits of most computers.

## PUBLICATIONS & DATA SETS

H. Mallinson and S. Lasher-Trapp, “An investigation of hydrometeor latent cooling upon convective cold pool formation, sustainment, and properties,” *Mon. Wea. Rev.*, vol. 147, no. 9, pp. 3205–3222, Sept. 2019.

D. H. Moser and S. Lasher-Trapp, “Cloud spacing effects upon entrainment and rainfall along a convective line,” *J. Appl. Meteor. Clim.*, vol. 57, no. 8, pp. 1865–1882, Aug. 2018.



DI

TN

FS

BI

PETASCALE POLAR TOPOGRAPHY PRODUCTION

**Allocation:** NSF PRAC/6,900 Knh  
**PI:** Paul Morin<sup>1</sup>  
**Co-PIs:** Claire Porter<sup>2</sup>, Ian Howat<sup>3</sup>, Myoung-Jong Noh<sup>3</sup>  
**Collaborators:** Charles Nguyen<sup>2</sup>, Cathleen Williamson<sup>4</sup>, Charles Crittenden<sup>4</sup>

<sup>1</sup>University of Minnesota  
<sup>2</sup>Polar Geospatial Center  
<sup>3</sup>The Ohio State University  
<sup>4</sup>National Geospatial-Intelligence Agency

EXECUTIVE SUMMARY

Surface topography is among the most fundamental earth science data sets, essential to a wide range of research activities that include hazard assessment and mitigation, hydrologic modeling, solid-earth dynamics, and many others. Change in surface topography provides critical information about surface processes such as plate tectonics, erosion and subsidence, glacier mass balance, and mass balance of woody vegetation. Over most areas of the globe, however, the spatial resolution (tens of meters) and accuracy (greater than 10 meters) of openly available topographic data sets are insufficient for many of these research activities. Even fewer areas have repeated high-precision elevation measurements for observing change.

RESEARCH CHALLENGE

There is a lack of high-resolution, consistent, high-quality elevation data available for the Earth. In 2000, the Shuttle Radar Topography Mission acquired Synthetic Aperture Radar data for the Earth that was processed into an elevation model with a 30-meter horizontal resolution. While invaluable for temperate regions, this mission was not repeated and had a limited resolution. The National Geospatial-Intelligence Agency, DigitalGlobe, and the



Figure 1: Map of the topography of the current production status of Australia. Each strip of elevation data is acquired in a single pass by Worldview-1, -2, or -3 and is processed on Blue Waters.

Polar Geospatial Center have built up a near-seamless archive of polar submeter stereo imagery that consists of hundreds of thousands of stereo pair strips from the Worldview-1, -2, and -3 satellites. Using photogrammetric algorithms, the research team is able to construct digital elevation models (DEMs) from the stereo pairs, allowing for mapping of surface features at the two-meter scale. These data are already being used by the research community to support activities that include transportation, national defense, land management, sustainable development, and scientific studies. Further, most geographic areas include repeated coverage with a frequency of months or even days. Such temporal coverage can be used for change detection with applications ranging from studies of land use to resource management and environmental change.

METHODS & CODES

The research team has spent seven years developing an efficient algorithm for constructing photogrammetric DEMs from satellite imagery with the objective of creating a fully automated pipeline capable of handling large amounts of data. The Surface Extraction from TIN-based Search-space Minimization (SETSM) algorithm, initially designed to extract elevation data over ice sheets, has been refined and optimized to handle stereoscopic imagery over any land cover [1,2]. Unlike other DEM extraction algorithms, SETSM's structure eliminates the need for an existing (*i.e.*, "seed") DEM for *a priori* constraints or any data-specific, user-defined search parameters, making it a truly automated algorithm. After an initial preprocessing step that corrects the source imagery for sensor-specific detector alignment artifacts, SETSM takes the two source images and derives increasingly detailed elevation models using its pyramid-based approach. The DEM extraction workflow runs on a single node for efficiency, and several thousands of these single-node tasks are bundled together using the Swift workflow management package to effectively submit jobs in 100- to 1,000-node batches.

RESULTS & IMPACT

Thus far, the research team has processed the stereo imagery into topography for the poles for all of 2018 as well as the first half of 2019. In addition, they have produced DEMs for 126,000,000 km<sup>2</sup> of the land surface of the Earth from 60°N to 60°S with ad-



Figure 2: Shaded relief image of the region around Paris, France. Note the meandering of the Seine River and the layout of the streets. If you look closely, the roundabout surrounding the Arc de Triomphe can be seen.

ditional focus on North America, Australia, Western/Northern Africa, and Central Asia. On average, the poles are covered eight to ten times, with some areas having several hundred unique DEMs for a given location. These data are also processed into continuous mosaics for over 99% of the 20,000,000 km<sup>2</sup> Arctic and the 15,000,000 km<sup>2</sup> Antarctic. The polar data have been released to the science community and the public through ArcticDEM.org, and Esri has developed web services and an interactive viewer. These data are now being used by scientists, national institutions, and regional and local governments for a broad range of scientific, civil engineering, and mapping applications. Three hundred scientific publications have used ArcticDEM and 27 of the Reference Elevation Model of Antarctica data sets since the data were first released.

WHY BLUE WATERS

Currently, no other academic computer has comparable capacity or large allocations. Over 1.5 billion core hours were required to process the archive of stereoscopic imagery over the life of the Arctic, Antarctic, and nonpolar elevation projects. Additionally, the Blue Waters' staff were invaluable in adapting the system to handle the single-node, high-throughput ArcticDEM workflow. With their help, the researchers adopted a strategy that enabled ArcticDEM jobs to use primarily backfill nodes on a low-priority basis, increasing overall system utilization and minimizing impact on other projects.

PUBLICATIONS & DATA SETS

I. M. Howat *et al.*, "The reference elevation model of Antarctica," *Cryosphere*, vol. 13, no. 2, pp. 665–674, 2019.

I. M. Howat *et al.*, "The reference elevation model of Antarctica," Harvard Dataverse, V1, 2018, doi: 10.7910/DVN/SAIK8B.

M.-J. Noh and I. M. Howat, "Automatic relative RPC image model bias compensation through hierarchical image matching for improving DEM quality," *ISPRS J. Photogramm. Remote Sens.*, vol. 136, pp. 120–133, 2018.

M.-J. Noh and I. M. Howat, "The Surface Extraction from TIN based Search-space Minimization (SETSM) algorithm," *ISPRS J. Photogramm. Remote Sens.*, vol. 129, pp. 55–76, 2017.

I. M. Howat, "Terra cognita: The silicon age of Earth exploration," presented at the American Geophysical Union Fall Meeting, Washington, D.C., U.S.A., Dec. 10–14, 2018, Abstract No. G54A–03.

P. Morin *et al.*, "ArcticDEM; a publically available, high resolution elevation model of the Arctic," in *European Geosciences Union Gen. Assembly Conf. Abstracts*, vol. 18, Vienna, Austria, Apr. 17–22, 2016.

C. Dai and I. M. Howat, "Measuring lava flows with ArcticDEM: Application to the 2012–2013 eruption of Tolbachik, Kamchatka," *Geophys. Res. Lett.*, vol. 44, no. 24, 2017.

C. Dai *et al.*, "Estimating river surface elevation from ArcticDEM," *Geophys. Res. Lett.*, vol. 45, no. 7, pp. 3107–3114, 2018.

C. Porter *et al.*, "ArcticDEM," Harvard Dataverse, 1, 2018, doi: 10.7910/DVN/OHHUKH.

Polar Geospatial Center, "ArcticDEM Mosaic, release 5," Jun. 2017. [Online]. Available at [www.pgc.umn.edu/data/arcticdem](http://www.pgc.umn.edu/data/arcticdem).



# HIGH-RESOLUTION NUMERICAL SIMULATIONS OF CONVECTION INITIATION OVER THE SIERRAS DE CÓRDOBA MOUNTAINS IN ARGENTINA

**Allocation:** Illinois/500 Knh  
**PI:** Stephen W. Nesbitt<sup>1</sup>  
**Collaborator:** Itinderjot Singh<sup>1</sup>

<sup>1</sup>University of Illinois at Urbana–Champaign

## EXECUTIVE SUMMARY

Previous studies using satellite and radar data show that some of the most intense convective storms initiate over and near the Sierras de Córdoba (SDC) mountain range in Argentina. However, gaps in data and knowledge exist as to how these storms initiate, which contribute to poor predictability of these storms. This study aims to understand the mesoscale processes involved in the initiation of deep moist convection in the region. The study uses quasi-idealized numerical modeling of a real case along with data obtained during the RELAMPAGO–CACTI field experiments in Argentina.

## RESEARCH CHALLENGE

This study aims to understand how the background flow responds to the daytime heated terrain of the SDC, and how low-level convergence is generated. Improving the understanding of the mesoscale processes involved in the initiation of these storms has the potential to improve numerical weather forecasts in the region. This could help in providing timely warning for these storms, thus mitigating potential damage to life and property by severe storms. The issue is not only important from a theoretical and scientific point of view but also for many communities around the SDC and in the downstream regions where these storms propagate.

## METHODS & CODES

The research team is conducting quasi-idealized simulations of convection initiation using Cloud Model 1 (CM1) [1] to enhance

the understanding of the physical processes involved. CM1 is a three-dimensional, nonhydrostatic, nonlinear numerical model used for idealized studies of atmospheric flows. It contains several physics packages for parameterizing subgrid turbulence, radiation, cloud microphysics, and the like. The preconvective thermodynamic and kinematic environment for these simulations was taken from one of the Intensive Observation Periods (IOP) of the RELAMPAGO–CACTI field experiment, which was conducted in November–December 2018. This IOP involved complex interactions among the northerly South American low-level jet, thermally driven upslope flows, outflow boundaries from previous convection, and the SDC terrain. The team has conducted simulations of increasing complexity to parse the contributions of various physical processes to convection initiation. The researchers are also using data collected during RELAMPAGO to validate these model runs.

## RESULTS & IMPACT

The research team carried out three simulations: (1) with no background winds to isolate the contribution of daytime upslope flows to Convective Initiation; (2) with background winds; and, (3) with background winds with a cold pool. In the third simulation, a cold bubble with a temperature perturbation of -10K was dropped at the southern end of the SDC to mimic the cold pool generated by a previous thunderstorm. Radiation and surface fluxes were turned on in all the simulations. Preliminary results

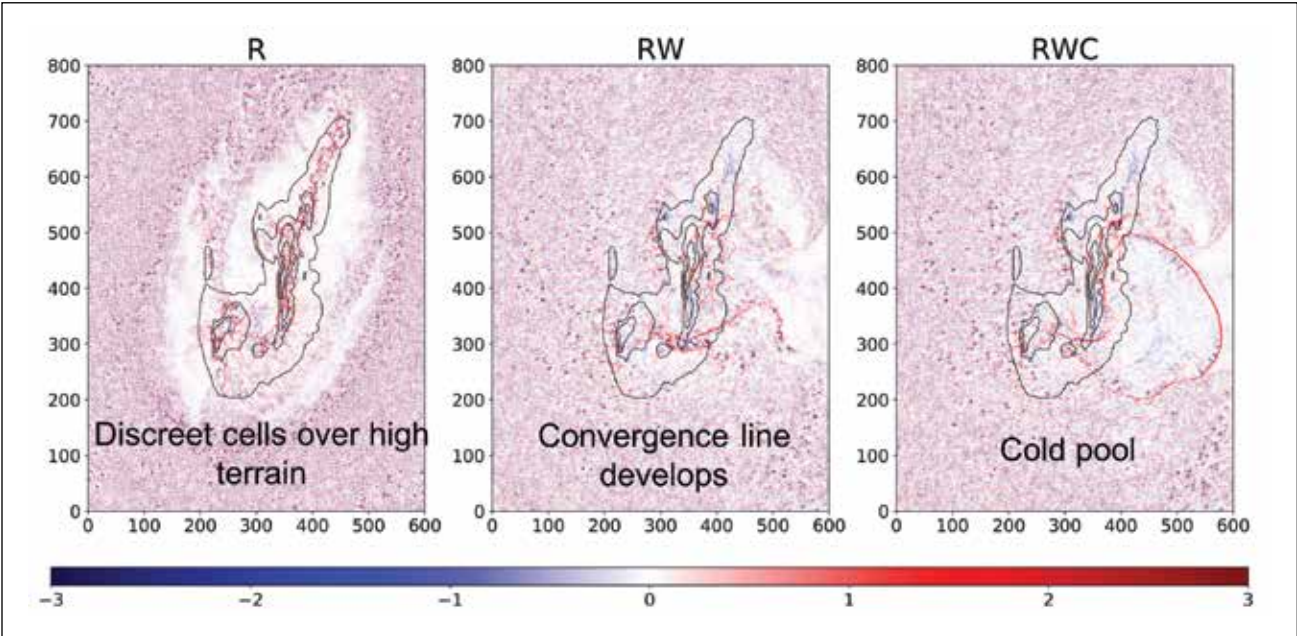


Figure 2: The vertical velocity field at model level 5 (500 meters above ground level) at 19:40 UTC (about eight hours after the start of the simulation).

indicate that the background winds advect the heat and moisture fluxes away from the terrain, causing a smaller increase in convective available potential energy and a smaller decrease in convective inhibition when compared to the simulation with no winds. As shown in Fig. 2, convergence lines and convective cells develop along the high terrain in the simulation with no winds. In the simulation with background winds, a convergence line develops to the south of the main SDC ridge, initiating convection upon interacting with the terrain. In the third simulation, no convection developed at the boundary of the cold pool.

Future simulations will involve parameterizing the effect of synoptic-scale heat and moisture advection and imposing it on the thermodynamic environment in the simulation.

## WHY BLUE WATERS

Blue Waters has the capacity to provide the sheer amount of computational power required to conduct these simulations. Additionally, the high-resolution model output needs tens of terabytes of space, which Blue Waters has provided. Further, the Blue Waters staff was very prompt in responding to all queries. The research team had some issues compiling the code on Blue Waters, but the staff helped the team through the process.

## PUBLICATIONS & DATA SETS

I. Singh and S. W. Nesbitt, “High-resolution simulations of orogenic convection initiation over the Sierras de Córdoba Mountains,” presented at the 18th American Meteorological Society Conf. on Mesoscale Processes, Savannah, GA, U.S.A., Jul. 29–Aug. 1, 2019, poster no. 40.

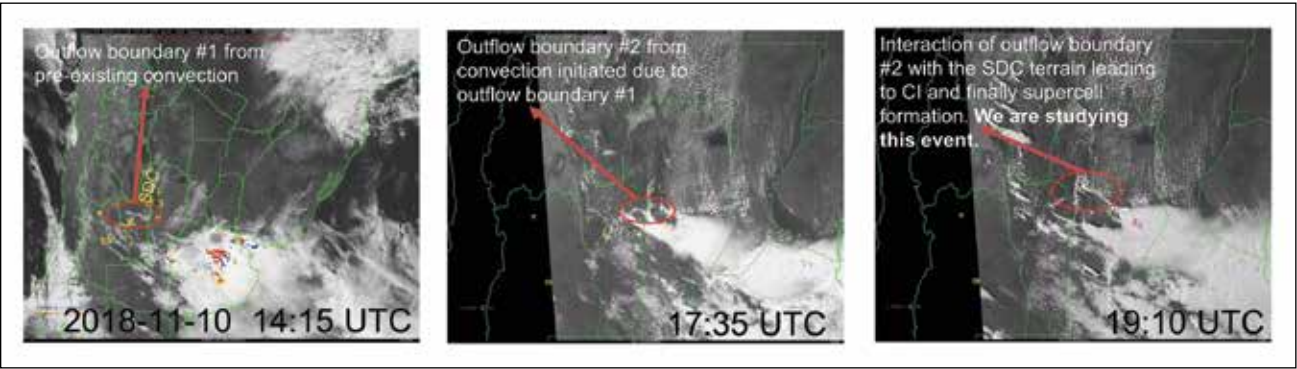


Figure 1: GOES-16 visible satellite imagery at different times, leading up to the initiation of a supercell. The location of the Sierras de Córdoba mountains is indicated in the figure.



DI

TN

MI

BW

# SIMULATIONS OF VIOLENTLY TORNADIC SUPERCELLS AND DAMAGING THUNDERSTORMS

**Allocations:** NSF PRAC/3,630 Knh (bagm)  
NSF PRAC/21,000 Knh (baye)  
GLCPC/360 Knh (baro)  
**PI:** Leigh G. Orf<sup>1</sup> (all three projects)  
**Co-PI:** Catherine Finley<sup>2</sup> (bagm)  
**Collaborators:** Kelton Halbert<sup>1</sup> (baye), Eric Savory<sup>3</sup> (baro)

<sup>1</sup>University of Wisconsin–Madison  
<sup>2</sup>University of North Dakota  
<sup>3</sup>University of Western Ontario

CI

FS

BI

## EXECUTIVE SUMMARY

Three projects on Blue Waters carried out by PI Leigh G. Orf have led to a better understanding of violently tornadic supercells (bagm and baye) as well as thunderstorms that produce damaging straight-line winds (baro).

Breakthrough simulations of violently tornadic supercells previously conducted on Blue Waters brought to light new features in these simulated storms that directly challenge long-held conceptual models of supercell morphology and tornado formation. A feature the research team dubbed the streamwise vorticity current that occurs in a part of the storm that has been long ignored by field meteorologists has been identified as playing a key role in tornado genesis and maintenance.

In new work (baye), the research team conducted a simulation study of a violently tornadic supercell, run at unprecedented scale and resolution. The simulation, requiring 20,000 Blue Waters nodes and one-quarter of a trillion grid zones, contains the genesis and maintenance of a long-track EF5-strength tornado exhibiting a multiple-vortex structure. Novel techniques of visualization recently developed by the team combine volume-rendered imagery of vorticity and cloud field with the track of the tornado as manifest by ground-relative traces of the vorticity, velocity, and pressure fields. These tracks bear a striking resemblance to damage swaths found in nature, clearly showing the cycloidal paths of rapidly moving suction vortices embedded within the wide, wedge-shaped tornado.

## RESEARCH CHALLENGE

Tornadoes are common in the United States, and the strongest tornadoes cause devastating damage and severe loss of life. Understanding the nature of the strongest tornadoes (rated EF4/EF5 on the Enhanced Fujita Scale) is of great societal interest. A better understanding of these storms will help improve forecasting of the events and provide accurate, targeted warnings, thereby reducing the false alarm rate of the National Weather Service.

Thunderstorms that produce downbursts are also of interest to atmospheric scientists. While “dry” downbursts are relatively understood to be forced by the evaporation of rain in a deep,

dry atmospheric boundary layer, “wet” downbursts, which are believed to be forced primarily by drag induced by falling rain and hail, are less understood.

## METHODS & CODES

The team used the CM1 model, developed at the National Center for Atmospheric Research, for all three projects. One of the PIs, Leigh Orf, has written an I/O driver for CM1 and a set of tools to read data, called LOFS or the “lack of file system.” LOFS is so named because it is a file-based file system that sits upon Lustre. Crucially, LOFS allows very high throughput of saved data; organizes it in an efficient, logical manner; and allows for the use of lossy compression on floating point data using ZFP. In addition, LOFS has a simple application programming interface for reading LOFS data directly as well as for converting it to the popular netCDF format, which can be used to share the data with colleagues or as input for visualization tools.

## RESULTS & IMPACT

*Simulations of violently tornadic supercells (bagm).* These simulations and the hours of high-definition animated video shared in near real time on YouTube have shaken up the field of meso-scale meteorology, directly challenging long-established conceptual models of supercells and tornadoes. The processes the research team has identified are novel, with one direct outcome of this work being a \$2.3-million National Science Foundation-sponsored program (TORUS) to search for evidence of these features in the real atmosphere.

*Quarter-trillion-zone tornadic supercell simulation (baye).* As this simulation is not yet complete and was only recently integrated to its current state, results are preliminary. Video sequences showing high-definition animations of cloud and vorticity fields from a recent talk, shared on social media, have been viewed hundreds of thousands of times. The severe storms community is very aware of this work on Blue Waters, primarily through less highly resolved simulations covered in another Blue Waters allocation.

This new 10-meter simulation and the features that are resolved will be of significant interest to scientists and to the general public. Owing to its sheer size, it will take years to analyze these data;

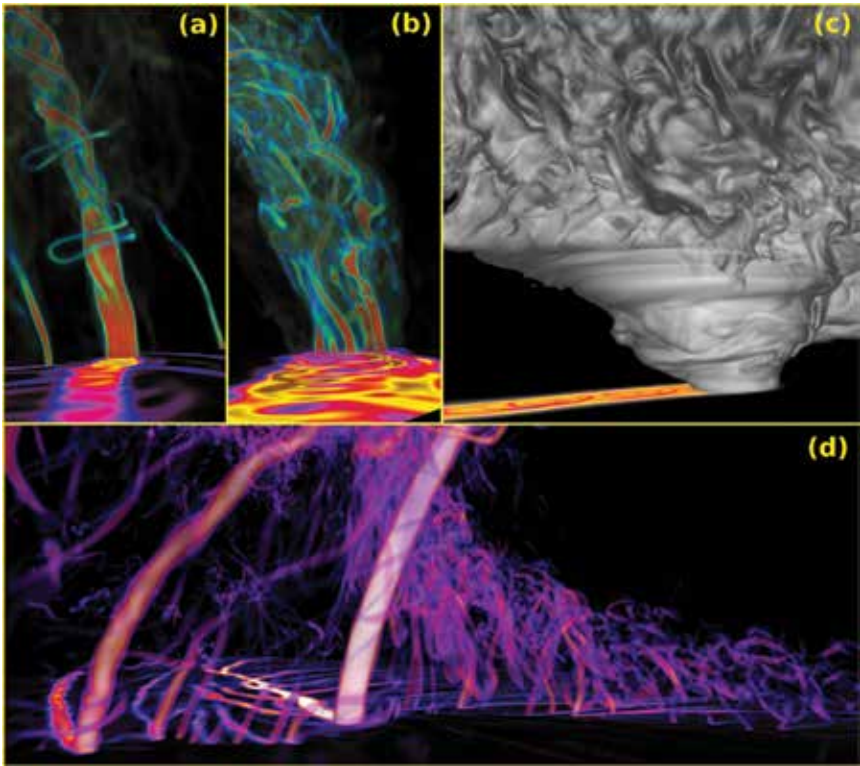


Figure 1: (a) Vertical vorticity field and vorticity tracks of the EF5 tornado shortly following tornadogenesis. (b) As in (a) but during the maintenance phase of the tornado. (c) The cloud field and negative pressure perturbation track during the maintenance phase. (d) A wide-angle view of the tornado shortly following tornadogenesis.

however, without Blue Waters, the simulation simply could not have been performed. The PI’s efforts in data management and compression enable the data to be saved at very high temporal resolution; data will be analyzed using new techniques on GPUs on the next-generation Frontera machine and in-house GPUs.

*Damaging thunderstorms (baro).* Both simulations produced complexes of downbursts as well as surface instantaneous winds exceeding 80 miles per hour. Studying the forcing that results in these strong surface winds will help scientists understand the interplay between falling precipitation and thermodynamic effects involving phase changes of water substance that lead to strong downdrafts and subsequent damaging near-surface horizontal winds.

## WHY BLUE WATERS

No other nonclassified machine available to U.S. scientists has Blue Waters’ capacity. In addition, NCSA has been exceptional in terms of staff support, helping the team over seemingly insurmountable hurdles and keeping the machine in a healthy state.

## PUBLICATIONS & DATA SETS

L. Orf, “The role of the streamwise vorticity current in tornado genesis and maintenance,” presented at the 29th Conf. Severe Local Storms, Stowe, VT, U.S.A., Oct. 22–26, 2018.

K. T. Halbert, “Cold pool horizontal streamwise vorticity during tornadogenesis and maintenance in a simulated supercell thunderstorm,” presented at the 29th Conf. Severe Local Storms, Stowe, VT, U.S.A., Oct. 22–26, 2018.

C. A. Finley, L. Orf, B. D. Lee, and R. Wilhelmson, 2018: “High-resolution simulation of a violent tornado in the 27 April 2011 outbreak environment,” presented at the 29th Conf. Severe Local Storms, Stowe, VT, U.S.A., Oct. 22–26, 2018.

M. A. Elmore, “Sensitivity of tornado evolution to changes in 0–500 m wind shear in high-resolution simulations,” presented at the 29th Conf. Severe Local Storms, Stowe, VT, U.S.A., Oct. 22–26, 2018.

L. Orf, “The use of ZFP lossy floating point data compression in tornado-resolving thunderstorm simulations,” presented at the 2017 AGU Fall Meeting, New Orleans, LA, U.S.A., Dec. 11–15, 2017. [Online]. Available: <https://agu.confex.com/agu/fm17/meetingapp.cgi/Paper/209055>.

L. Orf, “Petascale supercell thunderstorm simulations, and a new hypothesis for tornado formation and maintenance,” presented at the Joint Congress of the Canadian Society for Mechanical Engineering and the CFD Society of Canada (CSME-CFDSC Congress 2019), Western University, London, ON, Canada, June 2-5, 2019. [Online], available: <https://youtu.be/5cel1fLxR04?t=2291>.

S. Li, S. Jaroszynski, S. Pearse, L. Orf, and J. Clyne, “VAPOR: A visualization package tailored to analyze simulation data in earth system science,” *Atmosphere*, vol. 10, no. 9, p. 488, 2019.

L. Orf, “A violently tornadic supercell thunderstorm simulation spanning a quarter-trillion grid volumes: Computational challenges, I/O framework, and visualizations of tornadogenesis,” *Atmosphere*, vol. 10, no. 10, p. 578, 2019.



# PREDICTION OF GEOMAGNETIC SECULAR VARIATION WITH LARGE-ENSEMBLE GEOMAGNETIC DATA ASSIMILATION

**Allocation:** Innovation and Exploration/80 Knh  
**PI:** Nikolaos Pavlis<sup>1</sup>  
**Co-PI:** Weijia Kuang<sup>2</sup>

<sup>1</sup>National Geospatial–Intelligence Agency  
<sup>2</sup>NASA Goddard Space Flight Center

## EXECUTIVE SUMMARY

The geomagnetic field varies in time, mostly owing to the fluid motion in the Earth’s outer core. Geomagnetic data assimilation can provide accurate estimates of the core state for fundamental research into such questions as the Earth’s interior structure and its evolution. Geomagnetic data can also provide accurate secular variation (SV) forecasts for global geomagnetic models that are used for industrial and navigational applications. Accurate prediction of SV can be achieved via large-ensemble assimilation of geomagnetic observations and theoretical geodynamo models that investigate the self-sustaining process responsible for maintaining the Earth’s magnetic field. However, this re-

quires at least one thousand times more computing resources (in both CPU time and data storage) than those for pure geodynamo simulation, which alone is already computationally challenging. Blue Waters enables this research by reducing the research time from years to weeks and by increasing resolutions for geodynamo simulations with Earthlike parameters.

## RESEARCH CHALLENGE

The time-varying geomagnetic field is of fundamental importance for basic and applied scientific research: it provides key information about the Earth’s evolution over geological time scales; it plays a critical role in interactions between the Earth’s core and

other components of the Earth, which give rise to other geodynamic variations, *e.g.*, long-term variations in the Earth’s rotation; it protects the Earth’s surface and atmosphere from high-energy particles from coronal mass ejections and extreme ultraviolet fluxes from the Sun that are detrimental to life on Earth; and it has long been used by mankind for navigation, exploration, and other applications. Geomagnetic studies are very challenging because the dynamics vary over a broad spectrum in time (from subannual timescales to billions of years) and in space (from centimeters to 10,000 kilometers). Modeling and predicting such processes involve physical parameters (and thus model parameters) varying over  $10^{20}$  in magnitude, thus demanding extremely high spatial and temporal resolution, which lead to approximately  $10^{21}$  floating point operations for a typical simulation. Therefore, petascale supercomputing systems such as Blue Waters are instrumental for geomagnetic field research.

## METHODS & CODES

The research team’s system, called the Geomagnetic Ensemble Modeling System (GEMS), was developed exclusively in NASA’s Goddard Space Flight Center [4,9,10,12]. It comprises three major subsystems: a geodynamo model (called MoSST) to provide forecast results, an ensemble Kalman filter model (called EnKF) to provide analysis (initial state) for making forecasts, and a geomagnetic data assimilation driver (called GDAS) to manage the entire analysis–forecast cycle. MoSST [2,5,6], was written in Fortran 2003, is based on a hybrid spectral-finite difference algorithm to solve the magnetohydrodynamic state in the Earth’s core, and is computationally most demanding (> 90%). EnKF [9] utilizes an ensemble Kalman filter methodology to optimally integrate geodynamo model outputs and global geomagnetic field models derived from surface geomagnetic, paleo/archaeomagnetic data [1,3,7,8] to provide analysis for forecasts. GDAS is a shell-script-based system that manages the interactions between MoSST and EnKF, and controls the forecast cycles.

## RESULTS & IMPACT

The main objective of this project is to investigate the convergence of assimilation with different ensemble sizes and simulation resolutions for given physical parameters. In two months of work, the research team found that the ensemble size of approximately 256 is optimal for assimilation, based on the computational needs and the forecast accuracies. This result is very important as it establishes a quantitative correlation among the forecast accuracy requirements, computational resource needs, and time periods for progress. For example, with the spatial resolution of approximately  $256 \times 256 \times 256$ , a single geodynamo simulation (*i.e.*, an ensemble member of GEMS) could require a one-month (wall-clock) computation time with 256 processors/cores. Optimal ensemble sizes can greatly reduce the computational expense and research time without compromising research objectives. In addition, simultaneous 256-ensemble runs can make accurate forecasts of five-year geomagnetic SVs in one month that

would otherwise require 20 years if the ensemble runs were limited to sequential executions (one member at a time).

## WHY BLUE WATERS

Blue Waters provides the computing resources needed for the research team’s geomagnetic data assimilation research project. Further, the technical staff provide much-needed knowledge to improve and optimize GEMS.

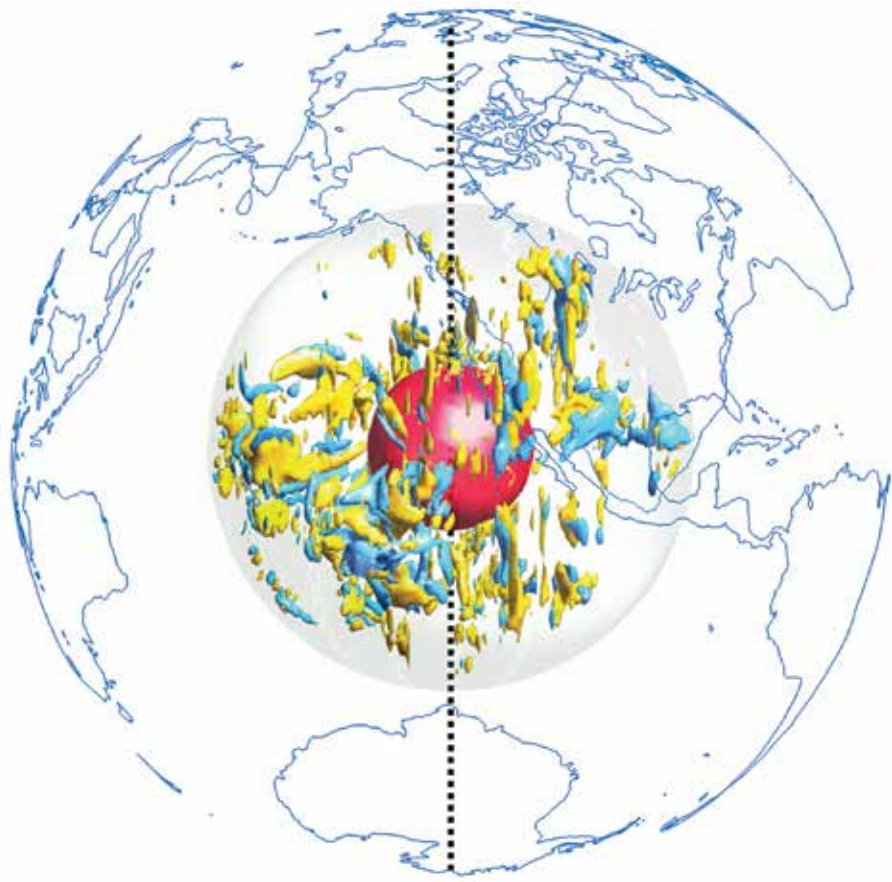


Figure 1: A snapshot of dynamo action in the Earth’s outer core, defined between the (red) inner core and the (transparent) core–mantle boundary. The net gain and loss of magnetic energy is described by the yellow and blue blocks, respectively. The dashed line is the mean rotation axis of the Earth.



MACHINE LEARNING FOR ERROR QUANTIFICATION IN SIMULATING THE CLIMATE IMPACTS OF ATMOSPHERIC AEROSOLS

Allocation: Innovation and Exploration/200 Knh  
PI: Nicole Riemer<sup>1</sup>

<sup>1</sup>University of Illinois at Urbana–Champaign

EXECUTIVE SUMMARY

Atmospheric aerosol particles influence the large-scale dynamics of the atmosphere and climate because they interact with solar radiation, both directly by scattering and absorbing sunlight and indirectly by forming clouds. Current climate simulations use highly approximate aerosol models with large and unquantified errors.

To address this uncertainty, the PI applied a new ultrahigh-detail spatially resolved and particle-resolved aerosol model, WRF–PartMC, which tracks size and composition information on a per-particle basis. The scientists then used its model output to train machine learning models to predict errors owing to a simplified representation of the aerosol. This allows error predictions merely on the basis of output from the simplified model without running a computationally expensive particle-resolved benchmark case. The scientific impact of this work is the development of a new, innovative method that changes the way aerosol impacts on climate are quantified in current regional and global climate models.

RESEARCH CHALLENGE

One of the largest uncertainties in global climate prediction involves aerosols and their impacts on the radiative budget, a topic of great societal relevance [1]. Aerosol interactions are influenced by both the size and composition of individual particles. Models provide important insights in the study of aerosols but experience a trade-off between the representation of physical detail and spatial resolution. State-of-the-art 3D weather- and climate-scale models focus on large-scale transport but assume a crudely simplified aerosol representation. Commonly applied simplifications assume that all particles look very similar within a particle population, which is typically not representative of reality. In contrast, current-generation box models capture the small-scale features of aerosol physics and chemistry but cannot resolve spatial heterogeneities of aerosol populations. As a result, simulating the evolution of aerosols and predicting their impacts remains a challenge owing to the multiscale nature of the system.

METHODS & CODES

To address these research challenges, the PI has developed the particle-resolved aerosol model WRF–PartMC, which is the first model explicitly resolving the evolution of individual aerosol par-

ticles within the grid cells of a state-of-the-art atmospheric fluid/meteorology model. This model makes no simplifying assumption in regard to aerosol composition. Therefore, WRF–PartMC is uniquely suited to fill the role of a benchmark model for simulating atmospheric aerosol composition. The particle composition of hundreds of billions of particles in the atmosphere is represented at a given time, and each particle’s composition evolves over time owing to coagulation with other particles and condensation of gas vapors.

WRF–PartMC simulations provide a wealth of data on aerosol mixing state and aerosol aging under different environmental conditions. The data generated give one training sample per grid cell per timestep, consisting of the global model state variables in that grid cell and the computed error for the grid cell. The PI uses these data combined with state-of-the-art machine learning techniques to train models for predicting errors in climate-relevant quantities such as optical properties and cloud-forming abilities, which occur when using simplified aerosol treatments. This will allow researchers to make error predictions merely on the basis of output from the simplified model without running a computationally expensive particle-resolved benchmark case.

RESULTS & IMPACT

The PI conducted several particle-resolved simulations for the domain of northern California. The output will serve as training data for the machine learning portion of the project. These stochastic simulations utilize realistic source-resolved emissions, capable of modeling different emission sectors such as diesel vehicles, gasoline vehicles, and power plants, which all have complex aerosol composition. Each simulation was initialized for June 17, 2010, and simulated 24 hours, utilizing 6,656 cores and 12 wall-clock hours. Simulations consisted of 5,000 computational particles per grid cell with a domain size of 170 × 160 and 40 vertical levels. Each simulation models the complex aerosol dynamics and chemistry for on the order of 5 billion individual particles, where each particle is represented as a vector of masses of 20 aerosol species. Simulations of this size were previously unfeasible. Now that such simulations can be conducted, errors owing to aerosol representation can be quantified.

The mixing state parameter  $\chi$ , as described in [2], quantifies the extent to which the particle population is internally mixed by examining how complex individual particles are and how simi-

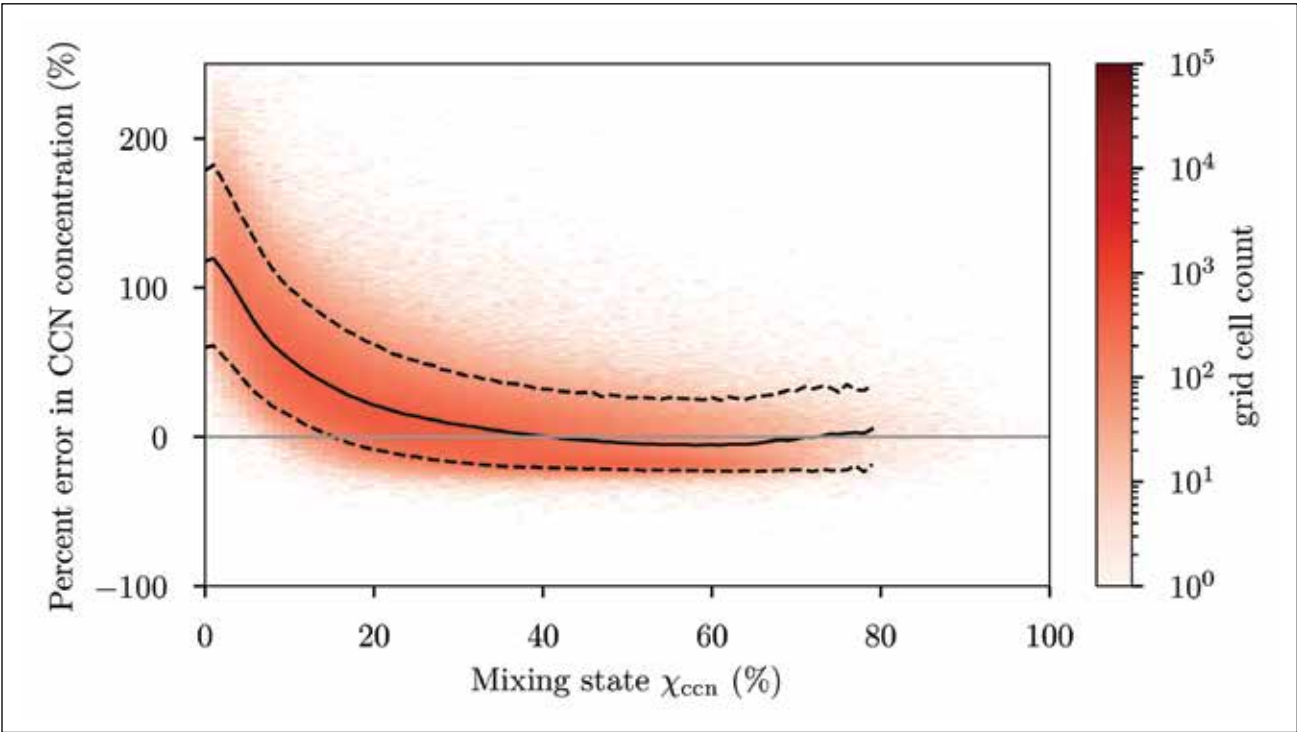


Figure 1: Relative error in cloud condensation number concentrations when assuming an internally mixed aerosol as a function of mixing state parameter  $\chi_{CCN}$ . The mixing state parameter is 100% for completely internal mixtures and 0% for completely external mixtures.

lar the particles are within the population. The mixing state parameter varies from 0% to 100%, ranging from all particles containing a single species to 100%, where all particles are identical in composition. Fig. 1 shows relative error in CCN (cloud condensation nuclei) number concentrations as a function of mixing state parameter. Particle populations are projected to fully internally mixed populations, a common representation in other models. When particle populations are further from the model assumption of internally mixed ( $\chi = 100\%$ ), models typically overestimate the number of CCN available for cloud formation. This has implications for cloud radiative properties that depend on droplet number and size.

WHY BLUE WATERS

Access to the computational power and storage space on Blue Waters allows for running simulations to produce data for machine learning. Simulations rely on sufficient memory per core to produce statistically powerful particle populations. Working alongside Blue Waters’ staff has alleviated a large portion of the performance issues regarding output by removing tiny writes of data and paying careful attention to what information is output. Reducing the cost of output allows more frequent output, which in turn provides more data for machine learning, as it requires the entire particle state.



DI

SIMULATING HYDROCLIMATE CHANGE IN SOUTHWEST NORTH AMERICA AT 21,000 YEARS AGO

**Allocation:** NSF PRAC/680 Knh  
**PI:** Clay Tabor<sup>1</sup>  
**Co-PI:** Isabel Montañez<sup>2</sup>  
**Collaborators:** Marcus Löfverström<sup>3</sup>, Jessica Oster<sup>4</sup>, Barbara Wortham<sup>2</sup>, Cameron de Wet<sup>4</sup>

<sup>1</sup>University of Connecticut  
<sup>2</sup>University of California, Davis  
<sup>3</sup>The University of Arizona  
<sup>4</sup>Vanderbilt University

MP

CI

FS

EXECUTIVE SUMMARY

How the hydroclimate in Southwest North America will change in the future remains an open question. Although models generally predict an increase in climate extremes in this region with both more severe droughts and more intense precipitation events, large uncertainties remain. By studying the past, we can improve our understanding of the mechanisms driving current and future hydroclimate change in the area.

Records suggest that the hydroclimate of this region was drastically different during much of the last glacial cycle. However, the various mechanisms that produced these signals are difficult to deconvolve and continue to be debated. In this project, the research team employed a high-resolution, water isotope-enabled Earth system model for simulations of the last glacial maximum and preindustrial climates. The combination of water isotope tracers and high resolution allows for more direct comparison of the model outputs with the proxy records.

RESEARCH CHALLENGE

Southwest North America has proven to be climatically sensitive over the observational record. However, confidence in model projections of that region's climate is limited by an incomplete understanding of the complex interactions driving regional climate variability and the unprecedented nature of future climate perturbation [1,2]. To better understand the range of possible climates in Southwest North America and provide additional sources of model validation, researchers can look to the past. Fortunately, this region has been particularly well surveyed, with some of the densest and most diverse coverage of paleoclimate proxy records available in the mid-latitudes. Of the many proxy archives gathered in Southwest North America, measurements of  $\delta^{18}\text{O}$  (a measure of the ratio of stable isotopes oxygen-18 and oxygen-16) in cave records, known as speleothems, have proven to be particularly valuable. These records suggest that in contrast to today, the hydroclimate of Southwest North America was cold and wet during much of the last glacial cycle [3,4]. However, the various mechanisms that produced these  $\delta^{18}\text{O}$  signals are difficult to deconvolve and continue to be debated.

Comparisons between climate model outputs and cave records are difficult. One common limitation relates to the heterogeneity of the land surface. Many proxy sites, especially those located in Southwest North America, are found in topographically diverse regions that create drastically different microclimates within a few hundred kilometers. In contrast, climate models typically do not resolve such fine spatial scales (Fig. 1), which limits the utility of models for mechanistically understanding this unique climate proxy. Another limitation comes from the indirect comparison of model variables such as temperature and precipitation with the  $\delta^{18}\text{O}$  values found in the cave records. To overcome these common limitations of model-proxy comparison, the research team used a high-resolution, water isotope-enabled Earth system model to simulate the last glacial maximum (21,000 years ago) and preindustrial climates. These experiments shed light on the driving mechanisms behind the dynamic hydroclimatic changes that have been suggested in Southwest North America from more than a century of paleoclimate research.

METHODS & CODES

For this project, the team used the Community Earth System Model (CESM) with prognostic water isotopologue tracking. CESM is a widely employed, Intergovernmental Panel on Climate Change-class Earth-system model with the ability to accurately simulate preindustrial and present-day climates [5]. This version of CESM, known as iCESM, can track  $\delta^{18}\text{O}$  and  $\delta^2\text{H}$  in atmosphere, ocean, land, and sea ice components. Previous studies demonstrated that the  $\delta^{18}\text{O}$  and  $\delta^2\text{H}$  distributions within this version of CESM compare favorably with other isotope-enabled models of similar complexity [6]. Further, this version of CESM can track the transport of water in the atmosphere, allowing for a precise understanding of the mechanisms responsible for changes in isotopic composition of precipitation [7]. Given the topographic heterogeneity of Southwest North America, the team will perform isotope-enabled atmosphere and land-only simulations at approximately 0.25° resolution using prescribed sea surface temperatures.

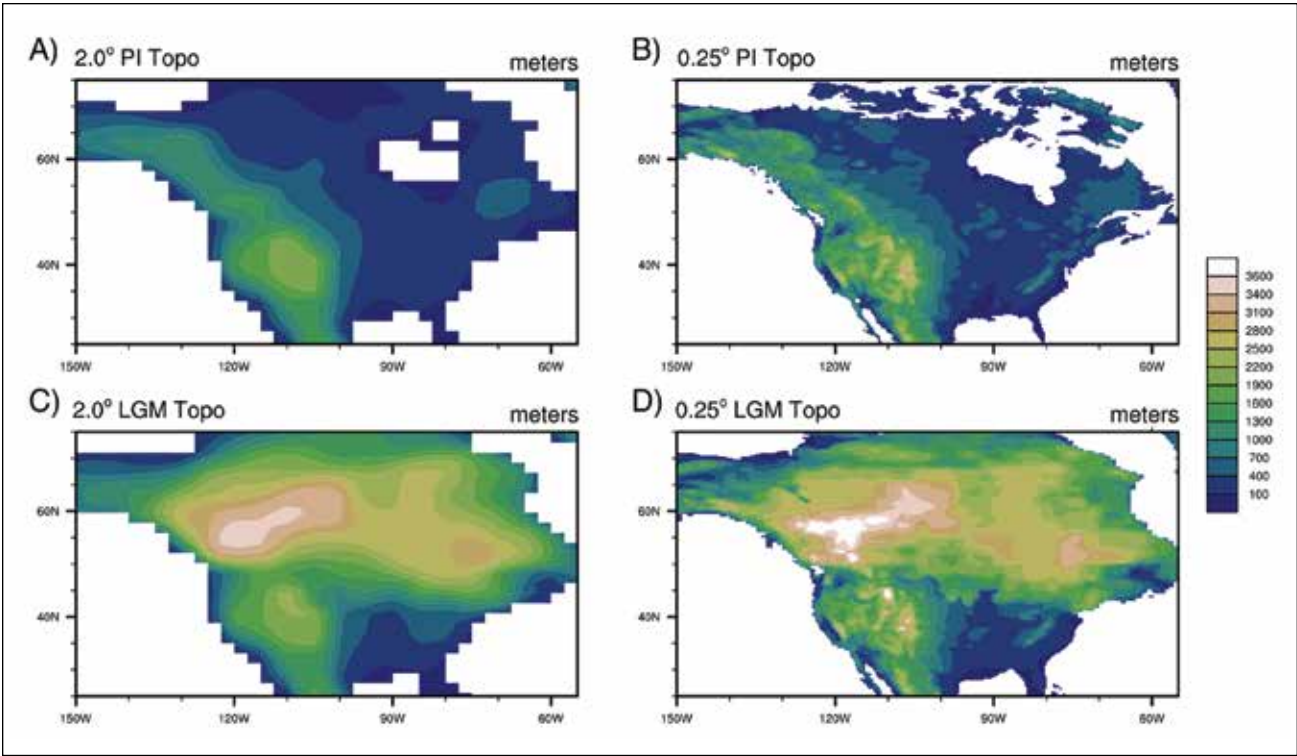


Figure 1: North American topography used in preindustrial and last glacial maximum (21 kiloannum) model simulations. (A) and (B) show preindustrial topography using the standard- (1.9 x 2.5°) and high- (0.23 x 0.31°) resolution model configuration, respectively. (C) and (D) show last glacial maximum topography using the standard- (1.9 x 2.5°) and high- (0.23 x 0.31°) resolution model configuration, respectively.

RESULTS & IMPACT

It is clear that further investigation is necessary to determine the most important mechanisms for driving the pattern of hydroclimatic change at the last glacial maximum. The combination of water isotope-enabled model experiments with speleothem records will allow the team to disentangle the influences of moisture source and transport, temperature, and precipitation amount on speleothem proxy records, an understanding that can be applied broadly to improve proxy interpretations across the region. From the results, the research group will be able to distinguish between several long-standing hypotheses of hydroclimate change in Southwest North America at the last glacial maximum, including: (1) a southward-displaced Pacific jet stream [8]; (2) a strengthening and meridional compression of the storm track [4]; (3) a thermodynamic control arising from a steepened humidity gradient [9]; and (4) an increase in moisture from a southwesterly, subtropical source [10] perhaps owing to increased contributions from atmospheric rivers [11]. These global simulations will also prove valuable for a wide range of paleoclimate questions. As the highest-resolution global simulations of the last glacial maximum ever performed, these outputs will be a valuable resource for the paleoclimate community and of particular interest to the paleoclimate model intercomparison project [12].

WHY BLUE WATERS

The large number of compute nodes on Blue Waters is ideal for these simulations. The combination of water isotope tracers and high resolution in CESM requires significant computing resources. Fortunately, CESM throughput scales well with a large number of processors. The research team expects high efficiency when employing well over 1,000 processors simultaneously per simulation. This research would not be possible without the computing power provided by Blue Waters.



IMPLEMENTATION AND USE OF A GLOBAL NONHYDROSTATIC MODEL FOR EXTENDED-RANGE WEATHER PREDICTION DURING THE RELAMPAGO FIELD CAMPAIGN

Allocation: Blue Waters Professor/240 Knh  
PI: Robert J. Trapp<sup>1</sup>

<sup>1</sup>University of Illinois at Urbana–Champaign

EXECUTIVE SUMMARY

The Model for Prediction Across Scales (MPAS) was implemented and then executed on Blue Waters during the RELAMPAGO (Remote sensing of Electrification, Lightning, And Meso-scale/microscale Processes with Adaptive Ground Observations) field campaign in 2018. MPAS is a new, nonhydrostatic weather and climate model that allows for local grid refinement. Because MPAS is also a global model, it is well suited for extended-range predictions, as was demonstrated by the four-day predictions made daily during the campaign. A configuration detail of par-

ticular relevance was the specification of 3-km gridpoint spacing over the entirety of South America, with 15-km gridpoint spacing elsewhere around the globe. The 3-km spacing is considered to be “convection permitting,” thus allowing for the explicit representation of thunderstorms over large domains.

Two other models were also employed for separate projects: the Weather Research and Forecasting (WRF) model, for studies of hail and tropical cyclones using a pseudo-global warming approach, and the Cloud Model 1 (CM1), for high-resolution idealized simulations of individual thunderstorms.

RESEARCH CHALLENGE

Thunderstorms and attendant phenomena such as hail, tornadoes, and extreme rainfall have high socioeconomic impact worldwide, thus motivating research to improve not only their predictions but also the understanding of their basic processes. One challenge is that their spatial and temporal scales are small relative to the scales of their meteorological forcing. To accurately predict and represent these storms and phenomena, an approach that can account for temporal scales ranging from days to seconds and spatial scales of thousands of kilometers to hundreds of meters is needed. In other words, very large geospatial domains that have fine gridpoint spacings and long time integrations with high rates of model output are required. The Blue Waters allocation is providing the research team with the resources needed to achieve this level of simulation.

METHODS & CODES

The team used MPAS with the aforementioned hybrid 3-km/15-km grid configuration and the convection-permitting suite of physics parameterizations. The model was integrated daily for four-day periods, using initial (and lower-boundary) conditions from the Global Forecast System model. This was done for the 45-day duration of the field campaign. The team also employed two other models for separate projects: the WRF model, under a pseudo-global warming approach, and the CM1 model, for high-resolution idealized simulations of individual thunderstorms.

RESULTS & IMPACT

The team concluded, based on the 45-day experiment, that MPAS offers some skill in extended-range, South American forecasts at convection-permitting scales. It would appear that the implied predictability is provided in part by the significant terrain-associated forcing, including that owing to the Andes Mountains and the Sierras de Córdoba Mountains. The researchers did notice, in some cases, that the predictability actually appeared to degrade rather than improve over shorter integration lengths. This could possibly be associated with model spinup; efforts to understand and quantify this are underway.

In the non-MPAS efforts, the team completed the simulation work on tropical cyclones and hailstorms (WRF, pseudo-global warming), and idealized thunderstorm simulations (CM1). The latter simulations were used to examine the strong connections, within storms, between the physical sizes of vertical drafts (“updrafts” and “downdrafts”) and the depth of the pool of cool air generated through these drafts. These physical sizes have impacts on tornado intensity and, more generally, have implications on how convective storms are parameterized in coarse-resolution models.

WHY BLUE WATERS

The project staff provided tremendous assistance in helping to install the model and associated software on Blue Waters. The staff also set up a daily reservation for the model runs. Compared to other high-end high-performance computing resources, Blue

Waters is incredibly stable and reliable, which ensured timely delivery of the daily forecasts.

PUBLICATIONS & DATA SETS

G. R. Marion and R. J. Trapp, “The dynamical coupling of convective updrafts, downdrafts, and cold pools in simulated supercell thunderstorms,” *J. Geophys. Res.-Atmos.*, vol. 124, 2019, doi: 10.1029/2018JD029055.

D. Carroll–Smith, L. C. Dawson, and R. J. Trapp, “High resolution real-data WRF modeling and verification of tropical cyclone tornadoes associated with Hurricane Ivan 2004,” *EJSSN*, vol. 14, no. 2, pp. 1–36. 2019.

R. J. Trapp, G. R. Marion, and S. W. Nesbitt, “Reply to ‘Comments on the regulation of tornado intensity by updraft width,’” *J. Atmos. Sci.*, vol. 75, pp. 4057–4061, 2018, doi: 10.1175/JAS-D-18-0276.1.

R. J. Trapp, K. A. Hoogewind, and S. Lasher–Trapp, “Future changes in hail occurrence in the United States determined through convection-permitting dynamical downscaling,” *J. Clim.*, 2019, doi: 10.1175/JCLI-D-18-0740.1.

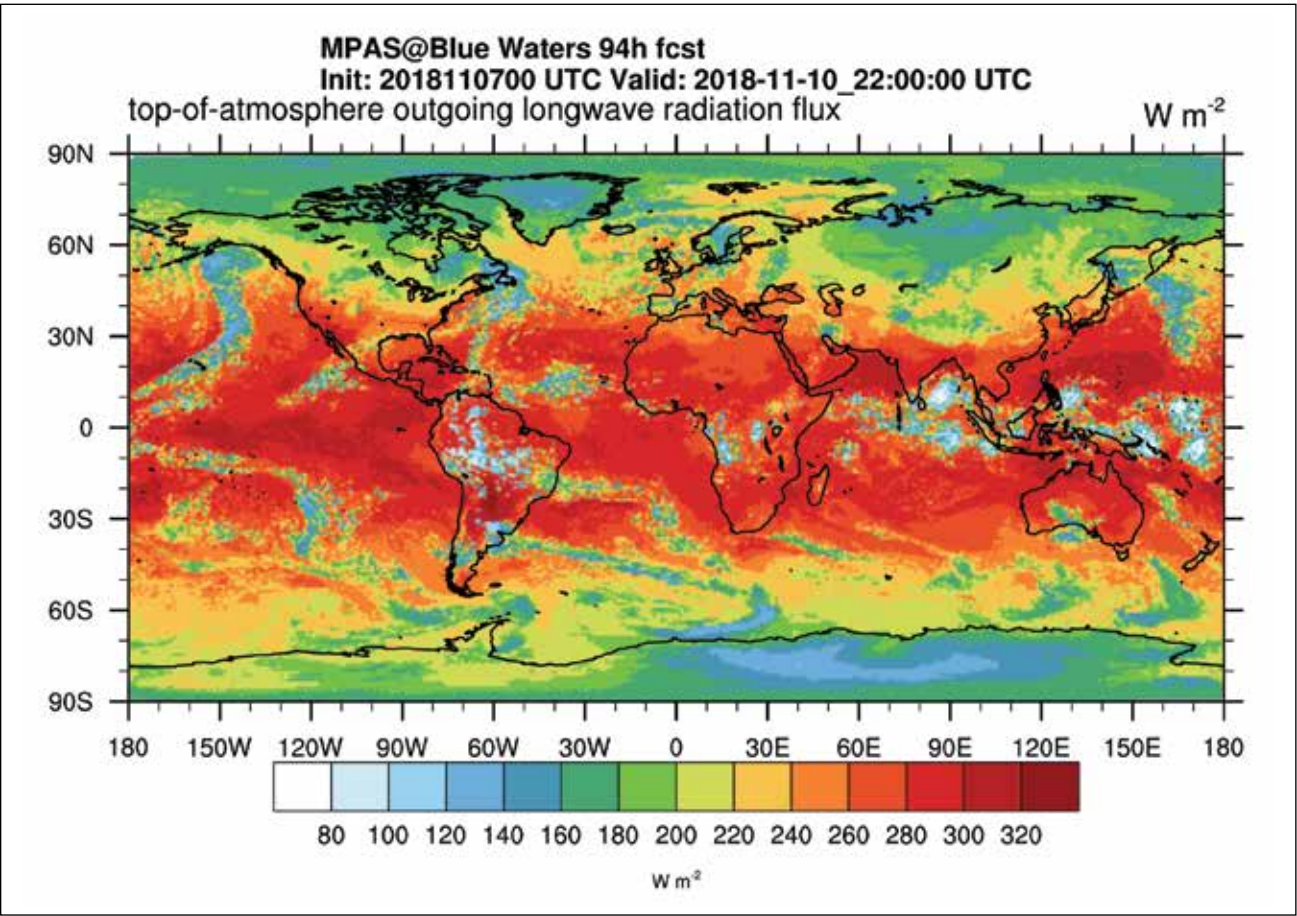


Figure 1: MPAS-simulated outgoing longwave radiation flux (W/m²) at the “top-of-atmosphere,” for a four-day prediction of severe thunderstorm activity over the Córdoba region of Argentina.



DI

GA

SIMULATING LARGE CALIFORNIA EARTHQUAKES BEFORE THEY OCCUR

**Allocation:** NSF PRAC/9,620 Knh  
**PI:** John Vidale<sup>1</sup>  
**Co-PIs:** Yifeng Cui<sup>2</sup>, Philip Maechling<sup>1</sup>, Kim Olsen<sup>3</sup>, Ricardo Taborda<sup>4</sup>  
**Collaborators:** Thomas H. Jordan<sup>1</sup>, Jacobo Bielak<sup>5</sup>, Alexander Breuer<sup>2</sup>, Scott Callaghan<sup>1</sup>, Jacquelyn Gilchrist<sup>1</sup>, Christine A. Goulet<sup>1</sup>, R. W. Graves<sup>6</sup>, Xiaofeng Meng<sup>1</sup>, Dawei Mu<sup>2</sup>, Dmitry Pekurovsky<sup>2</sup>, Daniel Roten<sup>3</sup>, William Savran<sup>1</sup>, Bruce Shaw<sup>7</sup>, Qian Yao<sup>8</sup>

<sup>1</sup> University of Southern California	<sup>5</sup> Carnegie Mellon University
<sup>2</sup> San Diego Supercomputer Center	<sup>6</sup> U.S. Geological Survey
<sup>3</sup> San Diego State University	<sup>7</sup> Lamont–Doherty Earth Observatory, Columbia University
<sup>4</sup> University of Memphis	<sup>8</sup> Microsoft Corp.

EXECUTIVE SUMMARY

The Southern California Earthquake Center (SCEC) used Blue Waters to develop the CyberShake probabilistic seismic hazard analysis (PSHA) method and to apply this method to all major urban areas in California. SCEC’s CyberShake [1] hazard models use detailed earthquake fault and seismic velocity models and high-performance software to calculate physics-based probabilistic ground motion forecasts. SCEC is actively collaborating with geo-scientific groups, national seismic hazard map developers [2], and civil engineering groups [3] to verify and validate the CyberShake California seismic hazard models for use in broad impact engineering and public seismic hazard applications and to apply the CyberShake method to other national and international regions.

RESEARCH CHALLENGE

PSHA earthquake forecast models [4] are the scientific basis for many engineering and social applications: performance-based design, seismic retrofiting, resilience engineering, insurance rate setting, disaster preparation and warning, emergency response, and public education. The U.S. Geological Survey (USGS) currently uses PSHA for promoting seismic safety engineering and disaster preparedness across the United States, including California, through its National Seismic Hazard Mapping Project

[5]. During the last year, researchers with the SCEC used the high-performance computing capabilities of Blue Waters to calculate physics-based PSHA models for northern California to better understand earthquake hazards and to better inform civil engineering organizations as they develop earthquake-resilient societal infrastructure.

METHODS & CODES

The SCEC earthquake system science program requires a collection of interoperable earth models and open-source scientific application programs including OpenSHA [6], UCVM [7], Hercules [8], and AWP–ODC [9]. SCEC’s CyberShake seismic hazard model calculations use a workflow system based on HT–Condor [10] and Pegasus–WMS [11] to perform large regional-scale seismic hazard studies. CyberShake extends existing PSHA methods to produce site-specific seismic hazard curves and other seismic hazard information such as duration of shaking, which is not available from earlier methods. In 2018, SCEC performed CyberShake Study 18.8, which used NCSA’s Blue Waters and OLCF’s Titan to calculate PSHA hazard curves up to 1 Hz for 869 locations in central and northern California, producing a physics-based PSHA hazard model for a large Northern California region that includes the San Francisco Bay Area.

RESULTS & IMPACT

Regional PSHA hazard models are used by engineers, seismologists, and governmental organizations in building design, urban planning, community earthquake awareness, and disaster preparation. During the last year, SCEC completed CyberShake Study 18.8, the first physics-based PSHA model for the San Francisco Bay region. This study used over 3.8 million Blue Waters node hours to calculate a PSHA seismic model for northern California, using deterministic wave propagation simulations in 3D seismic velocity models, combining estimates of hazard curves from 869 locations in California. CyberShake data products show the effects of basin structures and rupture directivity on hazard, improve upon standard attenuation-based methods of calculating seismic hazard, and identify research targets to further improve PSHA estimate accuracy. As a result, the scientific and compu-

tational advancements in CyberShake work can help reduce the total uncertainty in long-term hazard models, which has important practical consequences for the seismic provisions in building codes and especially for critical-facility operators.

PSHA users including scientific, engineering, and governmental agencies such as the USGS, are evaluating the new information provided by CyberShake results. For seismologists, CyberShake provides new information about the physics of earthquake ground motions, the interaction of fault geometry, 3D earth structure, ground motion attenuation, and rupture directivity. For governmental agencies responsible for reporting seismic hazard information to the public, CyberShake represents a new source of information that contributes to their understanding of seismic hazards, which they may use to improve the information they report. For building engineers, CyberShake represents a refinement of existing seismic hazard information that reduces some of the uncertainties in current empirical ground motion attenuation models.

CyberShake PSHA estimate simulations for Southern California are under review as inputs to a new Los Angeles urban seismic hazard map under development by the USGS [2]. The SCEC committee for Utilization of Ground Motion Simulations (UGMS) is working within the framework of the Building Seismic Safety Council activities to develop long-period, simulation-based response spectral acceleration maps for the Los Angeles region. CyberShake hazard maps are under consideration for inclusion in the National Earthquake Hazards Reduction Program and the American Society of Civil Engineers (ASCE) 7–10 Seismic Provisions, and for the Los Angeles City Building Codes. The UGMS group is using CyberShake simulations to quantify the effects of sedimentary basins and other 3D crustal structures on seismic hazard, information that is difficult to obtain with traditional empirical methods. Prototype risk-targeted maximum considered earthquake response spectra have been mapped using a combination of the empirical approach and the CyberShake model and are being integrated into the ASCE Project 17 recommendations for tall buildings [3].

WHY BLUE WATERS

SCEC used Blue Waters to perform large-scale, complex scientific computations involving thousands of large CPU and GPU parallel jobs, hundreds of millions of short-running serial CPU tasks, and hundreds of terabytes of temporary files. SCEC scientists and technical staff have worked closely with the Blue Waters staff to achieve a series of breakthroughs including integration of new physics into wave propagation software [12], optimization of production calculations using GPU code improvements [13], and optimization of the CyberShake runtime performance.

Using the well-balanced system capabilities of Blue Waters’ CPUs, GPUs, disks, and system software, together with scientific workflow tools, SCEC’s research staff can now complete CyberShake calculations in weeks rather than months, improvements that were made during years of Blue Waters access and opera-

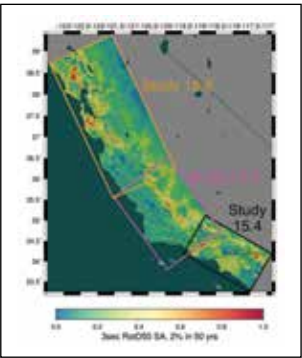


Figure 2: This California map shows three CyberShake probabilistic seismic hazard models calculated using Blue Waters including southern California (CyberShake Study 15.4), central California (CyberShake Study 17.3), and northern California (CyberShake Study 18.8), showing the CyberShake hazard estimates that have now been calculated for the most densely populated regions in the state.

tions. Blue Waters has enabled SCEC scientists to improve their seismic hazard methodology at a rapid pace.

PUBLICATIONS & DATA SETS

- A. Breuer, Y. Cui, and A. Heinecke, “Petaflop seismic simulation on elastic cloud clusters,” presented at ISC’19, June 16–20, Frankfurt, Germany, 2019.
- S. Callaghan, G. Juve, K. Vahi, P. J. Maechling, T. H. Jordan, and E. Deelman, “rvGAHP: push-based job submission using reverse SSH connections,” in *Proc. 12th Workshop on Workflows Support of Large-Scale Science (WORKS ’17)*, Denver, CO, U.S.A., Nov. 12–17, 2017, pp. 3:1–3:8, doi: 10.1145/3150994.3151003.
- C. Crouse, T. H. Jordan, K. R. Milner, C. A. Goulet, S. Callaghan, and R. W. Graves, “Site-specific MCER response spectra for Los Angeles Region based on 3D numerical simulations and the NGA West2 equations,” presented at the 11th Nat. Conf. Earthquake Eng., Los Angeles, CA, U.S.A., June 25–29, 2018, Paper 518.
- T. H. Jordan *et al.*, “CyberShake models of seismic hazards in southern and central California,” presented at the 11th Nat. Conf. Earthquake Eng., Los Angeles, CA, U.S.A., June 25–29, 2018, Paper 1458.
- D. Roten, K. B. Olsen, S. M. Day, and Y. Cui, “Quantification of fault-zone plasticity effects with spontaneous rupture simulations,” *Pure Appl. Geophys.*, vol. 124, no. 9, pp. 3369–3391, 2017, doi: 10.1007/s00024-017-1466-5.
- B. E. Shaw *et al.*, “A physics-based earthquake simulator replicates seismic hazard statistics across California,” *Sci. Adv.*, vol. 4, no. 8, 2018, doi: 10.1126/sciadv.aau0688.
- P. Small *et al.*, “The SCEC Unified Community Velocity Model software framework,” *Seismolog. Res. Lett.*, vol. 88, no. 6, 2017, pp. 1539–1552, doi: 10.1785/0220170082.
- R. Taborda, N. Khoshnevis, S. Azizzadeh–Roodpish, and M. Huda, “Influence of the source, seismic velocity, and attenuation models on the validation of ground motion simulations,” presented at the 16th World Conference on Earthquake Engineering, Santiago de Chile, Chile, Jan. 9–13, 2017.
- J. Tobin, A. Breuer, C. Yount, A. Heinecke, and Y. Cui, “Accelerating seismic simulations using the Intel Xeon Phi Knights Landing processor,” in *ISC’17*, pp. 1–19.

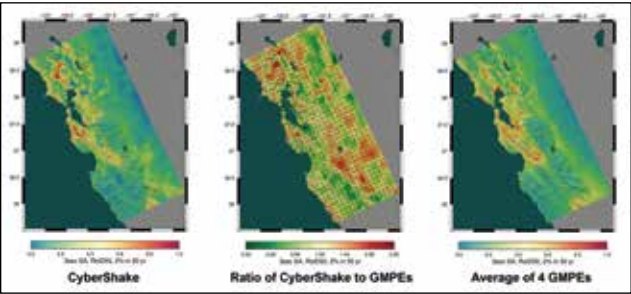


Figure 1: These maps show the CyberShake Study 18.8 model and compare the results to current standard methods based on ground motion prediction equations. This CyberShake seismic hazard model provides multiple layers of information that include hazard maps, hazard curves for selected sites, rupture models, seismograms, and site-specific seismic intensity and shaking duration measurements.



MATERIALS SIMULATIONS IN GEOPHYSICS

**Allocation:** Innovation and Exploration/200 Knh  
**PI:** Renata Wentzcovitch<sup>1</sup>  
**Collaborators:** Dong-Bo Zhang<sup>2</sup>, Kei Hirose<sup>3</sup>

<sup>1</sup>Columbia University  
<sup>2</sup>Beijing Computational Science Research Center  
<sup>3</sup>Earth-Life Science Institute, Tokyo Institute of Technology

EXECUTIVE SUMMARY

Thermodynamics of materials is a fundamental subject for understanding planetary processes and states. The research team has contributed to the development and application of *ab initio* methods for computing thermal properties of minerals and their aggregates (rocks). Thermodynamic and thermoelastic properties and thermal conductivity are among the main properties needed for modeling thermal convection in and interpreting the seismic tomographic structure of Earth's interior. Discovery of novel silicate and oxide phases at pressures (*P*) of tens of Mbar and temperatures (*T*) of approximately 10<sup>4</sup> K are also advancing understanding of the internal structure of recently discovered terrestrial-type exoplanets. Blue Waters has been used by the research group to perform high-throughput calculations of thermal properties of mineral phases required to model heat transport and interpret the 3D pattern of seismic wave propagation throughout the Earth's mantle and the inner cores of Earth and terrestrial exoplanets.

RESEARCH CHALLENGE

*Ab initio* materials simulations for geophysics applications are intrinsically high-throughput calculations since materials properties need to be computed vs. *P*, *T*, composition (*x*), strains (*ε*), atomic configuration, and the like. Such simulations are essentially a “phase space” sampling problem. Mineral phases are structurally and chemically complex, consisting of solid solutions with multiple components, including strongly correlated ones such as FeO, Fe<sub>2</sub>O<sub>3</sub>, etc., with primitive cells containing tens of atoms. In addition to properties of single phases, properties of polycrystalline aggregates—rocks—in thermodynamic equilibrium must also be addressed. The thermodynamic equilibrium problem in a multiphase system with multiple components, a challenge of its own, can only be addressed after thermodynamic properties of single phases are accurately obtained. To facilitate these HPC (high-performance computing) calculations, novel methods and workflows have been developed and implemented on Blue Waters for the calculations of single-phase thermal properties.

METHODS & CODES

Two novel methods based on the phonon quasiparticle concept [1] have been implemented and used on Blue Waters. The first is for computation of lattice thermal conductivity,  $\kappa_{\text{lat}}$ , and the second for thermodynamic properties of anharmonic systems, both

in the thermodynamic limit ( $N \rightarrow \infty$ , with  $N$  = number of atoms). Molecular dynamics (MD) is first used to obtain phonon quasiparticle properties, *i.e.*, phonon lifetimes ( $\tau_i$ ) and temperature-dependent phonon frequencies [ $\omega_i(T,V)$ ] at few-phonon wave vectors. Novel interpolation schemes are then used to obtain quasiparticle properties throughout the Brillouin zone (a uniquely defined primitive cell in reciprocal space), which are then used in conjunction with the Boltzmann transport equation (BTE) to compute  $\kappa_{\text{lat}}$ . *T*-dependent phonon dispersions are also used to

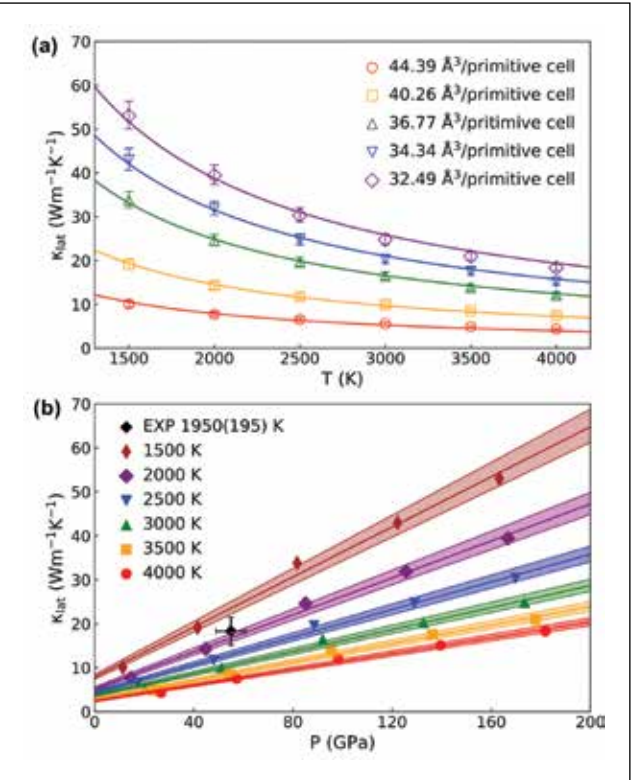


Figure 1: (a)  $\kappa_{\text{lat}}$  vs. *T* of Ca-Pv at five volumes. Solid curves show the 1/*T* dependence. (b)  $\kappa_{\text{lat}}$  vs. *P* of Ca-Pv at six temperatures. Solid curves show the linear fit. The error bars and the shaded areas indicate the uncertainties caused by fitting errors.  $\kappa_{\text{lat}}$  in both (a) and (b) are calculated from phonon quasiparticles properties sampled in a 20×20×20 q-point grid obtained using an interpolation of  $\tau_i(\omega)$  vs.  $\omega$  obtained by direct MD calculations sampling 2×2×2 (40 atoms), 3×3×3 (135 atoms), and 4×4×4 (320 atoms) q-point grids. The experimental datapoint was measured using laser-heated diamond anvil cell by K. Hirose and K. Ohta [Zhang *et al.*, 2019].

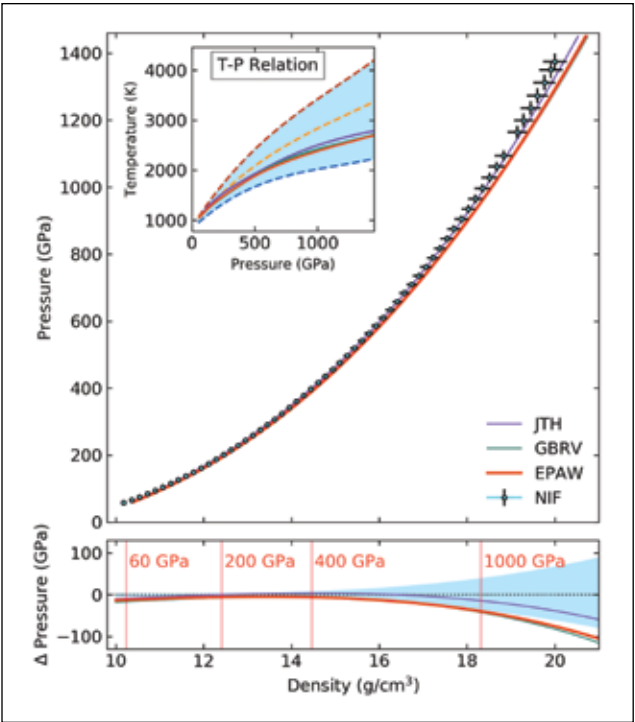


Figure 2: Isentropic *P*- $\rho$  EoS of  $\epsilon$ -Fe. The NIF data are shown as circles with uncertainty. The pink curve stands for this work, using a novel evolutionary PAW for iron validated up to 8 Mbar [2]. Other colored lines were obtained using a JTH PAW and a GBRV ultrasoft potential. The *T*-*P* plot in the inset indicates adiabatic temperature. In the NIF experiment, the temperature is uncertain (blue shaded area), and several profiles are possible depending on assumptions made, such as an intermediate-strength model (red dashed line) and low-strength model (yellow dashed line) for iron. Pink, purple, and green lines show the adiabatic *T*-*P* relation obtained with EPAW, JTH, and GBRV potentials, respectively. The bottom figure shows the predicted pressure difference,  $\Delta P=0$ , between that along the predicted adiabats and those predicted along NIF estimated temperature profiles [Zhuang *et al.*, 2019].

compute thermodynamic properties for weakly or strongly anharmonic systems in the thermodynamic limit. The alternative approach is thermodynamic integration, which carries uncertainties related to limited simulation cell sizes.

RESULTS & IMPACT

Applications of these methods on Blue Waters are illustrated here for two problems. The first is the calculation of  $\kappa_{\text{lat}}$  in Ca-SiO<sub>3</sub>-perovskite (Ca-Pv), an important phase (up to 10 vol%) of the Earth's lower mantle and a major phase in basaltic crust (approximately 25 vol%). Ca-Pv exists in the cubic structure in Earth's lower mantle. At low temperatures (*T* < approximately 700 K) it is mechanically unstable and becomes tetragonal. Calculations and measurements of its properties at relevant conditions (*T* > 2,000 K and *P* > 23 GPa) have been highly controversial because of its strongly anharmonic nature, preventing calculations of lifetimes using perturbation theory and of  $\kappa$  using BTE. Fig. 1 shows calculated (LDA) values compared to a single experimental datapoint obtained by collaborators (see Fig. 1 caption). Experimental verification of these results, still underway, is important be-

cause of the novelty of the approach and because Ca-Pv's  $\kappa_{\text{lat}}$  is three times larger than that of Mg-Pv, the major mantle phase. This result will change the estimated values of  $\kappa_{\text{lat}}$  in the lower mantle by up to 20% to 30%, with profound implications for convection style in the Earth, heat extraction from the core, and the age of Earth's inner core.

The second problem is the calculation of thermodynamic properties and high-*T* and adiabatic equations of state of  $\epsilon$ -iron, the solid phase of iron expected to exist in planetary cores. These calculations provided *T*-dependent phonon dispersion, which were subsequently employed to compute thermodynamic properties using a novel method [1] based on the phonon gas model. In metallic  $\epsilon$ -iron, electron thermal excitations and not anharmonicity are the main source of *T*-dependent phonon frequencies. These calculations were carried out up to 14 Mbar, a condition achieved only at the National Ignition Facility (NIF) at Lawrence Livermore National Laboratory. Temperature in the NIF data (Fig. 2) is expected to follow an adiabat. The research team's adiabatic equation of state agrees well with the NIF data up to 4 Mbar, deviating at higher pressures. This result suggests the unknown temperature in the NIF data might not quite follow an adiabat, with a cumulative effect causing deviations from adiabaticity in wide pressure ranges.

WHY BLUE WATERS

Such calculations require the execution of large numbers (approximately 10<sup>2</sup> to 10<sup>3</sup>) of HPC tasks (approximately 10 nodes each). The large number of Blue Waters' nodes allows the execution of independent tasks simultaneously. This is important for efficient execution of workflows with subsequent stages, each requiring the execution of large numbers of tasks that have dependencies across but are independent within different stages.

PUBLICATIONS & DATA SETS

G. Shukla, K. Sarkar, and R. M. Wentzcovitch, “Thermoelasticity of iron- and aluminum-bearing MgSiO<sub>3</sub> postperovskite,” *J. Geophys. Res.–Solid Earth*, vol. 124, pp. 2417–2427, 2019, doi: 10.1029/2018jb016379.

Z. Zhang, D. B. Zhang, T. Sun, and R. M. Wentzcovitch, “phq: A Fortran code to compute phonon quasiparticle properties and dispersions,” *Comput. Phys. Commun.*, vol. 243, pp. 110–120, 2019, doi: 10.1016/j.cpc.2019.05.003.

Z. Zhang *et al.*, “Thermal conductivity of CaSiO<sub>3</sub> perovskite at lower mantle conditions,” in preparation for *Science*, 2019.

J. Zhuang, H. Wang, Q. Zhang, and R. Wentzcovitch, “Thermodynamics properties of  $\epsilon$ -iron to 14 Mbar,” in preparation for *Phys. Rev. Materials*, 2019.



SIMULATING AEROSOL IMPACTS ON CLIMATE, ONE PARTICLE AT A TIME: A REGIONAL-SCALE, PARTICLE-RESOLVED AEROSOL MODEL TO QUANTIFY AND REDUCE UNCERTAINTIES IN AEROSOL-ATMOSPHERE INTERACTIONS

Allocation: Illinois/280 Knh  
PI: Matthew West<sup>1</sup>  
Co-PI: Jeffrey Curtis<sup>1</sup>

<sup>1</sup>University of Illinois at Urbana–Champaign

EXECUTIVE SUMMARY

This research aims to reduce key uncertainties in quantifying the impact of atmospheric aerosol particles on the Earth's climate. Aerosol particles can be brought into the atmosphere by a wide range of anthropogenic activities (resulting from the influence of human beings on nature) or by natural sources. They profoundly impact the large-scale dynamics of the atmosphere because they interact with solar radiation by scattering and absorbing light and by forming clouds. These impacts depend on the composition and size of the particles.

The uncertainties in quantifying these impacts originate from scale interactions and the high computational cost of modeling them. To tackle this problem, the research team developed the

particle-resolved 3D model WRF–PartMC, which has the unique ability to track size and composition information at a per-particle level. Particle-resolved simulations require efficient numerical algorithms and a computational resource with the capabilities of Blue Waters. Together, they allow for the ultrahigh-detail simulations needed to quantify the impact of aerosol particles on weather and climate at the regional scale.

RESEARCH CHALLENGE

Many of the greatest challenges in atmospheric modeling and simulation involve the treatment of aerosol particles, ranging from the prediction of local effects on human health [1] to understanding the global radiation budget via their indirect and direct effects [2]. Models provide important insights in the study of aerosols but experience a trade-off between the representation of physical detail and spatial resolution. Due to computational constraints, current models do not resolve individual particles and their microscale interactions. Instead, current methods of representing the high-dimensional and multiscale nature of aerosol populations apply large simplifications. While this makes computation much cheaper, it introduces unknown errors into model calculations. This has far-reaching consequences for the estimation of how aerosol particles impact regional and global climate, a topic of great societal relevance.

METHODS & CODES

To overcome the current limitations in representing aerosols and associated uncertainties, the particle-resolved model PartMC–MOSAIC [3] was coupled to the state-of-the-art 3D Weather Research and Forecast (WRF) model [4]. Aspects of these two models complement each other. The box model PartMC–MOSAIC is a highly detailed aerosol model that tracks the size and complex composition of individual particles in the atmosphere but is unable to resolve spatial heterogeneities of aerosol populations. The 3D regional WRF model is an advanced numerical weather model that captures the transport of chemical species in the atmosphere but assumes a crudely simplified aerosol representation. The resulting WRF–PartMC model uses a 3D Eulerian grid for the atmospheric flow, while explicitly resolving the evolution of individual aerosol particles per grid cell.

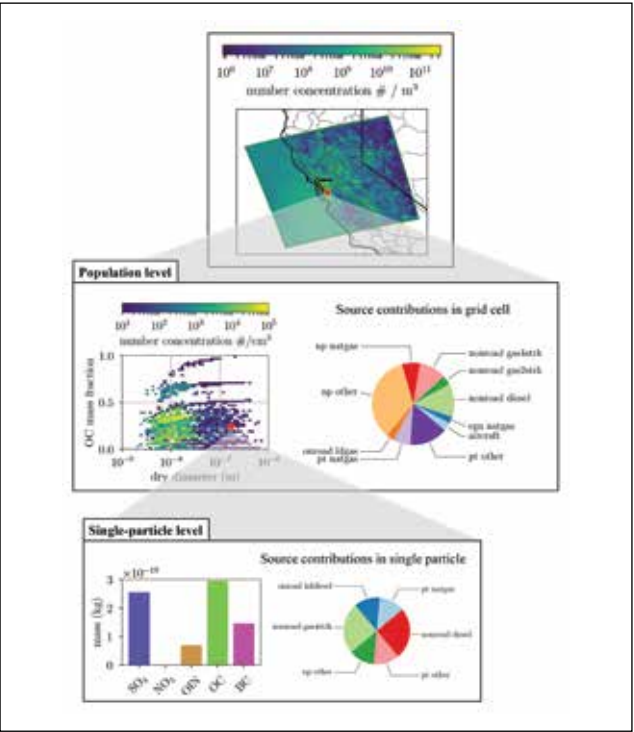


Figure 1: High-level detail obtained by WRF–PartMC: For any given simulated particle, the researchers store its chemical composition and the contributing sources. For each grid cell in the model domain, it is then possible to construct the full mixing state of the aerosol population as well as the full source apportionment profile.

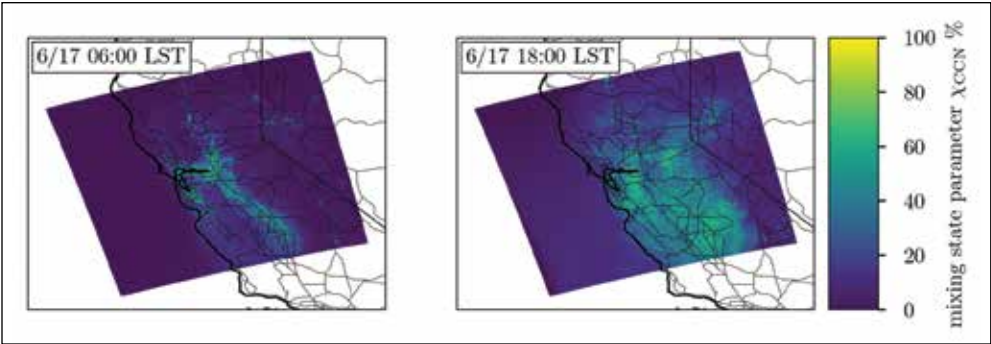


Figure 2: Horizontal distribution of mixing state parameter: A value of 100% indicates particles are fully internally mixed while 0% indicates fully externally mixed, two commonly applied assumptions in traditional models. WRF–PartMC is uniquely able to simulate a complex mixing state that varies both spatially and temporally.

RESULTS & IMPACT

Aerosol modeling is challenging because of the multiscale nature of the problem: The macroscale aerosol impact on climate is determined by microscale processes on the particle scale. The innovation of the WRF–PartMC model consists of representing many of these microscale processes explicitly on a per-particle level, which allows for an improved process-level simulation of the key interactions among aerosols, clouds, and radiation.

WRF–PartMC is the only model of its kind, and this research is changing the field of aerosol science because it provides the first benchmark for more approximate models commonly used in the field. It also provides a basis for rigorous coarse-graining to develop physically robust parameterizations for use in larger-scale models. By simulating at a much higher level of detail, particle-resolved models can help close the gap in understanding the effects of modeling choices in global models. Regional-scale particle-resolved simulations allow the quantification of the spatial heterogeneity that determines the conditions where highly detailed aerosol composition is necessary. This next-generation model captures the complex aerosol composition that current-generation models are unable to simulate.

The research team produced findings from a particle-resolved aerosol simulation for a realistic, spatially resolved three-dimensional domain in California, U.S.A. Aerosol and trace gas emissions were taken from the 2011 National Emission Inventory [5], and the meteorology corresponded to June 17, 2010, which coincides with the Carbonaceous Aerosol and Radiative Effects Study field campaign conducted during May–June 2010. On the order of 50 billion computational particles were tracked in this simulation, including their compositional changes owing to gas-to-particle conversion, their coagulation events, and their transport by both the wind and turbulence. The simulation was run on 6,636 cores. Most of the compute time was spent on particle coagulation and dynamic gas-particle partitioning on a per-particle basis.

Fig. 1 shows an example of the level of detail that can be obtained by WRF–PartMC for a simulation for California. In such a model run, the team is simulating the complex aerosol dynamics and chemistry for about five billion individual particles. For any given particle in the simulation, its chemical composition and the sources of the constituent particles that it is composed of (par-

ticles aggregate because they coagulate with each other during transport) are stored. Next, the full mixing state of the aerosol population as well as the full source apportionment profile can be constructed for each grid cell in the model domain.

The full aerosol particle state allows for investigation of the spatial and temporal distribution of the mixing state parameter [6]. Fig. 2 shows the mixing state parameter  $X_{ccn}$  at 06:00 LST (Local Standard Time) and 18:00 LST on June 17, 2010. The mixing state parameter  $X_{ccn}$  represents the extent that hydrophobic (tending to fail to mix with water) and hydrophilic (tending to mix with water) species are internally mixed. At 06:00 LST, these components remain mostly externally mixed with exceptions in the vicinity of urban areas and roadways. At 18:00 LST,  $X_{ccn}$  increased and is more spatially homogeneous. This indicates that the hydrophobic and hydrophilic material is becoming more internally mixed over the course of the day. The WRF–PartMC model allows for the full evolution of the aerosol mixing state, unlike traditional models. WRF–PartMC model results have been used to benchmark these more simplified models and evaluate the error incurred in models owing to simplifying assumptions.

WHY BLUE WATERS

Access to Blue Waters allows for a cutting-edge model formulation that pushes both science and computing by combining the large-scale features of state-of-the-art 3D models with the process-level physical representation of box models. Modeling 3D domains on the order of 100 billion tracked particles creates computational challenges owing to computationally intensive equations per particle and memory requirements to track high-dimensional particle composition. To enable simulations of aerosols at both a high spatial and compositional resolution, there is a need for tens of thousands of cores, fast interconnections among those cores, and sufficient memory per process.

Access to the Blue Waters staff was essential for addressing issues regarding the I/O challenge of outputting billions of particles, their chemical composition, and physical properties. Discussion and suggestions of how to change the model output code led to a two-orders-of-magnitude reduction in the time required to generate output. This removed output as a bottleneck to code performance.



EVOLVING AIR QUALITY UNDER THE CHANGING CLIMATE

**Allocation:** Illinois/589.958 Knh  
**PI:** Donald J. Wuebbles<sup>1</sup>  
**Co-PI:** Xin-Zhong Liang<sup>2</sup>  
**Collaborator:** Swarnali Sanyal<sup>1</sup>

<sup>1</sup>University of Illinois at Urbana–Champaign  
<sup>2</sup>University of Maryland

EXECUTIVE SUMMARY

This study investigates the effects of the changing twenty-first century climate on human health by evaluating projected effects on air quality exceedance events, focusing on particulate matter (PM<sub>2.5</sub>) and ozone (O<sub>3</sub>). The research team employed fully coupled global climate–chemistry modeling analyses using the Blue Waters system to simulate the historical and future time periods for multiple scenarios. The focus was on the United States, India, and China. The frequency of exceedance events increased in India for both scenarios and the resulting changes in climate, but the United States and China showed improvement in the lower-emissions scenario. The researchers also examined an ideal clean energy scenario, where mid-century fossil fuel emissions are reduced to zero. By eliminating the burning of fossil fuels, both PM<sub>2.5</sub> and O<sub>3</sub> concentrations reduce by 20% to 60% in high-pollution regions, greatly reducing future health risks.

RESEARCH CHALLENGE

Many studies have shown that projected climate change could affect air quality, but there is little known about the resulting effects on health. One way to look at health effects is to consider exceedances set by environmental policy for upper limits of

exposure. The objectives of this study are to better understand how global changes in climate and emissions will affect air quality, focusing on particulate matter and ozone; to project their future trends; to quantify key source attributions; and thus provide actionable information for environmental planners and decision makers to design effective dynamic management strategies, including local controls, domestic regulations and international policies to sustain air quality improvements in a changing world. The research team applied a state-of-the-science dynamic prediction system that couples global climate–chemical transport models to determine the individual and combined impacts of global climate and emissions changes from the present to 2050 under multiple scenarios.

The team has conducted three primary experiments using the dynamic prediction system: (1) historical simulations for the period 1990–2015 to establish the credibility of the system and refine the process-level understanding of U.S. regional air quality; (2) projections for the period 2030–2060 to quantify individual and combined impacts of global climate and emissions changes under multiple scenarios; and (3) sensitivity analyses to determine future changes in pollution sources and their relative contributions from anthropogenic and natural emissions, long-range pollutant transport, and climate change effects. The advanced state of the prediction system will produce a more complete scientific understanding of the challenges from global climate and emissions changes imposed on air quality management and a more reliable projection of future pollution sources and attribution changes.

METHODS & CODES

The research uses a state-of-the-science approach for advancing quantitative knowledge of the impacts of global changes in climate and emissions on U.S. air quality. The team used CESM1.2 default emissions, which represent surface emissions of approximately 30 species of speciated aerosols. The surface emission of each species is composed of all possible sources of emissions, including those from biomass burning, domestic sources, transportation, waste treatment, ships, industry, fossil fuels, and biofuels, and were composed from the POET, REAS, GFEDv2, and FINN emissions databases [1,2]. The results from the runs done using the current allocation have been presented at scientific conferences.

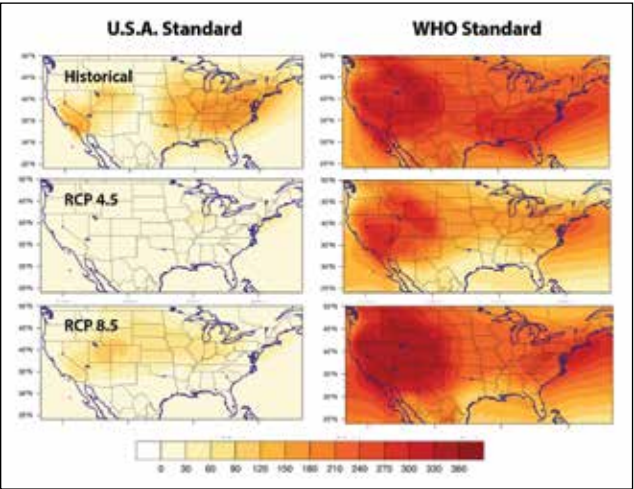


Figure 1: Total number of days annually when the 8-hour average ozone (O<sub>3</sub>) level exceeds the U.S. standard of 70 parts per billion (ppb) (left column) and exceeds 50 ppb (WHO standard) (right column) over the continental United States. This is for 1991–2014 (first row) and for 2031–2060 with the low (second row) and the high (third row) scenarios.

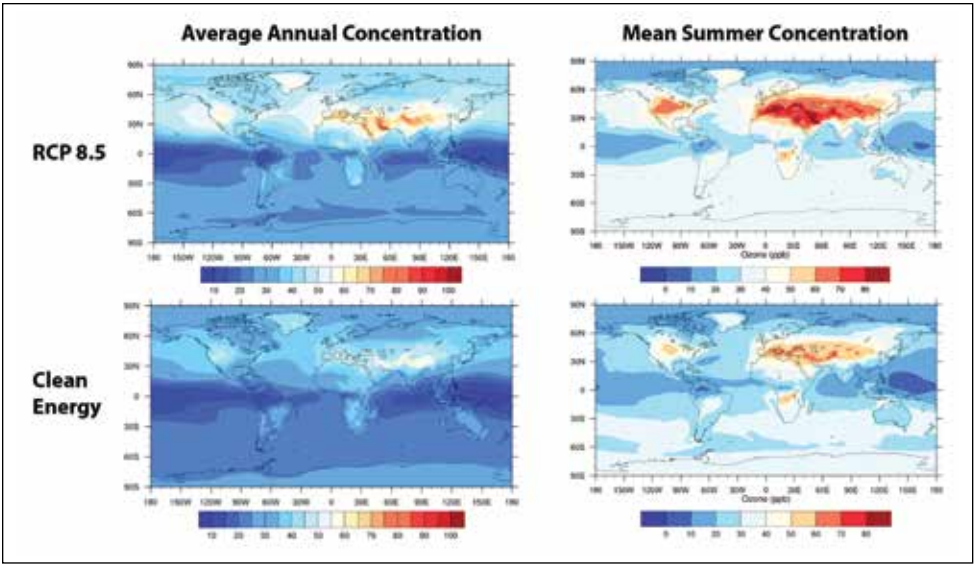


Figure 2: Annual average surface ozone concentration (2051–2060) (top panel) and summer average surface ozone concentration (bottom panel) for baseline scenario and the clean energy scenario (note the difference in the color scale).

RESULTS & IMPACT

The research team used the long-term climate chemistry runs done as a part of this project to examine exceedances for surface ozone and particulate matter concentration for two different climate projections using the lower emissions RCP4.5 and the higher emissions RCP8.5 scenarios. The results from two 30-year time periods in the future (2031–2060) were compared with the historical 25-year period (1990–2014). Cumulative distribution functions of surface ozone and particulate matter concentration and trend analysis of exceedance events annually and seasonally were analyzed over three major regions: the United States, India, and China as well as megacities within them. The results show that the frequency of exceedance events for ozone increases significantly at 90% confidence interval in India for both of the climate scenarios and in China for the high climate scenario, but decreases in the future in US especially western US and in China under RCP4.5 scenario. Along with the overall increase in ozone exceedance events, the study also showed a significant shift in seasonality of the events, with the number of episodes increasing during colder months, although ozone has primarily been considered as a summer problem. Unlike ozone, particulate matter concentration showed a significant increasing trend in all the regions in the future, with an overall increase in the number of particulate matter exceedance events annually.

As a part of the project, the research team also looked at a clean energy scenario and its impact on air quality in the mid-century. In this hypothetical scenario, emissions owing to fossil fuel are reduced to zero from 2050 in order to study the impact of a non-fossil-fuel-based energy system. The study showed a significant reduction in concentration of both ozone and particulate matter in all the current global hotspots, with a reduction of 20% to 60% in most high-pollution regions.

WHY BLUE WATERS

The computational demand of high-resolution climate models used in this project is extensive, particularly the fully coupled model of the Earth’s climate system with chemistry. Blue Waters, with its petascale-class computational resources, large number of nodes, and storage capability for the output from high-resolution model simulations, was essential for the project. Blue Waters’ staff have been critical in figuring out the various issues arising with the long-term, fully coupled climate chemistry runs using the Community Earth System Model (CESM). Staff members have also helped figure out and resolve various issues with the CESM1.2.2 models. In short, Blue Waters has given the team the computational resources, data management, and staff support to perform this research

PUBLICATIONS & DATA SETS

- S. Sanyal, “Evolving air quality and impact of climate change,” Ph.D dissertation, Dept. of Atmos. Sciences, University of Illinois at Urbana–Champaign, Urbana–Champaign, IL, U.S.A., 2019.
- S. Sanyal *et al.*, “Impact of changing climate on surface ozone exceedance events,” in preparation, 2019.
- S. Sanyal *et al.*, “Impact of changing climate on particulate matter exceedance events,” in preparation, 2019.
- S. Sanyal *et al.*, “Impact of clean energy on air quality in the mid-century,” in preparation, 2019.
- S. Sanyal and D. J. Wuebbles, “Particulate matter and ozone prediction and source attribution for U.S. air quality management in a changing world,” poster presented at the American Geophysical Union Fall 2018 Meeting; Washington, DC, U.S.A., Dec. 10–14.
- S. Sanyal and D. J. Wuebbles, “Particulate matter and ozone prediction and source attribution for U.S. air quality management in a changing world,” poster presented at the Midwest Student Conference on Atmospheric Research, Urbana, IL, U.S.A., Oct. 27–28, 2018.



SENSITIVITY OF ARCTIC SEA ICE THICKNESS DISTRIBUTION TO SEA ICE INTERNAL DYNAMICS IN A CHANGING CLIMATE

Allocation: Innovation and Exploration/678.5 Knh  
PI: Xiangdong Zhang<sup>1</sup>  
Co-PIs: Liran Peng<sup>1</sup>, Jing Zhang<sup>2</sup>

<sup>1</sup>University of Alaska Fairbanks  
<sup>2</sup>North Carolina A&T State University

EXECUTIVE SUMMARY

Changes in sea ice are a critical indicator of the climate system state. However, uncertainties exist in understanding, simulating, and predicting sea ice thickness distributions in the Arctic Ocean. In particular, the drastically thinned sea ice and more frequently occurring intense storms might have dramatically changed sea ice dynamic properties and air–ice momentum flux, raising further challenges to reducing the uncertainties. By conducting sensitivity experiments using the coupled sea ice–ocean component of the Community Earth System Model, the research team examined the interactive process between sea ice internal dynamics and thickness distribution. The results suggest that sea ice thickness distribution is highly sensitive to the treatment of its internal force in collaboration with air–ice momentum flux. A decrease in ice strength causes more energy conversion to potential energy, leading to an increase in the ridging process and thickness but a decrease in export via the Fram Strait. A decrease in air–ice momentum flux, however, demonstrates the opposite effect.

RESEARCH CHALLENGE

Arctic sea ice thickness is nonuniformly distributed in space, resulting from the complex interactive processes of dynamic and thermodynamic forcings across atmosphere, sea ice, and ocean interfaces and within sea ice itself. The realistic simulation and understanding of sea ice thickness distribution have been long-standing challenges. Along with amplified warming in the Arctic Ocean, dynamic and thermodynamic forcings across the atmosphere and sea ice interfaces have experienced pronounced changes. In particular, intense storms have more frequently occurred in the Arctic Ocean [3], causing even larger fluctuations or changes of atmospheric forcings on underlying sea ice. All of these further complicate the problem of how sea ice internal dynamics influence sea ice thickness distribution in the context of a changing climate and, in turn, contribute to the large-scale Arctic sea ice and climate system changes.

METHODS & CODES

The coupled sea ice–ocean component model of the National Center for Atmospheric Research’s Community Earth System Model was employed to conduct 25 sensitivity experiments with prescribed different sea ice strengths. The model experiments were initialized using the Polar Science Center hydrographic cli-

matology data [2] and forced by the monthly mean climatological forcing data constructed from the ERA-Interim reanalysis data set [1]. Each experiment covered a period of 100 years, allowing the sea ice and upper ocean to reach a quasi-equilibrium state.

RESULTS & IMPACT

Through examination and comparison of the results from the sensitivity experiments, the PI found that sea ice thickness distribution and sea ice motion are highly sensitive to perturbed sea ice strength prescribed in the model in collaboration with different air–ice momentum fluxes. Using a default sea ice strength defined as the ratio between total sea ice energy loss and potential energy changes, thick sea ice occurs along the Canadian Archipelago and greatly decreases toward the central Arctic Ocean and Eurasian Arctic shelf seas (Fig. 1). This pattern is similar to the observed sea ice thickness distribution, although the simulated sea ice is too thin in the central Arctic Ocean. At the same time, a basinwide, clockwise sea ice motion pattern appears with obviously large sea ice export via the Fram Strait. When sea ice strength decreases, sea ice thickness largely increases from the Canadian Archipelago to the Eurasian Arctic shelf seas. As a result, there is an increase in total sea ice volume for the entire Arctic Ocean throughout the year. All of these changes can be attributed to an enhanced conversion of kinetic energy to potential energy to build up sea ice ridges instead of frictional loss and a decreased sea ice export via the Fram Strait. A close examination also suggests that decreased sea ice strength causes a larger sea ice velocity.

To further investigate the upscaling impacts of the small-scale sea ice internal dynamics on shaping basinwide sea ice thickness distribution, the PI implemented tracers at the beginning of the modeling experiments with two of them in the Beaufort Sea: one in the East Siberian Sea and the other in the Laptev Sea. The PI then identified the pathways of the tracers, showing the origins of the tracked sea ice and its variation along the paths. The results indicate that decreased sea ice strength or increased air–ice momentum flux cause a clockwise rotation of the ice transpolar drift, resulting in a decrease in sea ice export via the Fram Strait and, therefore, an increase in the basinwide sea ice thickness. In contrast, counterclockwise rotation of the sea ice transpolar drift leads to less sea ice circulation and accumulation in the central/western Arctic, increasing sea ice export via the Fram Strait.

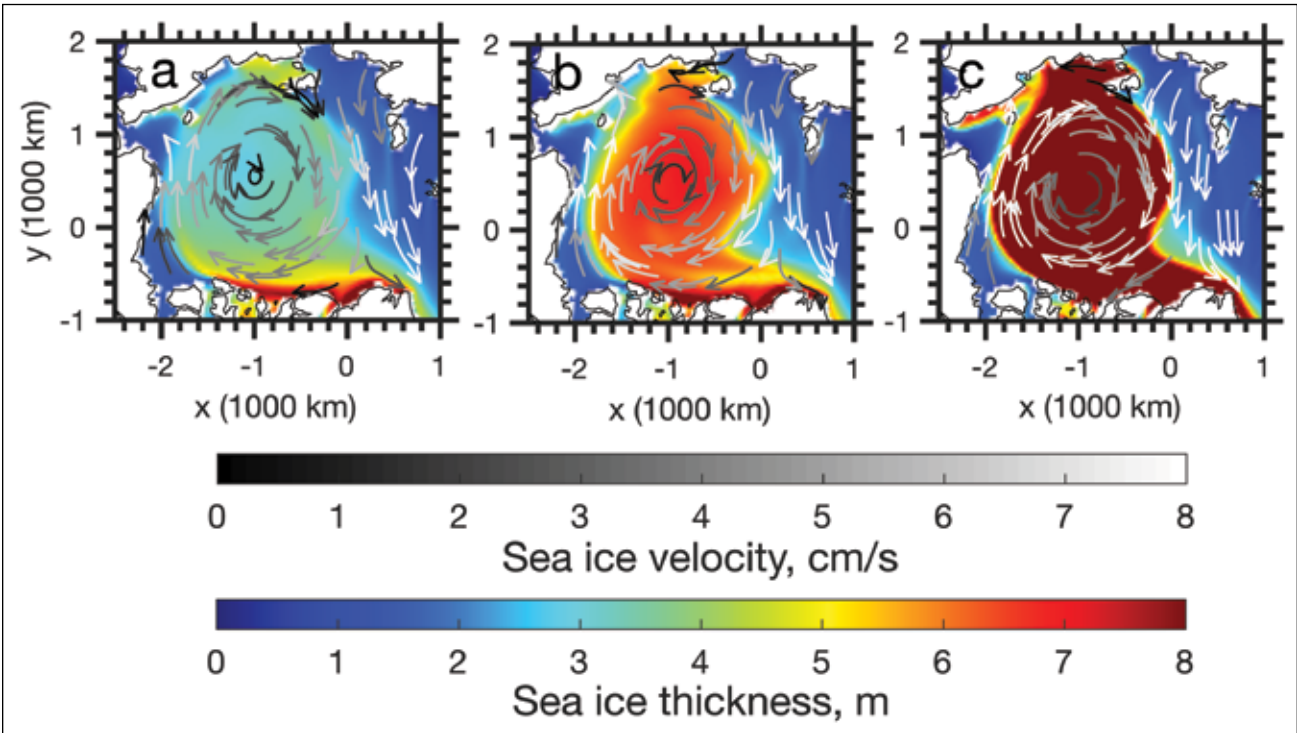


Figure 1: (a) Arctic sea ice thickness distribution and sea ice motion when the ratio between loss of total sea ice energy and change in sea ice potential energy (a nondimensional parameter) is defined as 17, which is a default value commonly used in ocean–sea ice climate modeling studies. (b) and (c) are the same as (a), but the sea ice strength is reduced to 60% and 20% of its default value, respectively. The sea ice strength represents the conversion ratio between kinetic energy and potential energy of sea ice due to its dynamic deformation, which is a measure of sea ice internal force (Peng and Zhang, 2019).

WHY BLUE WATERS

Blue Waters provided a unique opportunity to conduct modeling experiments at an ultrahigh resolution. Dynamic processes associated with sea ice thickness occur at small spatial and temporal scales. The only way to solve these problems with higher accuracy is through model simulation at superhigh resolutions. Furthermore, high resolution simulations also improve the understanding of upscaling impacts on basin scale sea ice thickness distribution. In addition, synoptic-scale intense storms and the resulting large fluctuation of forcings occur at small spatial and temporal scales. The state-of-the-art climate modeling studies generally do not take this into account. The Blue Waters system and staff paved the way for successful implementation of these model experiments.

PUBLICATIONS & DATA SETS

L. Peng *et al.*, “Impacts of intense Arctic storms on the melting process of sea ice in summer 2016,” to be submitted, 2019.  
L. Peng and X. Zhang, “Modeling study on sensitivities of Arctic sea ice thickness distribution to momentum flux and ice strength,” to be submitted, 2019.



# PHYSICS & ENGINEERING

FLUIDS

MATERIALS

NANOTECHNOLOGY

NUCLEAR ENGINEERING

NUCLEAR PHYSICS

PARTICLE PHYSICS

QUANTUM PHYSICS

SOLIDS & STRUCTURES

- 120

*The Mechanism of Proton Diffusion in ABO<sub>3</sub> Perovskite Oxides*
- 122

*Identification of Amino Acids with Sensitive Nanoporous MoS<sub>2</sub>: Toward Machine Learning-Based Prediction*
- 124

*Transfer-Learning-Based Coarse-Graining Method for Simple Fluids: Toward Deep Inverse Liquid-State Theory*
- 126

*High-End Visualization of Coherent Structures and Turbulent Events in Wall-Bounded Flows with a Passive Scalar*
- 128

*Design of Atomically Precise Nanoscale Negative Differential Resistance Devices*
- 130

*Using OpenMP Offloading to Run Code on Blue Waters' GPU Nodes*
- 132

*Numerical Investigation of Turbulence Suppression in Rotating Flows*
- 134

*An Efficient Method for Hypersonic Laminar-Turbulent Transition Prediction*
- 136

*Quantum Simulations: Properties of Dense Hydrogen*
- 138

*Role of Interfaces on the Shear Strength and Bending Properties of van der Waals Two-Dimensional Materials*
- 140

*Atypically Entangled Phases and New Methods for the Quantum Many-Body Problem*
- 142

*Direct Numerical Simulation of Pressure Fluctuations Induced by Supersonic Turbulent Boundary Layers*
- 144

*The Anomalous Magnetic Moment of the Muon: An Improved Ab Initio Calculation of the Hadronic Vacuum Polarization Contribution*
- 146

*Accelerating Thermoelectric Materials Discovery via Dopability Predictions*
- 148

*Machine-Learning Turbulence Models for Simulations of Turbulent Combustion*
- 150

*Turbulence-Resolving Modeling of Oscillatory Boundary Layer Flows*
- 152

*Machine Learning for Particle Physics: Employing Deep Learning for Particle Identification and Measurement at Colliders*
- 154

*Molten-Salt Reactors and Their Fuel Cycles*
- 156

*A Novel Crystal Structure with Spin-Protected Surface Electronic Conduction*
- 158

*Inertial Collapse of Individual Bubbles near Solid/Free Boundaries*
- 160

*Electronic Structure of Microscale Dielectric Barrier Discharges*
- 162

*Accelerating Virtual Prototyping and Certification in the Aerospace Industry with Scalable Finite-Element Analysis*
- 164

*Graphene Nanopore Transistor for DNA-Nick Detection*
- 166

*Compressibility Effects on Spatially Developing Plane Free Shear Layer*
- 168

*Outwardly Propagating Turbulent Flames*
- 170

*Deep Learning for Higgs Boson Identification and Searches for New Physics at the Large Hadron Collider*
- 172

*Designing Quantum Logic Gates on Silicon Chips with Large-Scale Multiphysics Simulations*
- 174

*Simulation of Rotating Detonation Engines*
- 176

*Mapping Proton Quark Structure: Looking Inside the Proton—How Do Quarks Spin?*
- 178

*Electron Dynamics of Ion-Irradiated Two-Dimensional Materials*
- 180

*Discovery of New Plasmonic Materials via High-Throughput Machine Learning*
- 182

*Turbulent Multiphase Thermal Flow Modeling of Defect Formation Mechanisms and Electromagnetic Force Effects in Continuous Steel Casting*
- 184

*Investigation of Sediment Transport Through Aquatic Vegetation Using Large-Scale High-Fidelity Turbulence Simulations*
- 186

*Machine Learning-Assisted High-Throughput Computational Design of Solvents for Liquid-Exfoliation*
- 188

*High-Throughput Materials Modeling Optimization*
- 190

*Detecting Neurotransmitters with DNA-Wrapped Nanotube Sensors*
- 192

*Accurate Effective Interactions in Quantum Materials*
- 194

*Supersonic Jet Noise Prediction Using High-Order Large-Eddy Simulation*
- 196

*Spin Spirals in Multiferroic Bismuth Ferrite and at Metal Surface: From Fully First Principles*
- 198

*Numerical Simulations of a Collapsing Cavitation Bubble Near an Elastically Deformable Object*
- 200

*Numerical Study on Shock Wave-Boundary Layer Interaction and Its Control*
- 202

*Free-Surface Flow Modeling of Multiple Tidal Turbines*
- 204

*New Insights on Intermittency and Circulation Statistics Obtained From a Massive Turbulence Simulation Database*
- 206

*Effects of Surface Defects on Hydrophobicity at Rare-Earth Oxide Interfaces Using Molecular Dynamics Simulations Driven By Ab Initio-Based Deep Neural Network Potentials*
- 208

*Constraining the Properties and Interactions of Dark Matter*



THE MECHANISM OF PROTON DIFFUSION IN ABO<sub>3</sub> PEROVSKITE OXIDES

Allocation: Strategic/500 Knh  
PI: Narayana R. Aluru<sup>1</sup>

<sup>1</sup>University of Illinois at Urbana–Champaign

EXECUTIVE SUMMARY

Perovskite oxides (ABO<sub>3</sub>) are well-known proton conductors. However, the role of the A-site ion on proton diffusion in perovskite oxides is not clear. By performing detailed density functional theory (DFT) calculations, the PI investigated the effect of A ion vacancy on proton transfer in yttrium-doped BaZrO<sub>3</sub>, which is one of the widely studied perovskite oxides.

The study showed that the presence of A ions reduces the barrier for proton diffusion in a perovskite oxide, demonstrating the significance of perovskite structures as proton conductors. This also showed that proton movement in a perovskite oxide is governed by hydroxide ion rotation and proton transfer from one oxygen to another. Both these motions are strongly coupled to lattice deformations. The PI identified the key physical mechanisms and the energy barriers associated with both hydroxide ion rotation and proton transfer by performing nudged elastic band (NEB) calculations. Finally, the PI calculated the bond strength and showed that the presence of an A ion can reduce the bond strength between O and B ions, thereby reducing the energy barrier for local lattice deformations.

RESEARCH CHALLENGE

Although the proton transport mechanisms in perovskite oxides have been explored [1–5], the mechanisms governing the lower activation energy of proton diffusion in perovskite oxides remain unclear. Therefore, the design and development of novel proton-conducting solid oxide electrolytes with high conductivity remain a significant challenge. In addition, another key question that has not been addressed regarding proton transport in perovskite oxides is the role of A-site ions on proton diffusion; *i.e.*, why do good proton conductors have the ABO<sub>3</sub> perovskite structure instead of, for example, the BO<sub>3</sub> structure with BO<sub>6</sub>-octahedrons?

Clear insights may be obtained by analyzing the role of A-site ions on the proton diffusion mechanism by considering the elementary steps underlying the proton transfer, hydroxide ion rotation, and the coupling of lattice dynamics to hydroxide ion rotation, which does not have a negligible barrier as shown in [6]. In this work, the investigator performed DFT calculations to address the two key questions above. This is the first systematic study on the origin of activation energy of proton transfer and hydroxide ion rotation, and the role of A-site ions, providing a deep understanding of the advantage of perovskite oxide as a proton conductor.

METHODS & CODES

All DFT calculations were performed using the Vienna Ab initio Simulation Package [7–9]. The Perdew–Burke–Ernzerhof [10] exchange–correlation functional was employed based on the projector augmented-wave method [9]. The cutoff energy for the plane-wave basis set was 500 eV for all calculations, which were nonspin polarized. The migration energy barriers were calculated using the climbing-image nudged elastic band method [11].

RESULTS & IMPACT

Based on DFT calculations, the origin of the activation energy of proton transfer and hydroxide ion rotation were revealed. Specifically, the outward O–B–O bending and A ion motions and hydroxide ion reorientation govern the hydroxide ion rotation process while the inward O–B–O bending motion and donor oxygen–proton–acceptor oxygen interactions govern the proton transfer process. The presence of A ions reduces the bond strength between O and B ions, thereby reducing the energy barrier for local lattice deformations such as the O–B–O bending motion. In addition, the presence of A ions decreases the bonding strength of protons with donor oxygen and increases the bonding strength of protons with acceptor oxygen, promoting proton motion from donor oxygen to acceptor oxygen. This work provides a thorough atomistic understanding of the role of A-site ions on proton diffusion in perovskite oxides and the results can enable design and discovery of novel materials with improved proton diffusion properties.

WHY BLUE WATERS

This project required large-scale *ab initio* simulations to obtain the origin of activation energy of proton diffusion in solid oxides. For the DFT calculations, using eight to 10 nodes (256 to 320 cores) for each job can achieve the best performance, which is attributed to the power of Blue Waters and the support of project staff. Thus, running *ab initio* simulations on Blue Waters was easy and quick, speeding up the research greatly.

PUBLICATIONS & DATA SETS

Y. Jing and N. R. Aluru, “The role of A-site ion on proton diffusion in perovskite oxides (ABO<sub>3</sub>),” *J. Power Sources*, vol. 445, p. 227327, Jan. 2020.

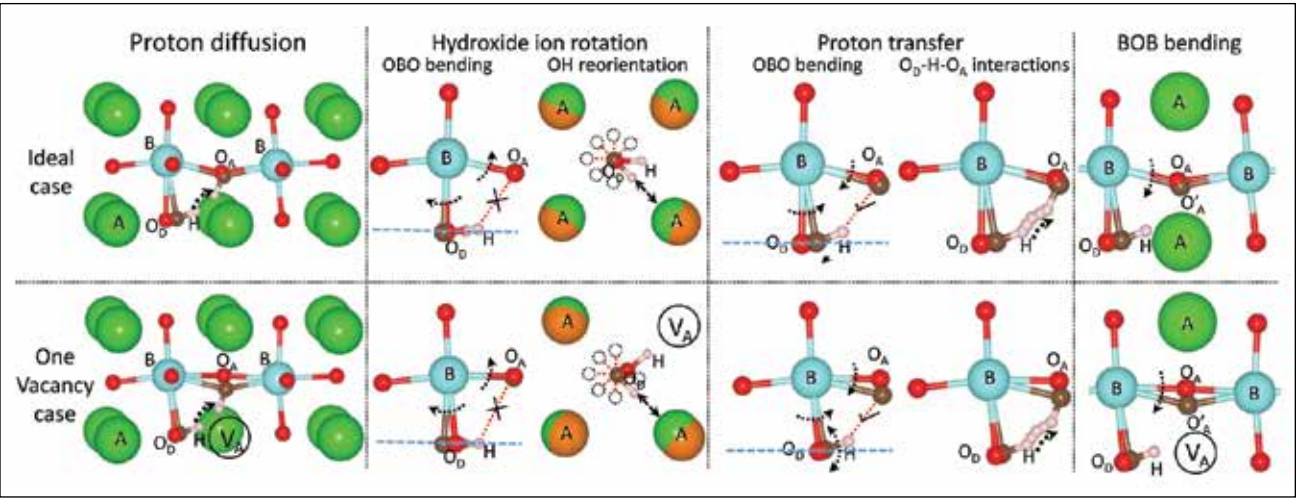


Figure 1: Schematic illustration of the proton diffusion mechanism in an ABO<sub>3</sub> perovskite oxide without (top row) and with (bottom row) one A ion vacancy.



IDENTIFICATION OF AMINO ACIDS WITH SENSITIVE NANOPOROUS MoS<sub>2</sub>: TOWARD MACHINE LEARNING-BASED PREDICTION

Allocation: Illinois/135 Knh  
PI: Narayana R. Aluru<sup>1</sup>

<sup>1</sup>University of Illinois at Urbana–Champaign

EXECUTIVE SUMMARY

Identifying a chain of amino acids can enable breakthrough advances in early diagnosis of disease and the health status of the human body. Many diseases, including cancer, diabetes, and digestive disorders, are caused by malfunctioning ribosomes leading to defective proteins. Therefore, sequencing an amino acid chain helps diagnose diseases at early stages. In this study using petascale-based molecular simulations with a total aggregate simulation time of 66 microseconds ( $\mu s$ ), the re-

searcher demonstrated that a nanoporous single-layer molybdenum disulfide ( $MoS_2$ ) can detect individual amino acids in a polypeptide chain with high accuracy and distinguishability. With the aid of machine learning techniques, the PI featurized and clustered the ionic current and residence time of the 20 amino acids and identified the fingerprints of the signals. In addition, using advanced machine learning classification techniques, the PI was able to predict the amino acid type of over 2.8 million hypothetical sensors.

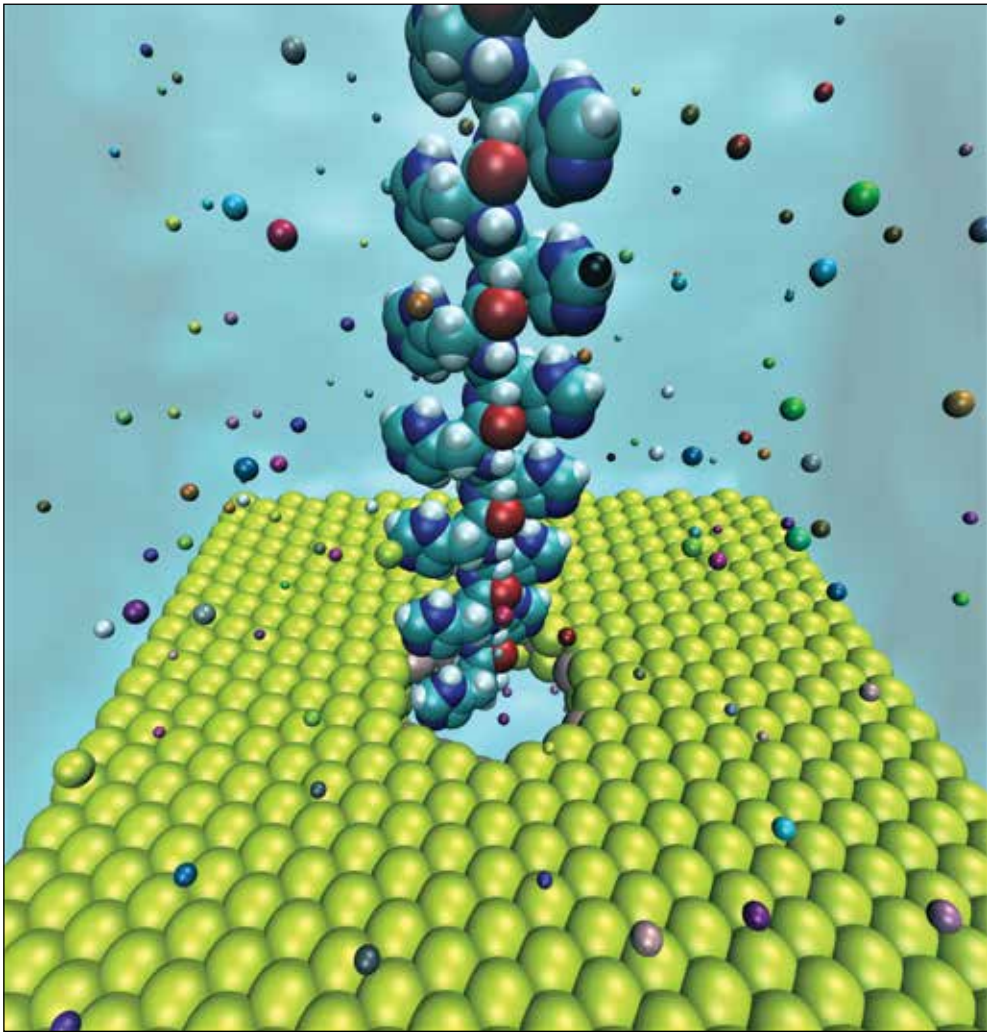


Figure 1: A snapshot of proline polypeptide translocation through the  $MoS_2$  nanopore.

RESEARCH CHALLENGE

DNA sequencing using nanopore technology has evolved significantly during the last few years. Oxford Nanopore Technologies Ltd. is currently fabricating a USB stick-sized device that can sequence DNA in several hours. In recent years, both biological and synthetic nanopores have been used for “label-free,” high-resolution DNA sequencing. In addition to DNA sequencing, detection of proteins can lead to advances in improving the health status of the human body.

The challenges posed by biological molecule detection using nanopore technology are the low signal-to-noise ratio, pore degradation owing to multiple uses, the identification of single bases in real time, and the high speed of translocation [1,2]. Engineering the translocation of molecules through biological/synthetic nanopores has been defined as one of the challenging problems of biotechnology. This project has shown that an ultrathin  $MoS_2$  nanopore is capable of detecting and identifying all 20 standard amino acids in proteins by using extensive molecular dynamics simulations.

METHODS & CODES

Molecular dynamics (MD) simulations were performed using the large-scale atomic/molecular massively parallel simulator (LAMMPS). LAMMPS is an open source classical MD code for simulation of liquid, solid, and gas phases. The LAMMPS-based simulations involved three different interatomic potentials: Tersoff potential, Lennard–Jones potential, and long-range Coulombic via particle–particle particle–mesh. Each simulation box contained about 32,000 atoms consisting of a monolayer  $MoS_2$ , an amino acid chain, water molecules, and ions. The amino acid chain was pulled through the nanopore using an external force. Fig. 1 shows a proline chain translocating through the nanopore of  $MoS_2$ .

RESULTS & IMPACT

This study has shown that a single-layer  $MoS_2$  nanopore can detect individual amino acids in a polypeptide chain with high accuracy and distinguishability. Using extensive MD simulations (with a total aggregate simulation time of 66  $\mu s$ ) the ionic current and residence times of each residue of amino acid types was characterized and featurized. The amino acids were clustered into different groups based on their physical properties (e.g., size, polarity, and hydrophobicity). In addition, the type of amino acid was

classified using machine learning techniques for any future ionic current and residence time sensor readings. Logistic regression, nearest neighbor, and random forest machine learning classifiers resulted in the prediction of amino acid types with an accuracy of 72.45%, 94.55% and 99.6%, respectively.

Identifying protein chains is necessary for diagnostic purposes and early-stage detection of cancer and other diseases. In fact, the data acquired from proteomic fingerprints can be as crucial as the genome in defining the health status of humans. The proposed high-precision, single-base resolution, and fast biological molecule sequencing using nanopore technology can lead to fabrication of inexpensive personal health diagnostic devices, improving the health status of individuals. This will enable the rapidly emerging fields of predictive and personalized medicine and will mark a significant leap forward for clinical genomics and proteomics.

WHY BLUE WATERS

This project involved 4,103 extensive MD simulations with up to 50,000 atoms and an aggregate simulation time of 66  $\mu s$ . These expensive computations would not have been possible without a petascale supercomputer. Also, the LAMMPS MD package scales almost linearly with the number of cores up to 100 on Blue Waters.

PUBLICATIONS & DATA SETS

A. B. Farimani, M. Heiranian, and N. R. Aluru, “Identification of amino acids with sensitive nanoporous  $MoS_2$ : Towards machine learning-based prediction,” *npj 2D Mater. Appl.*, vol. 2, no. 1, May 2018.



TRANSFER-LEARNING-BASED COARSE-GRAINING METHOD FOR SIMPLE FLUIDS: TOWARD DEEP INVERSE LIQUID-STATE THEORY

Allocation: Illinois/377.8 Knh  
PI: Narayana R. Aluru<sup>1</sup>

<sup>1</sup>University of Illinois at Urbana–Champaign

EXECUTIVE SUMMARY

Machine learning is an intriguing method to circumvent difficulties faced with the development and optimization of force field parameters. In this project, a deep neural network (DNN) was used to solve the inverse problem of the liquid-state theory—particularly, to find the relationship between the radial distribution function (RDF) and the Lennard–Jones (LJ) parameters at various thermodynamic states. Using molecular dynamics (MD), many observables, including RDF, are uniquely determined by specifying the interatomic potentials. However, the inverse problem (*e.g.*, determining the potential using a specific RDF) is not feasible through MD simulations unless it is combined with a complementary method.

In this project, the PI developed a framework integrating DNN with 1.5-terabyte MD trajectories (26,000 distinct systems and a total simulation time of 52 microseconds) to predict LJ parameters. The results show that DNN was successful not only in parameterization of the atomic LJ liquids but also in parameterizing the potentials for coarse-grained models of multiatom molecules.

RESEARCH CHALLENGE

The Lennard–Jones (LJ) potential form is one of the widely used interatomic potentials to study atomistic systems [1]. By specifying the potential parameters and the thermodynamic state, MD can compute various quantities of interest such as the radial distribution function (RDF). However, given a specific RDF, MD

cannot directly predict the potential parameters; *i.e.*, the inverse problem is considered a difficult task [2]. As per Henderson’s theorem [3], the relationship between the pair potential and RDF is unique at a given thermodynamic state, implying that the potential parameters can be determined uniquely based on the RDF. In this work, the PI explored the feasibility of force field development based on Henderson’s theorem with a data-driven and deep learning-based approach.

This inverse problem can also be viewed as a coarse-graining problem where the objective is to develop a pair-potential between coarse-grained particles such that the RDF of the original system is reproduced. While different frameworks such as the fundamental measure theory and integral equations have been developed to address this problem, both accuracy and generalizability to more complex systems are still issues. Alternatively, reproducing a given RDF relies on MD to refine the potential parameters. For example, MD data are either integrated with a theoretical framework (*e.g.*, iterative Boltzmann inversion) or used to optimize statistical errors with given RDFs (*e.g.*, simplex and relative entropy). These approaches require thousands of simulations for a specific system and the data are often not reused. The main bottleneck in reusing data to parameterize a new system originates from the complexity of physics-based model development and long-term storage of the data. To overcome these issues, recent data-driven approaches, known as the physics of big data [4], use surrogate models.

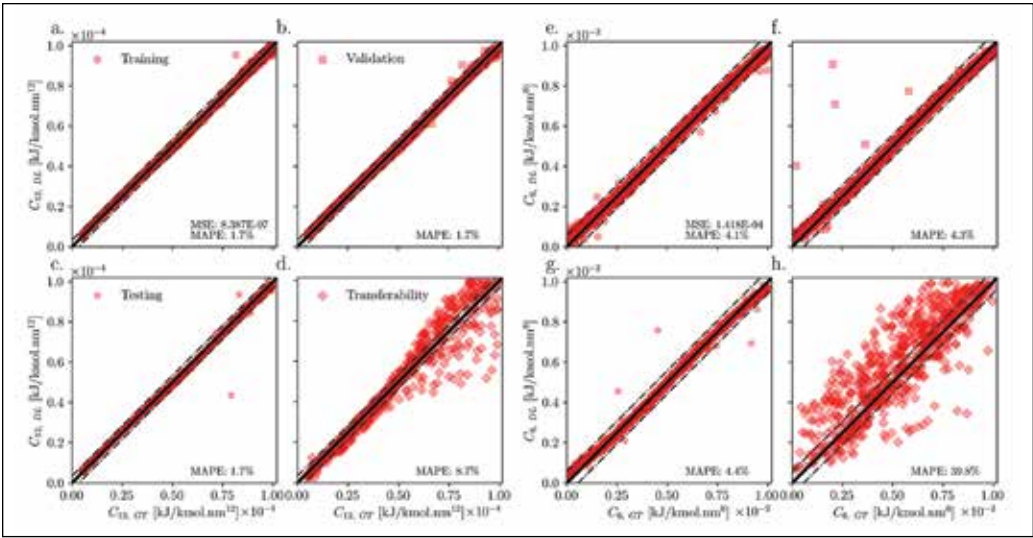


Figure 2: Predictive capability of the model.

METHODS & CODES

The PI performed MD simulations using the GRONingen Machine for Chemical Simulations package (GROMACS). GROMACS uses the well-known LJ potentials with the Verlet algorithm. The deep learning model development was performed using TensorFlow on Blue Waters. Fig. 1c shows the workflow of the data generation, training, inference, and assessment phases.

RESULTS & IMPACT

The PI assessed the performance of DNN by considering two cases. First, he investigated the generalizability and transferability of the interatomic potential for LJ particles (development of atomistic force fields for single-atom particles). Second, he considered transfer learning for coarse-grained (CG) force-field development. Generalizability points to using DNN to predict potential parameters for LJ particles for a given thermodynamic data set. Transferability refers to using DNN to estimate potential parameters with thermodynamic states outside the training data set. Transfer learning refers to the use of DNN to estimate potential parameters for CG representation of simple multiatom molecules.

Fig. 2 compares the predicted DNN and the ground truth potential parameters used in the MD simulations. All the points are distributed almost uniformly around the one-to-one mapping line (the line on which the ground truth and DNN parameters match exactly). While the thermodynamic states and RDFs vary for each point, the DNN is able to relate them correctly to the underlying potential parameters. The DNN results for the LJ particles exhibit no more than 4.4% mean absolute percentile error over the data set, which implies that the DNN is an efficient approach to solve the inverse problem of the liquid-state theory for the LJ particles.

To investigate the transferability as a new coarse-graining route, the PI developed single-bead CG models of simple multiatom

molecules such as carbon monoxide, fluorine, and methane. The center of mass RDFs of these systems were first obtained using all-atom MD simulations. Then, the center of mass RDFs and thermodynamic states were fed into the DNN, as shown in Fig. 1. The results indicated that DNN force fields are indistinguishable from the other available methods. Considering that DNN is a single-shot method, its speed in deriving the CG force field is faster compared to the simplex and relative entropy methods. Following the procedure shown in Fig. 1, the PI assessed the accuracy of CG models with two additional metrics: total deviation in the RDFs and Kullback–Leibler (KL) divergence. Both error metrics show a small deviation from the DNN-predicted LJ parameters with a distance less than 0.1% of the maximum error. As DNN does not have prior knowledge about the information theory (KL metric) or the statistical mechanics metric (error in the total variation of the RDF), the PI concluded that deep learning is a good coarse-graining strategy as it performs well on both metrics.

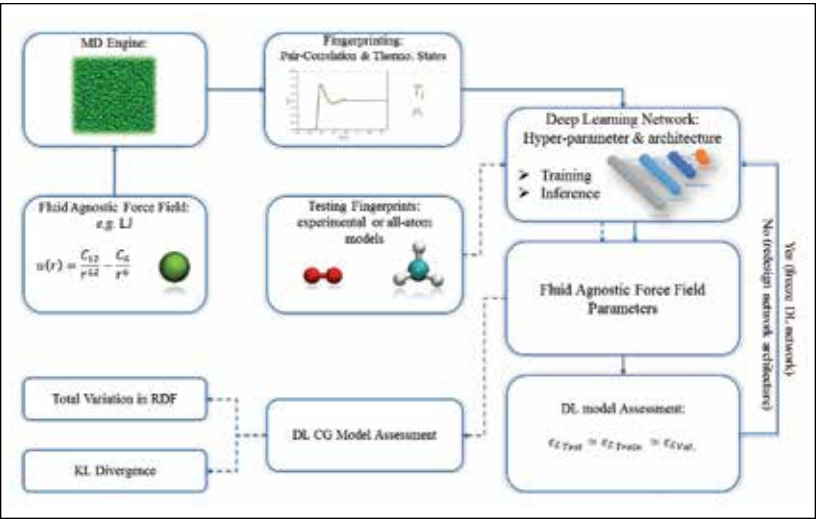
WHY BLUE WATERS

This project involved 40,000 MD simulations of up to 10,000 atoms with a total simulation time of 52 microseconds. The feasibility of such computations is a result of the petascale resources and GROMACS’ linear scaling on Blue Waters. Similarly, deep learning training using the TensorFlow module of Python relies heavily on access to the XK nodes on Blue Waters. Various neural network architectures were tried on Blue Waters (over 10 million training iterations) to select the optimized network and to avoid overfitting and underfitting.

PUBLICATIONS & DATA SETS

A. Moradzadeh and N. R. Aluru, “Transfer-learning-based coarse-graining method for simple fluids: toward deep inverse liquid-state theory,” *J. Phys. Chem. Lett.*, vol. 10, no. 6, pp. 1242–1250, Feb. 2019.

Figure 1: Workflow of the current study. An MD engine (GROMACS) generated big data that is used in the training of a deep learning model.





DI

TN

HIGH-END VISUALIZATION OF COHERENT STRUCTURES AND TURBULENT EVENTS IN WALL-BOUNDED FLOWS WITH A PASSIVE SCALAR

**Allocation:** Innovation and Exploration/90.25 Knh  
**PI:** Guillermo Araya<sup>1</sup>  
**Co-PIs:** Guillermo Marin<sup>2</sup>, Fernando Cucchietti<sup>2</sup>

<sup>1</sup>University of Puerto Rico, Mayaguez  
<sup>2</sup>Barcelona Supercomputing Center

EXECUTIVE SUMMARY

Despite the chaotic behavior of turbulence, investigations performed during the last six decades have conclusively demonstrated the presence of organized motions called “coherent structures” in turbulent boundary layers, which are responsible for transporting most of the turbulent kinetic energy. Investigations have also elucidated passive scalars, which are defined as diffusive contaminants that exist in a low concentration but sufficiently enough to provoke a significant impact on energy expenditures, heat transfer, and air pollution. Research has also found that the transport phenomenon in real-situation flows usually occurs under complicated external conditions such as favorable/adverse pressure gradients, local flow perturbations, and spatially developing boundary layers. Therefore, computational investigation through Direct Numerical Simulation (DNS) with millions of “flow and thermal sensors” and high temporal resolution may shed light on the unknown aspects of transport phenomena in accelerated/decelerated boundary layers. In addition, coherent structures in such a complex environment and their interactions (turbulent events) can be better identified and visualized by DNS.

RESEARCH CHALLENGE

These research efforts make use of tremendous computational resources, not only during the running stage but also in the visualization–animation stage. Therefore, state-of-the-art parallel computing and graphics processing unit (GPU) programming are essential. Furthermore, the high spatial/temporal resolution of DNS examines the physics behind turbulent events and pas-

sive scalar transport in highly accelerated/decelerated boundary layers for potential applications to flow/heat transfer control and turbulence modeling in aerospace applications.

The research team’s DNS study uses thousands of cores in a parallel computational environment. DNS is a numerical tool that aims to resolve all turbulent length and time scales, capturing the whole energy spectrum of the flow field. Consequently, the numerical approach requires high mesh resolution and, thus, very small physical timesteps or high temporal resolution. This results in highly costly numerical predictions with the payoff of getting the whole “picture” of the problem: no other numerical approach can supply such a level of information and accuracy in turbulent wall-bounded flows. Even experimental approaches fail to measure transport phenomena in the near-wall region of boundary layers. DNS is able to accurately predict flow parameters in regions up to one hundred times closer to the surface than in experiments. Traditionally, DNS has been limited to small computational domains (or low Reynolds numbers); however, the significant growth of petascale computing resources such as Blue Waters has allowed researchers to tackle higher Reynolds number problems in a reasonable time. To the best of the researchers’ knowledge, this project is the first time that numerical predictions of high-Reynolds-number boundary layers subject to extreme streamwise pressure gradients have been carried out, representing a formidable computational challenge with tremendous importance for the fluid dynamics and scientific visualization communities.

Figure 1: (left) Schematic of the spatially developing boundary layer in FPG flows with Inlet, Recycle, and Test planes. (right) Vortical structures emanating from the crossflow jet system.

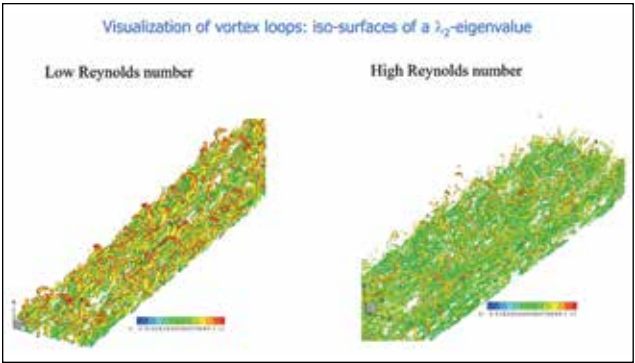
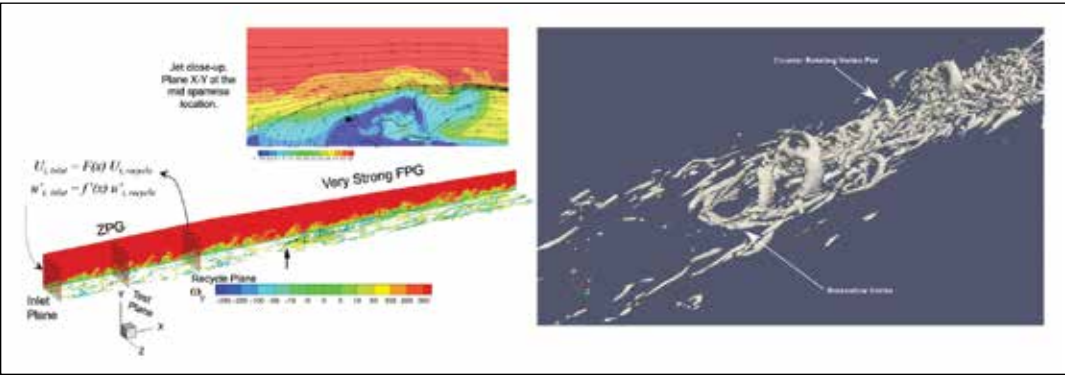


Figure 2: Iso-contours of a  $\lambda_2$  eigenvalue in low and high Reynolds numbers of zero-pressure gradient flows.

METHODS & CODES

Computationally speaking, it is very challenging to capture the physics of unsteady spatially developing turbulent boundary layers owing to the following requirements: (1) the high resolution necessary to resolve the small scales, (2) a computational box that is large enough to appropriately capture the influence of the large-scale motions, and (3) the realistic time-dependent inflow turbulent conditions that must be prescribed. Therefore, the research team used the inflow generation method devised by Araya *et al.* [1], which is an improvement on the original rescaling–recycling method of Lund *et al.* [2]. The seminal idea of the rescaling/recycling method is to extract the flow solution (mean and fluctuating components of the velocity and thermal field) from a downstream plane (called “recycle”) and after performing a transformation by means of scaling functions, the transformed profiles are reinjected at the inlet plane, as seen in Fig. 1.

To successfully perform the DNS, a highly accurate, efficient, and highly scalable computational fluid dynamics solver is required. PHASTA is an open-source, parallel, hierarchic (second- to fifth-order accurate), adaptive, stabilized (finite-element) transient analysis tool for the solution of compressible [3] or incompressible flows [4]. Combining minimal dissipation numerics and adaptive unstructured meshes, PHASTA has been applied to flows ranging from validation on DNS and large-eddy simulation benchmarks such as channel flow and decay of isotropic turbulence to cases of practical interest including incompressible and compressible boundary-layer flow control. PHASTA has been carefully constructed for parallel performance and scaling to 786,432 cores (on one, two, and four processes per core, which exceeds three million processes) in Mira (Argonne’s 10 petaflop supercomputer). PHASTA has also been ported and scaled well on GPU-based and Intel® Xeon Phi™-based machines.

RESULTS & IMPACT

This project, based on DNS big data, contributed to a better understanding of coherent structures, turbulent events, and passive scalar transport in boundary layers subject to streamwise pressure gradients and vertical jets, and pushed the boundaries

in terms of high-end visualization. It has also led to the improvement of flow control tools for mixing enhancement, drag reduction, and heat-transfer augmentation in aerospace applications such as the film cooling technique. Fig. 1 depicts iso-contours of the Q-criterion [5] in order to visualize coherent structures emanating from the crossflow jet. It has been observed that the counter-rotating vortex pair system experienced a quick attenuation owing to the strong flow acceleration prescribed (Favorable Pressure Gradient). Furthermore, DNS of high Reynolds number zero-pressure gradient flows has been performed in large-scale systems (approximately 52 million grid points). Fig. 2 shows iso-contours of a  $\lambda_2$  eigenvalue for coherent structure visualization in low and high Reynolds number flows. The presence of vortex loops in low Reynolds number flows (horseshoes) is evident, while the turbulence structure seems finer and more isotropic in high Reynolds number flows (hairpins). Since the lead institution of this project is the University of Puerto Rico at Mayaguez, the research impact has involved underrepresented minorities: six undergraduate students, two graduate students, and one early-career faculty.

WHY BLUE WATERS

This project makes use of state-of-the art DNS, which supplies the highest possible level of flow information but at a high computational cost. This endeavor would not be possible without the invaluable contribution and support of the Blue Waters resources.

PUBLICATIONS & DATA SETS

- G. Araya, G. Marin, F. Cucchietti, I. Meta, and R. Grima, “Visualization of a jet in turbulent crossflow,” in *High Performance Computing, 5th Latin American Conference*, Bucaramanga, Colombia, Sept. 27, 2018, pp. 174–178, doi: 10.1007/978-3-030-16205-4\_13.
- G. Araya and G. Torres, “Structural Reynolds analogy in laminar boundary layers via DNS,” *J. Vis.*, vol. 22, no. 3, pp. 529–540, June 2019, doi: 10.1007/s12650-019-00549-6.
- G. Saltar and G. Araya, “Turbulence modeling of boundary layers subject to very strong Favorable Pressure Gradient (FPG) with passive scalar transport,” in *Proc. 4th Thermal and Fluids Engineering Conf.*, Las Vegas, NV, U.S.A., Apr. 15, 2019, pp. 1819–1833, doi: 10.1615/TFEC2019.tfl.028426.
- J. Santiago, G. Araya, G. Marin, and F. Cucchietti, “Symbiosis of quasi-streamwise vortices and low-speed streaks in laminar boundary layers,” presented at the 71th Annual Meeting of the American Physical Society—Division of Fluid Dynamics, Atlanta, GA, U.S.A., November 18–20, 2018.
- G. Saltar and G. Araya, “Coherent structure assessment in crossflow jets subject to strong favorable pressure gradient (FPG),” presented at the 71th Annual Meeting of the American Physical Society—Division of Fluid Dynamics, Atlanta, GA, U.S.A., November 18–20, 2018.



DESIGN OF ATOMICALLY PRECISE NANOSCALE NEGATIVE DIFFERENTIAL RESISTANCE DEVICES

**Allocation:** NSF PRAC/4,200 Knh  
**PI:** Jerzy Bernholc<sup>1</sup>  
**Co-PIs:** Emil L. Briggs<sup>1</sup>, Miroslav Hodak<sup>1</sup>, Carl T. Kelley<sup>1</sup>, Wenchang Lu<sup>1</sup>  
**Collaborator:** Zhongcan Xiao<sup>1</sup>

<sup>1</sup>North Carolina State University at Raleigh

EXECUTIVE SUMMARY

Downscaling device dimensions to the nanometer range raises significant challenges to traditional device design owing to potential current leakage across nanoscale dimensions and the need to maintain reproducibility while dealing with atomic-scale components. The research team has investigated negative differential resistance (NDR) devices based on atomically precise graphene nanoribbons. This computational evaluation of the traditional double-barrier resonant tunneling diode NDR structure uncovers important issues at the atomic scale concerning the need to minimize the tunneling current between the leads while achieving high peak current.

The team has proposed a new device structure consisting of multiple short segments that enables high current by the alignment of electronic levels across the segments while enlarging the tunneling distance between the leads. The proposed structure can be built with atomic precision using a scanning tunneling microscope (STM) tip during an intermediate stage in the synthesis of an armchair nanoribbon. In addition, the team has conducted an experimental evaluation of the band alignment at the interfaces and an STM image of the fabricated active part of the device.

RESEARCH CHALLENGE

Designing band alignment to manipulate electronic transport behaviors across an interface is the key to achieving novel functionalities in semiconductor junctions and heterostructures (HSs). The recent development of graphene and graphene nanoribbons (GNRs) offers new opportunities to design nanoscale devices and to test NDR at the atomic scale. Following the bottom-up synthesis of atomically precise GNRs, HSs based on GNRs with sub-nanometer widths and various types of band alignment were designed and fabricated. In particular, the controllable polymer-to-GNR conversion reaction was demonstrated using charge injection from an STM tip. This advance has enabled the creation of atomically precise HSs and devices based on single ribbons.

In the simplest resonant tunneling diode (RTD) configuration, a quantum dot is separated from the leads by barriers, leading to confined electron level(s). When the bias initially increases, the source Fermi level moves closer to the confined level, leading to a current increase. At a certain bias, resonant transmission is achieved, and the current reaches a maximum. As the bias increases further, the source Fermi level moves above the reso-

nance and the current decreases. This results in an NDR region. When the bias increases further, the source Fermi level may approach another confined level and the current increases again. For an atomic-scale device, the conventional RTD design needs modification to become practical. First, the small size of the segments results in strong confinement, which limits the number of tunneling levels available at moderate voltages. Second, if the segments that are chosen are short, direct tunneling between leads may occur and thus wash out or eliminate the NDR. Conversely, if the segments are long enough to suppress direct tunneling, electron transmission across these regions with large electronic gaps results in a very small current, which may render the device impractical.

METHODS & CODES

The electronic structure calculations used the real-space multi-grid (RMG) and plane-wave self-consistent field codes. The quantum transport nonequilibrium Green's function calculations used the localized orbitals and nonequilibrium Green's function branches of the RMG code.

RESULTS & IMPACT

In this work, the research team has demonstrated a practical device structure based on armchair GNRs to deliver a strong NDR effect. The proposed GNR-based HSs consist of seven-carbon-atom-wide armchair GNRs (7-aGNRs) and an intermediate structure appearing in GNR synthesis. This intermediate structure consists of partially converted GNRs with one side of the polyanthrylene converted to the GNR structure while the other side remains in the polymeric structure. The GNR/intermediate HSs have been confirmed to have a type-I band alignment, and thus can be employed to design NDR devices. After illustrating the key issues of NDR device design at the atomic scale, the team has proposed an unconventional multipart device comprised of five short segments, which leads to a pronounced NDR with a current large enough for practical use. Such a multipart HS can be experimentally fabricated by using *in situ* growth from polyanthrylene precursors and STM manipulations, ensuring an atomically precise device with well-defined, reproducible characteristics.

The team uses the GNR/intermediate junctions and undoped bulk graphene as paradigmatic probes and first considers the conventional quantum dot device model. In the simple GNR/in-

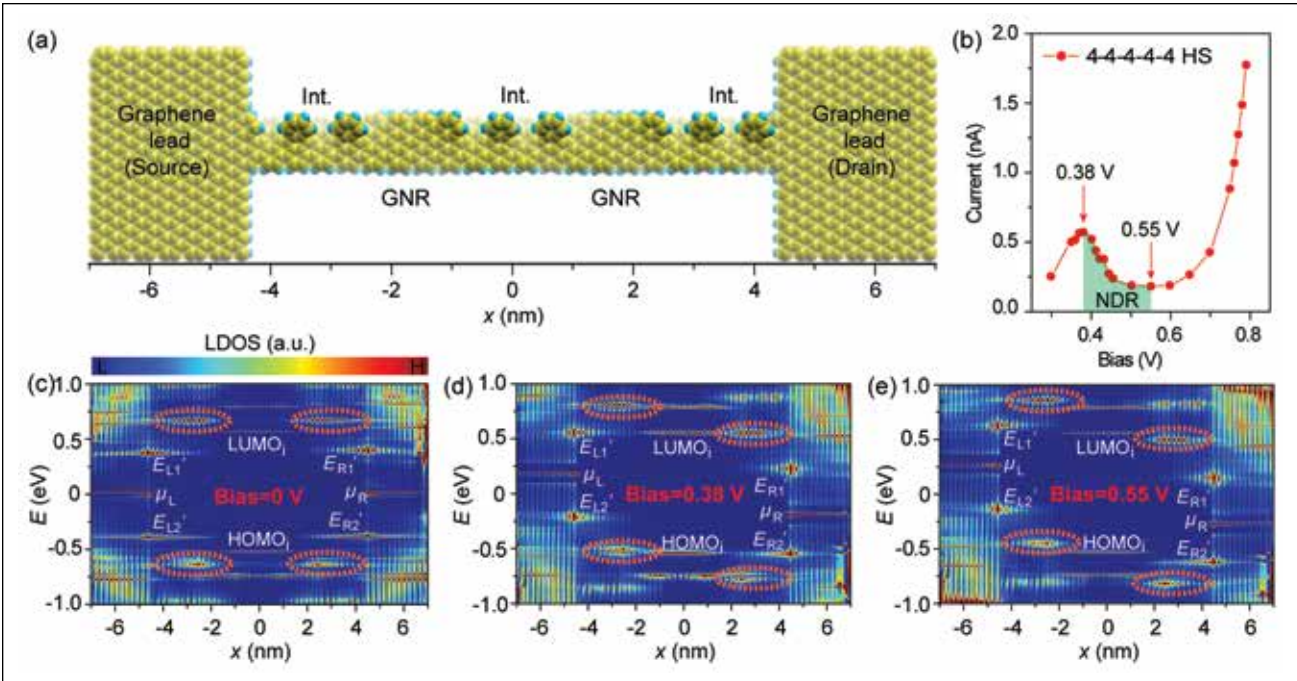


Figure 1: Atomic structure of a paradigmatic negative differential resistance (NDR) device consisting of three intermediate segments separated by two graphene nanoribbon segments. The NDR is marked with green shading. (c)–(e) Local density of states maps along the device showing alignment and misalignment of levels at zero bias, peak current, and current valley. (d) At 0.38 V bias, and (e) at 0.55 V bias. Note that the energy levels at  $-0.6$  eV and  $0.6$  eV are aligned at the bias of 0.38 V and misaligned at 0.55 V. Dashed orange ellipses mark the states emerging between the graphene nanoribbon and outer intermediate segments.

termediate/GNR double-barrier structure, two segments of the 7-aGNR, each with a length of four anthrylene units, approximately  $17.0$  Å, act as barriers, and they are directly connected to the bulk graphene leads. An intermediate structure of the same length, acting as a quantum dot, is sandwiched between the barriers, giving a structure labeled as a 4–4–4 HS. No obvious NDR is found, because direct tunneling can occur for this very short device. When much longer segments are used, NDR does occur, but the current decreases to the pA level, which is much too small to be used in devices.

To enlarge the magnitude of the current while enhancing the favorable NDR characteristics, the research team has proposed a new device design based on five short segments (Fig. 1). The team uses two 7-aGNR barriers and three intermediate parts, sandwiched between two graphene leads. Instead of connecting the 7-aGNR barriers directly to the graphene leads, the researchers inserted intermediate segments between the barriers and the leads to better align the energy levels on the opposite sides of the barriers. The use of five segments extends the active region of the device, preventing direct tunneling between the leads while allowing the barriers to be short, thereby enlarging the current. In the paradigmatic example, the team chose each segment to have the length of four anthrylene units, giving a 4–4–4–4–4 HS active region. In the calculated I–V curve in Fig. 1, the NDR appears at a relatively small bias with a large peak-to-valley ratio, which well satisfies the practical requirement for electronic circuit applications.

WHY BLUE WATERS

This application requires a very large parallel supercomputer with a high-speed interconnect between the nodes owing to the frequent exchange of substantial amounts of data between nodes. The evaluation and design of atomically precise structures required many runs to explore various scientific issues, with a substantial amount of analysis between the runs. High availability and quick turnaround are thus also very important for timely progress in this research.

PUBLICATIONS & DATA SETS

- Z. Xiao *et al.*, “Design of atomically precise nanoscale negative differential resistance devices,” *Adv. Theory Simul.*, vol. 2, p. 1800172, 2019, doi: 10.1002/adts.201800172.
- C. Ma *et al.*, “Direct writing of heterostructures in single atomically precise graphene nanoribbons,” *Phys. Rev. Materials*, vol. 3, p. 016001, 2019, doi: 10.1103/PhysRevMaterials.3.016001.



# USING OPENMP OFFLOADING TO RUN CODE ON BLUE WATERS' GPU NODES

**Allocation:** Blue Waters Professor/100 Knh  
**PI:** Daniel Bodony<sup>1</sup>  
**Collaborators:** Matthias Diener<sup>1</sup>, Sandeep Murthy<sup>1</sup>

<sup>1</sup>University of Illinois at Urbana–Champaign

## EXECUTIVE SUMMARY

Heterogeneous systems containing a CPU+GPU pair on a single node (such as Blue Waters' XK nodes) are becoming more common. Programming such systems effectively is a difficult challenge, especially for applications that have not been developed from the ground up to support such systems. In this context, OpenMP has emerged recently as an interesting solution, by leveraging OpenMP offloading capabilities in existing code. In this work, OpenMP offloading was added to PlasCom2, a multiphysics simulation application, with the Hydra framework. Hydra enables concurrent execution of application code on a CPU+GPU pair, resulting in efficient resource usage and high-performance portability. Performance results on Blue Waters show gains of up to three times on a single XK7 node [2,3].

## RESEARCH CHALLENGE

Programming heterogeneous systems is a challenging task, as few programming models support executing code on accelerators, leading to the use of specialized solutions such as CUDA, OpenCL, Legion, or Kokkos for these devices. Such specialized languages have the advantage of being able to provide the best performance in many cases since they can often provide support for special device features and offer good code generation for specific device types. However, existing application code can often not be reused and must be rewritten in a new language. Code is often not portable between devices (for example, between CPUs and GPUs), such that distinct devices may require different implementations, leading to duplicated code and an increased difficulty of code maintenance.

For PlasCom2, the research team needed an approach that was able to use the existing code (in C++ and Fortran) on several types of accelerators without having separate implementations for different device types. Furthermore, the team wanted to be able to support different hardware and software environments with this approach, and run concurrently on the host and offloading devices to use all available computing resources efficiently.

## METHODS & CODES

Based on the offloading support available in recent OpenMP versions, the team developed Hydra, which is a library to support concurrent execution on host and accelerator devices. For PlasCom2, Hydra [1–3] measures the relative performance of the host and accelerator at startup and determines the best work distribution based on these data. During execution, Hydra handles the data movement between host and offloading device as well as the actual code execution [2,3]. Hydra requires an OpenMP 4.5 compiler and runtime but has no additional dependencies [1].

## RESULTS & IMPACT

OpenMP offloading enables simple and efficient execution of a single code base on different types of devices, with minimal changes to existing code. Hydra builds on top of OpenMP offloading to add support for fully heterogeneous execution; that is, running parts of the problem concurrently on different device types. Using Hydra with PlasCom2 resulted in a speedup of three times compared to CPU-only execution on a Blue Waters XK7 node. Compared to running only on the GPU, performance from heterogeneous execution was improved by 17%. Some of the computationally intensive kernels of PlasCom2 showed a speedup of up to four times (Fig. 1). OpenMP offloading and Hydra show that existing codes can be enabled to run on heterogeneous systems with a low number of changes and high performance efficiency.

## WHY BLUE WATERS

Blue Waters was essential to the research by providing a stable, high-performance platform with easy access to modern accelerators. Hydra and its integration into PlasCom2 could be developed directly on Blue Waters, allowing the team to evaluate and compare different implementation possibilities on a real system.

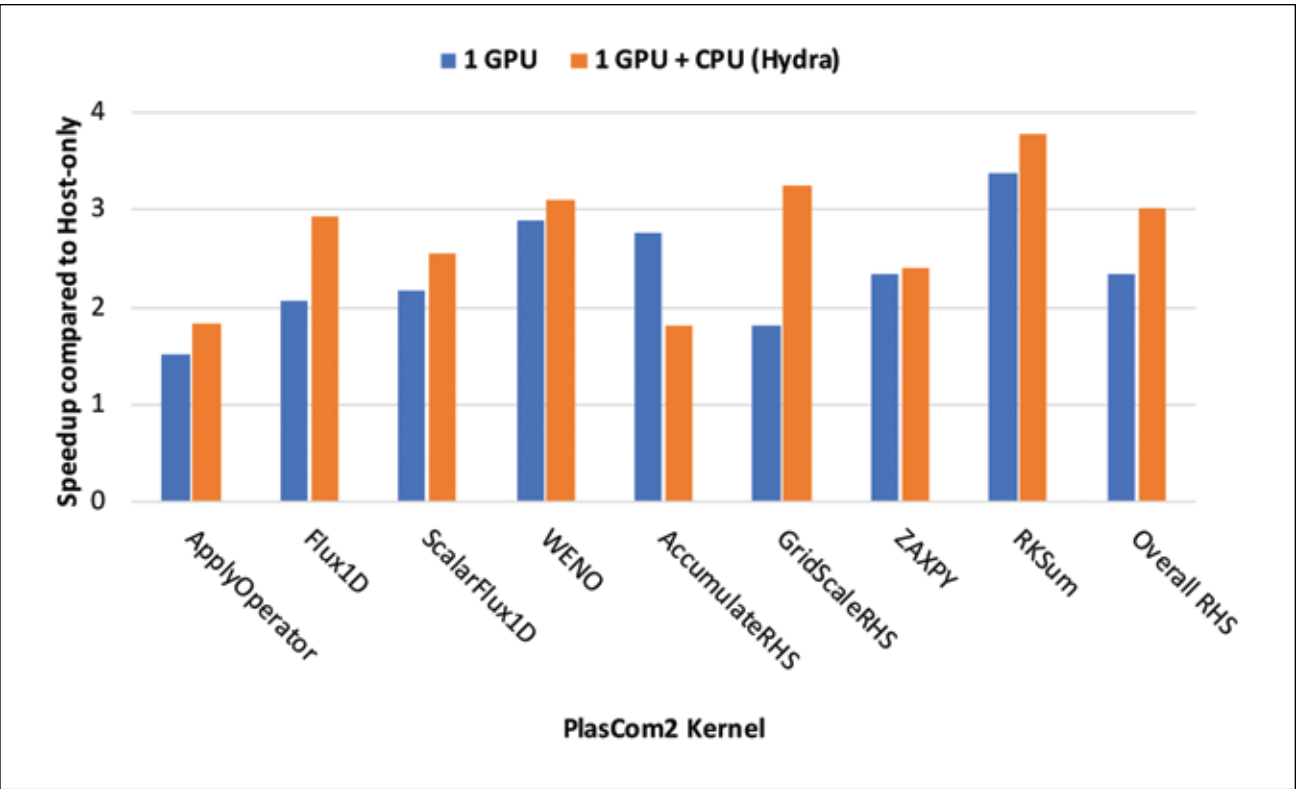


Figure 1: PlasCom2 kernel results on the Blue Waters XK7 system.



NUMERICAL INVESTIGATION OF TURBULENCE SUPPRESSION IN ROTATING FLOWS

Allocation: NSF PRAC/10,400 Knh  
PI: Christoph Brehm<sup>1</sup>  
Co-PI: Sean Bailey<sup>1</sup>  
Collaborator: Neil Ashton<sup>2</sup>

<sup>1</sup>University of Kentucky  
<sup>2</sup>University of Oxford

EXECUTIVE SUMMARY

In past experiments, simulations, and theoretical analyses, rotation has been shown to dramatically affect the characteristics of turbulent flows, such as by causing the mean velocity profile to appear laminar, leading to an overall drag reduction, as well as by affecting the Reynolds stress tensor (the total stress tensor in a fluid). The axially rotating pipe is an exemplary prototypical model problem that exhibits these complex turbulent flow physics. For this flow, the rotation of the pipe causes a region of turbulence suppression that is particularly sensitive to the rotation rate and Reynolds number (the ratio of inertial forces to viscous forces within a fluid). The physical mechanisms causing turbulence suppression are currently not well understood, and a deeper understanding of these mechanisms would be of great value for many practical applications involving swirling or rotating flows, such as swirl generators, wing-tips, axial compressors, hurricanes, and the like.

The research team conducted direct numerical simulations (DNS) of rotating turbulent pipe flows at different Reynolds and rotation numbers. The main objectives of this work were to analyze the effects of rotation on turbulence considering turbulence budgets and higher-order statistics as well as to quantify turbulence suppression for rotating turbulent pipe flow.

RESEARCH CHALLENGE

Swirling flows are an important class of flows, not only because of the complex flow physics but also because of their relevance to many industrial applications such as combustion, heat exchange-

ers, cyclone separation, mixing, etc. To design the next generation of ever-more-efficient cars, aircraft, or energy systems, it is important to understand turbulent swirling and rotating flows in order to predict and manipulate them. Thus, enhanced understanding of the physical mechanisms for swirling and rotating flows and improved prediction capabilities for these types of flows are highly beneficial for key U.S. industries where swirling and rotating flows appear, such as in the oil and gas, biomedical, energy harvesting, and aerospace sectors. When rotating a turbulent flow, a phenomenon known as turbulence suppression has been observed (see flow visualization in Fig. 1), which can cause a reduction in wall-shear stress or skin friction [1–5], making the understanding of turbulence suppression highly valuable for engineering applications. A key objective of this research is to obtain high-quality simulation data (in conjunction with ongoing experiments) that can be used to study the nature of the highly complicated flow physics of turbulent rotating flows.

METHODS & CODES

The research team solved the incompressible Navier–Stokes equations in a reference frame rotating with the pipe walls where the centrifugal and Coriolis forces (inertial forces that act on objects that are in motion within a frame of reference that rotates with respect to an inertial frame) were added as source terms. For these simulations, the researchers assumed fully developed turbulent flow and periodic boundaries in the streamwise direction. Sufficient temporal and spatial resolution is required to thoroughly study the intricate nature of turbulence and the relevant temporal and spatial scales. The computational meshes ensure that the wide range of turbulent scales are well resolved with grid spacing close to  $\Delta y^+ = 1$ . Simulations were conducted using the spectral-element solver Nek5000, developed by Fischer *et al.* [6]. Nek5000 is a higher-order accurate, open source, spectral-element solver used to solve the incompressible Navier–Stokes equations and is well known for its (spectral) accuracy, favorable dispersion properties, and efficient parallelization [7]. The spectral-element method is based on a weighted-residual approach for spatial discretization. For parallel computations, Nek5000 utilizes the MPI protocol and has shown excellent scaling characteristics on high-performance computing systems—making it well-suited for large-scale turbulence flow computations.

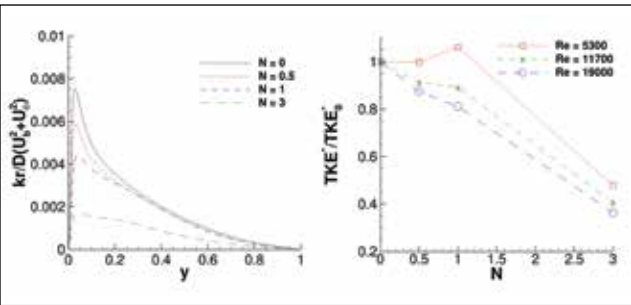


Figure 1: (a) Turbulent kinetic energy multiplied by radial position and normalized by total mean kinetic energy for  $Re = 19,000$  and (b) ratio of total turbulent kinetic energy for rotating and stationary turbulent pipe flow.

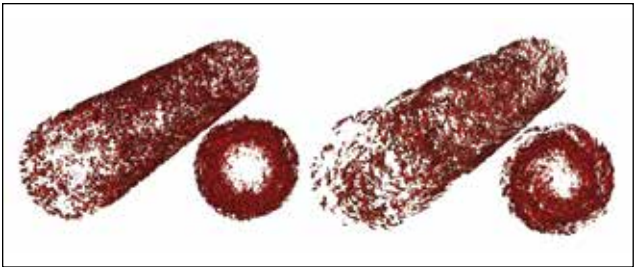


Figure 2: Iso-contour surface of Q-criterion with  $Q = 65$  for stationary and rotating turbulent pipe flow at  $N = 1$  and  $Re = 19,000$ .

RESULTS & IMPACT

The current research is concerned with axially rotating pipe flow. In this flow, the axis of rotation is parallel to the mean flow direction and, therefore, for the laminar case the axial mean flow is not directly affected by the rotation. Hence, the parabolic mean flow profile observed in laminar nonrotating pipes also describes the axial (laminar) velocity profile for the rotating case. The axially rotating pipe can be described by two nondimensional parameters: the Reynolds number  $Re = UD/\nu$  based on the mean bulk flow velocity  $U$ , the pipe diameter  $D$ , and the kinematic viscosity  $\nu$  as well as the rotation number  $N = V_w/U$  of the pipe, which is sometimes also referred to as swirl rate. The rotation number characterizes the angular velocity through the azimuthal velocity of the pipe inner wall,  $V_\theta(r = R) = \Omega D/2$  (in nonrotating reference frame).

While models such as Reynolds-averaged Navier–Stokes (RANS) and wall-resolved large-eddy simulation fail to accurately reproduce the flow physics involved in turbulence suppression, DNS can be used to effectively study rotation effects on turbulent structures. Ashton *et al.* [8] demonstrate some of the severe limitations of several advanced RANS models when compared to the DNS data presented here. Existing DNS studies of rotating pipe flows have been restricted to relatively low Reynolds numbers [9–11], and a strong dependence on rotation number has been observed. Thus, one of the goals of the current work is to provide detailed DNS data at large Reynolds numbers.

The streamwise velocity profiles (not shown here) illustrate how the turbulent flow is affected by rotation in the mean. The velocity profiles are plotted versus the distance from the wall  $y = 1 - r/R$ , where  $r$  is the local radius and  $R$  is the total radius of the pipe. It can be seen that the streamwise velocity profile tends toward the laminar profile as the rotation number  $N$  is increased. Near the wall, the wall-normal velocity gradient is reduced, which leads to a reduction in skin friction and a speed-up of the flow toward the center of the pipe.

Fig. 1a displays the turbulent kinetic energy versus wall distance compensated by the local radius (accounting for the area contribution in the integrand) and normalized with the total mean kinetic energy ( $MKE = 1/2(\langle V_z^2 \rangle + \langle V_\theta^2 \rangle)A$ ). A clear reduction in turbulent kinetic energy can be observed throughout the cross-section for  $Re = 19,000$ . The total turbulent kinetic energy for all three

Reynolds numbers versus rotation number is illustrated in Fig. 1b, and a reduction in turbulent kinetic energy was obtained for all rotation numbers at Reynolds numbers of  $Re = 11,700$  and  $19,000$ . Interestingly, an initial increase in turbulent kinetic energy can still be observed until  $N = 1$  for  $Re = 5,300$  and, thus, considering turbulence as being suppressed may not appropriately describe the characteristics of this flow at these conditions.

In summary, turbulence suppression is occurring (for large enough rotation numbers) for all three Reynolds numbers used in this study. The results for the turbulent kinetic energy also seemingly display a change in trends at  $N = 1$  for all three Reynolds numbers, which is discussed in more detail in [12,13].

WHY BLUE WATERS

As described in the research results discussion above, sufficient temporal and spatial resolution is required for these DNS to thoroughly study the intricate nature of turbulence, turbulence suppression, and relaminarization. In particular, to obtain data that are suitable for the description of the entire statistical distribution of the dissipative scales of turbulence, the DNS require sub-Kolmogorov scale grid resolution. Access to Blue Waters’ resources was essential to accomplish this computationally demanding research.

PUBLICATIONS & DATA SETS

- J. Davis, S. Ganju, and C. Brehm, “Direct Numerical Simulations of turbulence suppression in rotating pipe flows,” in *Parallel CFD Conf. Proc.*, Indianapolis, IN, U.S.A., May 14–17, 2018.
- C. Borchetta, C. Brehm, and S. Bailey, “On the development of an apparatus to examine rotating pipe flow at high rotation numbers,” APS DFD, Atlanta, GA, U.S.A., Nov. 18–20, 2018.
- J. Davis, S. Ganju, and C. Brehm, “Direct Numerical Simulations of rotating turbulent pipe flows at moderate Reynolds numbers,” APS DFD, Atlanta, GA, U.S.A., Nov. 18–20, 2018.
- C. Brehm, J. Davis, S. Ganju, and S. Bailey, “A numerical investigation of the effects of rotation on turbulent pipe flows,” 11th Inter. Symp. Turbulence and Shear Flow Phenomena (TSFP11), Southampton, England, U.K., July 30–Aug. 2, 2019.
- J. Davis, S. Ganju, N. Ashton, S. Bailey, and C. Brehm, “A DNS study to investigate turbulence suppression in rotating pipe flows,” in *AIAA Aviation 2019 Forum*, Dallas, TX, U.S.A., June 17–21, 2019, doi: 10.2514/6.2019-3639.
- N. Ashton, J. Davis, and C. Brehm, “Assessment of the elliptic blending Reynolds stress model for a rotating turbulent pipe flow using new DNS data,” in *AIAA Aviation 2019 Forum*, Dallas, TX, U.S.A., June 17–21, 2019, doi: 10.2514/6.2019-2966.



# AN EFFICIENT METHOD FOR HYPERSONIC LAMINAR-TURBULENT TRANSITION PREDICTION

**Allocation:** Innovation and Exploration/140 Knh  
**PI:** Oliver M. F. Browne<sup>1</sup>  
**Collaborator:** Christoph Brehm<sup>1</sup>

<sup>1</sup>University of Kentucky

## EXECUTIVE SUMMARY

A thorough understanding of the laminar-turbulent transition process of high-speed boundary layers is of paramount importance when designing supersonic or hypersonic vehicles. These high-speed vehicles are among the most difficult and challenging to design owing to significant aerothermal loads experienced on the vehicle during the transition to turbulence. An accurate prediction of the boundary-layer state can help reduce design margins and, ultimately, guide the development of novel, innovative, high-speed vehicles.

In this project, the research team has developed and validated a new efficient stability and transition prediction method for hypersonic boundary layers; namely, the AMR-WPT (adaptive mesh refinement wavepacket tracking) method, which was up to 10 times more efficient when compared with static mesh approaches. The AMR-WPT method was validated against conventional stability and transition methods (LST—linear stability theory, DNS—direct numerical simulation, etc.).

## RESEARCH CHALLENGE

Conducting numerical transition prediction investigations for complex geometries is either generally too computationally expensive (e.g., DNS) or unable to capture all flow physics (e.g., non-

linear receptivity, nonlinear instability mechanisms, nonlinear breakdown, and the like). The latter is due to the assumptions that these conventional methods rely upon (e.g., LST—linear and parallel flow assumption; parabolized stability equations—mostly linear, although they can be weakly nonlinear, etc.) The AMR-WPT method attempts to capture all flow physics as DNS can but also to be competitive with LST in terms of computational efficiency.

In many cases, the instabilities in high-speed boundary layers are convectively unstable and appear as localized convecting wavepackets. This inherent nature of the wavepacket can be exploited with AMR, which can be employed to track the wavepacket as it convects downstream and ultimately reduce the number of grid points that are required for the simulation.

## METHODS & CODES

The governing equations of the AMR-WPT method are the disturbance flow formulation of the 3D compressible Navier-Stokes equations. This formulation decomposes the flow state vector and fluxes in the Navier-Stokes equations into steady base flow (or mean flow) components and unsteady disturbance components. From the disturbance flow equations, the higher-order terms can either be dropped to solve the linear disturbance flow equations or included to solve for the nonlinear disturbance flow equations—

Figure 1: Overview of overset mesh approach for AMR-WPT method (adjusted from [1]).

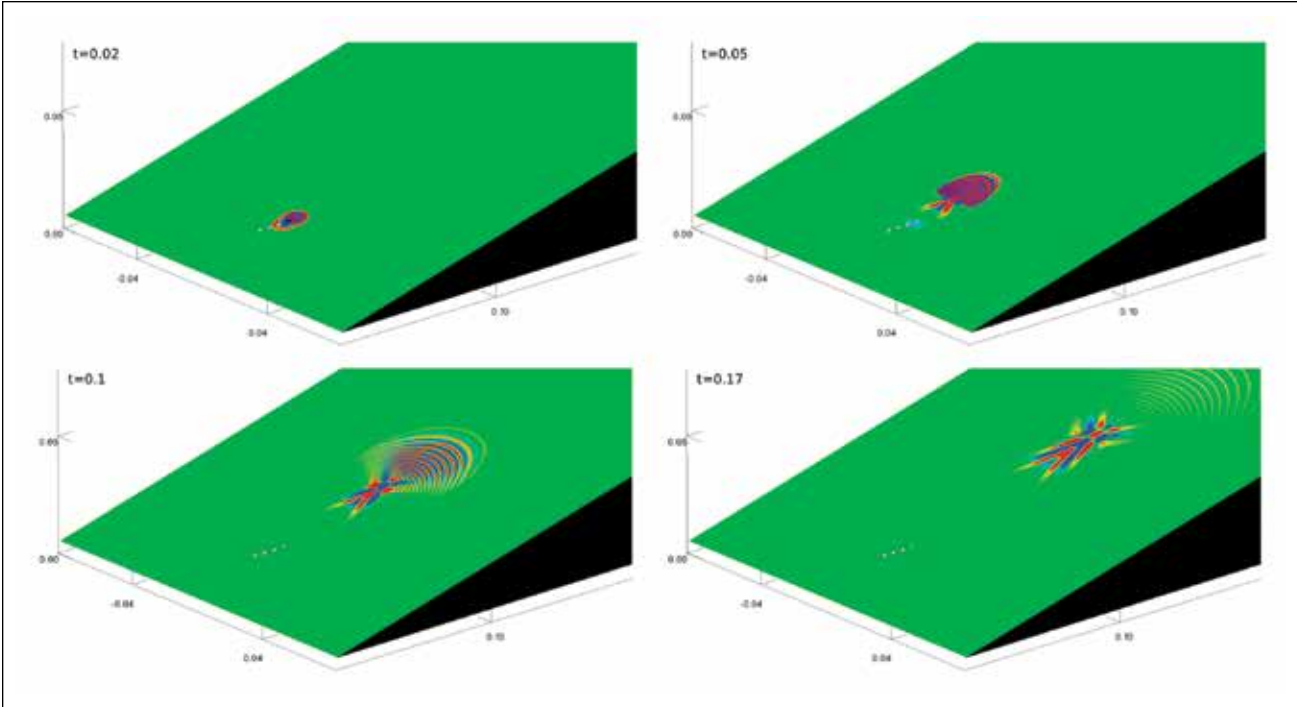
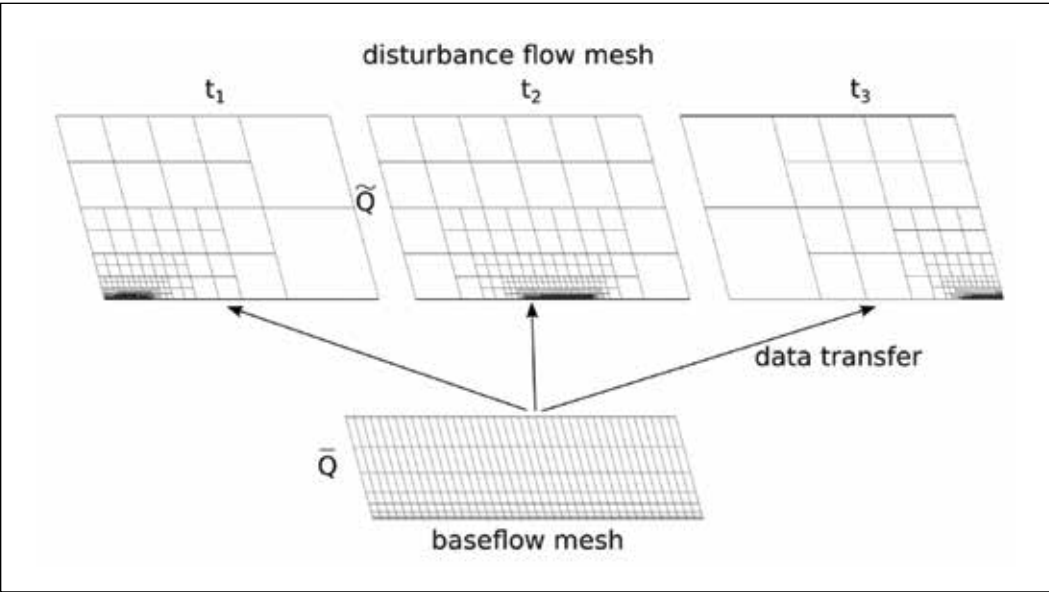


Figure 2: Instantaneous visualization of particulate trajectory and collision (contour showing pressure disturbance) with 3D  $M = 4$  14° straight wedge at four time instances (adjusted from [3]).

this research considered both. The base flow terms are computed by interpolating from a base flow solution on a base flow mesh.

The disturbance is introduced into the flow field via wall forcing (blowing/section) [1,4], which subsequently develops into a wavepacket. Disturbance generation via particulate collision was also investigated [3]. At predefined intervals, a refinement/derefinement step is performed to redistribute the grid points to track the wavepacket as it convects downstream (Fig. 1).

The AMR is handled by the external library PARAMESH [5]. Higher-order prolongation and restriction operators are employed for transferring the information between different refinement levels in the block-structured Cartesian mesh. Dynamic load balancing and Morton ordering is used to redistribute the loads among the different processors after each refinement/derefinement step is performed.

## RESULTS & IMPACT

The AMR-WPT method has been validated against a number of test cases including: (1) a 2D  $M = 5.35$  ( $M$  is Mach number) flat-plate boundary layer [1,2]; (2) an axisymmetric  $M = 10$  7° straight cone [1,2]; (3) a 3D  $M = 5.35$  flat-plate boundary layer [4]; and (4) a 3D  $M = 4$  14° straight wedge [3]. Fig. 2 shows an example of the AMR-WPT method for a particulate collision.

The AMR algorithms were able to successfully track the wavepackets in all cases and significantly reduce the number of grid points that were required when compared with static mesh methods. In static mesh approaches, the computational cost scales

with the size of the domain whereas with the AMR-WPT method the cost scales with the size of the wavepacket. Different criteria were tested to determine the most appropriate parameter(s) for tracking the wavepackets (i.e., refining/derefining the mesh). This ultimately comes down to a compromise between efficiency and accuracy [1–4]. The next step will be to examine more complex geometries.

## WHY BLUE WATERS

The Blue Waters supercomputer and its staff were essential for this research. The large amount of computational resources and storage facilities available on the system played a considerable part in enabling this research to move forward at the pace it did. The expertise of the project staff allowed the research team to maximize the efficiency of its code and parallelization capacity on the Blue Waters system.

## PUBLICATIONS & DATA SETS

O. M. F. Browne, S. M. A. Al Hasnine, and C. Brehm, “Towards modeling and simulation of particulate interactions with high-speed transitional boundary-layer flows,” presented at the AIAA Conf., Dallas, TX, U.S.A., 2019.

O. M. F. Browne, A. P. Haas, H. F. Fasel, and C. Brehm, “An efficient nonlinear wavepacket tracking method for hypersonic boundary-layer flows,” under consideration, 2019.



QUANTUM SIMULATIONS: PROPERTIES OF DENSE HYDROGEN

Allocation: Blue Waters Professor/200 Knh  
PI: David Ceperley<sup>1</sup>  
Collaborator: Carlo Pierleoni<sup>2</sup>

<sup>1</sup>University of Illinois at Urbana–Champaign  
<sup>2</sup>University of L'Aquila

EXECUTIVE SUMMARY

The research in this project on Blue Waters is related to the Materials Genome Initiative, the federally supported cross-agency program to develop computational tools for designing new materials. In the past year, the research group has been running calculations for dense hydrogen and lithium in order to make predictions that have been tested experimentally. In particular, the team performed coupled electron–ion quantum Monte Carlo calculations of the transition between molecular and atomic hydrogen to determine the changes in optical properties across the transition. During the past year, their prediction of a transition was verified by a new experiment.

RESEARCH CHALLENGE

The phase diagram of high-pressure hydrogen is of great interest both for fundamental research, such as in astrophysics, and for energy applications such as high-temperature superconductivity in hydrides.

METHODS & CODES

The team employs quantum Monte Carlo calculations that provide nearly exact information on quantum many-body systems. This is the most accurate general method capable of treating electron correlation; thus, it needs to be in the kernel of any materials design initiative. The method is able to use Blue Waters effectively because there are several pathways to find parallel performance.

RESULTS & IMPACT

The research group’s recent computations on Blue Waters included the following two projects:  
*Liquid–liquid phase transition in hydrogen by coupled electron–ion Monte Carlo simulations.* The phase diagram of high-pressure hydrogen is of great interest both for fundamental research, such as in planetary physics, and for energy applications. The existence and precise location of a phase transition in the liquid phase between a molecular insulating fluid and a monoatomic metallic fluid is relevant for planetary models. Recent experiments reported contrasting results about the location of the phase transition. Theoretical results based on density functional theory are also very scattered. Before 2018, the research team performed highly accurate coupled electron–ion Monte Carlo calculations of this transition, finding results that lay between the two experimental predictions, close to that measured in Diamond Anvil Cell experiments but at 25–30 gigapascal higher pressure. The transi-

tion along an isotherm is signaled by a discontinuity in the specific volume, a sudden dissociation of the molecules, a jump in electrical conductivity, and in electron localization. During 2016–17, the research team’s prediction of a transition was verified by three separate experiments. In particular, a new experiment [1] verified the results of the team’s calculation, which was covered by *The New York Times* [2].

Recently, the team has continued the calculations to control various computational approximations and quantify their errors. In the papers listed below, the researchers published additional information about dense hydrogen as the system changes from a molecular liquid to an atomic liquid. These findings are valuable in understanding and reconciling the different experiments.

*More accurate predictions of the Compton profile in liquid and solid lithium.* The research group has performed very accurate calculations of the momentum distribution of electrons in liquid and solid lithium to compare with new scattering experiments done on the latest experimental light source. These calculations removed many of the assumptions and approximations of earlier calculations. For example, the team simulated both solid and liquid lithium at the experimental conditions of temperature and included core electrons in the simulation. They were able to agree with experiment on an order of magnitude better than previous calculations. The work will be published in 2019 together with the experimental data.

WHY BLUE WATERS

Access to Blue Waters was necessary because of the computational demands of the calculation involving hydrogen and lithium. The systems typically had more than one hundred electrons treated as quantum particles and the ions (protons or lithium ions) were dynamic and disordered.

PUBLICATIONS & DATA SETS

G. Rillo, M. A. Morales, D. M. Ceperley, and C. Pierleoni, “Optical properties of high-pressure fluid hydrogen across molecular dissociation,” *Proc. Nat. Acad. Sci. U.S.A.*, vol. 116, no. 20, pp. 9770–9774, 2019.  
V. Gorelov, C. Pierleoni, and D. M. Ceperley, “Benchmarking vdW–DF first principle predictions against coupled electron–ion Monte Carlo for high pressure liquid hydrogen,” *Contrib. Plasma Phys.*, vol. 59, no. 4–5, p. e201800185, 2019.

C. Pierleoni, G. Rillo, D. M. Ceperley, and M. Holzmann, “Electron localization properties in high pressure hydrogen at the liquid–liquid phase transition by coupled electron–ion Monte Carlo,” *J. Phys.*, Conf. Ser. 1136, p. 012005, 2018.  
C. Pierleoni, M. Holzmann, and D. M. Ceperley, “Local structure in dense hydrogen at the liquid–liquid phase transition by coupled electron–ion Monte Carlo,” *Contrib. Plasma Phys.*, vol. 58, no. 2–3, pp. 99–106, 2018.



ROLE OF INTERFACES ON THE SHEAR STRENGTH AND BENDING PROPERTIES OF VAN DER WAALS TWO-DIMENSIONAL MATERIALS

Allocation: Illinois/230 Knh  
PI: Huck Beng Chew<sup>1</sup>  
Co-PI: Soumendu Bagchi<sup>1</sup>

<sup>1</sup>University of Illinois at Urbana–Champaign

EXECUTIVE SUMMARY

Two-dimensional van der Waals materials such as graphene and hexagonal boron nitride (h-BN) are of significant technological importance owing to their unconventional properties at the nanoscale. Using Blue Waters’ computational resources, density functional theory calculations and molecular statics simulations were used to quantify the interfacial strength properties of graphene–titanium (Ti) and graphene–aluminum (Al) nanocomposites, as well as the bending rigidity of layered graphene and hexagonal boron nitride nanostructures. The results show that the strongly chemisorbed graphene–Ti interface is drastically weakened by the formation of a metal oxide phase, while the weakly physisorbed graphene–Al interface is significantly strengthened through metal oxide formation. The research team’s simulations for the bending rigidity of two- to six-layered boron nitride nanosheets show substantially higher bending stiffness than that of multilayered graphene. These computational results are in excellent agreement with recent experiments.

RESEARCH CHALLENGE

Single-atom sheet nanostructures such as graphene and hexagonal boron nitride offer interesting design capabilities owing to their strong in-plane covalent bonding, with highly tunable but weak van der Waals interactions at the interfaces. For example, graphene or its rolled counterpart, carbon nanotube, can be used as reinforcements in lightweight metal matrix composites (e.g., Ti and Al), leading to high stiffness and strength as well as stability

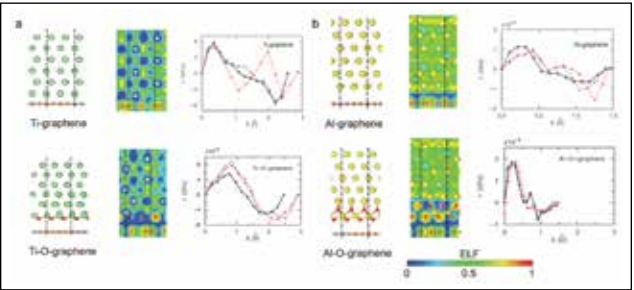


Figure 1: Effect of oxidation on the interfacial binding and shearing properties in titanium–graphene (a) and aluminum–graphene (b) interfaces. The figures in the upper and lower panels of (a) and (b) show (from left to right) the atomic structure, electron localization function contours, and shear stress response with relative sliding of the interfaces.

at high temperatures. This makes them ideal candidates for manufacturing lightweight damage-tolerant structures relevant for the aerospace and automobile industries. Such a unique combination of properties relies on effective bonding along the graphene–metal interface. The local binding properties of graphene on surface-oxidized metals can differ profoundly from those on bare metals, and the associated mechanisms can affect the overall strength of the nanocomposite. On the other hand, the bending rigidity of monolayer 2D materials is simply governed by the underlying chemical bond stretching/compression/rotation behaviors of atoms within the single atomic sheet. However, for multilayer 2D materials, the bending rigidity becomes more complicated owing to the added contributions of interlayer interactions. While monolayer h-BN and graphene have bending rigidities of a similar order of magnitude, the stronger interlayer interaction at h-BN interfaces suggests a potentially different modulus for their multilayer counterparts. Simulations of such local behavior with atomistic details can uncover key mechanisms that are otherwise inaccessible through state-of-the-art experiments.

METHODS & CODES

The research team performed density functional theory (DFT) calculations for graphene–metal interfaces using the Vienna Ab initio Simulation Package. They used Projector Augmented Wave-based pseudopotentials to represent the interaction between ionic cores and valence electrons, while the local density approximation was adopted for exchange and correlation. Electronic energy of the structures was minimized by the conjugate gradient method. The team determined the binding properties at the interfaces of graphene–metals, with and without oxidation, by computing electron localization functions that describe the probability of finding one electron in the neighborhood of another electron. Interfacial shear strength was extracted by computing the sliding potential energy surface as a function of relative interfacial displacements.

To investigate the bending and interlayer shear rigidities of h-BN and graphene nanosheets, the research team conducted molecular static simulations using the classical molecular dynamics code LAMMPS. The team adopted a recently developed Tersoff parametrization to represent the intralayer interactions in h-BN. The intralayer atomistic interactions between C atoms in graphene were governed by a second-generation reactive em-

pirical bond order potential. The interlayer interactions for both graphene and h-BN were described by registry-dependent Kolmogorov–Crespi-type force fields. The group performed stiffness calculations of folded h-BN and graphene sheets with varying numbers of interlayers by first constructing sinusoidally modulated initial geometries and minimizing their energies by the conjugate gradient method.

RESULTS & IMPACT

In Fig. 1, the DFT simulations suggest opposite changes to the binding and shear strength properties of graphene on these two types of metals when surface oxidized [1]. The researchers observed distinctly higher electron localization function values of approximately 0.3 across Ti–graphene versus approximately 0.1 across Al–graphene because of hybridization of the unoccupied *d*-orbitals in transition Ti metal with the 2*p*-orbitals of C atoms in graphene. For graphene on Ti–O and Al–O substrates, the Ti–C and Al–C interactions are weakened by O atoms on the metal surface. The high interfacial shear strength of approximately 5 GPa along pure Ti/graphene is reduced by two orders of magnitude to approximately 20 MPa along Ti–O/graphene, while the interfacial shear strength of approximately 10 MPa along pure Al/graphene remains similar in magnitude along Al–O/graphene. These fundamental insights on graphene–metal interfaces have important implications for graphene-reinforced metal matrix composites and the fabrication of graphene–metal contacts in graphene transistors, as well as the transfer printing and epitaxial growth of graphene on metallic substrates.

The research group’s simulations showed that the bending properties of multilayer h-BN and graphene are significantly different from their monolayer counterparts [2]. While the bending stiffness of monolayer h-BN and graphene is associated with pure bending of the single atomic layer, the stiffness of folded two- to six-layered h-BN and graphene has added contributions from interlayer sliding and out-of-plane deformation. Figs. 2b and 2c show the relative change in the normal ( $\delta_n$ ) and shear ( $\delta_s$ ) separations of the three interlayers ( $L_1$ – $L_3$ ) for the folded four-layer h-BN and graphene sheets depicted in Fig. 2a. The interlayer deformation localizes near the steepest slope of the folded configurations, as marked by red dashed lines in Fig. 2a, and results in significant distortions in stacking sequences for h-BN and graphene (insets of Fig. 2a). For h-BN, the relatively stronger binding between interlayers and the higher barrier energies for interlayer sliding result in smaller normal (Fig. 2b) and shear separations (Fig. 2c) compared to graphene. This explains the consistently higher effective stiffness of folded multilayer h-BN versus observations in recent experiments.

WHY BLUE WATERS

Blue Waters’ resources were required for the evaluation of multiple graphene–metal/metal oxide nanostructures (approximately 20 different structures) with several hundred atoms per structure in DFT. Bending stiffness computations of folded h-BN and

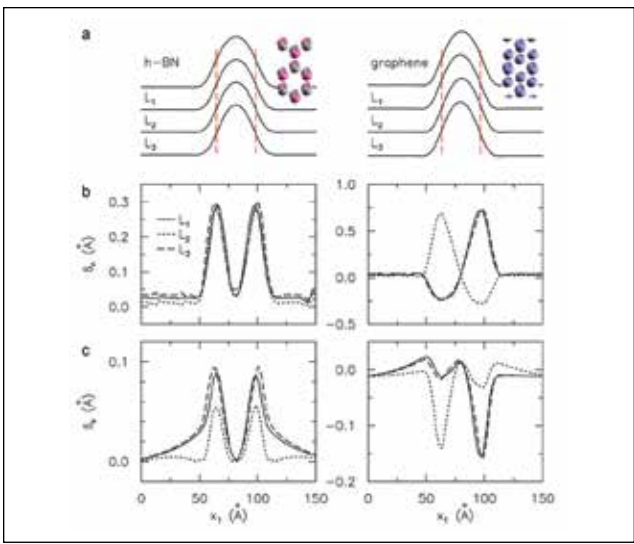


Figure 2: Folded geometries (a), with interlayer normal (b) and shear (c) separations in four-layer h-BN (left) and graphene (right) nanostructures. Stacking deviations, as viewed from the top, in inset (a) at locations of maximum normal and shear separations indicated by red dashed lines; pink, grey, and blue denote boron, nitrogen, and carbon atoms, respectively.

graphene nanostructures (comprised of hundreds of thousands of atoms per structure) using classical molecular dynamics simulations based on many-body potentials necessitated the massively parallel computing capabilities of Blue Waters.

PUBLICATIONS & DATA SETS

C. Yi *et al.*, “Direct nanomechanical characterization of carbon nanotubes–titanium interfaces,” *Carbon*, vol. 132, pp. 548–555, 2018.

S. Bagchi, A. Harpale, and H. B. Chew, “Interfacial load transfer mechanisms in carbon nanotube–polymer nanocomposites,” *Proc. Royal Soc. Lond. A*, vol. 474, Art. no. 20170705, 2018.

C. Yi, S. Bagchi, F. Gou, C.M. Dmuchowski, C. Park, C. C. Fay, H. B. Chew, and C. Ke, “Direct nanomechanical measurements of boron nitride nanotube–ceramic interfaces,” *Nanotechnol.*, vol. 30, Art. no. 025706, 2018.

S. Bagchi, C. Ke, and H. B. Chew, “Oxidation effect on the shear strength of graphene on aluminum and titanium surfaces,” *Phys. Rev. B*, vol. 98, Art. no. 174106, 2018.

S. S. Sawant, P. Rao, A. Harpale, H. B. Chew, and D. A. Levin, “Multi-scale thermal response modeling of an AVCOAT-like thermal protection material,” *Int. J. Heat Mass Transf.*, vol. 133, pp. 1176–1195, 2019.

Y. Cui, and H. B. Chew, “A simple numerical approach for reconstructing the atomic stresses at grain boundaries from quantum-mechanical calculations,” *J. Chem. Phys.*, vol. 150, Art. no. 144702, 2019.

W. Qu, S. Bagchi, X. Chen, H. B. Chew, and C. Ke, “Bending and interlayer shear moduli of ultrathin boron nitride nanosheet,” *J. Phys. D: Appl. Phys.*, vol. 52, no. 46, pp. 465301, 2019.



ATYPICALLY ENTANGLED PHASES AND NEW METHODS FOR THE QUANTUM MANY-BODY PROBLEM

**Allocation:** Blue Waters Professor/250 Knh  
**PI:** Bryan Clark<sup>1</sup>  
**Collaborators:** Dmitrii Kochkov<sup>1</sup>, Eli Chertkov<sup>1</sup>, Xiongjie Yu<sup>1</sup>, Di Luo<sup>1</sup>, Greg MacDougall<sup>1</sup>, Hitesh Changlani<sup>2</sup>, Rebecca Flint<sup>3</sup>, Jyotisman Sahoo<sup>3</sup>

<sup>1</sup>University of Illinois at Urbana–Champaign  
<sup>2</sup>Florida State University  
<sup>3</sup>Iowa State University

EXECUTIVE SUMMARY

This work focuses on: (1) developing new machine-learning approaches to simulate the quantum many-body problem, and (2) elucidating novel phenomena in atypically entangled phases. (The many-body problem, in brief, involves understanding the collective behavior of large numbers of interacting particles. Entanglement deals with the phenomenon of particles that remain correlated even when separated by great distances.) One new machine-learning approach [1] simulates difficult quantum many-body electronic systems by combining deep neural networks with ideas originally developed by Feynman [2] about backflow. Using this new approach, the research team extrapolated to near exact energies on difficult quantum many-body systems and competed favorably with other state-of-the-art methods. One important atypically entangled phase is the spin-liquid; the team has discovered an expanded spin-liquid regime in the phase diagram of the stuffed honeycomb model.

RESEARCH CHALLENGE

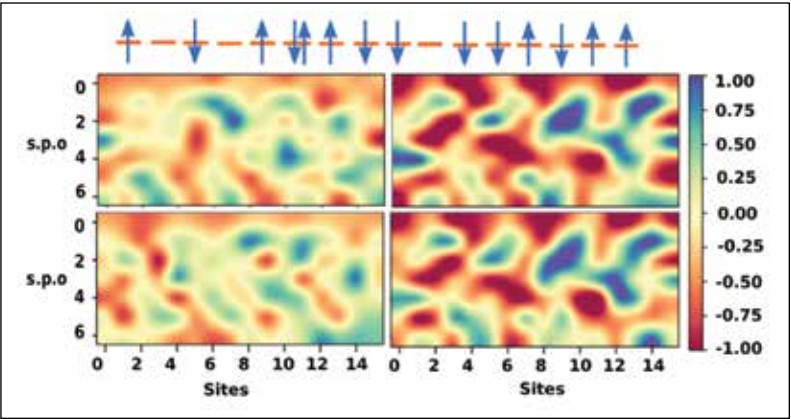
Entanglement makes quantum mechanics both interesting and difficult to simulate. Einstein once described entanglement in quantum mechanics as “spooky action at a distance,” and phases of matter ranging from spin-liquids to eigenstate phases have exotic properties because of their atypical entanglement. Entanglement is also responsible for causing the cost of exactly simulating quantum many-body systems to scale exponentially with system size; every two years researchers can simulate one more electron.

The research team’s problem is twofold: to identify, classify, and find phases with interesting entanglement, as well as to develop new methodologies based on ideas from machine learning to overcome the barriers to efficient simulation. The two phases of atypical entanglement the group is most interested in are spin-liquids and eigenstate phases.

Spin-liquids [3] are phases of matter whose entanglement is so complicated that they can’t be constructed with short quantum circuits. In spin-liquid materials, the electron “fractionalizes,” splitting into multiple pieces. Spin-liquids support anyons, which are important for constructing quantum memories and quantum computers. While the theory of spin-liquids is well established, the key question is to bring these spin-liquids into the real world by finding physical systems that support them.

Eigenstate phases [4] of matter are a recently discovered class of physical systems whose eigenstates have atypical entanglement. The entanglement of a typical eigenstate is boring; all particles are uniformly entangled with each other. On the other hand, in eigenstate phases, the entanglement in each eigenstate is highly structured. In addition to weird entanglement, these states of matter never equilibrate—the equivalent of a never-cooling cup of coffee. Eigenstate phases might be the key to allowing quantum computers to run at higher temperatures. The key question the team is addressing here is to increase the number of known eigenstate phases.

Figure 1: Matrix to diagonalize without (left) and with (right) neural network added to the wave function for spin-up (top) and spin-down (bottom) electrons. The neural network wave function restores the symmetry between the spin-up and spin-down electrons, which is known to exist in the exact wave function. Reproduced from [6].



METHODS & CODES

To find spin-liquids, the team has taken a two-fold approach. In one case, the researchers have mapped out a very general phase diagram of the stuffed honeycomb lattice that interpolates between a triangular and honeycomb lattice. To accomplish this, they have to find the lowest eigenvector of matrices, which are 68 billion by 68 billion using a parallel exact diagonalization code the team developed. In an alternative approach, they have searched for a numerical Hamiltonian that best fits the collaborators’ experiment on a spin-ice material [5].

The team has approached better quantum mechanical simulations in two ways: by using deep neural networks and also by using an inverse approach. In the first case, they use deep neural networks to represent the quantum wave function. Wave functions map electron positions to scalar amplitudes. To generate this amplitude, they take a configuration and have a deep neural net generate a matrix whose determinant is then evaluated [6]. This builds on other approaches [7] that have neural networks directly generate the amplitude. In the second case, the researchers have tackled the quantum many-body problem using an inverse approach that avoids the exponential cost of the forward method [8]. The team’s algorithm takes a targeted set of properties (encoded as a wave function) and outputs the Hamiltonians that might have generated it. This turns out to be extremely useful because it is easy to write down wave functions with interesting and exotic physics.

RESULTS & IMPACT

The research group has determined that the phase diagram of the stuffed honeycomb lattice supports nine different phases. One of these phases is a spin-liquid phase that significantly expands the known spin-liquid regime on the triangular lattice [9]. This increases the chance that experimentalists might be able to find spin-liquid behavior in real materials.

In eigenstate phases, the team has discovered an entirely new type of eigenstate phase [10], making it the second nontrivial concrete example of this type of phase. This phase has eigenstates with two different types of intermixed eigenstates: some of the eigenstates have entanglement that grows logarithmically with system size, and some are constant. The implications of this discovery is that whether the system equilibrates (*i.e.*, the coffee cup cools) depends sensitively on the starting conditions of the system.

The two algorithms the team has developed have significantly improved the regime of simulatable systems. The machine learning methodology extrapolates to the correct answer on difficult Hubbard systems in the regime where the researchers believe superconductivity exists and competes favorably with other state-of-the-art methods including the density matrix renormalization group. The inverse approach the team has developed has allowed them to find a whole class of Hamiltonians that support a spin-liquid-like state.

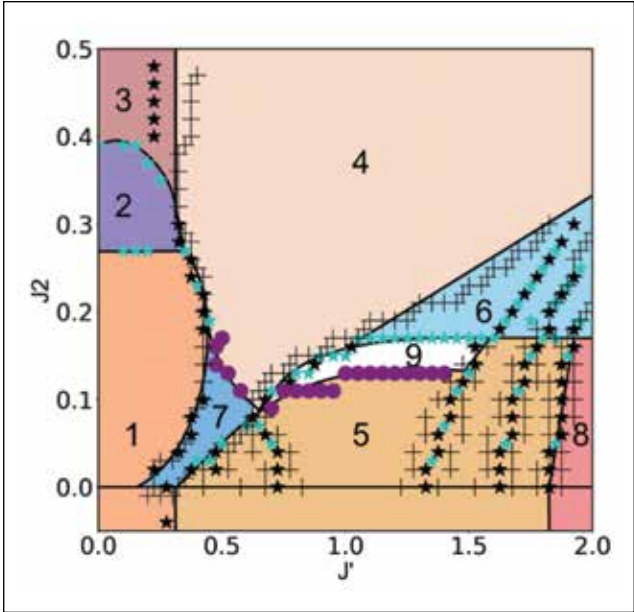


Figure 2: Phase diagram of the Heisenberg model on the stuffed honeycomb model including all nine phases. Phase nine is the spin-liquid phase. Dots and stars indicate different types of fidelity dips signaling a change in the phase.

WHY BLUE WATERS

Without Blue Waters, the research team would not have been able to perform these calculations. The stuffed honeycomb simulations required diagonalization of huge matrices at over 200 different phase points. For the eigenstate phases, the researchers needed ~100 different realizations for each of five different disorder strengths. Even testing and benchmarking the new algorithms was a significant undertaking computationally; for example, the deep neural network approaches scales of  $N^4$ , where  $N$  is the size of the system. Without being able to run these simulations in parallel, the team would not have been able to obtain results.

PUBLICATIONS & DATA SETS

- E. Chertkov and B. K. Clark, “Computational inverse method for constructing spaces of quantum models from wave functions,” *Phys. Rev. X*, vol. 8, no. 3, p. 031029, 2018.
- X. Yu, D. Luo, and B.K. Clark, “Beyond many-body localized states in a spin-disordered Hubbard model with pseudo-spin symmetry,” *Phys. Rev. B*, vol. 98, no. 11, p. 115106, 2018.
- J. Sahoo *et al.*, “Classical phase diagram of the stuffed honeycomb lattice,” *Phys. Rev. B*, vol. 98, no. 13, p. 134419, 2018.
- H. J. Changlani *et al.*, “Resonating quantum three-coloring wave functions for the Kagome quantum antiferromagnet,” *Phys. Rev. B*, vol. 99, no. 10, p. 104433, 2019.
- D. Luo and B. K. Clark, “Backflow transformations via neural networks for quantum many-body wave functions,” *Phys. Rev. Lett.*, vol. 122, no. 22, p. 226401, 2019.
- D. Reig-i-Plessis *et al.*, “Deviation from the dipole-ice model in the spinel spin-ice candidate  $\text{MgEr}_2\text{Se}_4$ ,” *Phys. Rev. B*, vol. 99, p. 134438, 2019.



# DIRECT NUMERICAL SIMULATION OF PRESSURE FLUCTUATIONS INDUCED BY SUPERSONIC TURBULENT BOUNDARY LAYERS

**Allocation:** NSF PRAC/200 Knh  
**PI:** Lian Duan<sup>1</sup>  
**Collaborator:** Meelan Choudhari<sup>2</sup>

<sup>1</sup>Missouri University of Science and Technology  
<sup>2</sup>NASA Langley Research Center

## EXECUTIVE SUMMARY

When air flows over solid surfaces at high speeds, the very thin region near the surface that is referred to as the boundary layer can become chaotic and turbulent; these turbulent motions can then generate intense, outward-propagating sound waves. In particular, for fast flows inside a supersonic wind tunnel, the turbulent boundary layer over the tunnel wall radiates outward-propagating sound waves and causes the generation of freestream acoustic noise in the wind tunnel. This research exploited the cutting-edge computational power of the Blue Waters to advance fundamental understanding of the generic statistical and spectral features of acoustic radiation from high-speed turbulent boundary layers. Such an understanding helps define the freestream disturbances environment in supersonic/hypersonic wind tunnels and allows more accurate extrapolation of experimental measurements from noisy wind tunnels to free flight.

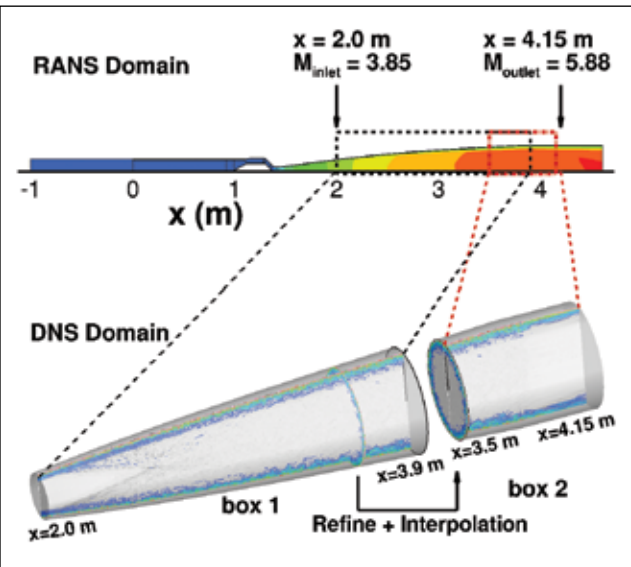


Figure 1: Hypersonic Ludwig Tube at the Technical University of Braunschweig. The DNS domain covers the origin of most of the acoustic sources responsible for generating freestream noise in the test section.

## RESEARCH CHALLENGE

Testing in conventional (noisy) wind tunnels has been an important means of characterizing and understanding when, where, and how a high-speed boundary layer flowing over an aerodynamic body transitions to turbulence, which causes a large increase in skin-friction drag and surface heating. Because the existing low-disturbance (*i.e.*, quiet) facilities operate only at Mach 6, moderate Reynolds numbers (the ratio of inertial forces to viscous forces within a fluid), fairly small sizes, and low freestream enthalpy (total heat content of a system), conventional facilities will continue to be employed for testing and evaluation of high-speed vehicles, especially for ground testing involving other Mach numbers, higher freestream enthalpies, and larger models. To enable better use of transition data from conventional facilities and more accurate extrapolation of wind tunnel results to flight, one needs an in-depth knowledge of the broadband disturbance environment in those facilities, which is dominated by acoustic radiation from tunnel wall turbulent boundary layers.

## METHODS & CODES

Direct numerical simulations (DNS) were conducted using HyperWENO, an in-house high-order finite-difference solver that solves the compressible Navier–Stokes equations describing the evolution of the density, momentum, and total energy of the flow. The governing equations can be described and solved in either general curvilinear coordinates or cylindrical coordinates, depending on the flow configuration. The inviscid fluxes of the governing equations were computed using a seventh-order weighted essentially nonoscillatory (WENO) scheme. Compared with the original finite-difference WENO introduced by Jiang and Shu [1], the present scheme is optimized by means of limiters [2] to reduce the numerical dissipation; WENO adaptation was limited to the boundary-layer region for maintaining numerical stability while the optimal stencil of WENO was used outside the boundary layer for optimal resolution of the radiated acoustic field. A fourth-order central difference scheme was used for the viscous flux terms, and a third-order low-storage Runge–Kutta scheme [3] was employed for time integration, which significantly relieved the memory requirement and is well suited for time-accurate simulations such as DNS. The turbulent inflow can be generated using either a recycling/rescaling method [4] or a digital filtering method [5]. On the wall, no-slip con-

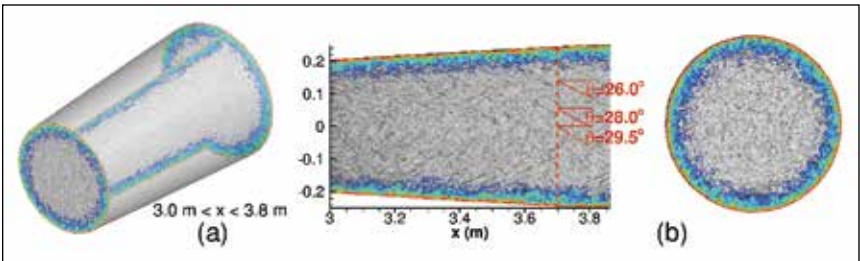


Figure 2: Numerical schlieren images (*i.e.*, density gradient contours) of radiated acoustic waves within the nozzle of the Hypersonic Ludwig Tube. The vertical dashed line indicates the axial location of the selected cross-section visualized in the right panel of (b).

ditions were applied for the three velocity components and an isothermal condition was used for the temperature. At the top and outlet boundaries, unsteady nonreflecting boundary conditions were imposed. Periodic boundary conditions were used in the spanwise or azimuthal direction.

## RESULTS & IMPACT

The current work advanced the state-of-the-art knowledge of the global pressure field induced by supersonic turbulent boundary layers across a wide range of Mach numbers. The study represents the first-ever attempt to exploit the advances in high-performance computing to overcome the difficulties in experimental measurements and to provide access to both flow and acoustic quantities that are difficult to obtain otherwise. In particular, the study led to an unprecedented simulation of a full-scale nozzle of a hypersonic wind tunnel (Fig. 1) and allowed the first successful comparison between numerical predictions and measurements of pressure fluctuations over the nozzle wall. The simulations also captured all major features of the freestream disturbance spectra and structures (Fig. 2) and helped clarify the physics of the noise generation process in supersonic/hypersonic wind tunnels. The characterization of wind tunnel freestream disturbances paved the way for extrapolation to flight from the boundary-layer transition data obtained in noisy wind tunnels.

## WHY BLUE WATERS

DNS are used to capture both the broadband turbulence field within the boundary layer and the near-field acoustic disturbances radiated by the boundary layer. In such simulations, extremely fine meshes are required to fully resolve all the turbulence scales in order to obtain the pressure spectra in the high-frequency/large-wave-number range. In the meantime, the simulations need large domain sizes to locate very-large-scale coherent structures in the pressure field as well as to accommodate the eddy decorrelation length and to minimize inlet transience as a result of inflow boundary conditions. A large number of timesteps are also required for the study of the low-frequency behavior of the pressure spectrum. As such, the proposed computational efforts cannot be conducted without the world-class computing capabilities of Blue Waters.

## PUBLICATIONS & DATA SETS

L. Duan *et al.*, “Characterization of freestream disturbances in conventional hypersonic wind tunnels,” *J. Spacecraft and Rockets*, vol. 56, no. 2, pp. 357–368, 2019, doi: 10.2514/1.A34290.

C. Zhang, L. Duan, and M. M. Choudhari, “Direct numerical simulation database for supersonic and hypersonic turbulent boundary layers,” *AIAA J.*, vol. 56, no. 11, pp. 4297–4311, 2018, doi: 10.2514/1.J057296.

“DNS: Supersonic/hypersonic zero-pressure-gradient plate flows,” webpage in *NASA Langley Turbulence Modeling Resource*, Nov. 2018. [Online]. Available: [https://turbmodels.larc.nasa.gov/Other\\_DNS\\_Data/supersonic\\_hypersonic\\_flatplate.html](https://turbmodels.larc.nasa.gov/Other_DNS_Data/supersonic_hypersonic_flatplate.html)



DI

GA

THE ANOMALOUS MAGNETIC MOMENT OF THE MUON: AN IMPROVED *AB INITIO* CALCULATION OF THE HADRONIC VACUUM POLARIZATION CONTRIBUTION

**Allocation:** Illinois/225 Knh  
**PI:** Aida X. El-Khadra<sup>1</sup>  
**Co-PIs:** Zech Gelzer<sup>1</sup>, Ruth Van de Water<sup>2</sup>  
**Collaborators:** Shaun Lahert<sup>1</sup>, Carleton DeTar<sup>3</sup>, Steve Gottlieb<sup>4</sup>, Andreas Kronfeld<sup>3</sup>, Doug Toussaint<sup>5</sup>

<sup>1</sup>University of Illinois at Urbana-Champaign  
<sup>2</sup>Fermi National Accelerator Laboratory  
<sup>3</sup>University of Utah  
<sup>4</sup>Indiana University  
<sup>5</sup>University of Arizona

CI

EXECUTIVE SUMMARY

A central goal of high-energy physics is to search for new particles and forces beyond the Standard Model of particle physics. The muon anomalous magnetic moment is sensitive to contributions from new particles and is also one of the most precisely measured quantities in particle physics, with an experimental uncertainty of 0.54 parts per million. At present, the measurement disagrees with Standard Model theory expectations by more than three standard deviations.

The Fermilab Muon g-2 Experiment will ultimately reduce the experimental error by a factor of four; first results are expected in the fall of 2019. To identify definitively whether any deviation observed is due to new particles or forces, the theory error must be reduced to a commensurate level. An ongoing project by this collaboration uses numerical lattice quantum chromodynamics (QCD) to target the hadronic vacuum polarization contribution, which is the largest source of theory error. This Blue Waters project targets one of the largest sources of error in the current calculation—the statistical errors at large Euclidean times—and com-

plements the ongoing project by calculating additional correlation functions to quantify contributions from two-pion states, which become important at large Euclidean times. While the ensemble to be used in this project is relatively small, the large number of propagator inversions needed to compute the additional correlation functions requires a petascale machine such as Blue Waters.

RESEARCH CHALLENGE

The muon anomalous magnetic moment (g-2) enables a very precise test of the Standard Model of particle physics and a probe of new particles and forces beyond. In the Standard Model, the anomaly arises owing to quantum-mechanical loop contributions. Virtual contributions from new particles could therefore lead to an observable deviation between measurements and the expected Standard Model value for the muon g-2, given sufficiently precise theory and experiment. The most recent measurement of the muon g-2 has a precision of 0.54 parts per million [1] and disagrees with Standard Model theory expectations by more than three standard deviations. The Muon g-2 Experiment at Fermilab (with team members from the University of Illinois at Urbana-Champaign) began running in 2019, and expects to reduce the experimental error by a factor of four [2]. To leverage the anticipated reduction in experimental errors and determine unambiguously whether or not the present disagreement is due to effects from new particles or forces, the theoretical errors on the Standard Model prediction must be brought to a commensurate precision on the experimental timescale.

The dominant source of uncertainty in the Standard Model prediction of the muon g-2 is from the hadronic vacuum polarization contribution owing to virtual quarks and gluons [3] (see Fig. 1), which is the target of this work. Numerical lattice-QCD simulations provide the only method for calculating the nonperturbative hadronic contributions to the muon g-2 with controlled uncertainties that are systematically improvable.

The basic quantity from which the hadronic vacuum polarization correction is calculated is a two-point function with vector-current operators at the source and sink [4]. The hadronic

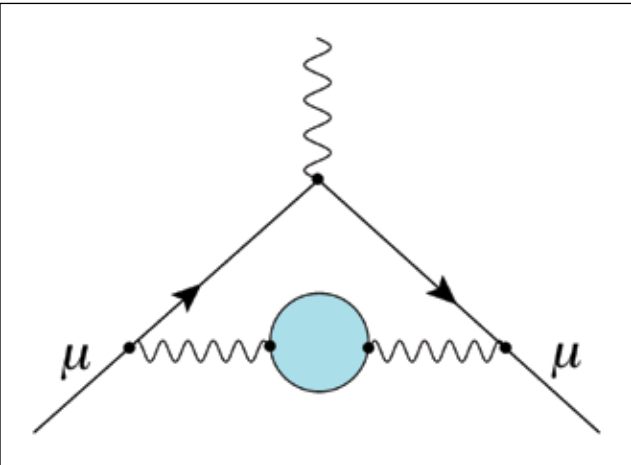


Figure 1: Hadronic vacuum polarization contribution to the muon’s anomalous magnetic moment.

vacuum polarization is obtained from integrals of this correlation function over Euclidean time. However, at large Euclidean time the vector-current correlation function receives contributions from two-pion states that cannot be resolved from a statistical analysis of the vector-current correlation function alone, because its statistical errors increase rapidly with the Euclidean time. In the joint project described above, these contributions are currently estimated phenomenologically from a scalar quantum electrodynamics calculation.

The goal of the project is a direct calculation of the two-pion contribution from correlation functions of two-pion operators, which will provide us with a better description of the contributions to the hadronic vacuum polarization from the large Euclidean time region. This requires the computation of two-, three-, and four-point functions, as illustrated in Fig. 2, including local and smeared source and sink operators, a range of momenta and source sink combinations, and up to three combinations of pion species, resulting in a large number of correlation functions.

METHODS & CODES

This project was carried out using the MIMD lattice computation (MILC) collaboration code [5], which has been in production use on a wide variety of massively parallel computers for over 20 years, with continual improvements to address the community’s evolving science goals and to accommodate changing hardware. The MILC code, which currently consists of about 300,000 lines, makes extensive use of the libraries of the U.S. Lattice-QCD collaboration’s QCD applications programming interface for CPUs ([6] and the QUDA framework for lattice-QCD on GPUs [7,8]). It was part of the Blue Waters acceptance test and one of the applications launched at its dedication and has been running on Blue Waters for over five years with significant im-

provement in performance over this period. Porting the MILC code to Blue Waters’ XE nodes was straightforward, and with the integration of the QUDA libraries, the MILC code also runs efficiently on the XK nodes.

RESULTS & IMPACT

The computations for this project are ongoing. The research team has already generated all of the needed two-point functions and is currently computing the three- and four-point functions, for which the team is reusing the propagators that were generated for the two-point functions. After the computations are finished, the researchers will perform a statistical analysis to determine contributions of the two-pion states to the hadronic vacuum polarization. This work is the first computation of its kind with staggered fermions. If successful, it will open up a new set of observables to be studied on the large library of ensembles with “highly improved staggered fermions” that has been generated by the MILC collaboration.

WHY BLUE WATERS

This project requires the computation of a large (approximately 100) number of propagators on each available configuration in order to obtain the desired correlation functions. While the small ensemble the research team used in this project allowed them to compute each propagator on only three or four nodes, performing all the propagator inversions needed on each of the 10,000 configurations in the ensemble required a petascale resource such as Blue Waters.

Another essential aspect is that the code used in this project had already been adapted and optimized to Blue Waters (with the help of Blue Waters’ staff) in the course of prior projects funded under PRAC allocations.

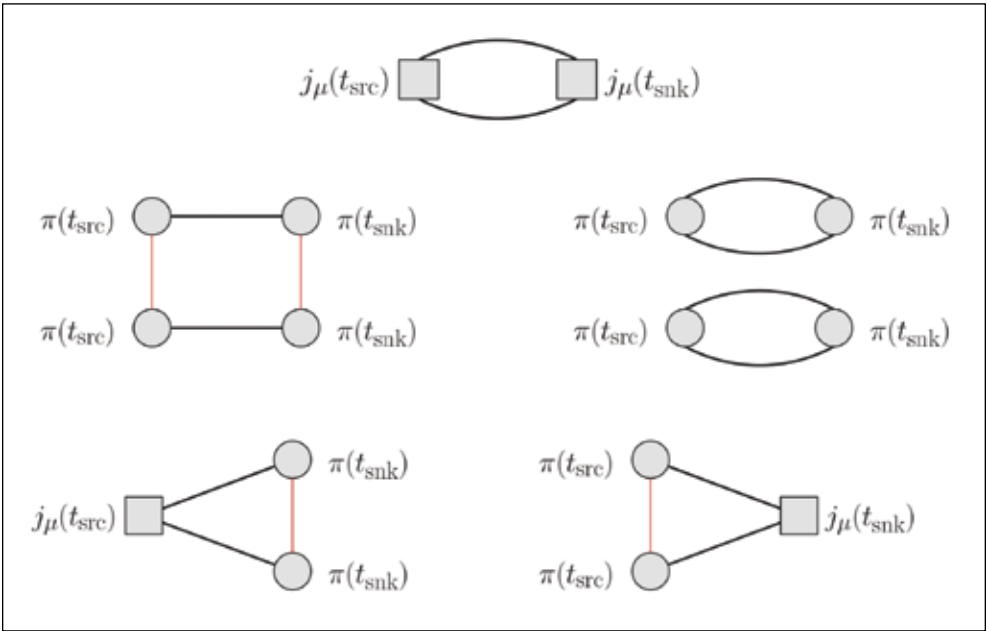


Figure 2: A representative subset of the correlation functions to be computed in this work. A single square at the source or sink indicates a fermion bilinear vector-current operator, while two circles with (without) a red line represent connected (disconnected) two-pion operators. The temporal locations are abbreviated as  $t_{src}$  and  $t_{snk}$  and the black lines indicate propagation in time.



# ACCELERATING THERMOELECTRIC MATERIALS DISCOVERY VIA DOPABILITY PREDICTIONS

**Allocation:** Illinois/185 Knh  
**PI:** Elif Ertekin<sup>1</sup>  
**Collaborators:** Vladan Stevanovic<sup>2</sup>, Eric Toberer<sup>2</sup>, G. Jeffrey Snyder<sup>3</sup>, Michael Toney<sup>4</sup>

<sup>1</sup>University of Illinois at Urbana–Champaign  
<sup>2</sup>Colorado School of Mines  
<sup>3</sup>Northwestern University  
<sup>4</sup>Stanford University

## EXECUTIVE SUMMARY

Nowadays, the experimental realization of new thermoelectric materials—which convert wasted heat to useful electricity—is greatly accelerated through computational guidance, with first-principles calculations providing relevant materials properties such as electronic and phonon transport and thermodynamic stability. Semiconductors are called semiconductors because they exhibit intermediate electrical conductivity (better than typical insulators but not as good as metals). For thermoelectrics, researchers need to improve their electrical conductivity by “doping” the semiconductor. Doping refers to the intentional introduction of defects in the material that help to improve the electrical conductivity—typically by several orders of magnitude.

The current bottleneck in the discovery of new thermoelectric materials, however, is that greater than 90% of the materials identified as promising on a computer cannot be doped in the laboratory. To overcome this, the research team uses Blue Waters to carry out dopability assessments through defect calculations on a diverse set of candidate semiconductors including Zintl compounds, sulvanites, and diamondlike semiconductors. The systems simulated on Blue Waters will be incorporated into TEdesignLab, an online open platform for easy, searchable data exchange to enable data-driven approaches to thermoelectrics design and discovery. This work would not be possible without Blue Waters, which to date has allowed the calculation of full defect properties of a class of approximately 30 candidate thermoelectric semiconductors.

## RESEARCH CHALLENGE

Of the 101.2 quadrillion BTUs of energy consumed in the United States in 2018, only 32.3% was used to carry out useful work while the other 67.7% was wasted, largely as heat loss to the environment [1]. This wasted energy provides a huge opportunity to reduce carbon emissions and improve carbon efficiency, if only it could be captured and harnessed for further use. A thermoelectric material is one that responds to a temperature gradient by creating electricity; these materials have the potential for large-scale conversion of waste heat to usable energy. So far, however, thermoelectric materials are not yet efficient enough for large-scale deployment and utilization. New materials need

to be discovered that are good at thermoelectric energy conversion, but thermoelectric energy conversion requires a strict and rare set of properties: the materials need to exhibit high electrical conductivity and low thermal conductivity.

Over the last several years, members of the team have used high-throughput first-principles calculations of materials properties (electronic, thermal, and thermodynamic stability) to classify the thermoelectric conversion potential of a wide set of materials (Fig. 1a) [2]. Approximately 50,000 high-throughput calculations were conducted; distilled results of these calculations are available at the TEdesignLab website ([www.tedesignlab.org](http://www.tedesignlab.org)) [3]. This initial effort uncovered a vast array of new candidates but also revealed a new limitation: most of the semiconductors identified (~90%) have proven not to be dopable in the laboratory—thus limiting their electrical conductivity. The current bottleneck to discovering high-performance thermoelectrics, then, is that experimentally, it has been found that only 10% of the candidate materials identified are readily dopable to the desired concentration.

The ability to control the carrier concentration of semiconductors represents both a scientific and a practical challenge. Scientifically, general understanding of the factors controlling the dopability of semiconductors is currently largely phenomenological and qualitative. This knowledge gap leads to wasted experimental efforts on compounds that cannot be optimized for thermoelectric performance. The research challenge then is to use high-performance computing to accurately select the materials that can successfully be doped, opening up rapid routes to thermoelectric materials design and discovery.

## METHODS & CODES

To overcome the bottleneck of identifying high electrical conductivity semiconductors as candidate thermoelectrics, the research team used Blue Waters to carry out high-throughput quantum mechanical calculations of semiconductor dopability and predict the best candidate materials before they are even synthesized in the laboratory. The dopability of a semiconductor is a result of its defect chemistry: some semiconductors resist doping via formation of intrinsic defects that counteract the effect of the dopant whereas others do not. Therefore, assessing dop-

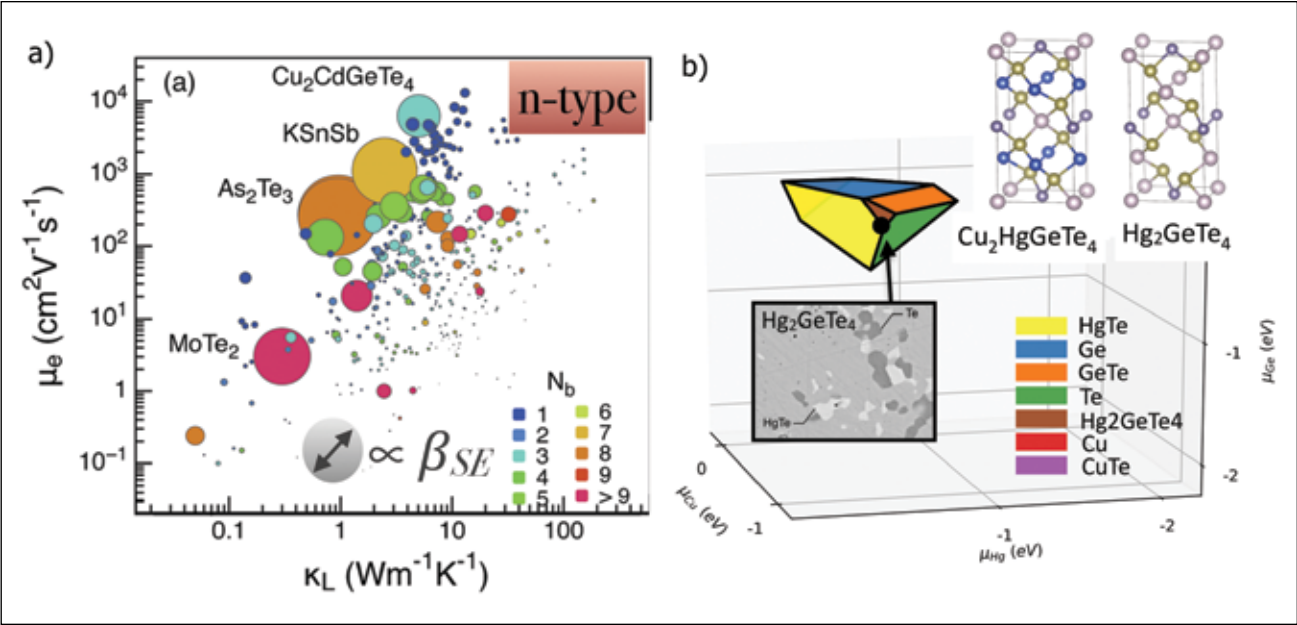


Figure 1: (a) Data set of thermoelectric materials: low thermal conductivity  $\kappa_L$  and high electron mobility  $\mu_e$  is desired. However, the dopability of these materials is generally not known and experimentally intensive to establish. After [2]. (b) High-throughput screening of materials on Blue Waters enables large-scale dopability predictions to accelerate thermoelectric materials discovery.

ability requires a comprehensive analysis of the possible intrinsic defects that could form in a material in response to any attempts to dope it. Identifying low-energy defects requires solving the quantum mechanical Schrödinger equation to determine the defect formation energy, which governs how readily the defect may form. For this, the team used first-principles quantum mechanical methods based on the framework of density functional theory, which approximates the Schrödinger equation to make it solvable.

As the new candidate materials being considered are structurally complex ternary and quaternary semiconductors, a large number of intrinsic defects are possible, requiring a large set of quantum mechanical simulations to accurately predict dopability and the laboratory growth environments that will maximize it. To streamline the simulation process, the team has developed an automated simulation workflow and protocol on Blue Waters to facilitate the large-scale analysis. This has enabled us to establish a closed feedback loop between simulations and experimental validation, enabling rapid screening of candidate materials in a manner not previously demonstrated.

## RESULTS & IMPACT

To date, using Blue Waters the team has carried out systematic investigations of a set of approximately 30 new candidate thermoelectric materials. The group has helped to predict several new ternary and quaternary materials that exhibit the desired degree of dopability, which have subsequently been synthesized in the laboratory or are soon to be synthesized [4–6]. Over the course of this project, the researchers expect this work to create

the largest data set of semiconductor dopability focused on complex ternary and quaternary thermoelectric materials. Selected results highlighting some new materials identified using this approach are shown in Fig. 1b.

## WHY BLUE WATERS

Practically, the key hurdle to this computation-driven approach is the need to calculate a large number of intrinsic and extrinsic point defects to directly assess dopability of semiconductors. The large computational costs historically had limited dopability calculations to case-by-case studies in which one material was considered at a time in a serial manner. Harnessing the predictive power of computation to model dopability on a large scale is critical in accelerating the transition from novel semiconductors to functional devices. Only Blue Waters offers the large-scale computational resources that are essential to the research team's integrated simulation/experimental approach to the discovery of new thermoelectric materials.

## PUBLICATIONS & DATA SETS

T. Zhu and E. Ertekin, “Mixed phononic and non-phononic transport in hybrid lead halide perovskites: glass-crystal duality, dynamical disorder, and anharmonicity,” *Energy Environ. Sci.*, vol. 12, pp. 216–229, Jan. 2019, doi: 10.1039/C8EE02820F.



MACHINE-LEARNING TURBULENCE MODELS FOR SIMULATIONS OF TURBULENT COMBUSTION

Allocation: Strategic/75 Knh  
PI: Jonathan Freund<sup>1</sup>  
Co-PIs: Justin Sirignano<sup>1</sup>, Jonathan MacArt<sup>1</sup>

<sup>1</sup>University of Illinois at Urbana–Champaign

EXECUTIVE SUMMARY

Predictive simulations of turbulent combustion are crucial to the design of energy-conversion systems in the transportation, power, and defense sectors, among others. Due to their multi-scale, multiphysics nature, these systems are typically intractable to simulation at full resolution. Lower-resolution simulations are possible but require closure models; currently, state-of-the-art closure models fail to capture key dynamics in certain regimes of turbulent combustion, and the use of these models can lead to incorrect predictions.

The research team has developed a novel approach for the design of turbulence closure models utilizing machine learning (ML) techniques and large-scale, fully resolved computational data sets. The team trains and refines neural network-based models across the problematic regimes, then refines the models using an adjoint solution, which is analogous to an on-the-fly error-control procedure. The resulting ML-based models are more accurate than the most commonly used models and can be generalized to different applications. Alternatively, the deep learning model can be harnessed to produce predictions of similar accuracy to those of existing models at reduced computational cost.

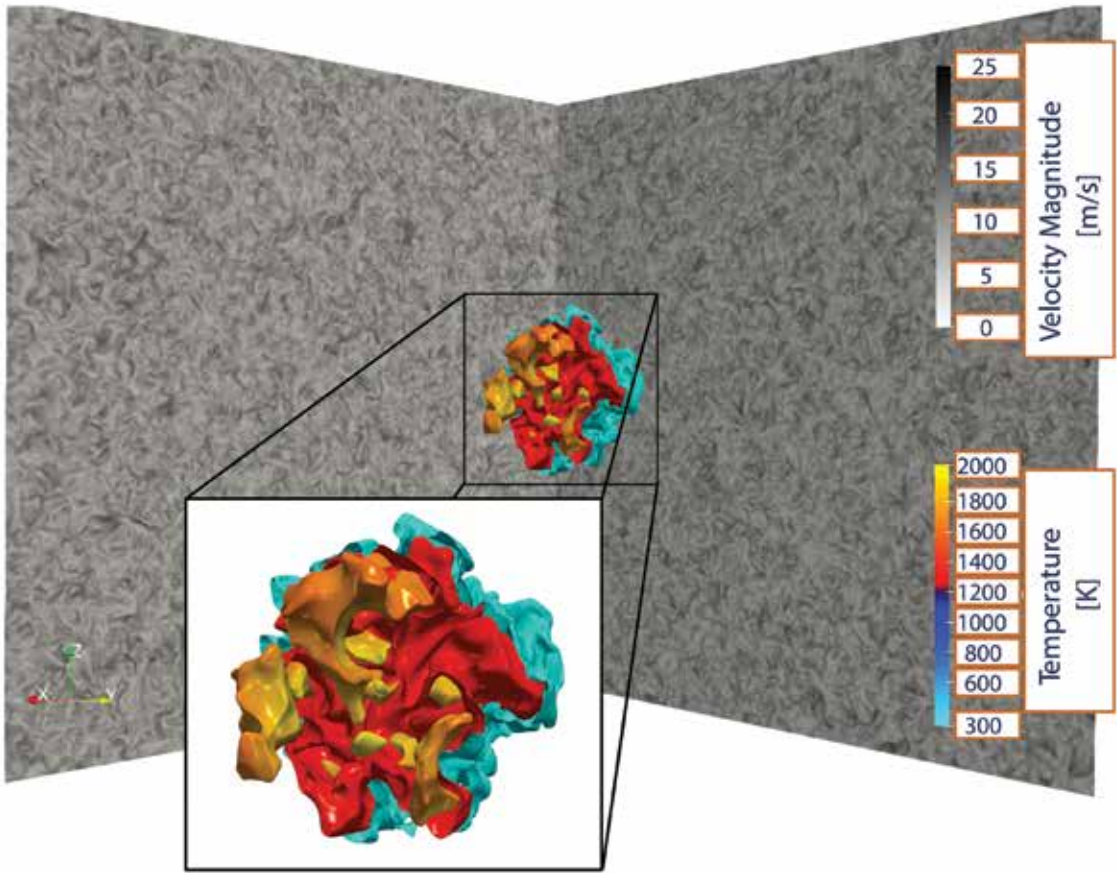


Figure 1: Illustration of a direct numerical simulation (DNS) of an expanding hydrogen–air premixed flame kernel (colored surfaces) in isotropic turbulence (grayscale background). The simulation is discretized using  $1,024^3$  grid points and utilizes up to 16,384 cores. Machine learning models trained on many DNS databases are tested in analogous large-eddy simulations (LES).

RESEARCH CHALLENGE

Turbulent combustion is an inherently multiscale, multiphysics phenomenon relevant to virtually every sector in which chemical energy is converted to mechanical energy. Modern combustors for gas turbines, internal combustion engines, rocket propulsion, and hypersonic flight all require turbulent flow to enhance fuel–oxidizer mixing and to increase burning rates. Accurate prediction of the performance, efficiency, and emissions of these devices is an essential aspect of the engineering design and test cycle.

However, the most common closure models used to make these predictions computationally tractable are based on nonreacting turbulence theory and are known to fail in regimes relevant to the next generation of clean combustors. One key challenge is the prediction of turbulence–combustion interactions in premixed flames. These interactions are necessarily precluded by the use of nonreacting turbulence models [1–3]. Linearly coupled reacting-turbulence models have been developed [3,4] but are of limited use owing to their exclusion of nonlinear interactions. The research team therefore focuses on developing nonlinear models that are capable of capturing such interactions.

METHODS & CODES

The researchers have developed a machine learning approach to turbulence model development with a focus on predicting nonlinear turbulence–combustion interactions. Numerical databases from fully resolved direct numerical simulations (DNS) of turbulent combustion form the basis of these model-development efforts. The team generates DNS databases across a range of turbulent combustion regimes using the semi-implicit, second-order, energy-conservative code NGA [5,6]. These databases are subsequently downsampled using a low-pass filter to obtain flow fields comparable to those obtained from a large-eddy simulation (LES; *i.e.*, modeled) calculation. Because the filtered fields originate from full-resolution data, the “true” model outputs are also available. Using deep neural networks as a nonlinear statistical model, the team approximates these “true” outputs from input variables that are available in LES. The resulting models are tested *a priori* using out-of-sample DNS data and *a posteriori* by implementation in analogous LES calculations. As a novel approach to model development, they have developed a new DNS/LES code, PyFlow, that is capable of refining the model during the *a posteriori* step using an adjoint solution. This code is GPU-accelerated and has potential to address modeling challenges over a wide range of flows.

RESULTS & IMPACT

The initial results show an encouraging ability of the machine learning-based models to predict turbulence–combustion interactions more accurately than traditional turbulence models. These findings are obtained from *a priori* testing and demonstrate that ML-based models have the potential to supplant traditional turbulence models in LES. However, in *a posteriori* testing, the re-

searchers found that numerical errors (including finite-difference errors) owing to the reduced LES resolution accumulate in neural network inputs and reduce the models’ predictive accuracy. The issue of numerical error is unavoidable in LES but can potentially be mitigated using the team’s *a posteriori* model refinement step. By training *in situ*, models can be developed that are less sensitive to numerical error, for example, by reducing weights associated with error-prone inputs and hidden parameters. The *a posteriori* training step, therefore, represents a potentially significant contribution to ML-based model development and LES of turbulent combustion.

WHY BLUE WATERS

This research relies on both CPU-only and GPU-accelerated compute nodes for different tasks. Because Blue Waters offers both types of nodes on a common file system, data access and sharing are greatly streamlined among tasks and project members. Additionally, the team utilizes Blue Waters-specific Python installations and packages that are supported by NCSA project staff. These installations are optimized for Blue Waters and offer substantially improved application performance compared to user-installed libraries.



# TURBULENCE-RESOLVING MODELING OF OSCILLATORY BOUNDARY LAYER FLOWS

**Allocation:** Illinois/993.725 Knh  
**PI:** Marcelo García<sup>1</sup>  
**Co-PI:** Paul Fischer<sup>1</sup>  
**Collaborators:** Dimitrios K. Fytanidis<sup>1</sup>, Jose M. Mier<sup>1</sup>

<sup>1</sup>University of Illinois at Urbana–Champaign

## EXECUTIVE SUMMARY

Oscillatory boundary layer flows play an important role in coastal engineering, offshore engineering, and coastal sediment transport. However, current state-of-the-art models fail to accurately predict the complex interaction of turbulent oscillatory flow with sediment transport, highlighting the existing knowledge gaps regarding the complex interactions among the oceanic flow, the coastal bottom, and sedimentation processes. Recent experimental and numerical studies conducted by the research team indicate the presence of a phase-lag between the time instance when the maximum bed shear stress occurs with respect to the maximum free-stream velocity in transitional oscillatory boundary layer flows. However, the effect of turbulent coherent structures and turbulence modulation under severe acceleration remain unknown. The current work is the first computational effort to simulate the effect of flow coherent structures and turbu-

lent events such as sweeps and ejections on the maximum bed shear stress phase difference compared to the maximum free-stream velocity value.

## RESEARCH CHALLENGE

Recent experimental studies conducted in the Large Oscillatory Water-Sediment Tunnel at the Ven Te Chow Hydrosystems Laboratory at the University of Illinois at Urbana–Champaign examine the transition between the laminar and turbulent flow regimes with a smooth bed. The results indicate a significant change in the widely used phase difference diagram between the maximum bed shear stress and the maximum free-stream velocity [1]. This observation is extremely important for the field of environmental fluid mechanics and coastal sediment transport. Nevertheless, due to the limitation of the applied pointwise experimental technique (Laser Doppler Velocimetry), it was not pos-

sible to explicitly associate this finding with the development of three-dimensional flow turbulence structures usually referred to as turbulence coherent structures.

This work is the first computational effort to quantify the effect of the three-dimensional turbulent flow structures on the phase difference between the maximum bed shear stress and the maximum free-stream velocity. Also, it will be the first numerical study that will quantify the effects of flow regime and bed characteristics on the turbulent characteristics and quadrant analysis under oscillatory flow conditions. It will also be among the first studies that will study the momentum exchange between the free-stream oscillatory flow and the seabed unsteady flow conditions.

## METHODS & CODES

In this work, the research team developed a direct numerical simulation (DNS) model capable of simulating the complex oscillatory boundary layer flow using the spectral element method framework provided by the highly scalable open-source code Nek5000 [2]. Except for the analysis of turbulence characteristics of oscillatory boundary layer flow over different flow conditions, the present work requires use of a proper model for the simulation of the suspended sediment using an Eulerian approach and proper boundary conditions for the sediment mass exchange between the coastal bed and the free-stream flow (*e.g.*, [3-4]).

Due to the data-rich outputs of our simulations, the team is collaborating with the Data Analytics and Visualization (DAV) group of the National Center for Supercomputing Applications in an effort to find an efficient way to visualize the coherent flow structures and the numerical results. The team plans to apply high-performance, data-intensive visualization and analysis techniques by means of producing high-quality, interactive visualizations of simulation results in an effort to uncover new knowledge through the efficient analysis of the information-rich data. Special attention will be given to the application of efficient methods for the estimation of the geometric characteristics of coherent flow structures as well as the quantification of their effect on flow behavior.

## RESULTS & IMPACT

DNS results for mean flow and turbulent statistics were compared against previous experimental and numerical observations [1,5] and the comparison agrees well both qualitatively and quantitatively. Quadrant analysis shows that turbulent events such as sweeps and ejections dominate for most of the period. Phase difference results agree well with the previous experimental findings.

This work is the first in the literature that explores the effect of turbulence characteristics of the flow—and particularly the turbulent flow structures, such as turbulent spots (Fig. 1)—on the phase difference between maximum bed shear stress and free-stream velocity. The team identified hairpin vortices for the first time in the oscillatory boundary layer in addition to other flow structures previously reported in the literature (vortex tubes and turbulent spots). They studied the effect of these structures on the turbulence statistics and found that vortex tubes seem to have minimal effect. On the contrary, turbulent spots, which are spatially and temporarily sporadic, lambda-shaped, highly energetic structures, have a significant effect on turbulence characteristics.

## WHY BLUE WATERS

The present work pushes the limit of the turbulent-resolving flow modeling of oscillatory flows. The dimensions of the computational domain were chosen based on the prior knowledge of experimental observation of turbulent spots [6,7] to ensure that the computational domain is big enough to allow these turbulent structures to develop. This size is larger than any of the previous domains reported in the literature, and together with the increased number of computational points (on the order of 0.8 billion) make this study the first of its kind in terms of the computational resources and the high-performance computing facilities it requires. Thus, it can be materialized only on a petascale supercomputer such as Blue Waters.

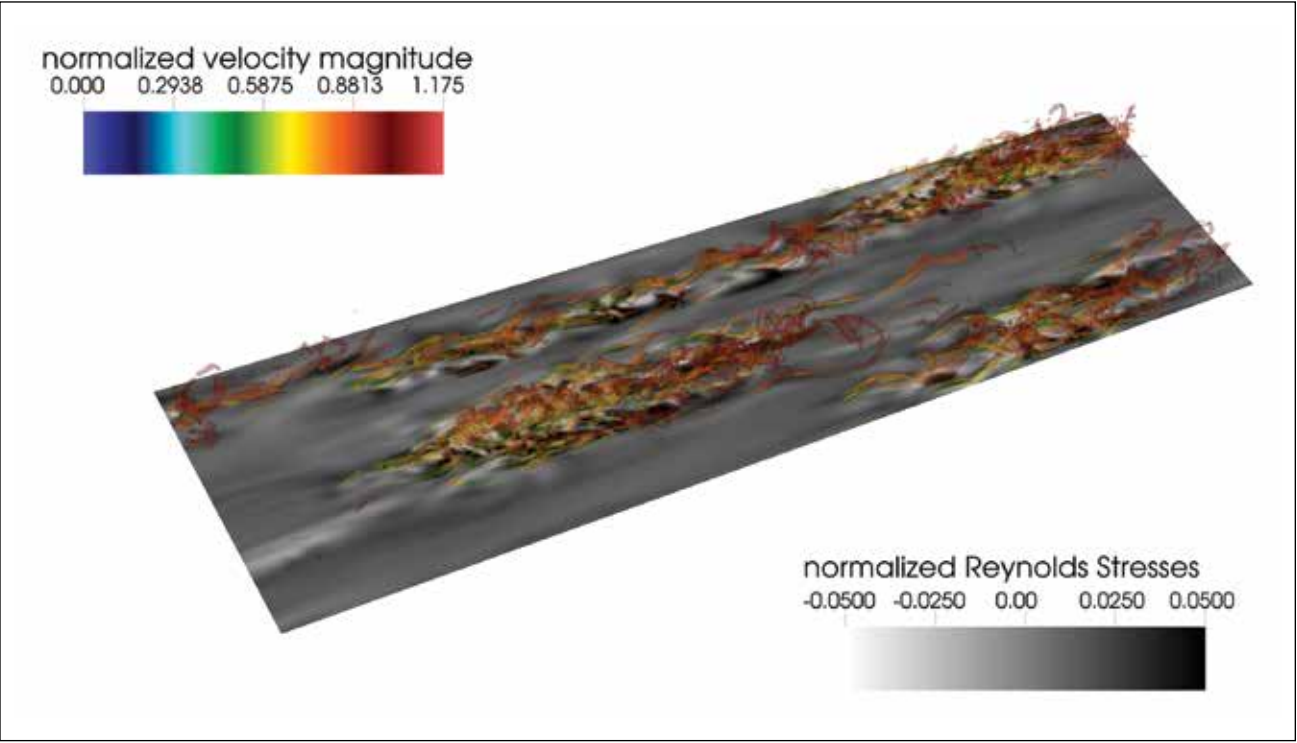


Figure 1: Turbulent spot visualization under oscillatory flow conditions.



MACHINE LEARNING FOR PARTICLE PHYSICS: EMPLOYING DEEP LEARNING FOR PARTICLE IDENTIFICATION AND MEASUREMENT AT COLLIDERS

Allocation: Illinois/125 Knh  
PI: Benjamin Hooberman<sup>1</sup>  
Co-PI: Amir Farbin<sup>2</sup>  
Collaborator: Matt Zhang<sup>1</sup>

<sup>1</sup>University of Illinois at Urbana–Champaign  
<sup>2</sup>University of Texas at Arlington

EXECUTIVE SUMMARY

The Large Hadron Collider (LHC) at CERN, in Switzerland, is the world’s most powerful particle accelerator. The LHC recreates the conditions of the Universe one tenth of a nanosecond after the Big Bang by colliding together protons traveling at 0.99999997 times the speed of light 40 million times every second. Each proton–proton collision creates up to several hundred particles that pass through one of four detectors situated at the LHC interaction points.

Reconstructing the collisions requires identifying these particles using their signatures in the detector. Recent advances in machine learning and artificial intelligence, known as deep learning, have made it possible to apply learning networks to many kinds of problems. In particular, identifying particles from their energy deposition in calorimeter cells bears a strong resemblance to problems in machine vision, in which objects are reconstructed from intensity values in pixel arrays. The research team has exploited deep learning techniques to identify and measure particles produced at colliders and have found that they provide improvements in performance with respect to conventional methods.

RESEARCH CHALLENGE

The LHC recreates the conditions of the Universe a tenth of a second after the Big Bang by colliding together high-energy protons. In 2012, the Higgs boson was discovered in LHC data, completing the Standard Model of particle physics and leading to the Nobel Prize in Physics in 2013. This discovery transformed our understanding of the building blocks of matter and the fundamental forces by explaining the origin of the masses of subatomic particles.

However, the Standard Model is not capable of resolving key open questions and thus cannot be the final theory of nature. In particular, it cannot explain the origin of dark matter, which comprises about five times as much total mass in the Universe as visible matter but whose nature is not understood. Various beyond-the-standard-model scenarios, including supersymmetry and extra dimensions of spacetime, have been posited to resolve these problems. These scenarios generically predict the existence of exotic new particles, which may be produced at LHC. Searching for these particles to understand the nature of physics beyond

the Standard Model is now the highest priority of the LHC physics program and the focus of this project.

Analyzing LHC data to search for physics beyond the Standard Model requires identifying and measuring the particles produced in proton–proton collisions. Particles produced in collisions traverse detectors, depositing their energy in calorimeters consisting of a granular array of detecting elements (pixels). The resulting image can be analyzed to distinguish among the six species of stable particles (electrons, photons, charged hadrons, neutral hadrons, and muons) and infer their energies. Electrons and photons are expected signatures of a wide variety of interesting new physics scenarios but may be mimicked by charged and neutral hadrons, which are produced at rates that are higher by several orders of magnitude. Since each collision contains typically thousands of particles, discriminating signals from electrons and photons from hadronic backgrounds is complicated by the presence of additional overlapping particles. Identifying and measuring electrons and photons, especially those with low energy, is thus a major challenge of high-energy physics.

METHODS & CODES

Recent advances in machine learning and artificial intelligence, known as deep learning, have made it possible to apply learning networks to many kinds of problems. These techniques are driven by the emergence of large data sets, powerful graphical processing unit (GPU) processors, and new techniques to train billion-neuron multilayer artificial neural networks (NN). In computer vision, deeply connected neural networks (DNN) and convolutional neural networks (CNN) have provided dramatic improvements in performance and speed with respect to conventional algorithms and require minimal engineering.

The research team employed DNNs and CNNs to distinguish among signals from electrons and photons and hadronic backgrounds and measure particle energies. The team simulated samples of individual electron, photon, charged hadron, and neutral hadron images in a simple high-granularity calorimeter detector implemented with the Geant4 simulation toolkit. These images were used to train NNs, using PyTorch, that distinguish between electrons vs. charged hadrons and photons vs. neutral hadrons and to measure the energies of the four particle species. To optimize the network architectures, the scientists varied the NN hyperparameters, including the number of NN layers (depth), number of neurons per layer (width), and the learning and dropout rates.

RESULTS & IMPACT

The research team evaluated the performance of DNNs and CNNs trained on particle images and compared the results to the current state-of-the-art algorithms widely used in particle physics. These algorithms employ NNs and boosted decision trees (BDTs) to analyze a precomputed set of particle features such as the calorimeter shower depth and width. For both classification and energy measurement using regression, the team found that the deep NNs provided significant improvements compared to

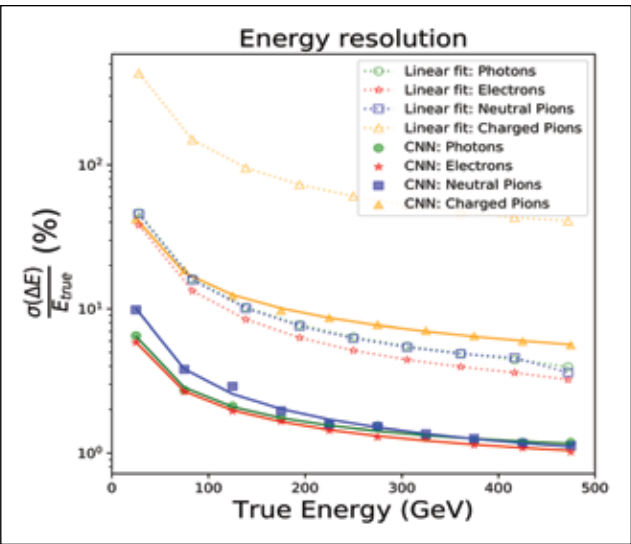


Figure 2: The relative energy resolution of four types of particles vs. the true particle energy for a (dashed line) simple linear fit to the total calorimeter energies vs. a (solid line) convolutional neural network.

the conventional methods. These results serve as a first step toward implementing deep learning for particle identification and measurement at the LHC.

WHY BLUE WATERS

Optimizing the network performance using hyperparameter scans requires retraining NNs hundreds or thousands of times, which is especially challenging for memory-intensive networks such as GoogLeNet or ResNet. The 4,228 GPU-enabled XK nodes with 25 TB of GPU accelerator memory available on Blue Waters enable training and optimization of neural networks beyond what has previously been achieved, allowing for detailed investigations of their behavior for both particle physics and general applications.

PUBLICATIONS & DATA SETS

B. Hooberman *et al.*, “Calorimetry with deep learning: particle classification, energy regression, and simulation for high-energy physics,” *Proc. Deep Learning for Physical Sciences Workshop at the 31st Conf. on Neural Information Process*, Long Beach, CA, U.S.A., Dec. 8, 2017. [Online.] Available: [https://dl4physicsciences.github.io/files/nips\\_dlps\\_2017\\_15.pdf](https://dl4physicsciences.github.io/files/nips_dlps_2017_15.pdf)

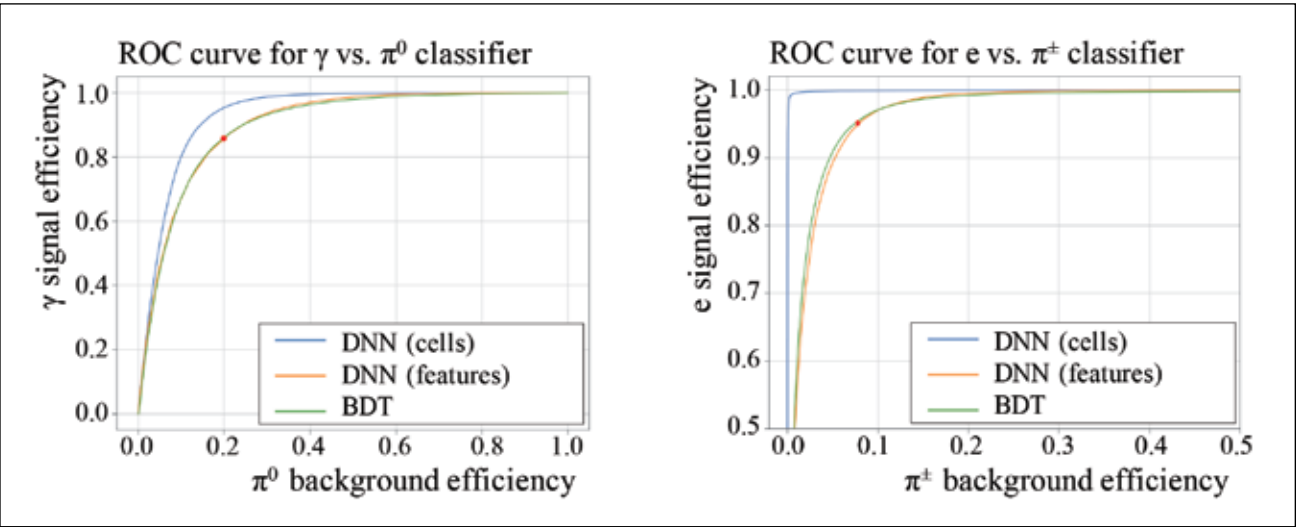


Figure 1: Signal vs. background efficiency receiver operating characteristic (ROC) curves for (left) photon vs. neutral pion and (right) electron vs. charged pion discrimination, using a boosted decision tree and a cell-based or feature-based deep neural network.



MOLTEN-SALT REACTORS AND THEIR FUEL CYCLES

**Allocation:** Blue Waters Professor/100 Knh  
**PI:** Kathryn Huff<sup>†</sup>  
**Collaborators:** Alexander Lindsay<sup>2</sup>, Andrei Rykhlevskii<sup>1</sup>

<sup>†</sup>University of Illinois at Urbana–Champaign  
<sup>2</sup>Idaho National Laboratory

EXECUTIVE SUMMARY

The Advanced Reactors and Fuel Cycles Group (ARFC) at the University of Illinois at Urbana–Champaign models and simulates nuclear reactors and fuel cycles with a goal of improving the safety and sustainability of nuclear power. In the context of high-performance computing, this work couples multiple physics at multiple scales to improve the design, safety, and performance of advanced nuclear reactors. In particular, thermal–hydraulic phenomena, neutron transport, and fuel reprocessing couple tightly in nuclear reactors. Detailed spatially and temporally resolved neutron flux, temperature distributions, and isotopic compositions can improve designs, characterize performance, inform reactor safety margins, and enable validation of numerical modeling techniques for those unique physics. In this work conducted on Blue Waters, ARFC has demonstrated the capability to simulate coupled, transient neutronics, thermal hydraulics, and fuel reprocessing in multiple advanced, molten-salt-fueled nuclear reactor designs.

RESEARCH CHALLENGE

Nuclear power is an emissions-free, safe source of electricity with unparalleled energy density, baseload capacity, and land-use efficiency, so the world’s energy future increasingly depends on improved safety and sustainability of nuclear reactor designs and fuel cycle strategies. The current state of the art in advanced nuclear reactor simulation focuses primarily on traditional light-water reactor designs. ARFC’s work extends that state of the art by enabling similar modeling and simulation capability for more advanced reactor designs that have the potential to improve the already unparalleled safety and sustainability of nuclear power. High-fidelity simulation of dynamic advanced reactor performance requires development of models and tools for representing unique materials, geometries, and physical phenomena. In particular, insights coupling feedback between the reactor scale and the fuel cycle scale are essential to the deployment and scalability of these innovations in the real world.

METHODS & CODES

Blue Waters has enabled ARFC to develop and test two new significant software solutions: Moltres and SaltProc. Moltres is a first-of-its-kind finite-element application simulating the transient neutronics and thermal hydraulics in a liquid-fueled molten-salt reactor design [1–4]. SaltProc is a highly capable Python tool for fuel salt reprocessing simulation [5–7].

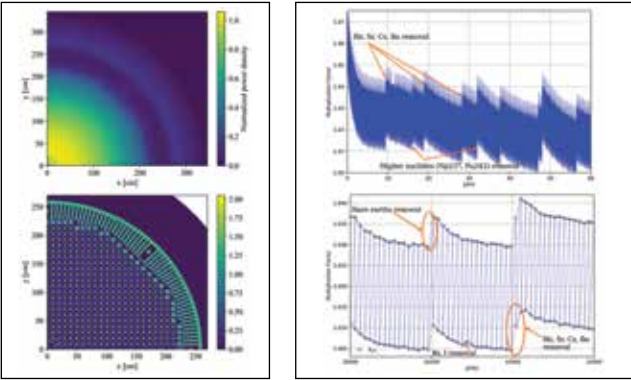


Figure 1 (left): Normalized power density (above) and <sup>232</sup>Th neutron capture rate normalized by total flux (below), both at the equilibrium fuel salt composition in the molten-salt breeder reactor [6]. Figure 2 (right): Reactor criticality dynamics for a full-core molten-salt breeder reactor over a 60-year reactor operation lifetime (above) and a zoomed time interval (below) at three-day time resolution. These were obtained using SaltProc and SERPENT 2, which simulate reprocessing and online refueling [5–7].

Moltres is a collection of physics kernel extensions and material definitions for the highly scalable, fully implicit, multiphysics, object-oriented simulation environment (MOOSE) framework from Idaho National Laboratory [9]. These physics kernels enable Moltres to solve arbitrary-group neutron diffusion, temperature, and precursor governing equations in up to three dimensions and on an arbitrary number of processing units.

Moltres is built on top of the MOOSE framework, which itself leans on the LibMesh [9] finite-element library and the PETSc [10] toolkit for solving systems of discretized partial differential equations (PDEs). MOOSE and LibMesh handle translation of Moltres-defined weak PDE forms into residual and Jacobian functions that PETSc solves simultaneously via Newton–Raphson routines. All codes use MPI for parallel communication and are easily deployed on Blue Waters. MOOSE applications such as Moltres use monolithic and implicit methods that are ideal for closely coupled and multiscale physics.

SaltProc includes fission product removal, fissile material separations, and refueling for time-dependent analysis of fuel-salt evolution. It relies on full-core high-fidelity Monte Carlo simulations to perform depletion computations that require significant computational time and memory. This capability enables unprecedented fidelity in dynamic fuel cycle analysis for liquid-fueled reactors. The analysis requires thousands of full-core simulations for relevant timescales. With Blue Waters, the research team ac-

complished this high-fidelity fuel cycle simulation and provided the field with unprecedented detail in the dynamics of fuel processing for molten-salt reactors.

Figs. 1 and 2 show the effective multiplication factors (criticality) obtained using SaltProc and SERPENT 2. The criticality values were calculated after removing fission products listed and adding the fertile material at the end of cycle time (three days for this work) over the full 60-year lifetime of this reactor concept. The simulation demonstrated in detail how the effective multiplication factor fluctuates significantly as a result of the batch-wise nature of this online reprocessing strategy.

The workflow begins when SERPENT calculates the effective multiplication factor for the beginning of the cycle (there is fresh fuel composition at the first step). Next, it computes the new fuel salt composition at the end of a three-day depletion. The corresponding effective multiplication factor is much smaller than the previous one. Finally, SERPENT calculates  $k_{\text{eff}}$  for the depleted composition after SaltProc applies feeds and removals. The  $k_{\text{eff}}$  increases accordingly since major reactor poisons (e.g., Xe, Kr) are removed, while fresh fissile material (<sup>233</sup>U) from the protactinium decay tank is added.

RESULTS & IMPACT

Current interest in advanced nuclear energy systems and molten-salt reactor (MSR) concepts has illuminated a need for tools that model these systems. By developing such applications in the open, ARFC enables both transparency and distributed collaboration on promising nuclear reactor concepts. Detailed spatially and temporally resolved neutron fluxes, temperature distributions, and changing isotopic compositions can improve designs, help characterize performance, inform reactor safety margins, and enable validation of numerical modeling techniques for unique physics.

ARFC has conducted simulations of multiple molten-salt reactor designs. Steady state, transient, and fuel cycle analysis simulations have been run in 2D as well as 3D by leveraging the Moltres and SaltProc tools developed in the research group.

Recently, fuel cycle dynamics were obtained from depletion and SaltProc reprocessing simulations for a 60-year time frame. A molten-salt breeder reactor full-core safety analysis was performed at the initial and equilibrium fuel salt compositions for various reactor safety parameters such as effective multiplication factor (shown in Fig. 2), neutron flux distributions, temperature coefficients, rod worths, power, and breeding distributions [5–7].

WHY BLUE WATERS

Simulations that faithfully capture this coupling at realistic spatial and temporal resolution are only possible with the aid of high-performance computing resources. To assess nuclear reactor performance under a variety of conditions and dynamic transients, the ARFC group must conduct myriad 2D and 3D finite-element simulations using the MOOSE framework and their in-house-developed modules. Such simulations commonly occupy tens of thousands of CPU cores at a time and vary in completion time. The MOOSE framework has been shown to scale

very well up to 10,000 cores. The ARFC group has demonstrated appropriate scaling for MSR simulation above 20,000 CPU cores (600 Blue Waters nodes). Transient and multiscale simulations, which require greater capability per simulation, are on the horizon for this work. These may occupy up to 100,000 CPU cores at a time. Only a few of those larger simulations will be necessary to enable better understanding of the dynamics in these reactor systems.

PUBLICATIONS & DATA SETS

A. Rykhlevskii, J. W. Bae, and K. D. Huff, “Modeling and simulation of online reprocessing in the thorium-fueled molten salt breeder reactor,” *Ann. Nucl. Energy*, vol. 128, pp. 366–379, Jun. 2019 doi: 10.1016/j.anucene.2019.01.030.

J. W. Bae, J. L. Peterson-Droogh, and K. D. Huff, “Standardized verification of the Cyclus fuel cycle simulator,” *Ann. Nucl. Energy*, vol. 128, pp. 288–291, Jun. 2019, doi: 10.1016/j.anucene.2019.01.014.

A. Lindsay, G. Ridley, A. Rykhlevskii, and K. Huff, “Introduction to Moltres: An application for simulation of molten salt reactors,” *Ann. Nucl. Energy*, vol. 114, pp. 530–540, Apr. 2018, doi: 10.1016/j.anucene.2017.12.025.

A. Rykhlevskii, “Advanced online fuel reprocessing simulation for thorium-fueled molten salt breeder reactor,” M.S. thesis, Univ. of Illinois at Urbana–Champaign, Apr. 2018. [Online]. Available: <http://hdl.handle.net/2142/101052>

A. Rykhlevskii, J. W. Bae, and K. Huff, “arfc/saltproc: Code for online reprocessing simulation of molten salt reactor with external depletion solver SERPENT,” *Zenodo*, Mar. 2018, doi: 10.5281/zenodo.1306628.

A. Lindsay and K. Huff, “Moltres: finite element based simulation of molten salt reactors,” *J. Open Source Softw.*, vol. 3, pp. 1–2, Jan. 2018, doi: 10.21105/joss.00298.

Ridley, G., “Multiphysics analysis of molten salt reactor transients,” Undergraduate Report No. UIUC-ARFC-2017-01, Advanced Reactors and Fuel Cycles Group Report Series, Univ. of Illinois at Urbana–Champaign, Urbana, IL, 2017. [Online]. Available: <https://github.com/arfc/publications/tree/2017-ridley-msrTransients>

A. Rykhlevskii, A. Lindsay, and K. D. Huff, “Full-core analysis of thorium-fueled molten salt breeder reactor using the SERPENT 2 Monte Carlo code,” presented at the 2017 ANS Winter Meeting and Nuclear Technology Expo, Washington, DC, U.S.A., Oct. 29–Nov. 2, 2017.

G. Ridley, A. Lindsay, and K. Huff, “An Introduction to Moltres, an MSR multiphysics code,” presented at the 2017 ANS Winter Meeting and Nuclear Technology Expo, Washington, DC, U.S.A., Oct. 29–Nov. 2, 2017.

G. Ridley, A. Lindsay, M. Turk, and K. Huff, “arfc/uiuc-arfc-2017-01: uiuc-arfc-2017-01,” *Zenodo*, Jan. 2018, doi: 10.5281/zenodo.1145437.

A. Lindsay, K. Huff, and A. Rykhlevskii, “arfc/moltres: Initial Moltres release,” *Zenodo*, Jun. 2017, doi: 10.5281/zenodo.801823.



# A NOVEL CRYSTAL STRUCTURE WITH SPIN-PROTECTED SURFACE ELECTRONIC CONDUCTION

**Allocation:** Illinois/40 Knh  
**PI:** Prashant K. Jain<sup>1</sup>  
**Co-PIs:** Sudhakar Pamidighantam<sup>2</sup>, Daniel Dumett Torres<sup>1</sup>

<sup>1</sup>University of Illinois at Urbana–Champaign  
<sup>2</sup>Indiana University at Bloomington

## EXECUTIVE SUMMARY

Newly discovered nanostructures can fill the need for materials in energy and optoelectronic technologies. Using a kinetically driven method of nanostructure synthesis, the research team accesses novel crystal structures of ionic semiconductors. For example, the group synthesized mercury selenide (HgSe) nanocrystals with a wurtzite structure quite unlike the natural zincblende form. The scientists computed the bulk band structure of the wurtzite form, which showed a finite band-gap and band inversion—prerequisites for a 3D topological insulator (TI) property. To determine if wurtzite HgSe is indeed a 3D TI, the team needed to elucidate the surface electronic structure. Using the current allocation, the researchers determined band structures of wurtzite HgSe slabs with specific surface facets. These computations showed the spin-protected nature of surface states in wurtzite HgSe, allowing its conclusive designation as a 3D TI. Spin-protection allows resistanceless electron transport along the surface. Thus, the newly discovered HgSe crystal structure can lead to logic devices capable of operating efficiently with minimal heat dissipation.

## RESEARCH CHALLENGE

Engineered nanocrystals are often utilized for making new functional electronic and optical materials such as superionic solids and battery electrodes. The research team makes use of unconventional methods that enable manipulation of the chemical composition and crystal structure of nanocrystals. These techniques often produce novel compositions and crystal phases that are often not found in the bulk phase diagram. Further, computational electronic structure investigations allow the researchers to explore the properties of these new, unconventional materials. In addition, the team is also elucidating chemical trends in heterostructures and alloys and developing solid-state principles from these trends. The results from these investigations will enable the rational design of new phases and compositions with targeted applications for resolving longstanding challenges of energy storage and device efficiency. With advances in nanotechnology and chemical synthesis, materials are becoming ever more complex. Computations can uncover chemical principles that will ultimately allow prediction of the properties of tomorrow’s indispensable materials—an exist-

ing Grand Challenge. However, these studies require extensive calculations spanning a range of physicochemical parameters. As opposed to a single large and expensive calculation, this work requires a library of moderately expensive calculations. The net cost for generating such a library of data is feasible only with a resource such as Blue Waters, with considerable payoff for future scientific advances. When solid-state principles such as those resulting from this project become known, the rate at which new materials can be discovered or designed will be greatly expedited, because the community is no longer limited to a time- and energy-consuming trial-and-error approach.

## METHODS & CODES

The research group used the open source Quantum Espresso software suite [1] to run density functional theory calculations of the electronic band structure. To study the effects of chemical composition, crystal structure, and crystallite size on the electronic properties, numerous calculations must be run. Each calculation is distinguished from the others by the crystal structure, chemical formula, or crystallite size. Through analysis of orbital energies, electronic character (orbital and spin), and band structures, the nature of surface electronic states and topology can be determined.

## RESULTS & IMPACT

Prior electronic structure calculations by the research team demonstrated that bond elongation in a novel wurtzite polymorph of mercury selenide (HgSe) and mercury–cadmium selenide (Hg<sub>x</sub>Cd<sub>1-x</sub>Se) is responsible for the opening of a band gap. Combined with the inverted band structure of HgSe, it was thought that these polymorphs would exhibit 3D topological insulator (TI) behavior [2]. Three-dimensional TIs are of interest because electrons at their surface states are spin-protected from back scattering. This protection allows 3D TI materials to conduct electrons along their surface without resistance. For this reason, 3D TIs are of interest as components of energy-efficient logic devices that can operate at high capacity with minimal heat dissipation. While the team’s past work suggested the possibility of the wurtzite phase being a 3D TI, a definitive conclusion was not possible without information on the surface electronic states. Therefore, the researchers went beyond band structure calculations of bulk HgSe and CdSe crystals to slab geometries with

well-defined exposed surface facets. As compared to electronic structure calculations of fully periodic bulk crystals, calculations of slab geometries are more expensive; but the latter system allows access to information of surface electronic states, which is missing in bulk crystal calculations. The most recent spin-resolved calculations of the surface band structures of faceted slabs have shown the existence of spin-protected electronic states on a wurtzite HgSe surface (Fig. 1). Spin protection, however, is missing from analogous wurtzite CdSe surfaces. These calculations now provide a firm basis for designating wurtzite HgSe as a 3D TI material. In addition, the research team performed all-atom calculations of nanocrystals. While computationally expensive, this work was undertaken to understand how crystallite size and nanoscale confinement influence surface electronic structure and topology. Unlike slab geometries, which are periodic in the *x* and *y* directions, the nanocrystal geometries are finite in all dimensions. Nanocrystals of wurtzite-structure HgSe and CdSe of different size (2 nm, 2.5 nm, and 3 nm) were studied. The team identified nanocrystal surface states based on the spatial character of the electronic states. This systematic set of calculations shows that surface states have a much richer character in these nanoscale-confined geometries as compared to that in the bulk. Single calculations of a slab and even all-atom calculations of a nanocrystal may be accomplished using computational resources other than Blue Waters. However, a large number of single calculations of these geometries with varying elemental composition and crystal dimensions are required for the research team’s study of size effects and chemical trends. The computational expense of such an effort would be prohibitive were it not for a Blue Waters allocation. Furthermore, the specialized hardware of Blue Waters allows the Quantum Espresso code to run even more efficiently. This is because Quantum Espresso’s parallelization schemes involve sizable and frequent communication among CPUs, which rely on the speed of the Blue Waters communication hardware. **PUBLICATIONS & DATA SETS** D. D. Torres, P. Banerjee, S. Pamidighantam, and P. K. Jain, “A non-natural wurtzite polymorph of HgSe: A potential 3D topological insulator,” *Chem. Mater.*, vol. 29, no. 15, pp. 6356–6366, 2017.

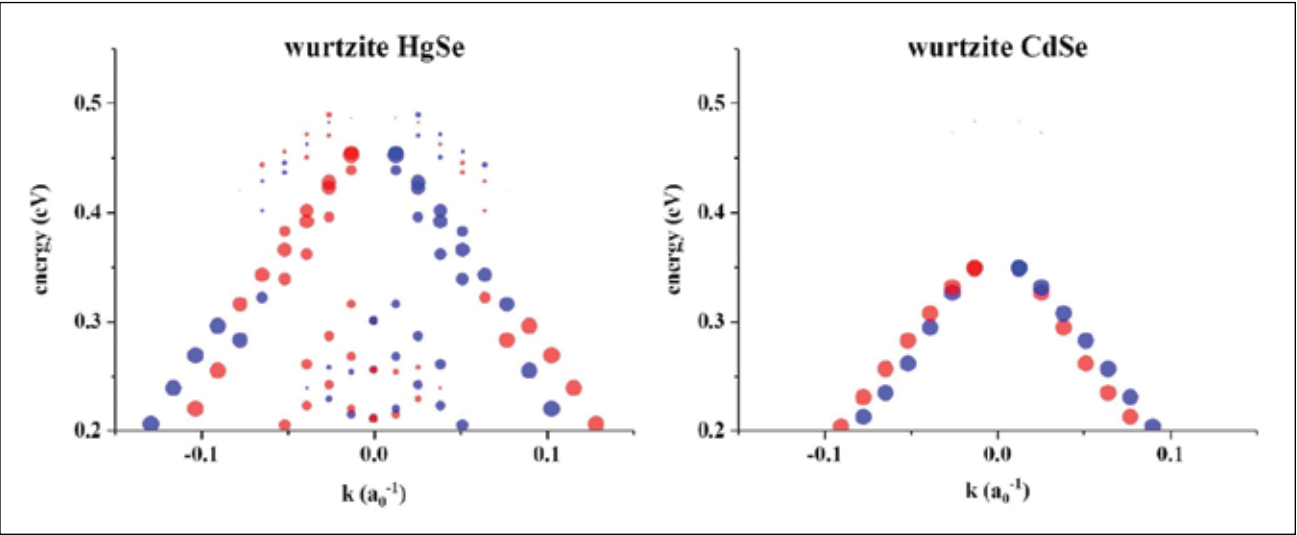


Figure 1: Calculated valence band structure of the (001) surface facet of a wurtzite HgSe (left) and CdSe (right) slab. Spins of opposite sign are distinguished by the color (red or blue). The size of the dots in the plot represents the electronic character contribution from the top (001) surface of the slab. The wurtzite HgSe surface band structure shows the presence of spin-protected surface states, as manifested by the states of opposite spin segregating on either side of the  $\Gamma$ -point ( $k = 0$ ). In contrast, in wurtzite CdSe, such a demarcation is not seen between states of opposite spin.



# INERTIAL COLLAPSE OF INDIVIDUAL BUBBLES NEAR SOLID/FREE BOUNDARIES

**Allocation:** GLCPC/350 Knh  
**PI:** Eric Johnsen<sup>1</sup>  
**Co-PIs:** Shahaboddin Alahyari Beig<sup>1</sup>, Mauro Rodriguez<sup>1</sup>, Minki Kim<sup>1</sup>

<sup>1</sup>University of Michigan

## EXECUTIVE SUMMARY

Inertial collapse of cavitation bubbles is a particularly complicated multiscale (ranging from micro- to macrobubbles in space, and microseconds to hours in time) and multiphysics (compressible fluid mechanics, multiphase flows, heat transfer, and solid mechanics) problem with a range of applications from naval hydrodynamics to biomedical ultrasound.

One of the main consequences of cavitation is structural damage to nearby objects. Collapse of cavitation bubbles can concentrate energy into a small volume, produce high pressures and temperatures, and generate strong shock waves. When they are adjacent to a neighboring interface such as another bubble, a solid object, or a free surface, the collapse becomes asymmetric and a re-entrant liquid jet penetrates the bubble. This jet hits the opposite side of the bubble and generates a radially propagating water-hammer shock.

To investigate this phenomenon, the research team carried out high-resolution numerical simulations of the collapse of: (1) a single bubble near an interface, and (2) multiple bubbles near a solid surface. These simulations yielded the detailed nonspherical bubble morphology as well as the pressure and temperature fields based on the relevant nondimensional parameters entering the problem. These simulations will be used to model the collapse of bubble clouds, to comprehend the damage mechanisms, and potentially to mitigate erosion.

## RESEARCH CHALLENGE

Cavitation research is essential to a variety of applications ranging from naval hydrodynamics to medicine and the energy sciences. Vapor cavities can grow from submicron-sized nuclei to millimeter-sized bubbles, and they collapse violently in an inertial fashion [1]. This implosion, which concentrates energy into a small volume, can produce high pressures and temperatures, generate strong shock waves, and even emit visible light [2]. It is known that high pressures and temperatures as well as the corresponding shock waves produced by the collapse of cavitation bubbles are capable of damaging nearby objects. This damage is recognized as one of the main consequences of cavitation and is an essential research topic in a variety of hydrodynamic and acoustic/biomedical applications [3–5].

The combination of compressibility effects of high-impedance fluids (liquids), the propagation of shock/rarefaction waves in a multiphase medium, and their interactions with material in-

terfaces and nearby solid/free boundaries results in a complex multiscale problem that ranges from micro- to macrobubbles in space and microseconds to hours in time and a multiphysics problem dealing with compressible fluid mechanics, multiphase flows, heat transfer, and solid mechanics. Laboratory experiments of such flows are challenging owing to the wide range of spatial and temporal scales, difficult optical access, and limitations of measurement devices. Thus, highly resolved numerical simulations have emerged as an alternative tool to complement experimental studies. However, the most commonly used algorithms are incapable of simulating the flow around the cavitating bubbles and treating the material interfaces correctly. To overcome these issues, the research team developed a novel numerical model to carry out accurate and efficient simulations of compressible multiphase flows [6,7]. These simulations provide valuable insight into the detailed dynamics of inertial collapse of individual bubbles near solid/free boundaries.

## METHODS & CODES

To carry out the simulations, the team developed a novel numerical framework to solve the compressible Navier–Stokes equations for a binary, gas–liquid system [6,7]. This numerical approach prevents spurious pressure and temperature oscillations across the material interfaces. For discretization, the group developed a solution-adaptive central-differencing/discontinuity-capturing approach. The basic idea is to use a high-order accurate, nominally nondissipative central difference scheme in smooth regions, and to apply a more dissipative, computationally expensive,

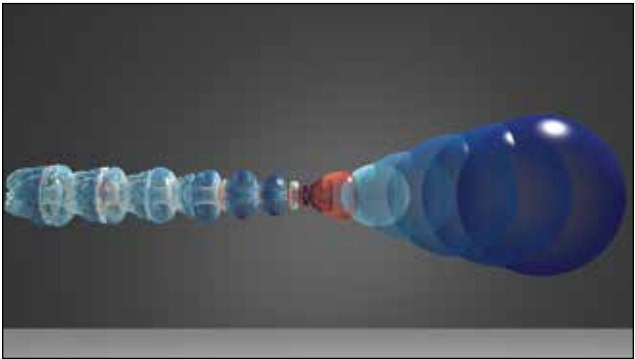


Figure 1: Volume rendering of the bubble interface colored by its temperature showing the re-entrant jet formation and vortex ring convection during the collapse of a single bubble near a solid surface.

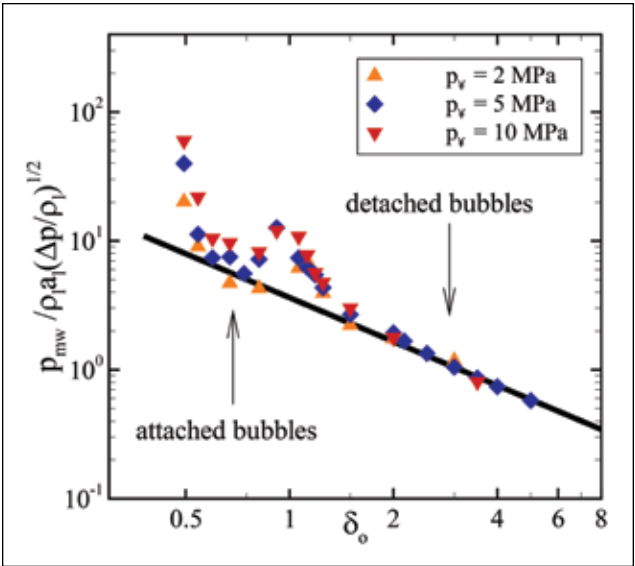


Figure 2: The scaling of the maximum pressure measured along a solid surface caused by the collapse of a nearby single bubble as a function of the initial stand-off distance from the surface and the collapse driving pressure.

high-order accurate, shock- and interface-capturing approach only to the regions with discontinuities. For this purpose, a discontinuity sensor discriminates between smooth and discontinuous (shocks, contacts, and interfaces) regions. The time marching is handled with a third-order accurate TVD Runge–Kutta scheme. To perform the three-dimensional numerical simulations of the problems of interest, an in-house code was developed in C++, which was parallelized using the MPI library and implemented the parallel HDF5 library for I/O. The code was verified and validated against a series of theoretical and experimental data. To better understand the detailed collapse dynamics, as well as to comprehend the damage mechanisms and potentially mitigate erosion, the team studied the collapse of individual bubbles near solid/free boundaries.

## RESULTS & IMPACT

When a cavitation bubble collapses near an interface, the collapse becomes nonspherical, leading to the formation of a high-velocity re-entrant liquid jet (Fig. 1). The impact of the jet on the opposite side of the bubble can generate a water-hammer shock wave that propagates in the surrounding media and thus can create high-pressure regions on the surface of neighboring solids (Fig. 2). Simulating the collapse of a single bubble near an interface delves into the details of the re-entrant jet formation and the corresponding shock propagation. The results will be used to provide scaling laws for important collapse parameters (e.g., jet velocity, bubble nonsphericity, collapse time, and shock pressure) that can be used to predict the single-bubble dynamics.

However, the disruptive effects of cavitation erosion in real flow problems are generally caused by the collapse of bubble clouds that include thousands of bubbles. To simulate such flows where

resolving every bubble is numerically impossible, a robust cloud model is required. Current models are typically based on spherical bubble dynamics, and thus neglect the nonspherical effects of the collapse owing to bubble–bubble/bubble–boundary interactions [8], which leads to an inaccurate estimation of the impact loads on nearby surfaces. To address this, the team simulated the collapse of a bubble pair near a solid surface to investigate and quantify the bubble–bubble/bubble–boundary interactions and their effects on the collapse dynamics. The simulation showed that such interactions intensify the overall collapse nonsphericity and may increase or decrease the collapse intensity depending on the initial flow configuration.

These results can be used to develop a more reliable cloud model that includes the nonspherical effects of the collapse, providing a more accurate prediction of the cloud dynamics and potential cavitation-induced erosion.

## WHY BLUE WATERS

Performing three-dimensional high-resolution simulations of up to 3.6 billion gridpoints that can effectively resolve the small-scale features of the flow, as well as handling postprocessing and visualizations of large data files, requires substantial computational power. A petascale computing resource such as Blue Waters makes these simulations possible and has been essential for the success of the present study. This project will help researchers to gain valuable insights and understanding of these complex flows, which was not previously possible.

## PUBLICATIONS & DATA SETS

S. A. Beig, B. Aboulhasanzadeh, and E. Johnsen, “Temperatures produced by inertially collapsing bubbles near rigid surfaces,” *J. Fluid Mech.*, vol. 852, pp. 105–125, 2018.

S. A. Beig, M. Kim, and E. Johnsen, “The role of compressibility in energy budget of spherical collapse of an isolated bubble,” *Phys. Rev. Lett.*, in preparation, 2019.

S. A. Beig and E. Johnsen, “Inertial collapse of a gas bubble near a rigid boundary,” *J. Fluid Mech.*, in preparation, 2019.

S. A. Beig and E. Johnsen, “On the effects of bubble–bubble interactions in inertial collapse of a bubble pair near a rigid wall,” *J. Fluid Mech.*, in preparation, 2019.

S. A. Beig, “Inertial collapse of individual bubbles near a rigid surface,” presented at the 10th Int. Conf. Mult. Flows, Rio de Janeiro, Brazil, May 19–24, 2019.

S. A. Beig and E. Johnsen, “Bubble–bubble interactions and wall pressures/temperatures produced by the collapse of a bubble pair near a rigid surface,” presented at the 10th Int. Cav. Sym., Baltimore, MD, U.S.A., 2018.



ELECTRONIC STRUCTURE OF MICROSCALE DIELECTRIC BARRIER DISCHARGES

Allocation: Illinois/350 Knh  
PI: Harley T. Johnson<sup>1</sup>  
Co-PI: Purnima Ghale<sup>1</sup>

<sup>1</sup>University of Illinois at Urbana–Champaign

EXECUTIVE SUMMARY

The research team conducted high-throughput simulations to predict rates of electron emission from dielectric surfaces under AC voltage. By spanning a large range of dielectric barrier discharge (DBD) operating conditions, the team produced a data set that is the first of its kind for DBD plasma generators. The work relies on a method the researchers developed previously for the accurate simulation of material systems containing up to ten million atoms and their electrons. The study also tested the feasibility of numerical integrators for slow, nonperturbative electron dynamics and showed that adaptive numerical integrators perform better than expected. This observation demonstrates that it is possible to leverage numerical engineering methods to optimize or design dynamic electron systems such as DBD plasma generators.

RESEARCH CHALLENGE

The research aims to extend the length and timescales of atomistic, quantum–mechanical materials simulation, particularly of systems containing millions and tens of millions of atoms, under slow, nonperturbative, realistic AC voltages. Materials and phenomena in these temporal and spatial domains are important in many different technologies, from microelectronics to medical science to energy technologies. Traditional electronic structure calculations at this scale are constrained by memory and communication overheads, as they were designed to minimize the number of FLOPs. The research team’s approach prioritizes memory and communication constraints as more limiting, and that is the reason they can simulate ten-million-atom systems on 100 nodes within a few minutes. The team has also taken a similar approach toward extending simulated timescales of electronic systems, with promising results. As a testbed application, the researchers have chosen to investigate dielectric surfaces in dielectric barrier discharge plasma generators, which are important in combustion, materials processing, and catalysis.

METHODS & CODES

The simulations use density-functional-based tight-binding, which is a semiempirical electronic structure method. The main computational workflow is to read the atomic coordinates, construct the Hamiltonian, and compute the required eigenspace that denotes the ground state of electrons.

*Extending system size.* In order to simulate million-atom systems, the team implemented its algorithm [1] using the PETSc

distributed matrix library [2]. They used the code they developed to compute the density matrix, which is a fundamental electronic structure quantity for a system of atoms and their electrons. This computational method [1] combines the advantages of two existing linear scaling methods—the kernel polynomial expansion (KPE) [3] and second-order spectral projection purification (SP2) [4] methods. The KPE method is computationally efficient and can be easily expressed in terms of sparse matrix–vector multiplications (SpMV) but cannot satisfy one of the required constraints on its own. On the other hand, the SP2 method is highly accurate but can be prohibitively costly in terms of memory and communication when expressed in terms of sparse matrix–matrix multiplications. When expressed in terms of SpMVs, the SP2 method scales exponentially with the number of iterations required to converge. An advantage of SP2, however, is that the method converges quadratically near the correct solution  $P$  [4]. Thus, the researchers have constructed a hybrid method that takes the inexpensive KPE solution and purifies  $P_{kpe}$  with a few SP2 iterations that are more expensive.

*Extending timescales.* The research team tested implicit–explicit time-steppers based on [5] within PETSc and found them to be appropriate for simulating 400-site electronic systems driven by slow, nonperturbative AC fields. This is significant because quantum–mechanical forward-time simulations are usually done using specialized explicit integrators [6–8]. Numerically, the team found that appropriately tuned implicit–explicit time-steppers conserve the invariants of electronic motion given in [9]. However, these temporal simulations currently scale quadratically with the number of sites, and future algorithmic work will be necessary to improve performance.

*High-throughput simulations of small systems.* The researchers built upon their electronic structure code, producing a data set for physical systems corresponding to pressures ranging from 10 Pa to 10 GPa and for electric fields between 10 V/m to 1 MV/m. The main computational bottleneck here was the size of data generated as well as constraints on wall-time. The constraint on wall-time was not prohibitive, but it limited the number of physical AC cycles that were averaged to produce the data set.

RESULTS & IMPACT

The team’s recent work expands the scope of atomistic electronic structure simulations, extending the boundaries of the realm of feasible simulations. This allows researchers to simulate larger,

more heterogeneous systems that are essential to understanding, designing, and optimizing complex devices and phenomena, directly affecting research on surface and dielectric properties for combustion, catalysis, and materials processing.

WHY BLUE WATERS

Access to the Blue Waters system made this work possible by permitting studies on a single platform that offered both large parallelism and resources for memory-intensive computations. Depending on the tuning of the method, it is possible to either carry out the entire computation on a single node, in a very memory-intensive approach, or, by distributing the computation over many processors, to perform the algorithm in a massively parallel way. The Blue Waters system allowed the team to study these approaches and to plan for future studies with more optimized tuning. The scale of Blue Waters also allowed them to extend the boundaries of possible electronic systems and phenomena that can be simulated in the future. In addition, they benefitted from the smooth, transparent, and dependable access to various system modules, as well as helpful and timely responses from Blue Waters staff.

PUBLICATIONS & DATA SETS

P. Ghale and H. T. Johnson, “Dielectric surface boundaries for plasma generators—parameterization of rates of electron emission from quartz under AC voltage,” in preparation, 2020.

P. Ghale and H. T. Johnson, “Benchmarking implicit-explicit integrators for direct evolution of many-electron systems under slow, non-perturbative dynamical loading,” in preparation, 2020.



ACCELERATING VIRTUAL PROTOTYPING AND CERTIFICATION IN THE AEROSPACE INDUSTRY WITH SCALABLE FINITE-ELEMENT ANALYSIS

**Allocation:** Director Discretionary/150 Knh  
**PI:** Seid Koric<sup>1</sup>  
**Co-PIs:** Robert F. Lucas<sup>2</sup>, Erman Guleryuz<sup>3</sup>

<sup>1</sup>University of Illinois at Urbana–Champaign  
<sup>2</sup>Livermore Software Technology Corp.  
<sup>3</sup>National Center for Supercomputing Applications

EXECUTIVE SUMMARY

The aerospace industry increasingly relies on physics-based modeling and simulation for the design and analysis of systems in complex engineering products. Advanced modeling and simulation techniques such as finite element analysis (FEA) are being used to replace physical testing with virtual testing and mitigate the costs and risks of certification. Improved fidelity in simulation models generally requires explicit modeling of smaller physical details, which leads to larger models with finer grids. To further the state of the art in computationally intensive implicit FEA and to reduce time-to-solution, the research group studied real-life jet engine models with LS–DYNA software and removed performance scaling barriers. The team solved an implicit finite-element model with over 200 million equations in-core on Blue Waters. This work demonstrated the feasibility of efficiently solving extreme-size real-world multiphysics problems at peta- and potentially exascale levels, thus adding novel value to engineering research by enabling the creation of high-fidelity models that yield detailed insight into the performance and safety of proposed engineering designs.

RESEARCH CHALLENGE

The imperative to design higher-performing products pushes technological advances in advanced computational analysis techniques such as FEA. FEA allows gas-turbine engine designers to develop systems that continue to increase thrust and fuel efficiency while meeting a complex set of requirements, including min-

imizing weight, maintaining part life, and reducing manufacturing costs. In practice, large finite element method (FEM) models today use hundreds of cores, and large original equipment manufacturers are evaluating the use of thousands of cores. More realistic models with higher fidelity will require tens of thousands of cores and even, in the near future, hundreds of thousands. The growth in high-performance computing (HPC) resources required to run larger, more realistic models is complemented by the industry’s need to reduce the time for each simulation so they can be included in the design cycle rather than remain a research function.

Several barriers exist today to increasing realism and adding optimization to the design cycle: (1) single simulations are not yet realistic enough, using too few cores; (2) commercial software developers have not had sufficient access to HPC to modernize their codes; and (3) commercially accessible HPC of sufficient scale is not available yet to foster stochastic optimization. This project addresses the technical barriers of scale and access for developers.

METHODS & CODES

The aerospace industry relies on advanced commercial FEA codes that are used for nonlinear, quasistatic, and dynamic deformations in computational structural models. LS–DYNA is a prominent FEA code used widely in the industry. This research has been performed with LS–DYNA and with modules of the LS–DYNA code, including a graph-partitioning scheme and the factorization kernel used in the direct solver. The aerospace community uses LS–DYNA for extreme-event modeling for gas-turbine engines such as bird-strikes, fan-blade-off, and whole engine thermomechanical analysis. Bird-strike and fan-blade-off testing is mandated by the Federal Aviation Administration and European Aerospace Safety Agency to certify the safety of jet engines. These tests must demonstrate that the engine casing can contain the debris from the rotors and that engine mounts do not fail.

To further the development of FEA to reduce time-to-solution, Rolls-Royce developed representative models of a gas-turbine engine for research (Fig. 1). As part of a multiyear research effort, the multidisciplinary research team formed by collaborators from NCSA, Rolls-Royce, Livermore Software Technolo-

gy Corp., and Cray has worked to improve the scalability of LS–DYNA using real-world engine models developed specifically for scaling studies on supercomputers.

RESULTS & IMPACT

The computational time of implicit FEA is dominated by the analysis and solution of a sparse linear system of equations. A single solution of the system suffices for linear problems. However, for nonlinear problems, within each quasistatic timestep a system of nonlinear equations is linearized and solved with a Newton–Raphson iteration scheme, often requiring several linear solver solutions. Stiff or ill-conditioned linear systems arise in several application areas. These usually involve fine-grain, unstructured computational meshes on irregular geometries, and the coupling of multiple models (e.g., multiscale, multiphysics). For these types of challenging systems, implicit methods with direct solvers are often the most robust solution methods. LS–DYNA offers a multifrontal direct solver for these numerically challenging linear systems of equations.

The implicit analysis has the following main steps: matrix assembly, constraint processing, matrix reordering, symbolic factorization, numeric factorization, and triangular solution. Because they have different parallel efficiencies, each step consumes a varying fraction of the wall-clock time as the processor count grows. A remarkable increase in efficiency (Fig. 2b) has been achieved since the beginning of this multiyear effort, to the point where LS–DYNA can successfully run the largest engine models on tens of thousands of cores on Blue Waters. The effort included application- and hardware-level profiling, creation of novel scalable algorithms, memory management improvements, com-

putation workflow modifications, and improvements to reduce previously unknown Amdahl (*i.e.*, serial) fractions. The computational complexity of the numerical pipeline is centered on the sparse matrix factorization step.

The team’s previous research proved for the first time that a sparse matrix factorization algorithm can scale on tens of thousands of computational cores [1]. Floating-point arithmetic throughput rate and memory utilization amount are two major metrics that characterize performance scaling of factorization. Fig. 2a shows strong scaling results for the multifrontal solver’s factorization kernel, when factoring symmetric indefinite sparse matrices. An increase in computing operations rate and reduction in memory storage needed per Message-Passing Interface rank has been observed for up to 33,000 OpenMP threads. This sustained performance increase indicates potential for more performance growth. This work constitutes, to the best of the researchers’ knowledge, the largest implicit LS–DYNA calculations to date, and the results have been presented in technical conferences [2–4].

WHY BLUE WATERS

Blue Waters—with large amounts of distributed memory, thousands of multicore processors, a low-latency file system, and increased bandwidth of advanced interconnect technologies—enabled the research team to address parallel processing challenges and to demonstrate that implicit FEA of large-scale models could be performed in a timely manner using large-scale computing systems. This research on Blue Waters has allowed computer-aided engineering to have a greater impact on the design cycle for new engines, and is a step toward the long-term vision of digital twins, *i.e.*, computer-generated representations of complex engineering systems.

PUBLICATIONS & DATA SETS

D. Liu, S. Koric, and A. Kontsos, “Parallelized finite element analysis of knitted textile mechanical behavior,” *J. Eng. Mater. Technol.*, vol. 141, no. 2, p. 021008, Dec. 2018.

R. Borrell, J. C. Cajas, D. Mira, A. Taha, S. Koric, M. Vázquez, and G. Houzeaux, “Parallel mesh partitioning based on space filling curves,” *Comput. Fluids*, vol. 173, p. 264, Sept. 2018.

F. A. Sabet, O. Jin, S. Koric, and I. Jasiuk, “Nonlinear micro-CT based FE modeling of trabecular bone—Sensitivity of apparent response to tissue constitutive law and bone volume fraction,” *Int. J. Numer. Meth. Biomed. Engng.*, vol. 34, no. 4, p. 2941, Nov. 2017.

A. Gupta, N. Gimelshein, S. Koric, and S. Rennich, “Effective minimally-invasive GPU acceleration of distributed sparse matrix factorization,” in *Euro-Par 2016: Parallel Processing, Lecture Notes in Computer Science*, P.F. Dutot, and D. Trystram, Eds., vol. 9833. Cham, Switzerland: Springer, 2016, p. 672.

M. Vázquez *et al.*, “Alya: Multiphysics engineering simulation toward exascale,” *J. Comput. Sci.*, vol. 14, p. 15, May 2016.

V. Puzyrev, S. Koric, and S. Wilkin, “Evaluation of parallel direct sparse linear solvers in electromagnetic geophysical problems,” *Comput. Geosci.*, vol. 89, p. 79, Apr. 2016.

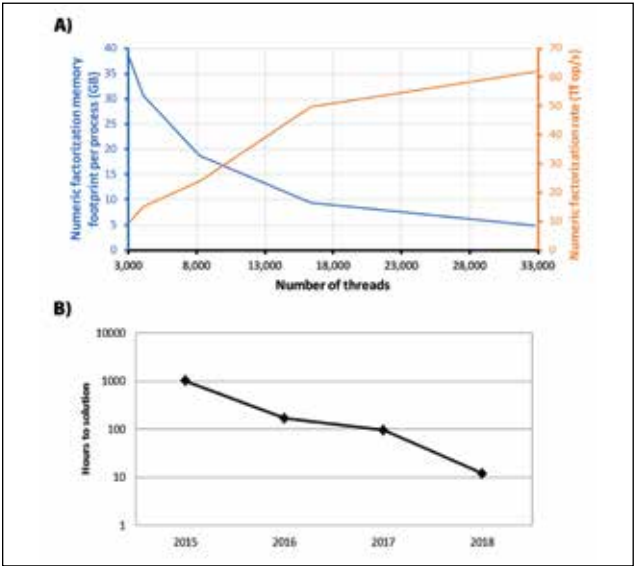


Figure 2: (a) Parallel scaling of arithmetic throughput and memory footprint for numeric factorization. A hybrid (MPI and OpenMP) parallelization model (eight threads per MPI rank) is used to solve the finite-element model with 200 million equations. (b) Aggregate improvement in time-to-solution during the course of the multiyear research effort.

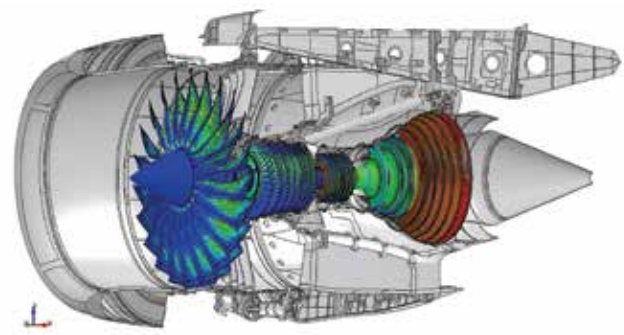


Figure 1: A cross-section of the Rolls-Royce jet engine model.



GRAPHENE NANOPORE TRANSISTOR FOR DNA-NICK DETECTION

**Allocation:** Illinois/510 Knh  
**PI:** Jean-Pierre Leburton<sup>1</sup>  
**Co-PI:** Olgica Milenkovic<sup>1</sup>  
**Collaborator:** Nagendra Athreya<sup>1</sup>

<sup>1</sup>University of Illinois at Urbana-Champaign

EXECUTIVE SUMMARY

Significant effort has been placed on advancing new generations of single-molecule detection technologies. Unlike other existing platforms, solid-state nanopores have the versatility to perform tasks that go beyond DNA sequencing, such as detecting epigenetic modifications, RNA and protein sequencing, and folding patterns. In such settings, it is crucial to develop computational tools that enable the identification of optimal solid-state membranes and pore geometries for the problems at hand. In this regard, the research team studies the implementation of innovative 2D solid-state nanopore devices to detect and map minute structural damages in the backbone of the double-strand DNA molecule, which have been suggested as a cause of cancer. The team's two-step objectives will be achieved through, first, understanding the physics behind the interactions of damaged DNA with 2D solid-state membranes and, second, detecting the variations in ionic current through the pore and the transverse electronic current across the membrane caused by such interactions.

RESEARCH CHALLENGE

A normal human cell is subjected to approximately 70,000 lesions per day. Of these, single-strand DNA (ssDNA) breaks, which are often converted into double-strand DNA (dsDNA) breaks, are a vast majority [1]. A single break in a critical gene can cause the cell to undergo apoptosis (programmed cell death). If the repair

mechanism fails, the dsDNA break can cause chromosomal instability, leading to tumorigenesis. Existing genome sequencing techniques are not suitable for detecting such changes directly. The only known means for ssDNA breakage positioning is to use immunoprecipitation followed by sequencing, which leads to highly limited resolution and low accuracy. Hence, as an alternative methodology, the research team has shown that graphene quantum point contact (g-QPC) nanopore transistors can be used efficiently to detect and map defects in the DNA backbone as miniscule as ssDNA breaks using electronic sheet currents obtained across the membrane.

METHODS & CODES

This research consists of a two-step process that first includes molecular dynamics (MD) simulations with the latest NAMD version and, second, the exploitation of MD data to calculate the current variations owing to DNA translocation through the nanopore via electronic transport modeling. The system is built, visualized, and analyzed using VMD [2]. The DNA is described by the CHARMM27 force field [3]. An external electric field is applied to the system to drive the DNA nicks through nanopores. For each frame of the trajectory files obtained from the MD simulations, the ionic current blockade is calculated instantaneously [4]. Further, the electrostatic potential induced by the biomolecule around the pore is obtained by using the self-consistent Poisson-Boltzmann equation (PBE) formalism. The PBE

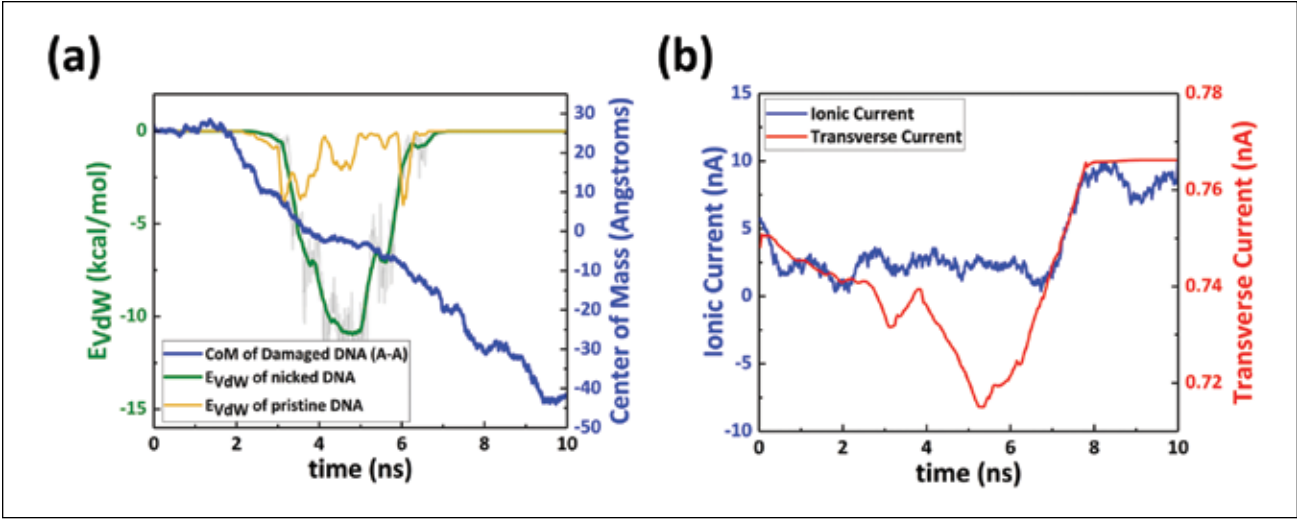


Figure 2: (a) Plot of van der Waals energies calculated between graphene membrane and nicked-/normal-DNA site showing strong attraction at the nicked site of damaged DNA. (b) Plot showing the ionic current and transverse sheet current indicating the detection of the nicked site by sheet currents (dip at ~5 ns).

is solved numerically using the multigrid method until the convergence criterion is met. Once the electrostatic potential coplanar to the membrane is obtained, the transverse conductance across the 2D nanopore, graphene, or MoS<sub>2</sub> membrane is computed by using the nonequilibrium Green's function formalism [5] or a semiclassical thermionic emission technique [6], respectively. The Leburton Group at the University of Illinois at Urbana-Champaign wrote and maintains the PBE and electronic transport code.

RESULTS & IMPACT

Real-time detection of damaged-DNA strands is nearly impossible and has not been reported by anyone using existing technologies. Fig. 1 illustrates the nanopore setup used for simultaneous calculation of the ionic and transverse sheet currents. Resolving the duration and magnitude of each dip of the ionic current signal gives an insight into the structure of part of the biomolecule inside the pore. Simultaneously, the variations in the electronic sheet current flowing from source to drain across the 2D membrane detect the change of the electric potential induced on the ridge of the nanopore by the translocating DNA, revealing the position of the damaged backbone.

The results obtained using this methodology are highly encouraging: In all the simulation runs, the molecule is halted in the pore at the nicked site owing to strong attraction between the graphene membrane and the damaged backbone. This behavior can be explained by the fact that the cleaved backbone of the dsDNA molecule is arrested in the pore owing to higher hydrophobic interactions between the DNA and the graphene membrane.

To validate this theory, the research team calculated the van der Waals energies between the normal DNA with graphene and the damaged DNA with graphene (Fig. 2a). It is clear that there is higher attraction at the nicked-site with the graphene atoms resulting in the molecule being arrested in the pore. The ionic

currents calculated for the translocation of 20 base-pair dsDNA with a break in the backbone show no distinct feature contributed by the nicked site on the signal. However, a clear dip is seen in the transverse sheet current signal corresponding to the location of the breakage, enabling the electronic detection of damages (Fig. 2b).

The research team has validated the methodology outlined above by detecting other sequence-specific dsDNA breaks along a randomly sequenced strand. This technique can be easily scaled to have a dense array of multiple g-QPC nanopores implemented on a complementary metal-oxide-semiconductor chip to detect multiple damaged DNA strands in a massively parallel scheme [7]. The researchers strongly believe such a detection mechanism can enable the development of versatile semiconductor electronics for early cancer detection caused by structural modification of the genome.

WHY BLUE WATERS

It is only possible to investigate the interactions of biomolecules with solid-state materials, to characterize the stochastic structural fluctuations of the DNA with nicks translocating through solid-state nanopores, and to further obtain the electronic response using all-atom MD simulations coupled with electronic transport calculations with petascale computing resources such as Blue Waters. The research team's systems consist of about 500,000 atoms, each requiring multiple MD simulation (NAMD) runs. With NAMD code efficiently deployed on XE/XK nodes to run highly parallel simulations of large biomolecular systems, Blue Waters is well-suited for the requirements of this research.

PUBLICATIONS & DATA SETS

N. Athreya, O. Milenkovic, and J.-P. Leburton, "Site-specific detection of DNA-nicks using ultra-thin nanopore membranes," in progress, 2019.

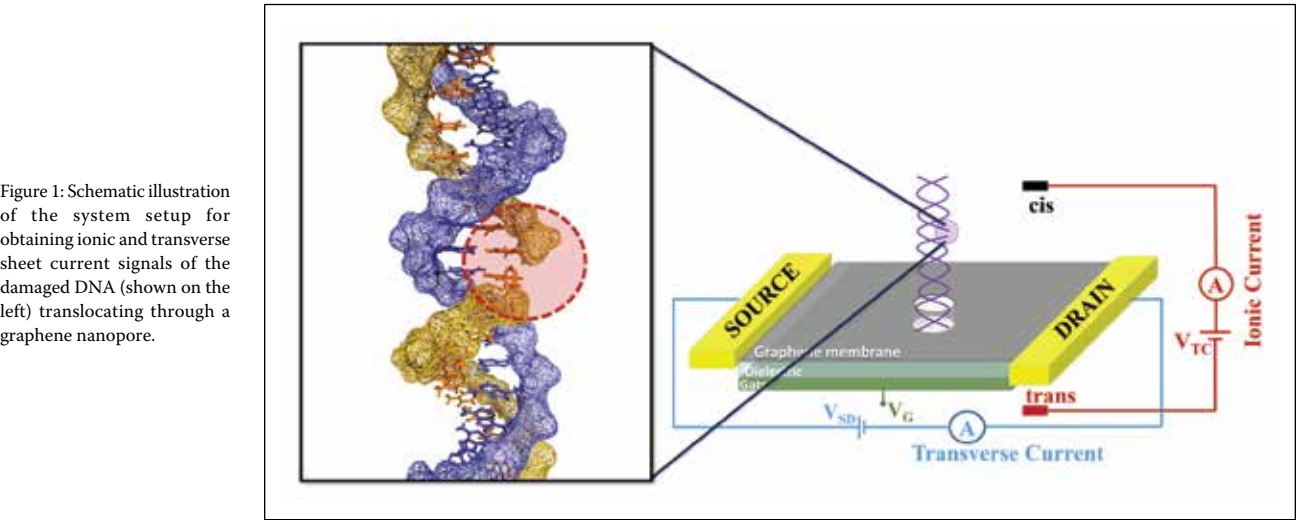


Figure 1: Schematic illustration of the system setup for obtaining ionic and transverse sheet current signals of the damaged DNA (shown on the left) translocating through a graphene nanopore.



# COMPRESSIBILITY EFFECTS ON SPATIALLY DEVELOPING PLANE FREE SHEAR LAYER

**Allocation:** Innovation and Exploration/190 Knh  
**PI:** Farzad Mashayek<sup>1</sup>  
**Co-PIs:** Jonathan Komperda<sup>1</sup>, Dongru Li<sup>1</sup>, Ahmad Peyvan<sup>1</sup>

<sup>1</sup>University of Illinois at Chicago

## EXECUTIVE SUMMARY

Compressible turbulent free shear flows have a wide variety of applications in modern technology and commonly occur in engineering systems such as gas turbines, scramjet engines, rocket exhausts, and the like. However, a fundamental understanding of the physics of such flows is limited by a lack of detailed information about the turbulent transition process and the turbulent quantities under the influence of compressibility. This work is the first computational effort to investigate the compressibility effects on the transition to turbulence and the turbulent energy exchange mechanisms in a three-dimensional, spatially developing turbulent plane free shear layer, via data produced by direct numerical simulation (DNS). The DNS was performed using a high-order discontinuous spectral element method for different convective Mach numbers with a naturally developing inflow condition. The location of the transition zone was predicted by the analyses of vorticities and the turbulent viscous dissipation rate. The energy exchange was examined via the analyses of the budget terms of turbulent kinetic, mean kinetic, and mean internal energy transport equations.

## RESEARCH CHALLENGE

DNS can generate all the information in a turbulent flow, which is impossible to accomplish experimentally. However, most of the DNS were performed on a temporally developing turbulent flow owing to its much lower computational cost compared to its counterpart—spatially developing flow. For the DNS of a highly compressible free shear layer (FSL), the spatially developing approach is the closest realization of a transition experiment. Therefore, the research team developed a computational model for a three-dimensional (3D), spatially developing turbulent plane FSL, with a naturally developing inflow condition, using DNS of the compressible full Navier–Stokes equations. This computational model employs a highly accurate numerical method, the high-order discontinuous spectral element method (DSEM) [1–3], to deal with the complicated issues arising in a highly compressible turbulent flow in an open computational domain for an FSL.

## METHODS & CODES

The research team performed DNS on Blue Waters using a high-order DSEM [1–3]. The DSEM used hexahedral nonoverlapping elements in an unstructured grid. A high-order local ba-

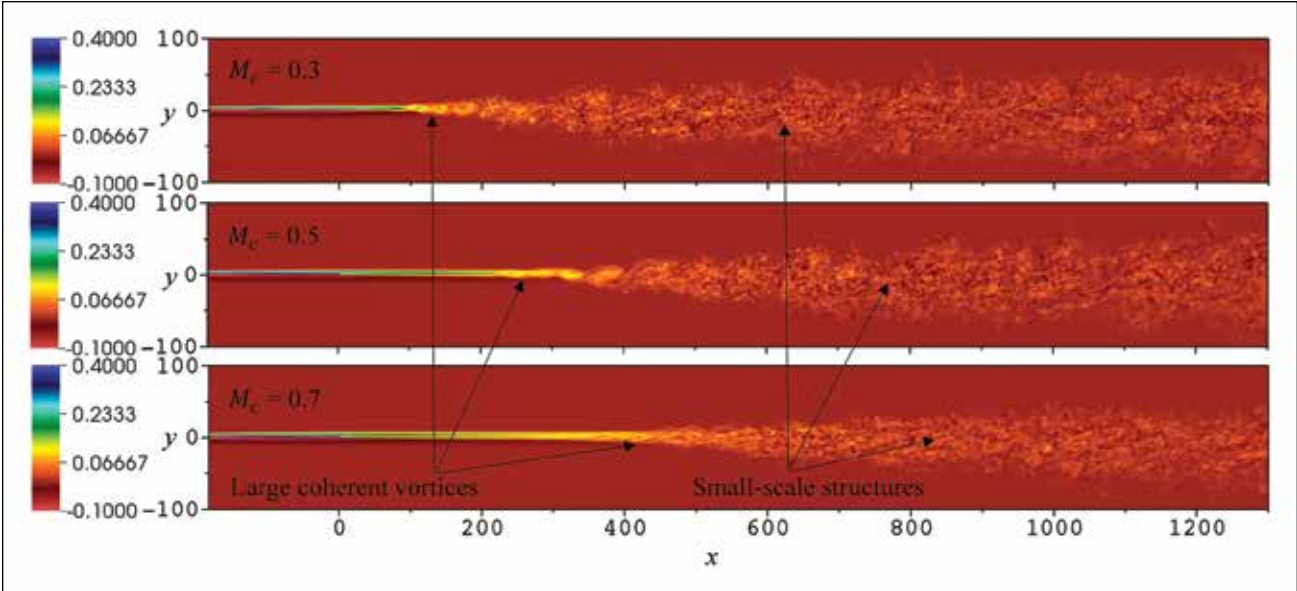


Figure 2: Spanwise vorticity contours for the cases with convective Mach numbers of 0.3, 0.5, and 0.7. The computed vorticity is normalized by  $\Delta U$ , which is the difference between the high- and low-speed streams.

sis function approximated the solution variables in each element [1,2]. Such features make the DSEM powerful in terms of handling complex geometries, local mesh refinement, efficient parallelization, and easy boundary condition implementation. Such discretization also enabled the team to interpolate the solution to a new local basis function with a different polynomial order within each element in order to change the grid resolution during simulation. The method introduces negligible diffusion and dispersion errors and is spectrally convergent for smooth solutions [1,2]. In the current work, the grid consists of 1,368,260 elements with a polynomial order of  $p = 5$ , resulting in a total of 295,544,160 solution points. The team’s scaling study indicated that using 2,048 cores is ideal for performing the simulations, considering that the efficiency is relatively high (82%) on Blue Waters.

## RESULTS & IMPACT

This work presents the first DNS results for a spatially developing, 3D compressible turbulent FSL with a naturally developing inflow condition for different convective Mach numbers. It shows a 3D representation of the instantaneous turbulent structures via iso-surface of the second invariant of velocity gradient tensor for convective Mach numbers of 0.3, 0.5, and 0.7 (Fig. 1). Further, it demonstrates the vortex stretching mechanism in the spanwise direction, such as the formation of secondary streamwise vortices and the breakdown of primary spanwise vortices, which are responsible for the onset of turbulence transition. Also, the contours of the instantaneous spanwise vorticity (Fig. 2) indicate the location of the turbulence transition region under the influence of compressibility for each case. The spiral-type roll-ups can only be observed in the flow with the highest convective Mach number, 0.7. The instantaneous variables and their statis-

tics were used to investigate the compressibility effects on turbulence transition and turbulent energy exchange and to calibrate turbulence models.

This work conducts the DNS of an FSL to generate detailed data for all flow field variables in both the transition and self-similar turbulent regions for different convective Mach numbers. It identifies the location of the turbulence transition zone under the influence of compressibility. It determines the energy exchange mechanisms responsible for energy redistribution among turbulent kinetic energy, mean kinetic energy, and mean internal energy and examines the influence of compressibility on such mechanisms. It provides calibrations for turbulent models.

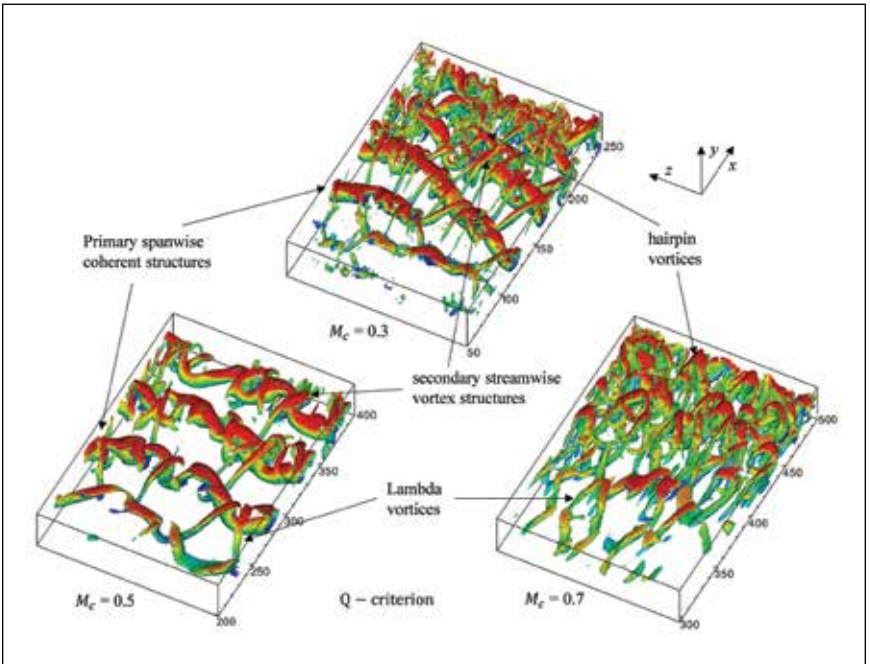
## WHY BLUE WATERS

To resolve all relevant scales of turbulent structures, the grid used in this work consisted of roughly 0.3 billion solution points. Also, the generation of full statistics for the cases required the computation and storage of 47 variables for fluctuations, nine for averages, five for instantaneous variables, and three for coordinates. The storage of these variables was necessary both on disk as well as in memory at runtime, which exceeds the available storage space and memory available per core on many supercomputers. Thus, this combination of computation and data storage is well suited for a leading-edge petascale high-performance computing system such as Blue Waters.

## PUBLICATIONS & DATA SETS

D. Li, J. Komperda, Z. Ghiasi, A. Peyvan, and F. Mashayek, “Compressibility effects on the transition to turbulence in a spatially developing plane free shear layer,” *Theor. Comput. Fluid Dyn.*, vol. 33, no. 3, 2019, doi: 10.1007/s00162-019-00507-w.

Figure 1: Iso-surface of the second invariant of velocity gradient tensor colored by streamwise velocity with convective Mach numbers of 0.3, 0.5, and 0.7. All terms are normalized by  $\Delta U$ , which is the difference between the high- and low-speed streams.





# OUTWARDLY PROPAGATING TURBULENT FLAMES

**Allocation:** Illinois/80 Knh  
**PI:** Moshe Matalon<sup>1</sup>  
**Co-PI:** Shikhar Mohan<sup>1</sup>

<sup>1</sup>University of Illinois at Urbana–Champaign

## EXECUTIVE SUMMARY

The research team’s objective is to simulate the evolution of centrally ignited flames in order to assess the roles played in their propagation of hydrodynamic and thermo-diffusive instabilities as well as background turbulence. The simulations were conducted within the framework of the hydrodynamic theory, which has been systematically derived using a multiscale asymptotic technique [1] and is thus free of turbulence-modeling assumptions and/or *ad hoc* adjustment parameters. The flame devolves into a well-defined moving interface that separates the burned products from the unburned mixture. The flame propagation relative to the flow depends on the local stretch rate, which combines the effects of surface geometry and hydrodynamic strain rate. The stretch rate is modulated by a mixture-sensitive parameter known as the Markstein number, which mimics the effects of diffusion and chemical reactions occurring inside the thin-flame zone. This eliminates the necessity for explicit reaction-chemistry modeling, thereby alleviating mesh size and timestep restrictions that render direct numerical simulation studies impractical.

## RESEARCH CHALLENGE

Combustion is and will in the foreseeable future remain the primary mode of power generation. The process is inherently complex, involving the interdiffusion of a large number of chemical species (nearly 5,000 for real fuels) that interact chemically as well as the generation of heat that affects the density of the mixture and modifies the underlying flow field.

Flames encountered in most applications are turbulent in nature. Turbulence adds a stochastic, time-dependent, three-dimensional aspect to this intricate problem. While fundamental processes such as reaction chemistry, molecular, and energy transport are reasonably well known, it is still a major challenge to integrate these components with a highly turbulent flow field to produce practical computational results for the description of turbulent combustion. Additional complexity arises from intrinsic flame instabilities associated with thermal expansion and/or the disparity between the diffusion rates of the different species and of mass and energy, which are known to distort the flame even under laminar conditions.

The hydrodynamic theory adopted in this work is formulated in an intrinsic coordinate system [2] that requires a parametric description of the flame surface. The flame’s evolution is described in terms of surface differential operators. The numerical solution of such an equation on an arbitrary, time-dependent

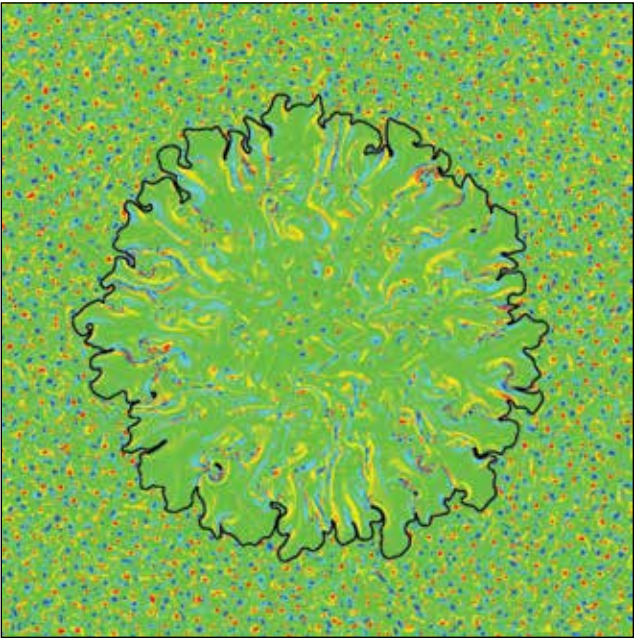


Figure 1: The structure of an expanding flame (black curve) in a turbulent medium; the wrinkled flame separates burned gas from the highly vortical flow of unburned gas (blue/red correspond to clockwise/counterclockwise vorticity). The highly wrinkled surface results from intrinsic combustion instabilities and the underlying turbulence; the lower-intensity turbulence in the burned gas is owing to gas expansion that results from the heat released during combustion.

surface is still an active area of research. Constructing an intrinsic coordinate system is nontrivial and computationally expensive, particularly for an evolving highly corrugated surface. The frequent need for remeshing renders this problem computationally intractable.

The research team uses a new class of embedding methods to extend the surface differential equation into Cartesian space such that when the solution is restricted to the surface, the solution to the original problem is recovered. This embedding method also allows the researchers to solve the partial differential equations using Cartesian operators instead of surface differential operators, further reducing execution time. The implementation appropriates concepts from computational geometry and has been used to study a diverse array of applications including image-processing.

## METHODS & CODES

The computations were made possible by coupling the embedded manifold approach with an adaptive mesh, variable-density incompressible Navier–Stokes solver. It is built on AMReX, an open source framework, to write massively parallel, block-structured adaptive mesh refinement applications [3]. The code underwent restructuring to improve efficiency and reduce memory demands for the nearly 30 million particles that constitute the turbulent spherical flame surface. Sample calculations illustrating a wrinkled flame front interacting with eddies of different sizes are depicted in Fig. 1.

## RESULTS & IMPACT

The major contribution of this work was to develop a scalable, hybrid partial differential equation-based Cartesian embedding method for moving surfaces. This method is capable of handling multivalued and disjointed flame surfaces to simulate a complex, turbulent premixed flame. The methodology was used to investigate influences of the hydrodynamic, or Darrieus–Landau, instability on early and long-time flame kernel development in turbulent flows. The hydrodynamic instability resulting from the gas expansion is responsible for the corrugated appearance of the flame surface even in the absence of significant perturbations such as boundaries, obstacles, and/or turbulence. The instability is known to induce acceleration and enhance the flame propagation speed. The most practical outcome of this work includes the derivation of scaling laws for the turbulent flame speed in terms of flow and physicochemical characteristics. Its importance is in estimating fuel-burning rates in internal combustion engines and other similar applications.

## WHY BLUE WATERS

Access to the Blue Waters system has been instrumental to this study; not only did it afford the research team access to extensive computing power, it also made available large memory nodes that were capable of processing the large number of particles on the flame surface.

## PUBLICATIONS & DATA SETS

M. Matalon, “The ramifications of the Darrieus–Landau instability in turbulent premixed flames,” in *SciTech Forum, Proc. AIAA SciTech 2019 Forum*, San Diego, CA, U.S.A., Jan. 7–11, 2019, doi: 10.2514/6.2019-0182.



DEEP LEARNING FOR HIGGS BOSON IDENTIFICATION AND SEARCHES FOR NEW PHYSICS AT THE LARGE HADRON COLLIDER

Allocation: Illinois/645.929 Knh  
PI: Mark Neubauer<sup>1</sup>  
Collaborators: Robert Gardner<sup>2</sup>, Dewen Zhong<sup>1</sup>

<sup>1</sup>University of Illinois at Urbana–Champaign  
<sup>2</sup>University of Chicago

EXECUTIVE SUMMARY

The Large Hadron Collider (LHC) is the world’s most powerful particle accelerator, designed to study the fundamental nature of matter and the forces that govern its interactions by colliding beams of protons at the highest-available energies. The research team is using Blue Waters to process, simulate, and analyze high-energy proton–proton collision data produced by the ATLAS experiment at the LHC and to improve researchers’ sensitivity to new phenomena by developing novel approaches to identifying Higgs bosons produced with high momentum at the LHC by using machine learning techniques.

RESEARCH CHALLENGE

The goal of particle physics is to understand the universe at its most fundamental level, including the constituents of matter, their interactions, and the nature of space and time itself. This quest is one of the most ambitious and enduring of human endeavors. The Standard Model (SM) of particle physics describes all known fundamental particles and their interactions, including the Higgs boson, which was discovered at the LHC in 2012 with significant contributions by the Illinois (Neubauer) Group. The discovery led to François Englert and Peter W. Higgs receiving the 2013 Nobel Prize in Physics. The SM has withstood the last 40 years of experimental scrutiny, with important exceptions being neutrino mass, dark matter, and dark energy. Recent developments in particle physics and cosmology raise the exciting prospect that we are on the threshold of a major step forward in our understanding. It is an enormous challenge to process, analyze, and share the 15 petabytes of data generated by the LHC experiments each year with thousands of physicists around the world. To translate the observed data into insights about fundamental physics, the

important quantum mechanical processes and the detector’s responses to them need to be simulated to a high level of detail and with a high degree of accuracy. A key thrust of this project is to use the Higgs boson to search for new physics in novel ways enabled by the Blue Waters supercomputer. The enormous energy available in proton–proton collisions at the LHC leads to the production of particles with very high velocity relative to the ATLAS detector (lab frame). Even massive particles like the Higgs boson can have a large momentum and, therefore, large Lorentz factor in the lab frame. When these “boosted” particles decay, their decay products are highly collimated (aka parallel rays) and not easily distinguished in the detector instrumentation (e.g., by calorimeters). This limits the sensitivity of searches for new physics such as  $X \rightarrow hS$ , where  $X$  is a new massive ( $\sim \text{TeV}/c^2$ ) particle,  $h$  is the Higgs boson, and  $S$  is a scalar particle that could be the  $h$  or a new particle yet to be discovered (Fig. 1). The research team has integrated Blue Waters into their production processing environment to simulate and analyze massive amounts of LHC data. Blue Waters resources are made available to the ATLAS computing fabric using a system called ATLAS Connect, which is a set of computing services designed to augment existing tools and resources used by the U.S. ATLAS physics community. Docker images are run via Shifter to create an environment on Blue Waters’ nodes that is compatible with the ATLAS job payload. Through the work of Dewen Zhong using Blue Waters and with collaborators at Indiana University and the University of Göttingen, the research team developed a novel four-prong jet tagger for boosted  $WW \rightarrow 4q$  identification based on machine learning techniques, including feed-forward deep, 2D convolutional,

Figure 1: (Left) Decay of a heavy hypothetical particle  $X$  to a Higgs boson and a scalar particle that decays to a  $W$  boson pair. The daughter particles from this decay are highly collimated in the lab frame and therefore difficult to reconstruct using traditional methods. (Right) Minimum distance between final-state quarks versus Higgs boson transverse momentum. The dashed line shows the typical jet clustering radius, indicating that the majority of decay products are in close proximity to each other.

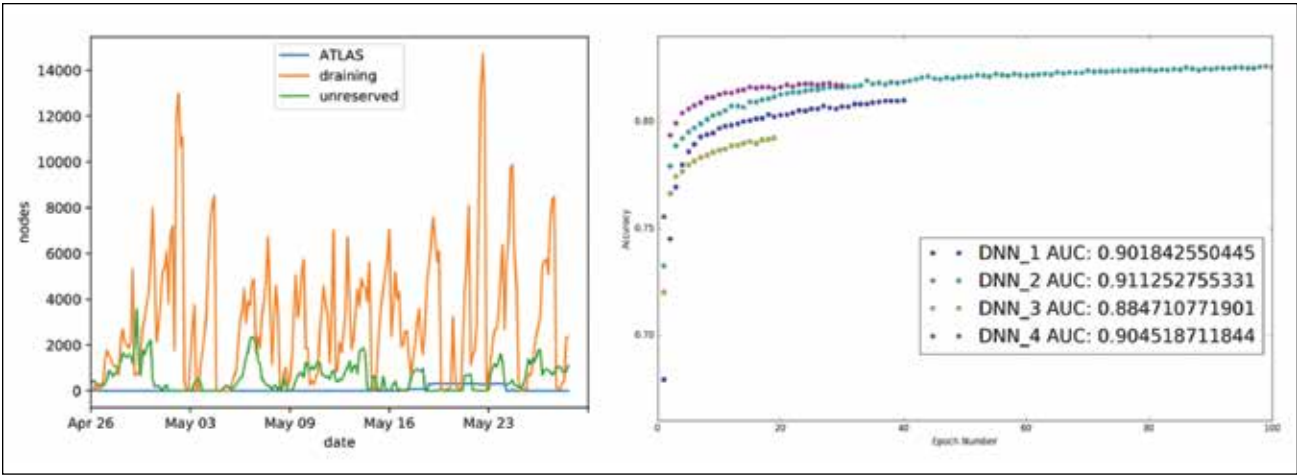
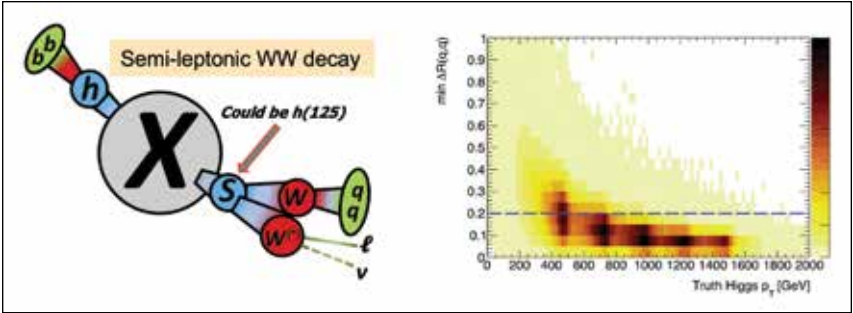


Figure 2: (Left) This shows a period of time during which 35 million ATLAS events were processed using 300 Blue Waters nodes. Utilization during this period averaged 81%, which is typical for Blue Waters. (Right) Higgs boson identification accuracy versus training epoch for the boosted Higgs boson tagger developed on Blue Waters.

and recursive neural networks. Among the best performing approaches is a deep neural network (DNN) that has been the primary study of Zhong. The research team is working to extend this approach to the single-lepton channel, which brings additional challenges, and to develop it as a general  $WW$  tagger for ATLAS that can be applied to searches involving boosted dibosons and not limited to the semileptonic Higgs+scalar searches described here. The team is using Blue Waters to implement this approach as a jet-mass decorrelated tagger to avoid learning the mass and training DNNs using both calorimeter and tracking information.

RESULTS & IMPACT

Fig. 2a shows a particular one-month period in 2018 in which 35,000 Blue Waters cores were utilized to process 35 million collision events. The top panel of this figure shows that this approach is cost-effective, boosting cluster utilization, and has no adverse effect on other high-performance computing workloads. The job output was made available to the rest of the ATLAS collaboration for use in analysis of the LHC data to improve SM measurements and to search for new physics beyond the SM. Fig. 2b shows the Higgs boson identification accuracy as a function of the number of training epochs for a variety of DNN configurations and hyperparameter settings. The team is also using Hyperas, a convenience wrapper using Hyperopt with Keras models, on Blue Waters to automate the scanning of hyperparameters in a variety of machine and deep learning approaches to improve the Higgs boson identification over backgrounds. The techniques show promise in addressing the challenges of boosted Higgs boson identification and improving the sensitivity of new physics searches at the LHC.

WHY BLUE WATERS

Blue Waters, as a large CPU and GPU resource with high data-throughput capabilities, greatly facilitated this research. The strong support for containers allowed the research team to deploy their science application on Blue Waters’ nodes. Also, Blue Waters provided a means for a highly parallelized and automated scanning of free parameters in the team’s machine learning configurations and, therefore, rapid optimization of the researchers’ boosted Higgs boson identifier.

PUBLICATIONS & DATA SETS

M. Belkin *et al.*, “Container solutions for HPC systems: A case study of using shifter on Blue Waters,” in *Practice and Experience on Advanced Research Computing 2018, Proc. PEARC ’18*, Pittsburgh, PA, U.S.A., July 22–26, 2018, pp. 43.1–43.8. E. A. Huerta, R. Haas, S. Jha, M. Neubauer, and D. S. Katz, “Supporting high-performance and high-throughput computing for experimental science,” *Comput. Softw. Big Sci.*, vol. 3, no. 5, 2019.



# DESIGNING QUANTUM LOGIC GATES ON SILICON CHIPS WITH LARGE-SCALE MULTIPHYSICS SIMULATIONS

**Allocation:** Innovation and Exploration/221.4 Knh  
**PI:** Rajib Rahman<sup>1</sup>  
**Collaborator:** Andrew S. Dzurak<sup>2</sup>

<sup>1</sup>Purdue University  
<sup>2</sup>University of New South Wales

## EXECUTIVE SUMMARY

Quantum computing is a revolutionary method of information processing that takes advantage of the quantum mechanical principles of superposition and entanglement. It is theoretically proven to provide a massive improvement of the complexity of certain algorithms used in cryptography, database search, and modeling correlated quantum systems such as atoms, molecules, and superconductors.

One of the most scalable designs of quantum information processing chips is based on the interactions between a few electrons

trapped in silicon chips under metallic electrodes. These trapped electrons can be selectively manipulated in a crystal containing hundreds of millions of other electrons and nuclei, defects, magnetic and electrical noise, and mechanical vibrations [1]. In this project, the research team developed a method for computing the interactions between the trapped electrons in a realistic environment and studied the effect of variations in the environment on the interactions. This study will help in accelerating the experimental and chip design processes of quantum computing systems.

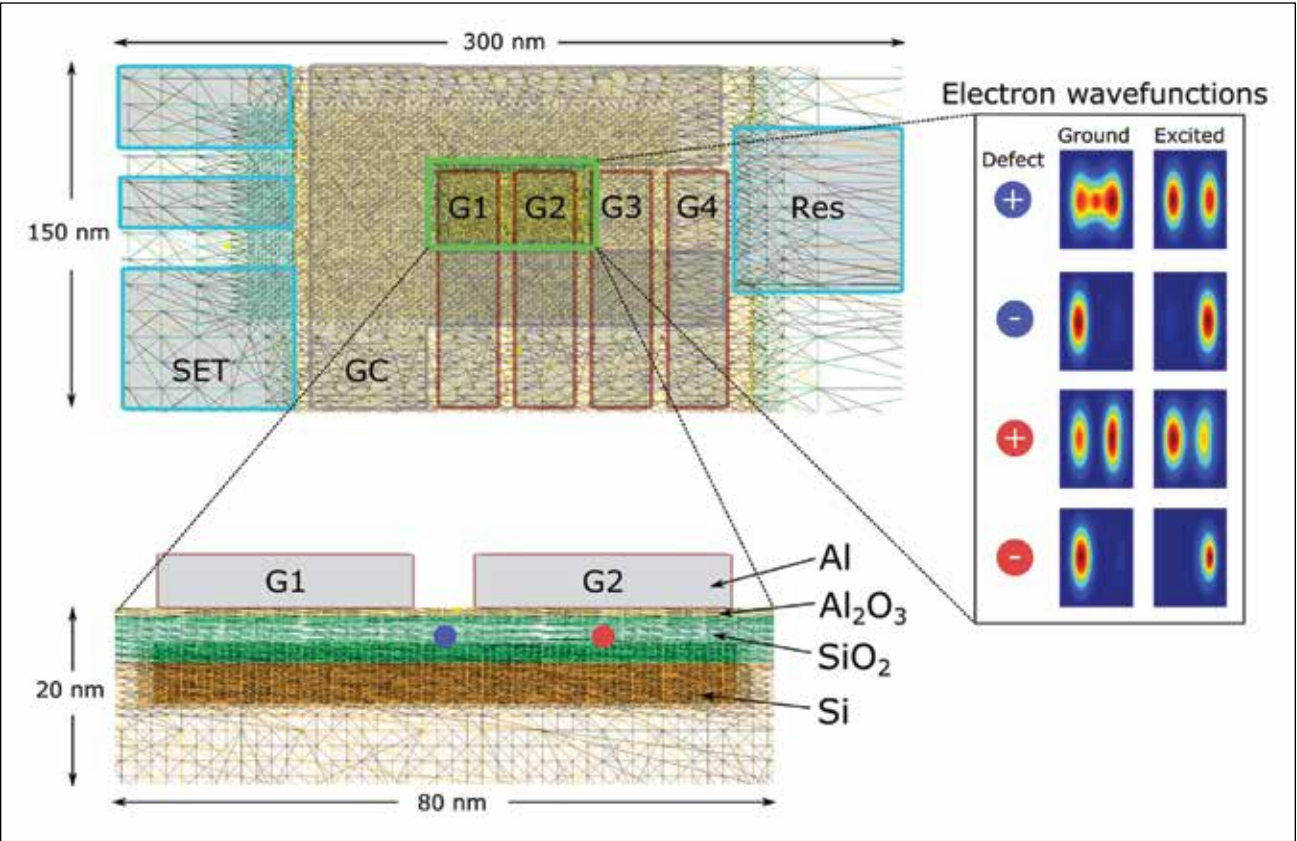


Figure 1: Top and side views of a multiqubit silicon device showing gate electrodes (G), a single-electron transistor (SET), a charge reservoir region, dielectrics ( $\text{Al}_2\text{O}_3$  and  $\text{SiO}_2$ ), and nonideal charge traps. Mixed inhomogeneous meshing was used to solve for the quantum electronic density of trapped electrons in continuum self-consistently. Atomistic quantum dot wave functions formed under electrodes G1 and G2 are shown for different defect locations and charges.

## RESEARCH CHALLENGE

The team computed the interaction between two electrons trapped under positively charged electrodes occupying an area of 25 nanometers (nm) x 25 nm and separated by 10 nm. The efficient working of these chips requires the interaction strength to be designed accurately. The interaction strength depends on multiple factors such as device geometry, electrode voltages, neighboring crystal atoms, proximity to other charged defects, electrical noise, and crystal strain. The biggest challenge is to efficiently address the impact of all these factors to provide insight and guidance into the design of quantum logic gates.

## METHODS & CODES

The factors that determine interaction strength can be grouped by the physics that efficiently describe them, *i.e.*, continuum, atomistic, and correlated. The team has developed three different modeling tools to address each of these: (1) finite-element method-based software for continuum physics, (2) large-scale atomistic tight binding software for atomic and solid-state physics, and (3) full configuration interaction method-based software for correlated physics. The tools consecutively process data to arrive at the final results. The methods are embedded in the NEMO package.

## RESULTS & IMPACT

The research team has shown the impact of gate lengths, gate spacings, and voltages on the quantum mechanical interaction strength between two qubits (the basic unit of quantum information). This “exchange” interaction has been shown to be sensitive to the design. The researchers have proposed the use of a special “exchange gate” for efficient control of two qubits. They also have shown that nonideal factors such as defects and strain can cause electrical distortions in the device. However, in most cases electrical tuning of gate voltages can be used to overcome nonideal behavior.

## WHY BLUE WATERS

All three methods used in this project require large processing power, memory, and storage facilities. Furthermore, the methods are often run sequentially with data being translated from one form to another. Blue Waters provided all the needed resources. Access to a large number of nodes and storage disks was particularly critical. Furthermore, the support staff constantly monitored the system and helped with libraries and compilation, which enabled the research team to focus on the science.

## PUBLICATIONS & DATA SETS

H. Sahasrabudhe and R. Rahman, “Computational modeling of exchange splitting in Si/SiO<sub>2</sub> based double quantum dots,” presented at the Amer. Phys. Soc. March Meeting, Boston, MA, U.S.A., March 4–8, 2019, id.E35.004.



SIMULATION OF ROTATING DETONATION ENGINES

Allocation type: GLCPC/480 Knh  
PI: Venkat Raman<sup>1</sup>

<sup>1</sup>University of Michigan

EXECUTIVE SUMMARY

Rotating detonation engines (RDEs) have received increased attention as possible replacements for conventional gas turbine systems owing to their expected high thermal efficiency. The compression across the detonation wavefront driven by chemical reactions induces additional pressure gain. Although extensive research has been performed by both the experimental and numerical communities to realize these devices, they are limited to hydrogen/air chemistry because of its high detonability and relatively low-complexity kinetic mechanism. To make RDEs more practical and applicable, simulations of the detailed physics with hydrocarbon chemistry need to be conducted. With the allocated time on Blue Waters in 2019, the researcher has performed a series of full-system simulations with ethylene/air chemistry. These simulations reveal that the slow wave propagation observed in experiments is due to a wave aided by the deflagration mode, with wave speeds within 5% of the experimental results.

RESEARCH CHALLENGE

Although RDEs are increasingly being studied, the concept originates from the 1960s. Recent advancements in materials science and a better understanding of detonative combustion enable the RDE system to be operable. Extensive research has been performed through experiments and simulations to understand and optimize the operation of RDEs [1–3]. However, most of the research has been conducted with hydrogen/air mixtures, while hydrocarbon chemistry is the more practical fuel [2,3]. Several experiments with ethylene/air mixtures reveal that the wave

speed is drastically slower than hydrogen/air mixtures, at nearly 50% of the ideal case [1]. Furthermore, the flame front is thicker than that of hydrogen/air chemistry. To understand the detailed physics of combustion in RDEs, simulations of the full system are necessary.

Full-system RDE simulations with ethylene/air chemistry raise three primary challenges owing to their unique geometry and dynamics. First, the fuel and oxidizer are separately injected using a nonpremixed injection scheme, which requires the use of unstructured grids. Secondly, ethylene/air chemistry is more complex and stiffer than hydrogen/air chemistry. In other words, the number of species being transported will increase the computational cost. Lastly, a finer resolution is necessary to resolve the reaction induced by the shock wave at the wavefront. Since the geometry is complex and three-dimensional, a finer resolution exponentially increases the computational cost. To perform the full-system simulation, a highly scalable solver and a large amount of allocation time on high-performance computing resources are necessary.

METHODS & CODES

The University of Michigan detonation solver, UMdetFoam, is based on OpenFOAM and Cantera. OpenFOAM provides the finite-volume tools, which are widely used in the turbulent and combustion community. UMdetFoam implements a MUSCL-based Harten–Lax–van Leer-contact scheme for space and second-order Runge–Kutta scheme for time integration. Diffusion terms are discretized by the Kurganov, Noelle, and Petrova method. In order to handle chemistry integration, this finite-volume meth-

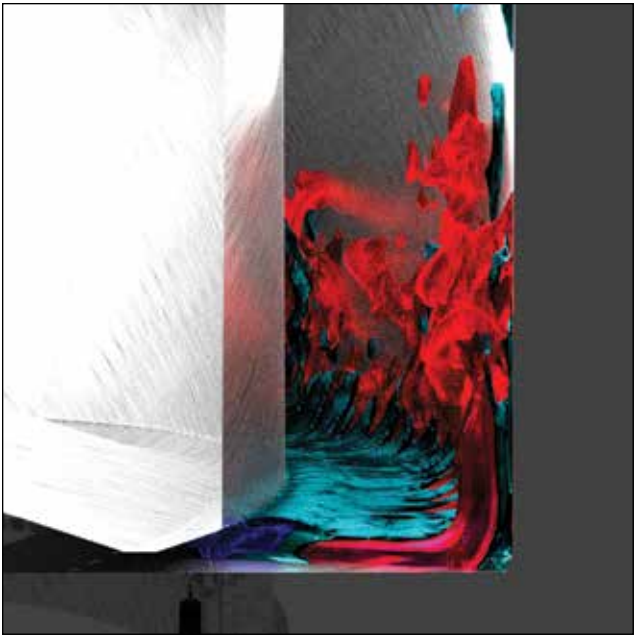


Figure 2: Injection dynamics of the fuel and oxidizer streams within the RDE annulus. The fuel stream is pushed to the outer wall by the oxidizer stream, imposing fuel/oxidizer stratification within the system.

od solver was coupled with Cantera, an open source framework for chemical kinetics, thermodynamics, and transport processes. UMdetFoam has MPI-based parallelism with linear scalability demonstrated up to 60,000 cores, with support for GPUs. For this study, a reduced-order two-step mechanism was employed to reduce the computational cost.

RESULTS & IMPACT

In this study, the ethylene/air RDE facility at the Air Force Research Laboratory was simulated. The channel width was 22.86 mm, and the dimensionless area ratio between the oxidizer inlet and the detonation chamber was 0.059.

The full-system simulation enabled extraction of extensive data on the physics in the combustion facility. Fig. 1 shows the general behavior of the system. The pressure wave propagates in the azimuthal direction, aided by the reaction. The wavefront of this RDE is thicker than that of hydrogen/air, as reported in the 2018 Blue Waters Annual Report. Furthermore, oblique shock waves are not created with ethylene/air mixtures, and the product gases appear at some distance from the chamber bottom. This wave standoff behavior is also observed in the experiment. Analysis suggests that the wave is not in the detonation mode but is in a deflagration mode strong enough to sustain the wave. The incomplete mixing and low detonability of the mixture cause this deflagration-dominant mode of operation. The simulation captures the slow wave speed within 5% of the experimental data, which is almost 50% of the ideal condition. The fuel and oxidizer injection behavior are extracted from the simulation as well. The fuel stream is pushed to the outer wall by the oxidizer stream, as

shown in Fig. 2. In other words, the mixing is not enhanced near the chamber bottom, causing the standoff of the flame.

The simulation also shows that the injection velocity was almost constant during RDE operation because the wave is too weak to affect the injection dynamics. This result can help researchers to optimize the design of RDEs and to understand the effect of low detonability and fuel-oxidizer stratification on the wave structure. Design improvements resulting from high-fidelity simulations and experiments bring RDE systems closer to realization.

WHY BLUE WATERS

The Blue Waters high-performance computing resource greatly accelerated this research for three primary reasons: (1) users are provided with a large number of cores to run cases and exploit the parallelizability of the solver, which is necessary to simulate full-system RDEs; (2) the working directory allows users to store the large amounts of data (50 TB) generated by numerical simulations without concerns over exceeding a storage quota; and (3) helpful support staff are available to professionally and promptly reply to user questions and concerns. In this study, nearly 10,000 cores were parallelized to perform full-system RDE simulations. The computing resources allowed the project to quickly implement, test, and deploy computational tools and to achieve the desired results.

PUBLICATIONS & DATA SETS

T. Sato and V. Raman, Detonation structure in ethylene/air-based non-premixed rotating detonation engine, submitted to *J. Propul. Power*, 2019.

S. Prakash, R. Fiévet, V. Raman, J. Burr, and K. H. Yu, “Analysis of the detonation wave structure in a linearized rotating detonation engine,” *AIAA Journal Special Issue on Continuous Detonation and its Applications* (in press), 2019, doi: 10.2514/1.J058156.

S. Prakash and V. Raman, “Detonation wave propagation through stratified fuel-air mixtures,” submitted to *Shock Waves*, 2019.

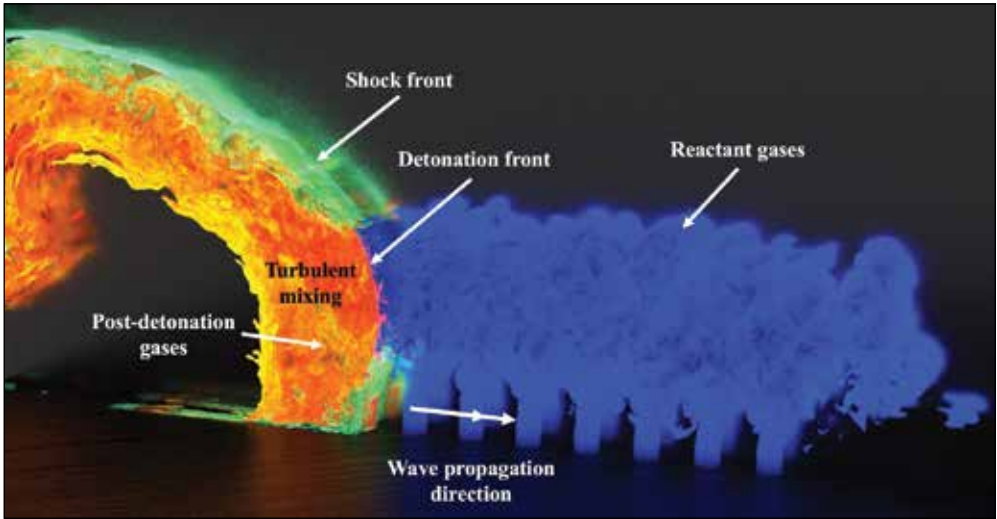


Figure 1: General behavior of the detonation wave through a linear array of injectors. A detonation wave propagates to the right, consuming an inflow mixture of premixed hydrogen/air reactant gases.



DI

TN

MAPPING PROTON QUARK STRUCTURE: LOOKING INSIDE THE PROTON—HOW DO QUARKS SPIN?

**Allocation:** NSF PRAC/9,440 Knh  
**PI:** Caroline Riedl<sup>1</sup>  
**Co-PIs:** Riccardo Longo<sup>1</sup>, Matthias Grosse Perdekamp<sup>1</sup>  
**Collaborators:** Marco Meyer<sup>1,2</sup>, Robert Heitz<sup>1</sup>, April Futch<sup>1</sup>, Charles Naim<sup>2</sup>, Angelo Maggiora<sup>3</sup>, Andrei Gridin<sup>4</sup>, Nicolas Pierre<sup>2</sup>, Anatolii Koval<sup>5</sup>, Celso Franco<sup>6</sup>, Po-Ju Lin<sup>2</sup>, Florian Kaspar<sup>7</sup>, Henri Pekeler<sup>8</sup>, Bakur Parsamyan<sup>3,9</sup>

<sup>1</sup>University of Illinois at Urbana–Champaign  
<sup>2</sup>CEA–Saclay  
<sup>3</sup>INFN Torino  
<sup>4</sup>JINR Dubna  
<sup>5</sup>University of Warsaw  
<sup>6</sup>LIP Lisbon  
<sup>7</sup>University of Munich  
<sup>8</sup>University of Bonn  
<sup>9</sup>CERN

EXECUTIVE SUMMARY

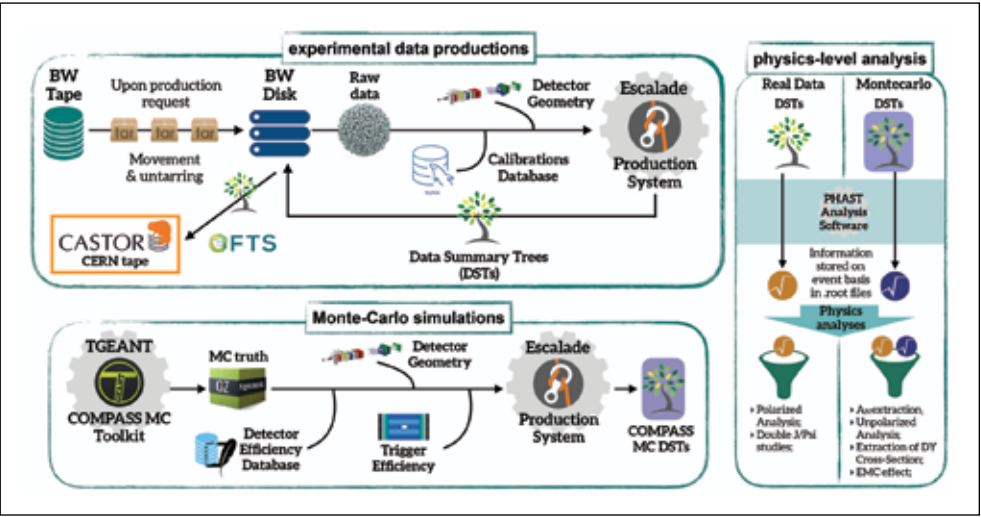
COMPASS (the Common Muon and Proton Apparatus for Structure and Spectroscopy experiment) probes proton substructures scattering high-energy pion and muon beams off nuclear targets at CERN. The experiment explores the momentum and coordinate phase space of quarks inside the proton. Observing correlations between proton spin and the transverse momentum of quarks will shed light on the quark dynamics inside the proton and will provide a critical test of fundamental predictions derived from quantum chromodynamics, the quantum field theory describing the nuclear force. The measurements produced 4.5 petabytes of experimental and simulated data. Raw COMPASS data were transferred to Blue Waters and converted into a format suitable for physics-level analysis. In parallel, precise detector efficiency maps were extracted from the data and simulations were carried out to correct the data for experimental imperfections. **Blue Waters enabled the research team to share the first**

**physics results of the 2018 COMPASS run with the nuclear physics community about one year earlier than usual. This fast turnaround time for the analysis was unprecedented.**

RESEARCH CHALLENGE

Observation of the sign change of the Sivvers quark distributions (“Sivvers functions”) in Drell–Yan scattering compared to existing measurements in semi-inclusive deep-inelastic scattering (SIDIS) is one of the few performance Nuclear Science Advisory Committee [1] milestones for DOE- and NSF-funded research in nuclear physics. The 2015 and 2018 Drell–Yan campaigns of the COMPASS experiment at CERN constitute the first measurements addressing this question [2]: the negative pion beam from the Super Proton Synchrotron was impinged on a target of transversely polarized protons. Sivvers functions arise from correlations between proton spin and quark transverse momentum and thus appear connected to quark orbital motion inside the proton.

Figure 1: Work flows for COMPASS experimental data productions (top), Monte–Carlo simulations (bottom), and physics-level analysis (right). (Courtesy R. Longo.)



METHODS & CODES

The team used Blue Waters (BW) for four major tasks: experimental data production to convert raw data into a format for physics-level analysis; extraction of detector efficiency maps from raw data as input to realistic simulations; Monte–Carlo simulations; and physics-level analysis. The various work flows are sketched in Fig. 1.

Three petabytes of raw COMPASS data collected at CERN were transferred to BW using the File Transfer System FTS3 [3], a bulk data mover created to globally distribute CERN–LHC data. The data were packed into tar files of 100 GB on BW Lustre and then stored on BW tape. Upon production request by the COMPASS analysis coordinator, they were then retrieved from tape.

For each triggered event in COMPASS, the information of the detectors was recorded by the Data AcQuisition (DAQ) system. The COMPASS Reconstruction Analysis Library (CORAL) software performed the conversion of raw data information to physical quantities. CORAL’s function was to reconstruct particle trajectories and momenta, and the position of vertices. The reconstructed information was stored in the form of Data Summary Trees (DSTs), which were read and analyzed using the COMPASS Physics Analysis Software Tools (PHAST). CORAL and PHAST jobs were submitted to the BW nodes using the production framework ESCALADE, which allows for a detailed bookkeeping of job status, failure, and output. Detector efficiencies were extracted from a sampled fraction of the experimental data. They required separate submissions to the BW grid for each of the 240 detector planes, which made the efficiency maps about a factor of seven more CPU-expensive compared to the standard data productions.

The production of Monte–Carlo data began with the generation of signal and background events with event-generator packages. For the simulation of the detector response to the physics event, a GEANT4 [4] toolkit was then used based on the description of the COMPASS apparatus. Lastly, simulated hits were subject to the same reconstruction CORAL and PHAST codes as experimental data.

RESULTS & IMPACT

Blue Waters enabled the research team to share the physics result of the COMPASS Drell–Yan 2018 data-taking campaign with the nuclear physics community in an unprecedentedly timely manner. Not only was it the first time in COMPASS history that full-scale productions were run before the end of the data-taking in quasi-online mode, the team was also able to release the data to the public half a year after the end of the data-taking, three times faster than for the previous 2015 Drell–Yan campaign. Results using data produced on BW were presented in four talks at an important conference in April 2019 [5]. A highlight—the Sivvers asymmetry from COMPASS Drell–Yan data—is shown in Fig. 2.

While the team reported last year about radiation simulations on BW to prepare the hardware setup for the 2018 run, this year the group ran data productions for the entire 2018 run on BW as

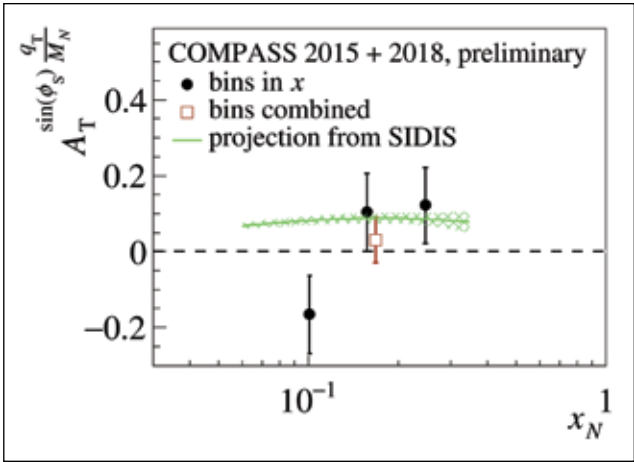


Figure 2: The Sivvers amplitude (qT-weighted) extracted from COMPASS Drell–Yan data. The green curve represents a projection assuming the sign change. The data are consistent with the predicted sign change of the Sivvers function.

well as second-pass productions with improved detector calibrations and alignments for the 2015 Drell–Yan data. Precise detector efficiency maps were extracted separately from the data, allowing for a novel treatment of simulated data. Blue Waters allowed the team to run simulations of COMPASS data at unprecedented precision with realistic, wide-open pile-up gates, which makes every scattering event CPU-expensive. Simulations and efficiency maps were also run for the complementary physics of Generalized Parton Distributions and multiplicities of hadrons (particles made of quarks) produced in deep-inelastic scattering of muons off protons from COMPASS 2016/17 data.

This Blue Waters project involved a dozen graduate students and young postdocs running data productions, simulations, and physics analysis codes. It thus offers outstanding educational potential toward building a community capable of using petascale computing.

WHY BLUE WATERS

Blue Waters allowed the research team to analyze the COMPASS data of the 2018 Drell–Yan campaign at unprecedented precision and speed. It also enabled the use of novel methods of data treatment. For the realization of an efficient data production flow, it was essential that BW provided major online disk space, which allowed the researchers to manage hundreds of terabytes of data between disk and tape in parallel, while still having some space available for DST output of the codes. In addition, BW staff helped to optimize job submission campaigns so that they would smoothly flow through the BW system.

PUBLICATIONS & DATA SETS

R. Heitz, “Transverse momentum dependent nucleon structure from pions impinged on a transversely polarized proton target,” Ph.D. dissertation, Univ. Ill. Urbana–Champaign, May 2019.



DI

TN

ELECTRON DYNAMICS OF ION-IRRADIATED TWO-DIMENSIONAL MATERIALS

Allocation: Illinois/200 Knh  
PI: André Schleife<sup>1</sup>  
Co-PI: Alina Kononov<sup>1</sup>

<sup>1</sup>University of Illinois at Urbana–Champaign

EXECUTIVE SUMMARY

Exploiting the unique properties of two-dimensional materials for next-generation electronic devices and other novel technologies depends on high-resolution techniques for nanoscale imaging and structure manipulation, which often employ focused ion beams. Along with radiation-induced degradation of thin materials and materials’ surfaces in space and nuclear applications, solving this technological challenge demands a detailed understanding of the response of thin materials to ion irradiation. In order to accurately study the subfemtosecond electron–ion dynamics in thin materials under ion irradiation, the research team performed first-principles simulations of few-layer graphene irradiated by charged particles ranging from protons to xenon ions. This research lays the groundwork for a predictive computational framework capable of determining optimal ion beam parameters for a desired imaging or patterning application and of indicating a material’s susceptibility to radiation-induced defects.

RESEARCH CHALLENGE

Two-dimensional materials have a variety of remarkable properties, making them promising candidates for a wide range of potential applications including flexible electronics, solar cells, nanoscale sensors, and other electronic devices [1]. However, the properties of atomically thin materials are often sensitive to de-

fects, nanopores, functionalization, and other types of nanostructure, which can either degrade performance when undesirable or enable an application when intentional [2,3]. Thus, precise techniques for imaging and patterning 2D materials, likely reliant on focused beams of energetic ions, are necessary for scalable and reliable manufacturing of devices based on 2D materials.

In addition, materials for space and nuclear applications must withstand constant bombardment by energetic ions, and mitigating erosion of nuclear cladding or radiation shielding materials fundamentally involves controlling surface behavior. Overcoming these engineering challenges requires a thorough understanding of the response of 2D materials and materials’ surfaces to ion radiation.

However, the vast majority of current knowledge about ion-irradiated materials pertains to bulk materials, leaving surface effects largely unknown. The research team’s first-principles approach provides unprecedented insight into electron dynamics in ion-irradiated materials, which is critical for developing both novel devices based on 2D materials and radiation-resistant materials for space and nuclear applications (see Fig. 1).

METHODS & CODES

The research team used Qbox/Qb@ll [4], their highly parallel implementation of real-time time-dependent density functional

theory [5], to perform accurate first-principles simulations of excited-electron dynamics. This approach treats nuclei as classical point charges interacting electrostatically with electrons. Electrons are treated quantum-mechanically; their quantum orbitals, represented in a plane-wave basis, are governed by the time-dependent Kohn–Sham (TDKS) equations, a system of coupled partial differential equations. The team used the common adiabatic (without transfer of heat or mass) local-density approximation for the exchange–correlation potential describing the quantum correction to the electron–electron interaction.

Starting with the lowest energy configuration of the material as the initial condition, the electronic orbitals were propagated in time by numerically integrating the TDKS equations. The simulations generated a time-dependent electron density that was further analyzed to extract the secondary electron yield and the charge captured by the projectile, among other quantities.

RESULTS & IMPACT

First, the team explored new numerical methods that could accelerate first-principles simulations of excited-electron dynamics, enabling study of longer timescales and larger systems. Specifically, the researchers interfaced Qb@ll with the PETSc library [6,7], allowing rapid testing of different numerical integration algorithms for propagating the electronic states over time. Thus far, the team has concluded that enforced time-reversal symmetry outperforms all available Runge–Kutta schemes in terms of stability and time-to-solution. Identifying an integrator that reduces computational cost without sacrificing accuracy would help make feasible extremely accurate simulations of defect formation in materials. This capability would deepen the understanding of defect formation mechanisms, provide transformative insight for nanostructure engineering techniques, and expedite the development of novel electronic devices for the benefit of society.

The team also implemented and applied new analysis techniques to extract the projectile charge state and the kinetic energy spectrum of emitted electrons. These methods provide additional information about the charge dynamics and energy dissipation mechanisms occurring as the ion interacts with the material, allowing more detailed insight into surface effects and predictions of projectile parameters most likely to induce defects. While models for bulk materials typically assume an equilibrated projectile charge and do not consider energy lost to electron emission, the team found that a highly charged projectile’s effective charge varies dynamically as it traverses graphene and does not reach an equilibrium value within a few-layer sample (see Fig. 2). Furthermore, the initial charge of a heavy projectile strongly influences both energy transfer and electron emission, and accounting for electron emission can change the projectile parameters that maximize energy deposited in the material. These findings not only demonstrate the disparity in behavior between few-layer material and bulk materials but also establish the need for further study of ion-irradiated 2D materials and special tuning of ion beam parameters for imaging and processing 2D materials.

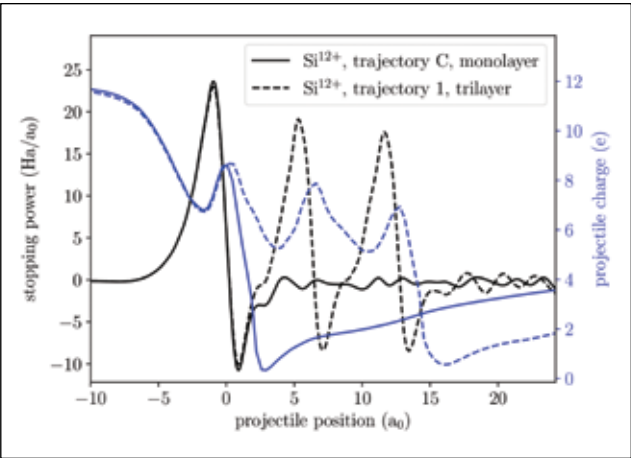


Figure 2: Instantaneous stopping power (black) and effective projectile charge (blue) as a Si<sup>12+</sup> ion traverses monolayer (solid line) and trilayer (dashed line) graphene. Both the stopping power and projectile charge evolve dynamically, without reaching an equilibrium value characteristic of bulk materials. (Units: a<sub>0</sub> is the Bohr radius of an electron; Ha is an energy unit equal to 27.2 eV.)

WHY BLUE WATERS

Blue Waters enabled the research team to conduct the long simulations of large systems involved in this project. In order to accurately model a few-layer material or a material surface under irradiation, scientists must evolve hundreds or thousands of electrons over thousands of timesteps in an elongated simulation cell containing a large vacuum outside the material. These aspects make the simulations computationally expensive and are only possible with a massively parallel implementation of the first-principles approach and a high-performance supercomputer.

PUBLICATIONS & DATA SETS

K. Kang *et al.*, “Pushing the frontiers of modeling excited electronic states and dynamics to accelerate materials engineering and design,” *Comp. Mater. Sci.*, vol. 160, pp. 207–216, Apr. 2019, doi: 10.1016/j.commatsci.2019.01.004.

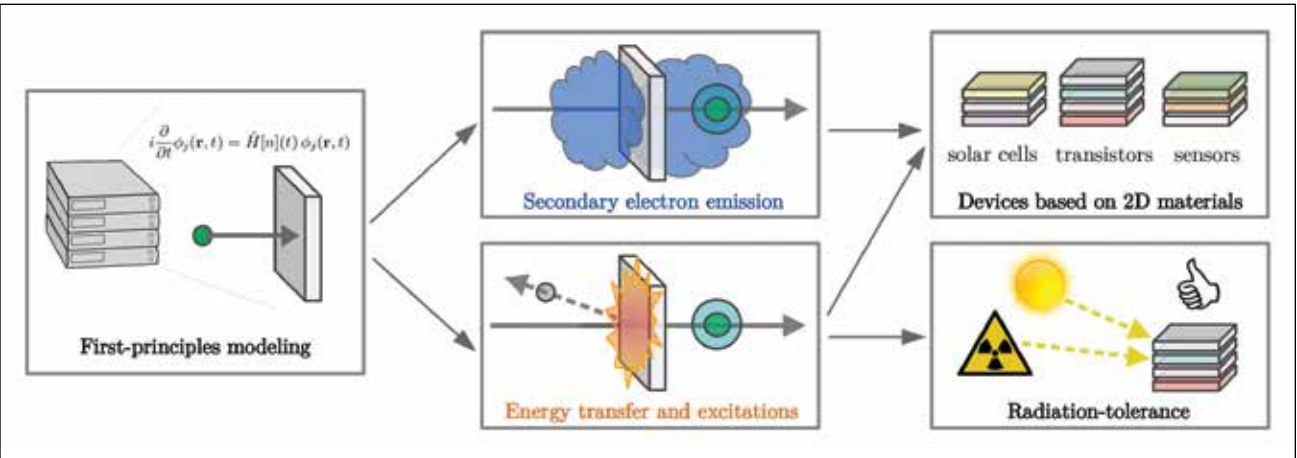


Figure 1: The first-principles simulations of ion-irradiated few-layer materials provide information about secondary electron emission, energy transfer, and excited electron dynamics. Understanding these processes will advance radiation tolerance and improve ion beam imaging and patterning techniques needed to create novel electronic devices based on 2D materials.



DI

DISCOVERY OF NEW PLASMONIC MATERIALS VIA HIGH-THROUGHPUT MACHINE LEARNING

Allocation: Blue Waters Professor/245 Knh  
PI: André Schleife<sup>1</sup>  
Co-PI: Ethan Shapera<sup>1</sup>

<sup>1</sup>University of Illinois at Urbana–Champaign

EXECUTIVE SUMMARY

Plasmonics aims to manipulate light through choice of materials and nanoscale structure. Finding materials that exhibit low-loss responses to applied optical fields while remaining feasible for widespread use is an outstanding challenge. Online databases have compiled computational data for numerous properties of tens of thousands of materials. Owing to the large number of materials and high computational cost, it is not viable to compute optical properties for all materials from first principles.

For this project, plasmonic quality factors for a training set of 1,000 metals and 2,000 semiconductors were computed using density functional theory (DFT) and the Drude model. The research team trained regressors to rapidly screen the Materials Project (MP) database to identify potential new plasmonic metals. Descriptors were limited to symmetry and quantities obtained using the chemical formula. The machine learning models filtered through 7,445 metals in the MP database. From this, the team predicted AlCu<sub>3</sub>, ZnCu, and ZnGa<sub>3</sub> as candidates and verified their quality factors with DFT.

RESEARCH CHALLENGE

As mentioned before, the field of plasmonics seeks to manipulate light at the nanoscale. Precise control over plasmon response enables many applications including subwavelength waveguides [1], nanoantennas [2], superlenses [3], subwavelength imaging [4], nanocircuitry [5], and biosensors [6]. Currently used plasmonic materials consist of noble metals and electron-doped semiconductors with high electrical conductivity [7]. Unfortunately, noble metals suffer from large losses in the visible spectrum owing to absorption while semiconductors require high electron doping concentrations. Further advances of plasmon-based technology require finding new high-performance materials.

METHODS & CODES

To quantify the response of a material to an applied electrical field, the team computed its dielectric function. Two dominant contributions to the dielectric function were considered: interband transitions of electrons from valence states to conduction states and intraband oscillations of electrons near the Fermi energy. Interband transitions were captured through DFT calculations performed using the Vienna Ab initio Software Package (VASP) [8]. The contribution to the dielectric function from intraband oscillations was described by the Drude conduction model with

a plasma frequency obtained through DFT. From the dielectric functions, plasmonic quality factors were computed based on analytic equations. DFT input files for 1,000 metals and 2,000 semiconductors to form the training sets were randomly chosen from the MP database with the pymatgen open source library [9,10].

Random forest regressors with adaptive boosting implemented in scikit-learn [11] related material descriptors to the DFT-calculated plasmonic quality factors. The descriptors consist of material properties readily obtainable from the chemical formulas and standard tables, *e.g.*, atomic masses and electronegativities. Constructed models were validated using an 80–10–10 validation scheme. With this approach, 10% of materials in the training set were randomly selected to form a testing set to be left fixed throughout the process. The remaining materials were randomly divided 80%–10% into fitting and validation sets. The fitting set was used to fit the model, with the model subsequently applied to the validation set. Model construction was then iterated over 100 random fitting–validation divisions. Each model was applied to all metals and semiconductors in the MP database to predict quality factors for all available materials. For the 200 materials with the largest predicted quality factors, the dielectric functions and quality factors were explicitly calculated with DFT.

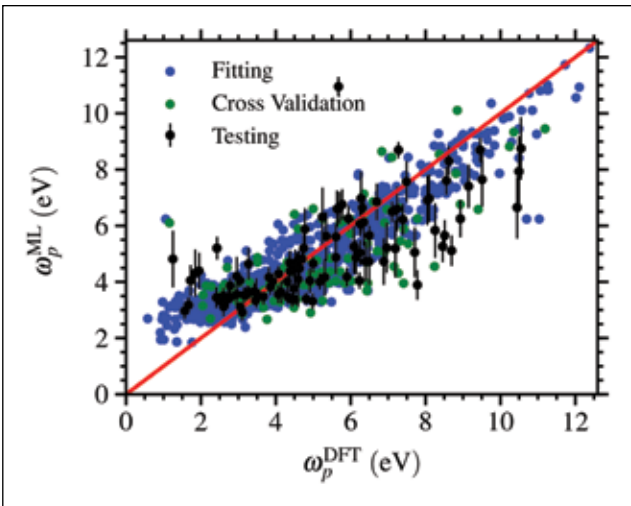


Figure 1: Comparison of metal plasma frequency predicted via machine-learned regressor vs. calculated with DFT. Plotted are the results of one fitting–validation iteration and average values and error bars for 100 iterations for testing. The red line is to guide the eye for machine learning predictions matching the DFT value.

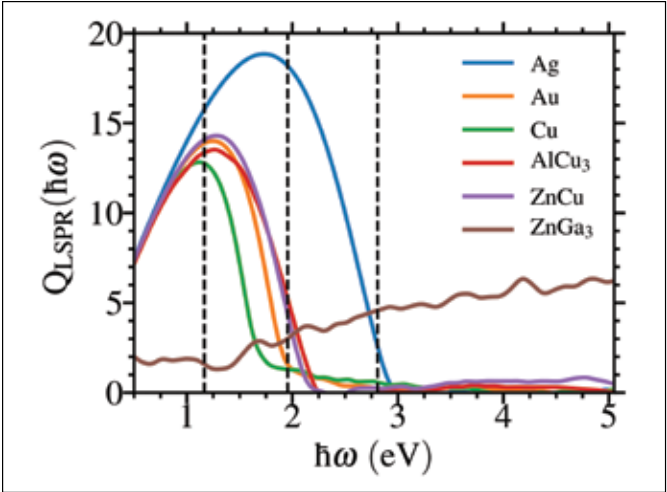


Figure 2: Computed plasmonic quality factor for currently used metals, Ag, Au, and Cu, and proposed new plasmonic metals, AlCu<sub>3</sub>, ZnCu, and ZnGa<sub>3</sub>. Vertical dashed lines correspond to operating energies of commonly used lasers. AlCu<sub>3</sub> and ZnCu are candidate plasmonic metals at low energy while ZnGa<sub>3</sub> shows large-quality factors in the UV spectrum.

RESULTS & IMPACT

We have used DFT calculations of materials’ dielectric functions to train and validate machine learning models which predict plasma frequencies and plasmonic quality factors for metals and semiconductors in the Materials Project database. The constructed models allow for rapid calculation of the optical properties necessary to predict the suitability of a material for plasmonic applications. By applying the models to a large database, we have been able to search for predicted high-quality factor metals and semiconductors without requiring explicit DFT calculations for all materials in the database, significantly reducing the computational cost. From the database metals, we have identified three potential new metals, AlCu<sub>3</sub>, ZnCu, and ZnGa<sub>3</sub>. AlCu<sub>3</sub> and ZnCu are predicted to match or outperform the commonly used Au, and Cu for energies between 1 eV and 2 eV, the infrared and low energy end of the visible spectrum. ZnGa<sub>3</sub> shows high-quality factors for energies between 2 eV and 3 eV, outperforming currently used metals for higher-energy plasmonics.

WHY BLUE WATERS

The present research requires DFT simulations for thousands of materials with up to dozens of atoms. The computational cost for DFT calculations is further increased by the need for accurately mapping the electronic energies near the Fermi surface. Dense 31 x 31 x 31 point meshes to sample momentum space are required to obtain converged calculations of the electronic properties near the Fermi energy. This imposes challenging requirements for CPU hours that cannot be met by a system such as the Campus Cluster. The computational cost of DFT scales as the cube of the number of electrons; for large system sizes such as in this project, available computing power can quickly become a limiting factor. Efficiently carrying out electronic structure DFT calculations for this work requires nodes with fast communication. In addition, this research requires multiple of these runs and, hence, it needs a machine such as Blue Waters that allows the research team to routinely carry out this work.



TURBULENT MULTIPHASE THERMAL FLOW MODELING OF DEFECT FORMATION MECHANISMS AND ELECTROMAGNETIC FORCE EFFECTS IN CONTINUOUS STEEL CASTING

**Allocation:** Illinois/75 Knh  
**PI:** Brian G. Thomas<sup>1</sup>  
**Co-PIs:** Seong–Mook Cho<sup>2</sup>, Hyunjin Yang<sup>1</sup>, Surya Pratap Vanka<sup>1</sup>, Matthew Zappulla<sup>2</sup>, Seid Koric<sup>1,3</sup>, Ahmed Taha<sup>3</sup>

<sup>1</sup>University of Illinois at Urbana–Champaign  
<sup>2</sup>Colorado School of Mines  
<sup>3</sup>National Center for Supercomputing Applications

EXECUTIVE SUMMARY

Multiphysics models have been developed to quantify the formation mechanisms of various defects related to flow instability, particle transport and capture, superheat transport, surface depression, and cracking in continuous steel casting. To simulate these complex and interrelated phenomena more accurately, the research team implemented a commercial computational fluid dynamics program, ANSYS Fluent high-performance computing (HPC), and the multi-GPU-based in-house CUFLOW codes on Blue Waters’ XE and XK nodes, respectively. Using these codes on Blue Waters’ resources, various turbulent flow models including large-eddy simulations and Reynolds-averaged Navier–Stokes models have been coupled with the volume of fluid method, discrete phase model, particle capture models, and heat transfer models. Finally, the team is applying these models with magnetohydrodynamics (MHD) models to investigate the effects of electromagnetic systems (static or moving magnetic fields) on defect formation, such as longitudinal cracks and hooks, and to explore practical strategies for reducing defects in the process.

RESEARCH CHALLENGE

Continuous casting is the dominant process used to solidify over 96% of steel produced in the world [1]. Thus, even small improvements can have tremendous industrial impact. Many defects in final steel products originate in the mold region of the process owing to transient phenomena, which include turbulent multiphase flow, particle transport and capture, heat transfer, solidification, and thermal–mechanical behavior. To reduce defect formation, various electromagnetic (EM) systems are often employed to control the transient turbulent flow and accompanying phenomena, according to the varying process conditions in the production facility [2]. Experiments and measurements to quantify these phenomena are extremely limited owing to the harsh environment and huge size of the process as well as the many process parameters. Therefore, the development of high-resolution computational models is an important tool to more accurately simulate and understand the process phenomena and defect formation and to find more practical ways to reduce defects and improve the process. Thus, the research team conducted multiphysics simulations on the Blue Waters supercomputer in order to quantify turbulent multiphase flow, argon gas bubble interaction and size distribution, particle transport and cap-

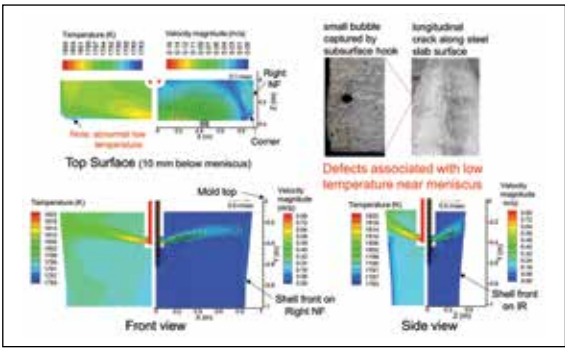


Figure 1: Mold flow pattern and resulting temperature distribution affect defect formation in continuous steel casting.

ture, and superheat transport, and to simulate the effects of moving magnetic fields on these phenomena, in order to investigate ways to reduce defects.

METHODS & CODES

The research team conducted large-eddy simulations (LES) coupled with the volume of fluid method to calculate transient behavior of three-dimensional surface slag/molten steel interface during continuous steel-slab casting [3]. The team used the hybrid multiphase turbulent flow model [4,5] that couples the Eulerian–Eulerian (EE) two-fluid model together with a discrete phase model (DPM) that was validated with lab-scale low-melting alloy experiments to simulate argon bubble interaction (gas pocket formation, gas expansion, breakup, and coalescence) and size distributions in a real slide-gate nozzle [6]. These models have been developed using a commercial computational fluid dynamics program, ANSYS Fluent HPC, on Blue Waters’ XE nodes (AMD 6276 “Interlagos” processor). In addition, transport and capture of the argon bubbles were calculated using LES coupled with the DPM and advanced capture models [7], which were implemented on Blue Waters’ XK nodes (NVIDIA GK110 “Kepler” accelerator) with the multi-GPU-based in-house code CUFLOW [8].

The team applied the magnetic-induction MHD model [2] together with the turbulent-flow models (LES/Reynolds-averaged Navier–Stokes models), the DPM of the transport and capture of inclusions and bubbles, and a heat transfer model, to investigate effects of moving magnetic fields, including EM level stabilizer, EM level accelerator, and mold EM stirring (M–EMS) [2] on the flow pattern and instability, gas bubble distribution, temperature distribution, and superheat at the shell solidification front in the slab mold during steady continuous casting.

RESULTS & IMPACT

Turbulent fluid flow, surface slag/molten steel interface instability, liquid-level fluctuations at the meniscus, slag entrainment, and entrapment were computed from the multiphysics model simulations. This allows understanding of slag defect formation mechanisms, especially the slag entrapment owing to sudden level drops near meniscus regions. From the validated EEDPM model [4,5] simulations, argon bubble behavior and size distributions in the turbulent molten-steel flow inside a slide-gate nozzle of the real caster [6] were revealed in detail. This calculation is expected to contribute to more accurate particle-capture results by being coupled with the advanced particle capture model [7].

In addition, the study investigated initial solidification-related defects such as meniscus freezing, hook formation [9,10], and longitudinal crack formation near the meniscus with the further aid of the coupled heat-transfer model. For example, in mega-thick slab casting, as shown in Fig. 1, the distribution of superheat flux around the mold perimeter was very nonuniform with the unoptimized mold flow pattern. In particular, superheat was unable to reach the meniscus corner, leading to deep hooks and/or longitudinal crack formation. In addition, the effect of the M–EMS on mold flow pattern, temperature, and superheat distribution was quantified from the magnetic-induction MHD model simulations. With M–EMS, the superheat flux at the shell front became more uniform owing to the rotating flow around the perimeter of the mold, resulting in higher superheat flux to the corners, as shown in Fig. 2. This effect is expected to lessen initial solidification defects, so long as the magnetic field strength is within an optimal range.

WHY BLUE WATERS

The high-resolution models used to more accurately simulate and better understand defect formation mechanisms in continuous steel casting are very computationally intensive. The many coupled governing equations need to be solved for turbulent flow, particle transport and capture, temperature, and MHD fields. Moreover, many computational cells are required to capture these complex and interrelated phenomena on micrometer and millisecond scale in the huge domain. Blue Waters enables such high-resolution simulations in a reasonable time frame by speeding up ANSYS Fluent HPC calculations by more than 3,000 times and CUFLOW calculations by 50 times. Furthermore, the Blue Waters

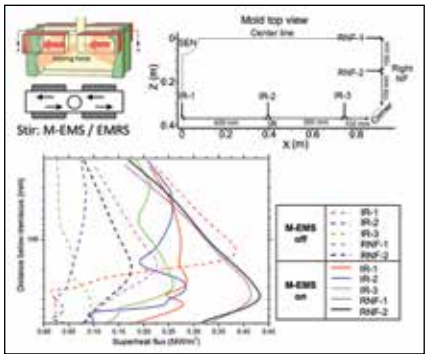


Figure 2: Effect of electromagnetic stirring on superheat flux profiles near the solidifying steel shell front around the perimeter of the mold.

parallel computing environment enables numerous cases to be calculated simultaneously with different process conditions for parametric studies essential to optimize this complex process. Thus, the Blue Waters supercomputer provides a great contribution to obtaining deep insights into complicated defect-related phenomena with high resolution in order to improve this important commercial process.

PUBLICATIONS & DATA SETS

M. L. S. Zappulla, S. Koric, S.–M. Cho, H.–J. Lee, S.–H. Kim, and B. G. Thomas, “Multiphysics modeling of continuous casting of stainless steel,” *J. Mater. Proc. Tech.*, vol. 278, p. 116469, 2020. doi: 10.1016/j.jmatprotec.2019.116469.

S.–M. Cho, M. Liang, H. Olia, L. Das, B. G. Thomas, “Multiphase flow-related defects in continuous casting of steel slabs” in *TMS 2020 149th Annual Meeting and Exhibition Supplemental Proceedings*, San Diego, CA, U.S.A., Feb. 23–27, 2020, pp 1161–1173. doi: 10.1007/978-3-030-36296-6.

S.–M. Cho, B. G. Thomas, and S.–H. Kim, “Effect of nozzle port angle on transient flow and surface slag behavior during continuous steel-slab casting,” *Metall. Mater. Trans. B*, vol. 50B, no. 1, pp. 52–76, 2019, doi: 10.1007/s11663-018-1439-9.

S.–M. Cho and B. G. Thomas, “Modeling of transient behavior of top-surface slag/molten steel interface in continuous slab casting,” presented at *STEELSIM 2019*, Toronto, Ont., Can., 2019.

M. L. S. Zappulla, S.–M. Cho, and B. G. Thomas, “Visualization of steel continuous casting including a new integral method for post-processing temperature data,” *Steel Res. Int.*, vol. 90, pp. 1–11, 2019. doi: 10.1002/srin.201800540.

M. L. S. Zappulla and B. G. Thomas, “Surface defect formation in steel continuous casting,” *Maters. Sci. Forum*, vol. 941, pp. 112–117, 2018, doi: 10.4028/www.scientific.net/MSF.941.112.

K. Jin, S. P. Vanka, and B. G. Thomas, “Large eddy simulations of electromagnetic braking effects on argon bubble transport and capture in a steel continuous casting mold,” *Metall. Mater. Trans. B*, vol. 49B, no. 3, pp. 1360–1377, 2018, doi: 10.1007/s11663-018-1191-1.

H. Yang, S. P. Vanka, and B. G. Thomas, “A hybrid Eulerian–Eulerian discrete-phase model of turbulent bubbly flow,” *J. Fluids Eng.*, vol. 140, no. 10, p. 101202, 2018, doi: 10.1115/1.4039793.

H. Yang, S. P. Vanka, and B. G. Thomas, “Modeling of argon gas behavior in continuous casting of steel,” *JOM*, vol. 70, no. 10, pp. 2148–2156, 2018, doi: 10.1007/s11837-018-2997-7.

H. Yang, S. P. Vanka, and B. G. Thomas, “Modeling of argon gas behavior in continuous casting of steel,” in *CFD Modeling and Simulation in Materials Processing 2018*, Warrendale, PA, U.S.A.: Springer, 2018, pp. 119–131. doi: 10.1007/978-3-319-72059-3\_9

S.–M. Cho and B. G. Thomas, “1) LES modeling of slag entrainment and entrapment and 2) Nozzle flow model validation with measurements of pressure-drop and bubble-size distribution,” *CCC Annual Report*, 2018.

L. Matthew, S. Zappulla, and B. G. Thomas, “Modeling and on-line monitoring with fiber-Bragg sensors of surface defect formation during solidification in the mold,” *CCC Annual Report*, 2018.



# INVESTIGATION OF SEDIMENT TRANSPORT THROUGH AQUATIC VEGETATION USING LARGE-SCALE HIGH-FIDELITY TURBULENCE SIMULATIONS

**Allocation:** Illinois/680 Knh  
**PI:** Rafael Tinoco Lopez<sup>1</sup>  
**Co-PI:** Paul Fischer<sup>1</sup>  
**Collaborators:** Som Dutta<sup>1</sup>, Pallav Ranjan<sup>1</sup>

<sup>1</sup>University of Illinois at Urbana–Champaign

## EXECUTIVE SUMMARY

Turbulence generated by aquatic vegetation in fluvial and coastal systems drives changes in the hydrodynamics and sediment mechanics within aquatic ecosystems. It is expected that plants extract energy from the flow, reducing near-bed flow velocities, thus damping erosion. However, turbulence generated by vegetation patches can actually enhance resuspension and transport under the right conditions. This study uses direct numerical simulations (DNS) and large-eddy simulations (LES) through various arrays of idealized vegetation, represented as cylinders, to increase the understanding of the interactions among vegetation, flow, and sediment. Simulations are conducted using the higher-order spectral element-based computational fluid dynamics (CFD) solver Nek5000. Staggered arrangement of cylinders mimicking an actual experimental setup have been simulated to shed light on the details of flow features and suspended sediment distribution as a function of flow, plant, and patch characteristics, thus increasing the understanding of different hydro- and morphodynamic processes.

## RESEARCH CHALLENGE

Vegetation patches in rivers and coastal areas act as ecosystem engineers to modify various aspects of their own ecosystem [1]. Aquatic vegetation provides a wide range of ecosystem services [2], from nutrient uptake and oxygen production to habitat creation to bed stabilization and even carbon sequestration and nutrient farming. Owing to computational limitations, past computational studies have mostly focused on using CFD models based on Reynolds averaged Navier–Stokes equations, which only provide an averaged approximation of the flow field. The few studies that have used high-fidelity LES have been limited by the number of vegetation elements that can be modeled. Further, the number of both experimental and numerical studies investigating vegetation–flow–sediment interactions is still very limited. Most of the existing data in this area come from laboratory experiments [3] that recreate conditions closer to nature but often lack the spatial and temporal resolution required to capture some fundamental processes in detail.

This study is coupled with experimental research currently being conducted at the Ven Te Chow Hydrosystems Laboratory

and the Ecohydraulics and Ecomorphodynamics Laboratory at the University of Illinois at Urbana–Champaign. The study conducts numerical simulations at an unprecedented scale, resolving details that in conjunction with the experiments will provide as yet unknown insights into the fundamental dynamics of flow and transport in the presence of aquatic vegetation [4]. These large-scale computations will help improve lower-order models of the processes while also informing better experimental design and measurement practices. The number of computation points required to model the whole domain is near 1.2 billion, making the study a unique opportunity because of the scale and complexity of the processes investigated. While such simulations are still tractable on a petascale platform such as Blue Waters, the computational cost is too high, resulting in a reduced number of modeled scenarios, which constrains the insights that a broader range of parameters could yield.

## METHODS & CODES

The research team conducted high-resolution LES and DNS of the flow at different configurations of the idealized vegetation using the open source, spectral element-based high-order incompressible Navier–Stokes solver Nek5000 [5,6]. Sediment trans-

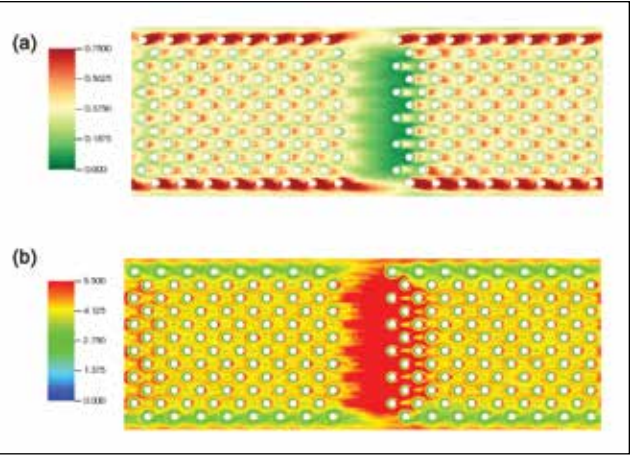


Figure 1: Near-bed nondimensional (a) turbulent kinetic energy field and (b) suspended sediment concentration distribution. The sediment deposition pattern on the channel bed is governed by the local level of turbulence in the region. Regions of low turbulence (gap) are marked by a higher concentration of suspended sediment.

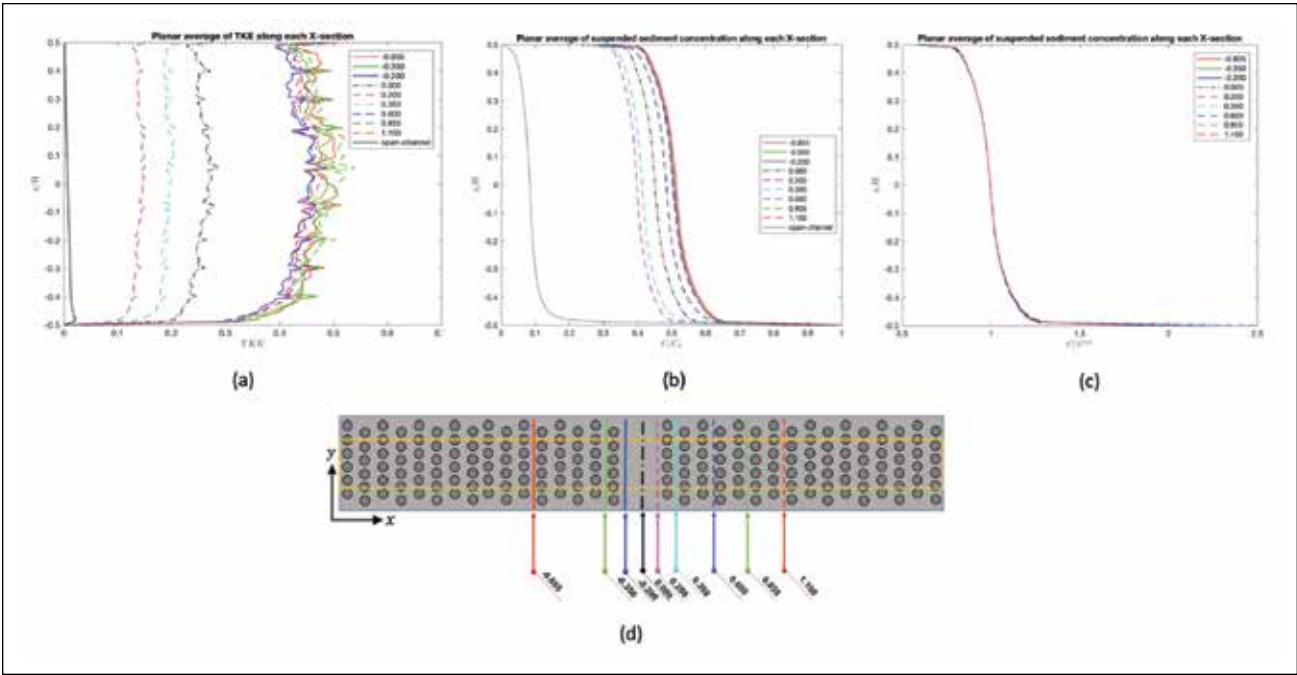


Figure 2: Time- and space-averaged vertical profiles of normalized (a) turbulent kinetic energy, (b) suspended sediment concentration (SSC) normalized using near-bed values, and (c) SSC normalized using volumetric concentration; (d) shows lateral cross-sections corresponding to vertical profiles in (a–c).

port was modeled under the Eulerian framework using the advection–diffusion equation [7], making this study one of the first to look at the complex interaction among sediment–flow–vegetation using high-fidelity CFD simulations.

## RESULTS & IMPACT

The research team conducted three-dimensional LES for a range of Reynolds numbers between 8,000 and 20,000 for a staggered array of rigid cylinders. The first part of the study focused on understanding hydrodynamic changes to the flow field owing to the presence of an emergent vegetation array compared to existing experimental data. To allow for access of measuring probes in laboratory experiments, a section of a vegetation patch is often cleared out, resulting in a small gap where the measurements are taken. The team has proven that measurements within such a gap may not be representative of flow within the patch (Fig. 1). An optimal dimension of a gap in the vegetation can be found such that experimental measurements of flow statistics within it are representative of the flow field within the array. Simulations focused on suspended sediment concentration highlighted that turbulent kinetic energy is the governing factor in erosion, resuspension, and transport of suspended sediments in vegetated flows. This suggests that the equilibrium boundary condition used for simulating suspended sediment transport in nonvegetated channel flows should be modified to account for local erosion and deposition based on turbulent kinetic energy rather than mean shear stress (Fig. 2).

These findings will improve the estimates of sediment transport in natural water bodies, allowing for better sediment man-

agement strategies for rivers, ports, and harbors. Simulations were conducted using up to 90 million computational points, running on 8,196 processors at a time, which would be unfathomable without the use of a petascale supercomputing facility such as Blue Waters.

## WHY BLUE WATERS

The study pushes the limit of the scale at which high-resolution simulations are used to study complex multiphase flow in environmental fluid mechanics, requiring computational resources with sustained computing power at an unprecedented scale such as Blue Waters. Simulations have been conducted for up to 296 million computational points, with the code scaling strongly up to 32,768 Message-Passing Interface ranks. Without access to petascale high-performance computing resources, completing the study within a realistic timeframe would be impossible. In addition, since visualization of a phenomenon is an effective way to understand and explain its mechanics, the team will continue to work with Blue Waters project staff to create animations using data from the simulations.

## PUBLICATIONS & DATA SETS

P. Ranjan, “High-resolution numerical investigation of hydrodynamics and sediment transport within emergent vegetation canopy,” *Diss.*, 2018.

P. Ranjan, S. Dutta, K. Mittal, P. F. Fischer, and R. O. Tinoco, “Flow and sediment transport through aquatic vegetation,” in preparation for *J. Fluid Mech.*, 2019.



DI

MACHINE LEARNING-ASSISTED HIGH-THROUGHPUT COMPUTATIONAL DESIGN OF SOLVENTS FOR LIQUID-EXFOLIATION

Allocation: Strategic/325 Knh  
PI: Kimani Toussaint<sup>1</sup>

<sup>1</sup>University of Illinois at Urbana–Champaign

EXECUTIVE SUMMARY

Computational study and machine learning are two advanced tools of scientific discovery. In this project, the research team performed high-throughput computational studies to understand and obtain data on the liquid-phase exfoliation (LPE) process with various solvents. Then, the team analyzed the data using various machine learning algorithms to obtain optimal solvent composition, which facilitates the LPE process. Considering the nanoscale of phenomena occurring in the LPE process, the combination of computational physics and machine learning is one of the best methods to optimize the solvent to further improve the LPE process. This can then guide experimentalists in initial steps by reducing the number of expensive experimental attempts, especially since commonly used solvents in the LPE process are room temperature ionic liquids (RTILs) that have up to several million possible solvents.

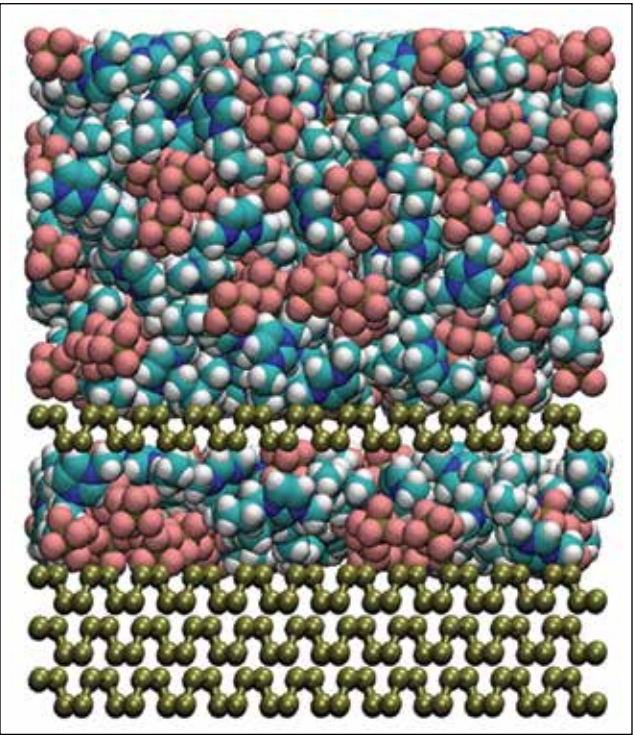


Figure 1: Schematic representation of the liquid-phase exfoliation process; solvent molecules move between layers of bulk materials, separating them from each other.

RESEARCH CHALLENGE

LPE is an advanced mass production method for 2D materials with applications in various technologies such as water desalination, energy storage, optoelectronic devices, DNA sequencing, and energy generation. Compared with other 2D material production methods such as mechanical exfoliation or chemical vapor deposition, LPE produces higher yields with lower costs. LPE is also a simple method that uses an external ultrasonic force on a bulk material immersed in a solvent [1,2]. The bottleneck of the LPE process is the design of a solvent that optimizes the yield, stability, and quality of produced 2D materials.

Because the LPE process occurs in the interface of a bulk material and a solvent, it has a nanoscale nature (Fig. 1). Therefore, knowledge of macroscopic properties of the solvent and bulk material does not allow experimentalists to select the optimal solvent. However, recent computational studies using molecular dynamics (MD) simulations have provided insight into LPE [3]. Even though these studies are crucial in understanding the key parameters of the LPE process, they do not contribute directly to the design and screening of all possible solvents. In this project, the research team combined high-throughput computational studies of various solvents with machine learning algorithms to relate interactions between 2D nanosheets with the composition of the solvents. Based on the applied machine learning algorithm, the team predicted the best solvent composition for the LPE process. The study, therefore, can significantly improve 2D material production, particularly since one of the objectives is to release a computational tool for users in industry and academia.

METHODS & CODES

The research team performed MD simulations using the GROMACs package. GROMACs is an open-source classical MD code for simulation of various physiochemical and biological systems. The GROMACs package supports various potential forms used in the MD simulation of both solvents (RTILs) and 2D materials. The machine learning model development was performed using the TensorFlow and scikit-learn packages, both of which are accessible through Python (bwp) on Blue Waters. Fig. 2 shows the workflow of the current study, in terms of both data generation and training as well as the inference and design phases.

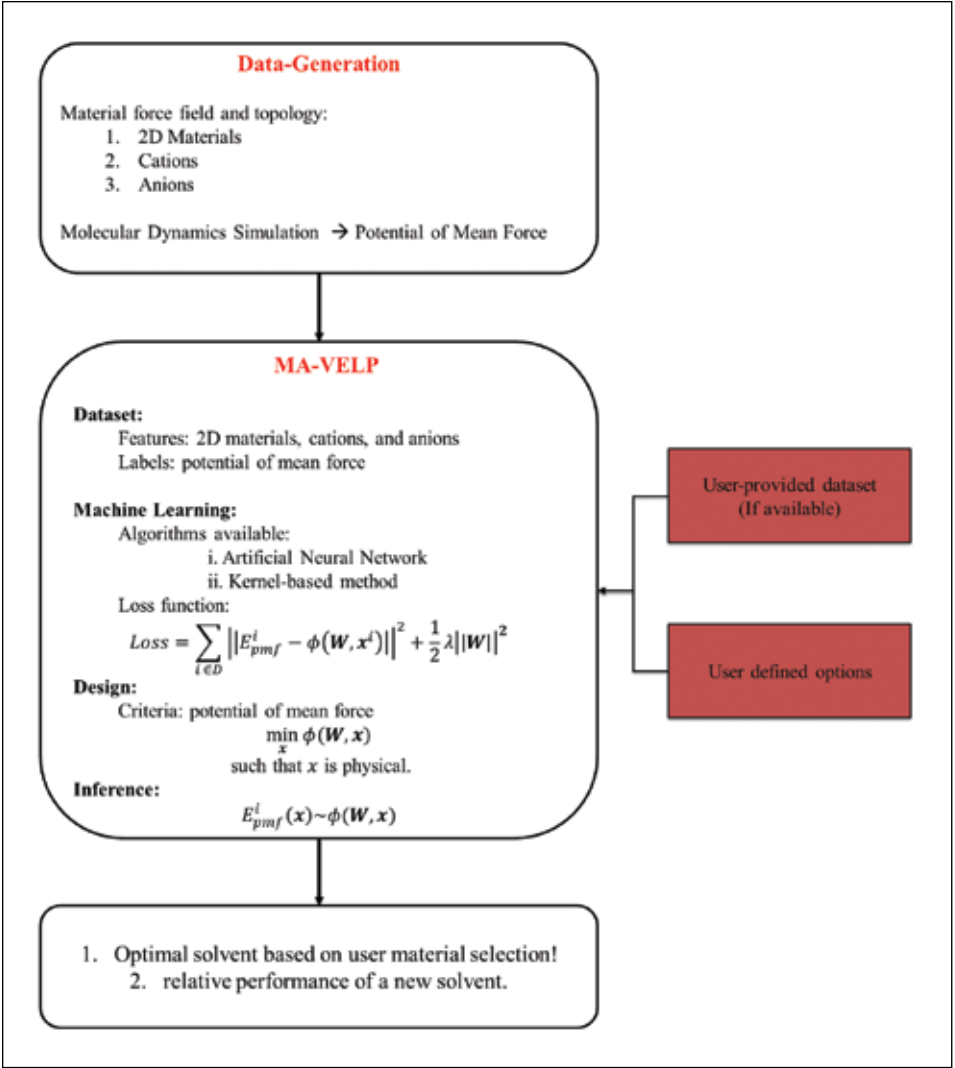


Figure 2: Workflow of current study for the tool development.

RESULTS & IMPACT

The current study uses MD simulations and machine learning methods to help users with computational or experimental background with the selection of an optimal solvent for 2D materials production with the LPE process. The most commonly used solvents for the LPE process are RTILs, also known as “designer solvents,” so named because more than a million RTILs could theoretically be synthesized. Considering the number of possible solvents and the costs and risks associated with carrying out experiments, it is necessary to guide experimentalists, especially in the initial steps. Even though MD simulations provide physically consistent and realistic data on the LPE process with a resolution impossible to obtain in experiments, MD data are used to understand the LPE process at the molecular level; the design of solvents requires further development. To make good use of the MD data, the team applied a machine learning algorithm with a high predictive capability to relate solvent composition with MD

data, with the objective of facilitating solvent design. The machine learning models can then be used later to optimize solvent compositions, narrowing the number of possible solvents and guiding experimentalists toward the most promising ones.

WHY BLUE WATERS

Increasing the accuracy of machine learning models requires a huge amount of data. Each specific solvent simulation must go through energy minimization, equilibrium in the NVT ensemble and, in the NPT ensemble, peeling and umbrella sampling, followed by data analysis. Such MD simulations are computationally expensive; Blue Waters makes such simulations possible.

PUBLICATIONS & DATA SETS

K. Toussaint, “Machine learning-assisted high-throughput computational design of solvent for liquid-exfoliation,” in preparation, 2019.



HIGH-THROUGHPUT MATERIALS MODELING OPTIMIZATION

**Allocation:** Illinois/796.196 Knh  
**PI:** Dallas R. Trinkle<sup>1</sup>  
**Co-PIs:** Dan Katz<sup>2</sup>, Joshua Vita<sup>1</sup>

<sup>1</sup>University of Illinois at Urbana–Champaign  
<sup>2</sup>National Center for Supercomputing Applications

EXECUTIVE SUMMARY

Empirical potentials are “course-grained” models of atomic interactions and are fundamental to materials modeling. They allow molecular dynamics simulations of processes involving 10<sup>6</sup>–10<sup>9</sup> atoms and timescales of nano- to microseconds or longer, and are necessary for both length- and time-bridging methods that span orders of magnitude in scale. Their optimization to reproduce computationally demanding quantum mechanics-based simulation methods is a significantly challenging problem.

Recently, the researchers developed a new approach that relies on a combination of Bayesian sampling of potential parameters [1] with the optimization of the fitting database [2]. The team’s algorithm optimizes the target structures and properties, as well as their “weight,” to guide the optimization of a potential to make accurate predictions [3]. This automated approach can work both for predictions where experimental or theoretical guidance is missing by including related structures and also to determine when an empirical potential form may be too limited to capture the predictions of interest.

RESEARCH CHALLENGE

The algorithm to optimize a coarse-grained empirical potential (Fig. 1) has recently been demonstrated in the team’s publications. Recently, they achieved the next step to reach increased complexity and, hence, significantly greater impact across materials science, physics, chemistry, and biology: improved parallelization of the algorithm to reach large scale. The algorithm relies on Bayesian sampling of parameter space to determine optimal parameters along with error estimates for predictions from the model. The parameters are optimized against a “fitting database”: a selection of structures with density functional theory (DFT) energies and force, and with relative weights capturing the importance of each entry. The database is optimized by using a genetic algorithm over the weights. At the center of this algorithm is their massively parallel evaluation engine that takes a list of structures (atomic positions and chemistries) and a large vector of parameters  $\theta$  (the spline values) for the empirical potential, and evaluates the energies and forces for each parameter. As the researchers include more structures—through prediction of new structures using both DFT and the empirical parameter optimization—and more parameters, the need for a massively parallel approach becomes apparent. For example, if the team were to consider 10<sup>6</sup> empirical potential parameters applied to

10<sup>3</sup> structures with 10<sup>3</sup> evaluations (energy, forces, and derivatives of predictions with respect to parameters), although each individual calculation requires less than a second to complete, there are approximately 10<sup>9</sup> calculations providing about 8 terabytes of data to be used simultaneously; the complete calculation requires ~10<sup>3</sup> cores for memory. Increasing the number of structures by one order of magnitude and the number of parameters by two orders of magnitude increases the scale to approximately 10<sup>6</sup> cores with about 8 petabytes of data. These calculations are embarrassingly parallel and are utilized within a master–worker approach, but the large amount of simultaneous data is spread across workers. By moving to these larger scales, the structural and energy landscape of an empirical potential can be fully analyzed for accurate and predictive empirical potentials. This transforms the problem of coarse-grained parameter optimization into a “Big Data” problem that is of the scale appropriate for a machine like Blue Waters, and provides for a big impact.

METHODS & CODES

This project uses the research team’s newly developed parallel evaluation engine, implemented in Python. The code continues to be in active development and is available through Github (<https://github.com/TrinkleGroup/s-meam>). The underlying parallel algorithm is worker–manager, where individual workers are tasked with evaluating forces or energies for a specific structure; sets of parameters can be passed to a given worker and the forces or energies sent back to the manager. At the beginning of a run, each worker analyzes its structure to convert the spline calculations into vector-matrix operations for efficient evaluation. This helps to keep each worker’s evaluation for one parameter set efficient but also permits even faster evaluation of sets of parameters through vectorized operations. The code uses NumPy and SciPy along with Message-Passing Interface for communication. Current runs of up to 512 cores have shown that the calculations—using a genetic algorithm for optimization—have been compute-bound rather than I/O-bound. All development of the algorithm has been on Blue Waters.

RESULTS & IMPACT

Generating a highly efficient and massively parallel materials modeling optimization engine will enable new approaches to the development of empirical potentials that leverage machines at Blue Waters’ scale. The research team is designing a general

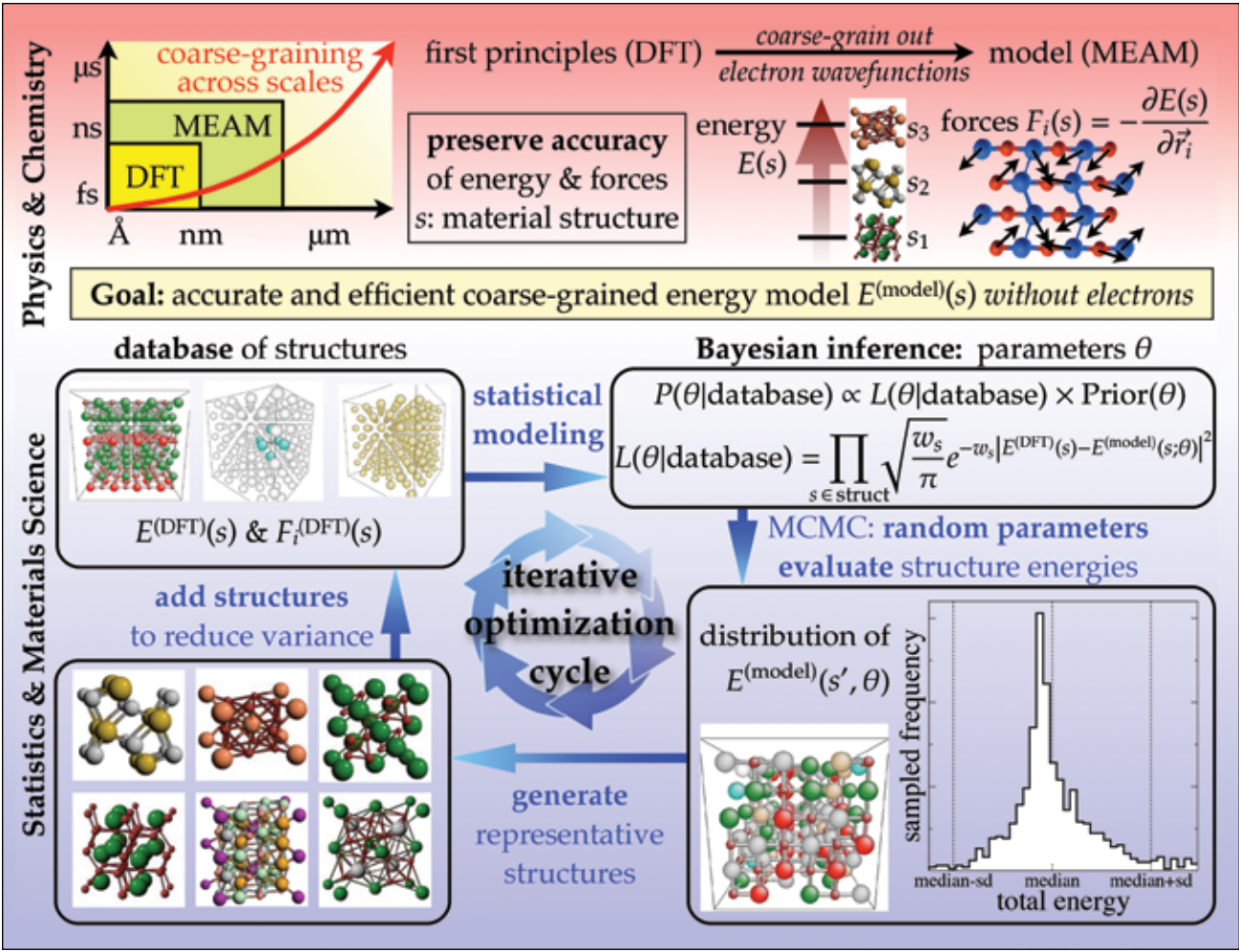


Figure 1: The parallel Bayesian optimization for coarse-graining complex interactions and directed knowledge discovery cuts across materials science, statistics, physics, and chemistry. While the research team’s first application of this method is for the modeling of materials at atomistic length scales—a general problem—the approach for predictive coarse-graining will be useful for other applications. The iterative optimization cycle produces a dual representation of the coarse-grained information and an accurate, predictive model of material properties.

framework that is maximally efficient for large sets of parameters to be evaluated against a fixed set of candidate structures. This computational problem is related, but distinct, from the evaluation typically needed for molecular dynamics calculations where a single parameter set is evaluated against a very large number of structures. The computational engine will be used to take advantage of genetic algorithms for parameter optimization, Monte Carlo evaluation of Bayesian estimates of uncertainty, and cyclic improvement of databases, but other optimization schemes (such as Pareto optimization) could be considered as well.

WHY BLUE WATERS

The computational engine is designed specifically to leverage massively parallel architecture with a worker–manager structure; access to Blue Waters has been instrumental for the implementation and testing of the code as well as preliminary runs. This work would have been impossible otherwise.



DETECTING NEUROTRANSMITTERS WITH DNA-WRAPPED NANOTUBE SENSORS

**Allocation:** Innovation and Exploration/200 Knh  
**PI:** Lela Vuković<sup>1</sup>  
**Collaborator:** Markita Landry<sup>2</sup>

<sup>1</sup>University of Texas at El Paso  
<sup>2</sup>University of California, Berkeley

EXECUTIVE SUMMARY

The rapid and efficient detection of modulatory neurotransmitter molecules stands to be transformative for studies of neurological diseases. Polymer-wrapped carbon nanotube (CNT) sensing platforms are well suited to address this critical need. Using the Blue Waters supercomputer, the research team performed extensive equilibrium and enhanced-sampling all-atom molecular dynamics simulations and obtained free energy landscapes that demonstrate that short DNA polymers can wrap CNTs in highly ordered ring conformations that can suppress the optical signal of the CNTs. In microseconds-long trajectories, dopamine neurotransmitters were shown to bind to DNA rings and disorder them on the CNT surface, which the research team associated with increased optical emission. The project’s experimental collaborator, Markita Landry at the University of California, Berkeley, demonstrated that these ring DNA-wrapped CNTs constitute

an ultrasensitive “turn on” nanosensor for the neurotransmitters dopamine and norepinephrine with a strong relative change in optical signal of up to 3,500%, appropriate for *in vivo* neuroimaging.

RESEARCH CHALLENGE

There is a critical need to develop neurotransmitter sensors that will eventually be able to probe the emergence, diagnosis, and treatment of multiple neurological diseases related to altered patterns of neurotransmission. However, a broadly utilized optical imaging technology to address the quantitative sensing of neurotransmitters does not exist. This project’s computational group has teamed up with an experimental lab to design and understand the functional mechanisms of novel sensors of neurotransmitters, based on carbon nanotubes wrapped by nucleic acid polymers.

METHODS & CODES

The research team performed microseconds-long equilibrium and enhanced-sampling replica exchange all-atom molecular dynamics simulations with the latest version of the NAMD software, a GPU-accelerated, highly parallelized code for high-performance simulations of biomolecules.

RESULTS & IMPACT

In this study, the research team performed multiscale simulations of short and long DNA polymers (12 to 30 nucleotides) with different sequences, wrapping (9,4) and (6,5) carbon nanotubes to disclose mechanisms responsible for a strongly quenched baseline fluorescence and a large nanosensor response to neurotransmitters observed in experiments. While longer 30-nucleotide DNA polymers remained in helical conformations in molecular dynamics (MD) simulations, shorter 12-nucleotide (GT)<sub>6</sub> DNA polymers rearranged from initial helical conformations into ringlike conformations in each of the five independent trajectories performed. To confirm that the ringlike conformation is a favorable adsorbed state of a (GT)<sub>6</sub> DNA on the (9,4) CNT, the team calculated the free energy landscape of the DNA (Fig. 1) on the (9,4) CNT surface at room temperature (T = 300K), using replica exchange MD. The landscape revealed two distinct stable conformations for (GT)<sub>6</sub>: a left-handed helix and a nonhelical ringlike conformation. The team next performed quantum calculations of the systems and developed a quantum model of an exciton at the CNT surface in the electrostatic environment generated by the DNA polymers, solvent, and neurotransmitter analytes. With the help of the simulations performed, the research team proposed the mechanisms behind the low optical signal of the ring-DNA-wrapped CNTs, and how the adsorbed dopamine neurotransmitter molecules distort ring-conformations of short DNAs and increase the optical signal of CNTs. Furthermore, the team has been screening short DNA polymers of different sequences that can form ring conformations on CNTs in order to identify novel polymer CNT wrappings for selective and sensitive detection of other neurotransmitter molecules.

WHY BLUE WATERS

To overcome the computational timescale limitations, the project required the use of replica exchange MD simulations, which can usually be performed only with access to the large resources of a petascale machine such as the Blue Waters supercomputer.

PUBLICATIONS & DATA SETS

A. G. Beyene *et al.*, “Ultralarge modulation of fluorescence by neuromodulators in carbon nanotubes with self-assembled oligonucleotide rings,” *Nano Lett.*, vol. 18, p. 6995, 2018, doi: 10.1021/acs.nanolett.8b02937.  
R. Nissler *et al.*, “Quantification of the number of adsorbed DNA molecules on single-walled carbon nanotubes,” *J. Phys. Chem. C*, vol. 123, p. 4837, 2019, doi: 10.1021/acs.jpcc.8b11058.

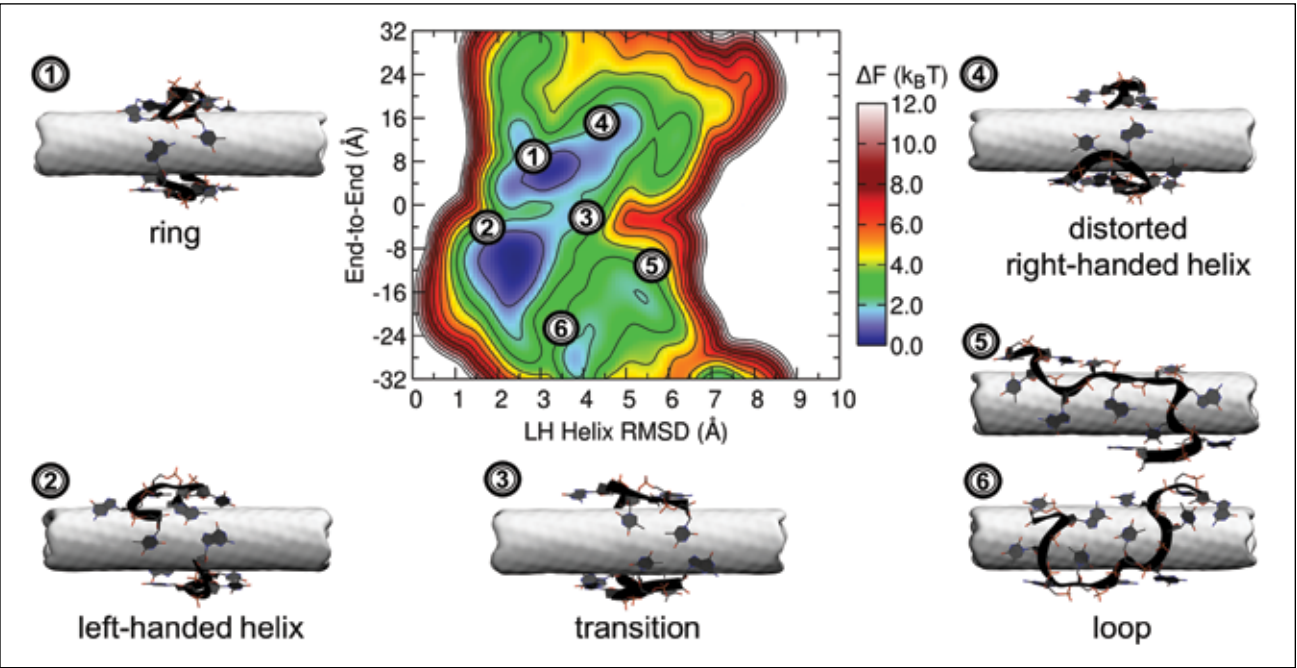


Figure 1: Free energy landscape of a short (GT)<sub>6</sub> DNA wrapping a (9,4) carbon nanotube. DNA favors ring (1) and left-handed helix conformations (2), shown in blue regions in the contour plot.



ACCURATE EFFECTIVE INTERACTIONS IN QUANTUM MATERIALS

Allocation: Illinois/585.93 Knh  
PI: Lucas Wagner<sup>1</sup>  
Collaborators: Daniel Shoemaker<sup>1</sup>, Greg MacDougall<sup>1</sup>

<sup>1</sup>University of Illinois at Urbana–Champaign

EXECUTIVE SUMMARY

In this project, the research team developed multiscale quantum models from first-principles (*ab initio*) calculations that explicitly consider the subatomic physics of electrons and nuclei. The work represents a new level in connection between the subatomic scale and up to nanometer- and micrometer-scale quantum effects. The research is enabled by a combination of high-performance computing to gather data and advanced data science techniques to analyze the resultant data. The ensuing multiscale-effective models may then be used to predict the behavior of materials at those length scales and to understand quantum materials, for which quantum effects are starkly visible in their behavior.

RESEARCH CHALLENGE

All materials that we deal with day to day are made up of the same ingredients—nuclei and electrons. The variety of different objects and devices emerges from the many-particle behavior of these systems. How that actually occurs is still not well understood in many cases. Part of the challenge is that the electrons and nuclei behave according to quantum mechanics, whose equations are notoriously difficult to solve for more than just a few particles. Researchers use the concept of effective models to de-

scribe the behavior without having to explicitly consider the details of the fundamental particles. These effective models are used to design devices and otherwise describe how the materials behave. However, for new and emergent materials, it is often not known how to connect these effective models to the underlying dynamics of electrons and how to predict the correct effective model without appealing to experiment.

METHODS & CODES

The research group has recently developed a way of exploring the quantum solution space, which in turn allows them to derive effective models from high-accuracy first-principles calculations. The method involves performing many Monte Carlo calculations to simulate the electrons, which can run in parallel on thousands of nodes, and to analyze their results. The team uses the open-source QWalk [1] code, which was developed at the University of Illinois at Urbana–Champaign.

RESULTS & IMPACT

Fig. 1 shows the application of the effective model ideas to real materials. The question the research group is trying to answer in this study is ultimately why some materials (called unconventional

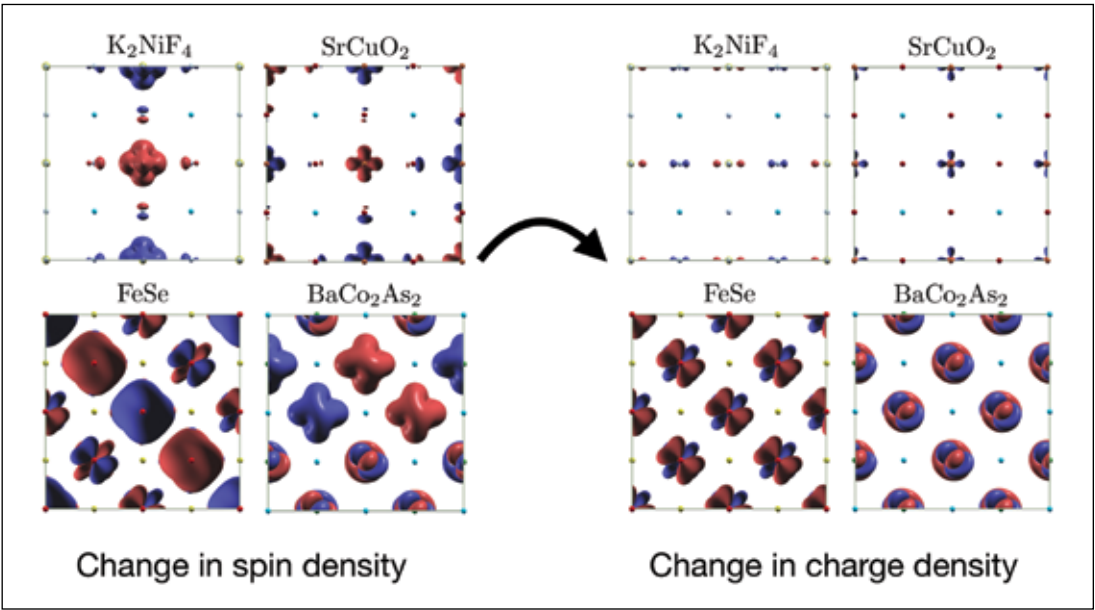


Figure 1: For some materials, spin and charge are linked together; a change in one will lead to a change in the other. This diagram shows that the superconductors SrCuO<sub>2</sub>, FeSe, and BaCo<sub>2</sub>As<sub>2</sub> all have coupling between charge and spin, while the simple antiferromagnet K<sub>2</sub>NiF<sub>4</sub> does not.

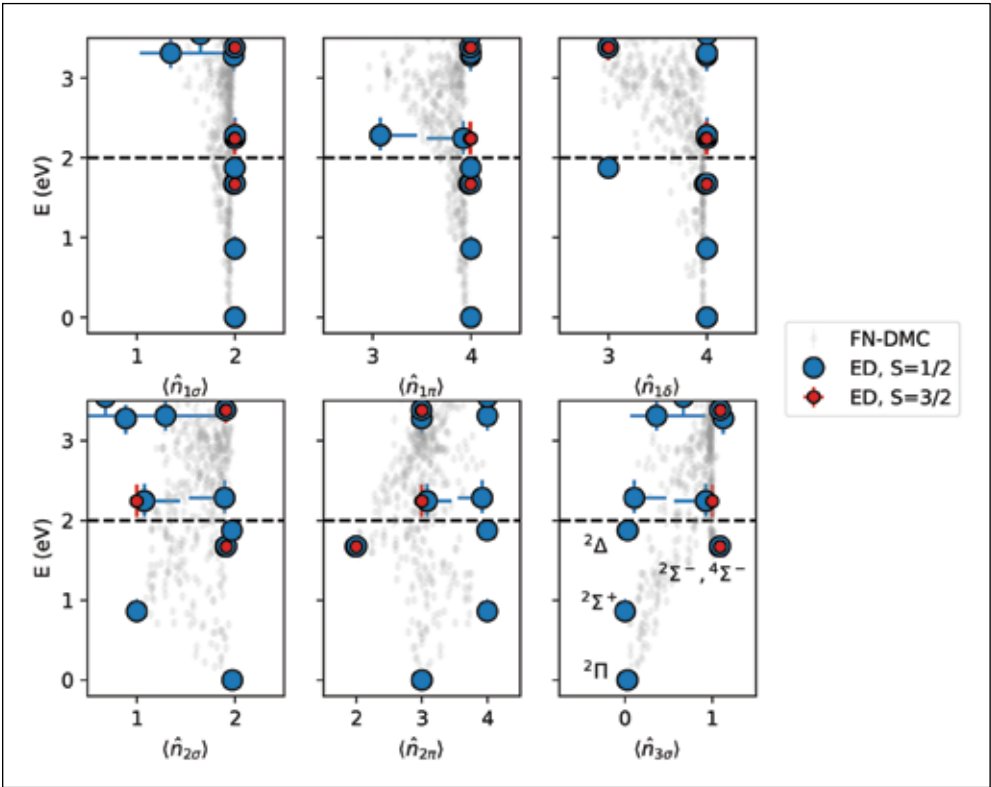


Figure 2: Data analysis of the CuO molecule. The gray dots are the simulation data generated on Blue Waters, while the blue and red dots correspond to the solution of the effective model that is fit to those data. The x-axis quantities are descriptors that are used to describe the effective model. The solution matches the experimental observations of the molecule.

superconductors) superconduct at high temperatures while some do not, which has been an open question in the field for decades.

One of the leading explanations is that the motion of charge in the system is coupled to the motion of small magnetic moments that surrounds each atom, called spin. In this study, the team explored the coupling between charge and spin by finding the lowest energy charge configuration while fixing the spin to different configurations. If they are coupled, then the charge configuration changes with the spin configuration. Using Blue Waters, the team performed a large-scale study of dozens of materials and found that all the unconventional superconductors have coupling between spin and charge. This advance may be useful in finding new similar materials using high-performance computation.

Fig. 2 shows the application of a more general version of the effective model-fitting technique. Again, the team performed highly accurate calculations on a small molecule, CuO, which were then analyzed using statistical techniques to determine an effective model for the system. While it is not possible to solve the full system exactly to find its excitation spectrum, the researchers were able to solve the reduced effective model, which agrees with experiment. In fact, many of the excitations that were found are not easily computed in the full system. During this work, the team was able to dramatically improve the efficiency of their model-generation algorithms because of the large amount of data gathered using Blue Waters.

There have been a number of other projects that have been enabled by this computational resource. These have been performed

in close collaboration with experimentalists, and several predictions have inspired new experiments on materials.

WHY BLUE WATERS

Blue Waters was critical for the performance of this work. The quick turnaround of the calculations enabled the research team to accelerate the development of challenging methods by a large, difficult-to-quantify amount. In fact, some of these problems would have been unapproachable without the power of the Blue Waters system.

PUBLICATIONS & DATA SETS

B. Busemeyer, G. J. MacDougall, and L. K. Wagner, “Prediction for the singlet-triplet excitation energy for the spinel Mg-Ti<sub>2</sub>O<sub>4</sub> using first-principles diffusion Monte Carlo,” *Phys. Rev. B*, vol. 99, p. 081118, 2019.

H. Zheng, H. J. Changlani, K. T. Williams, B. Busemeyer, and L. K. Wagner. “From real materials to model Hamiltonians with density matrix downfolding,” *Front. Phys.*, vol. 6, 2018.

A. R. Munoz, L. Kish, G. J. MacDougall, and L. K. Wagner, “Benchmarking computed and experimental neutron form factors for MnO and NiO,” in preparation, 2019.

J. N. B. Rodrigues and L. K. Wagner, “Charge-spin susceptibility as a useful probe of quantum materials,” in submission, 2019.

S. Pathak and L. K. Wagner, “A low-energy model for the CuO molecule using *ab-initio* quantum Monte Carlo density matrix downfolding,” in preparation, 2019.



DI

SUPERSONIC JET NOISE PREDICTION USING HIGH-ORDER LARGE-EDDY SIMULATION

Allocation: Innovation and Exploration/300 Knh  
PI: Zhi Jian Wang<sup>1</sup>

<sup>1</sup>University of Kansas

EXECUTIVE SUMMARY

The environmental impact of aviation is measured in emissions and noise. The communities in the vicinity of airports bear the brunt of aircraft noise during takeoffs, climbs, flyovers, approaches, and landings. Research has shown that exposure to loud noise is harmful to human physiological and psychological health and welfare.

The research team at the University of Kansas has identified a novel and powerful means of mitigating jet noise by inducing shear layer swirl through embedded vanes near the nozzle exit lip. The team has employed a newly developed high-order (up to sixth order) Navier–Stokes solver capable of handling mixed unstructured meshes to perform computational predictions of the aero-acoustic noise using implicit large-eddy simulations. Blue Waters enables the team to accurately compute the near- and far-field supersonic jet noise. As a result, the research group can perform computational evaluations of a new concept in supersonic jet noise mitigation, which has the potential to significantly reduce the jet noise generated by military aircraft.

RESEARCH CHALLENGE

The noise generated by supersonic aircraft is deafening. The communities in the vicinity of airports bear the brunt of the noise during their takeoff, climb, flyover, approach, and landing. Human physiological and psychological health and welfare are negatively affected by exposure to this noise. The problem in the military is even greater since there are servicemen and women working near advanced supersonic jets during takeoffs and landings. The long-term health risk has created a huge problem for the Department of Defense (DOD). However, finding a solution has been a challenge. This project evaluates a promising new concept to reduce such noise.

METHODS & CODES

The research team used swirl-generating vanes placed inside the nozzle to enhance mixing in order to reduce noise [1]. The team employed a state-of-the-art unstructured mesh-based high-order Navier–Stokes solver (up to sixth order in space and fourth order in time) called hpMusic [2] in combination with large-eddy simulation to directly compute the near- and far-field jet noise. The near-field simulation was coupled with the Lighthill acoustic analogy [3] to compute the far-field jet noise.

HpMusic has been developed with support from AFOSR (Air Force Office of Scientific Research), NASA (National Aeronautics and Space Administration), and ARO (Army Research Office), and has been used to perform Implicit Large-Eddy Simulations for real-world complex configurations. The solver is based on the flux reconstruction/correction procedure via reconstruction method [4] on mixed unstructured meshes [5]. In addition, hpMusic has both explicit and implicit time-marching algorithms such as the explicit SSP Runge–Kutta scheme, an optimized second-order backward difference formula. The solver has been successfully implemented on massively parallel supercomputers such as the petascale Blue Waters. A scalability study demonstrated excellent parallel efficiency with over 10,000 cores.

RESULTS & IMPACT

The research team has completed the simulations of one baseline supersonic nozzle and one model with vanes to generate swirls to enhance turbulent mixing and mitigate jet noise. In addition, the team has performed mesh refinement and p-refinement studies to assess the sensitivity of the mesh resolution and the order of the simulation.

Based on the results of simulations and data analysis, the group will design more refined vane configurations to reduce jet noise. The team anticipates multiple iterations will be required to arrive at a nearly optimal design configuration. Extensive data analysis and comparison with other simulations and experimental data is currently being performed. So far, the computational results are very promising.

WHY BLUE WATERS

Jet noise computations using large-eddy simulation are very computationally expensive even on supercomputers because the acoustic magnitude is several orders smaller than the mean flow scales [6]. The use of Blue Waters increases the computation speed by at least a factor of 20 compared with the team’s local cluster, thus enabling these very complex simulations to be carried out in a timely fashion. The DOD-funded jet noise mitigation project lasts for only one year. Access to Blue Waters is essential for the team to achieve its goals for this project along with the professional and timely support provided by the Blue Waters staff.

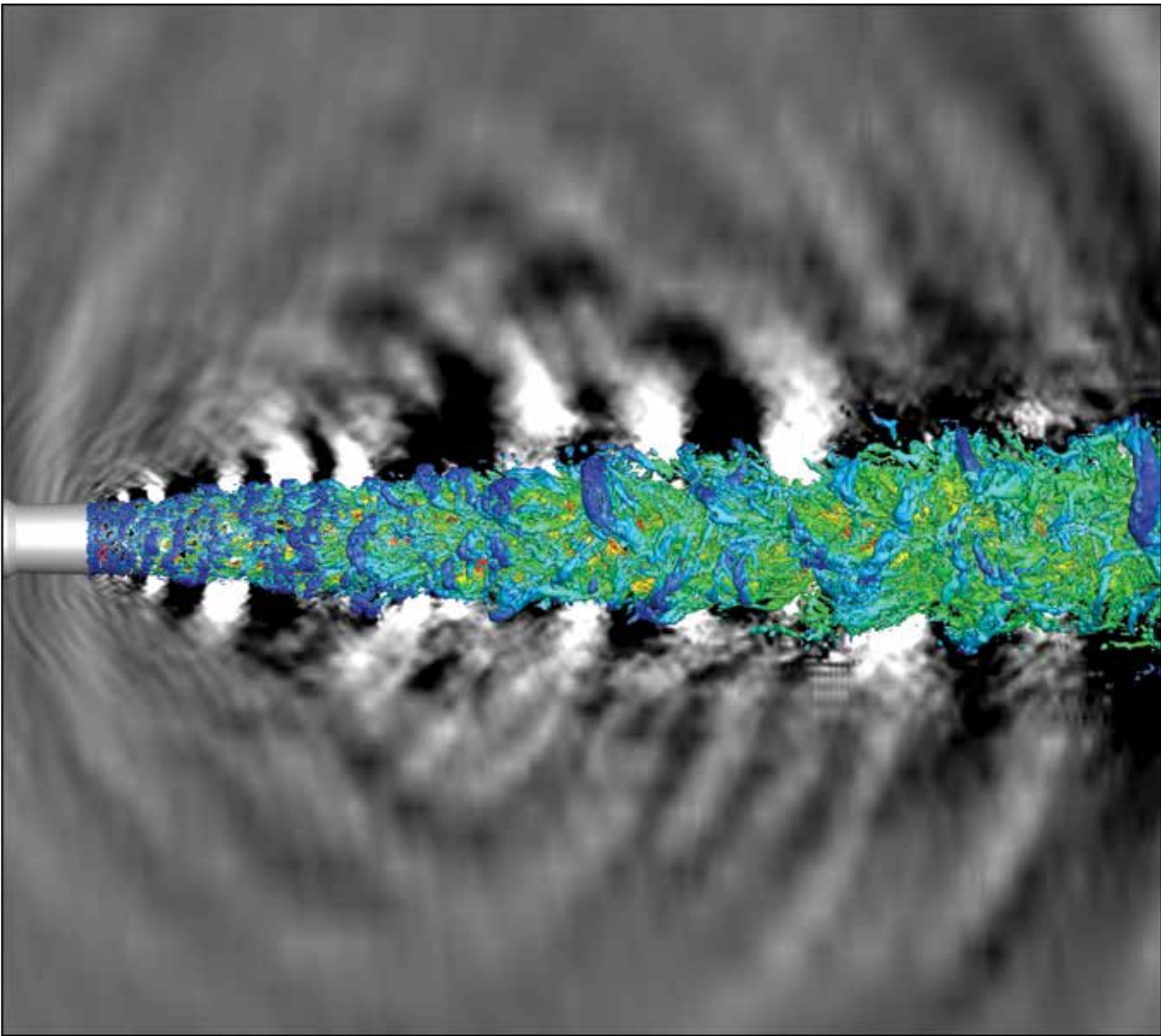


Figure 1: Key flow structures and acoustic waves captured in a large-eddy simulation of a supersonic jet (iso-surfaces of the Q-criterion colored by the streamwise velocity, with acoustic waves shown on the symmetry plane).



DI

TN

BW

# SPIN SPIRALS IN MULTIFERROIC BISMUTH FERRITE AND AT METAL SURFACE: FROM FULLY FIRST PRINCIPLES

**Allocation:** Innovation and Exploration/250 Knh  
**PI:** Bin Xu<sup>1</sup>  
**Co-PIs:** Laurent Bellaiche<sup>1</sup>, Bertrand Dupé<sup>2</sup>  
**Collaborator:** Wojtek Morawiec<sup>2</sup>

<sup>1</sup>University of Arkansas  
<sup>2</sup>Johannes Gutenberg University of Mainz

## EXECUTIVE SUMMARY

This work explores the long-ranged noncollinear magnetic textures, called spin spirals, in the multiferroic material BiFeO<sub>3</sub>, as well as the BiFeO<sub>3</sub>/metal interface. These spin spirals are of prime importance for technological applications owing to their potential for spintronics and low-power magnetoelectric devices. The presence of spin spirals is induced by the Dzyaloshinskii–Moriya interaction (DMI), which originates from spin-orbit coupling. This research investigates the magnetic interaction, and especially the DMI, in multiferroic materials and the interface between multiferroics and metals. With the computational resources of Blue Waters, the research team carried out fully first-principles or *ab initio* calculations in BiFeO<sub>3</sub> to reveal the stability of spin cycloids (the precession of spins along a propagation direction with a perpendicular rotational axis) under various strain conditions and to obtain the accurate magnetic interaction parameters that will be used to construct an effective Hamiltonian for combining with Monte–Carlo simulations to study finite-temperature properties.

## RESEARCH CHALLENGE

BiFeO<sub>3</sub> (BFO) is a multiferroic magnetoelectric material at room temperature. This means that a magnetic field can change its polarization and an electric field can change its magnetization. This property is very rare in nature, which explains the strong interest in BFO and why it is called the “holy grail of multiferroic physics.” BFO is also a noncollinear antiferromagnetic (AFM) material, exhibiting a magnetic cycloid at room temperature. Although this spin spiral is well characterized experimentally, several questions remain, such as the stabilization mechanism(s) of the different spin spirals and their interplay with structural distortion.

For instance, the type of cycloid in bulk BFO in the R3c phase is well characterized as being propagated along one of the [1–10] directions. However, a few recent experimental studies proposed that a type-II cycloid with a propagation direction along [11–2] may be favored in BFO thin films that are moderately strained, but the measurements could also be explained as a mixture of different type-I cycloids. Further, when BFO is deposited on a substrate with a large compressive strain (more than 4.5% in magnitude), it undergoes a phase transition toward the so-called super tetragonal phase, which is of great interest because of its very

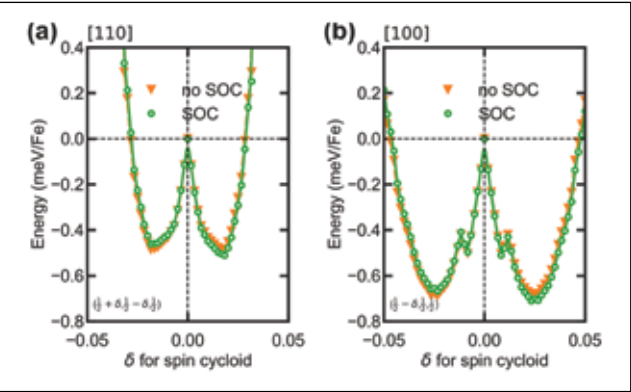


Figure 1: Spin spiral energy dispersion curves near the G-type antiferromagnetic state of BFO in the P4mm or super-tetragonal phase. (a) Cycloid propagation along the [1–10] direction; (b) cycloid propagation along the [100] direction.

large polarization. This phase may host exotic noncollinear magnetic structure.

In addition, one of the challenges in spintronics is to understand the interfacial DMI in sputtered films, which are commonly used in industry. In that case, the difficulty is to model the polycrystalline of the sample. Experimentally, the DMI can only be measured with a few techniques such as bubble expansion or via Brillouin light scattering (BLS). In the latter, only the averaged DMI over the entire sample can be measured. It is therefore of prime importance to understand the origin of the different contributions to the DMI. To that end, the research team has explored the magnetic exchange interaction and the DMI in Fe<sub>1-x</sub>Co<sub>x</sub>/Pt in collaboration with experimentalists.

## METHODS & CODES

The research team used first-principles calculations based on density functional theory (DFT) in order to study the magnetic spin spirals in BFO. They applied the full potential linearized augmented plane wave (FLAPW) method [1,2] as implemented in the FLEUR code (www.flapw.de). This method is one of the most accurate implementations within DFT, owing to the self-consistent treatment of all the involved electrons (core and valence) and the precise and “natural” expansion of the wave functions (*i.e.*, in spherical harmonics around the nuclei and in plane waves in the interstitial region). The choice of an accurate

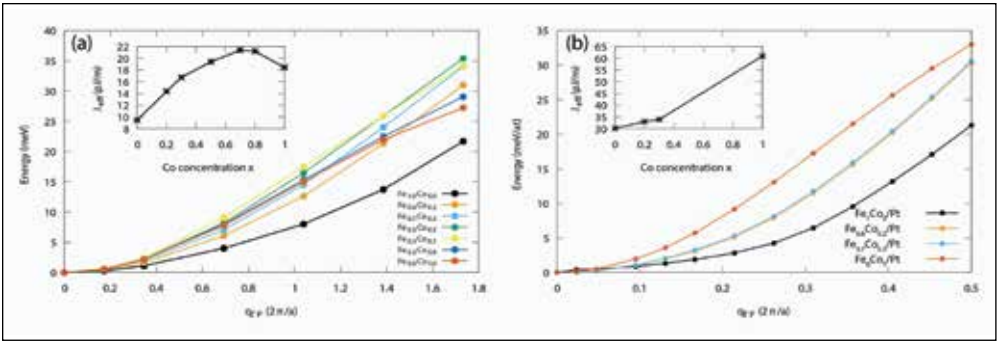


Figure 2: Comparison between the magnetic exchange interaction of Fe<sub>1-x</sub>Co<sub>x</sub>/Pt without the Pt substrate (panel a) and with the Pt substrate (panel b). The magnetic exchange interaction of an ultrathin Fe<sub>1-x</sub>Co<sub>x</sub> depends significantly on the substrate hybridization.

all-electron full-potential code was motivated by the fact that this project requires the treatment of noncollinear magnetism and spin-orbit coupling on the same footing as well as a highly accurate description of eigenfunctions and eigenenergies (leading to the need for an extremely good basis set). Therefore, the simulations of complex magnetic states in the presence of spin orbit coupling (SOC) represent a cutting-edge problem in supercomputing. FLEUR is to date the only existing code containing all of the above-mentioned features and that is, therefore, capable of calculating SOC effects in noncollinear systems fully *ab initio*.

Further, the team computed the total energies at various reciprocal-space points based on the generalized Bloch’s theorem. The ground-state magnetic configuration was determined and compared with experiments. Remarkably, this approach allowed the team to extract the exchange and DMI interaction coefficients by considering neighboring interactions over a long range.

## RESULTS & IMPACT

The research team explored for the first time different spin spirals in BFO at an *ab initio* level. The study revealed that the competition between the first- and second-nearest neighbor exchange interactions leads to a flat energy landscape near the G-type antiferromagnetic state, slightly favoring a spin cycloid propagating in the [1–10] direction. Its stability is enhanced by the polarization but suppressed by the antiphase oxygen octahedral tiltings. Surprisingly, a similar trend is found for the DMI energy, which further lowers the energy of the cycloidal state and reduces the cycloidal period. The predicted period is in good agreement with experiment.

For the super-tetragonal phase of BFO, the calculated energy dispersion in the vicinity of the G–AFM state suggests that cycloid formation is strongly favored. As illustrated in Fig. 1, the cycloid is likely to propagate along the [100] direction, as it has a noticeably lower energy than that of propagation along [1–10]. To our knowledge, this is the first work that predicts the existence and type of cycloid in T–BFO, which can then be verified by future experiment and may lead to promising spintronic applications.

To shed light on the open issue of slightly strained BFO, the research team performed a systematic study in the Cc phase with strain from compressive (–2%) to tensile (+2%). Comparing the energetics of cycloids with [1–10] and [11–2] propagation direc-

tions, the researchers found that type-II cycloid propagating in the [11–2] direction is not favored over type-I cycloid propagating in the [1–10] direction, implying that the experimentally observed type-2 cycloid is likely to be a mixture of two type-1 domains.

In parallel, the team explored the magnetic exchange and the DMI in Fe<sub>1-x</sub>Co<sub>x</sub>/Pt as shown in Fig. 2. Counterintuitively, the magnetic exchange interaction changes dramatically as a function of the concentration. The magnetic exchange shows a maximum for Co-rich Fe<sub>1-x</sub>Co<sub>x</sub>/Pt. The value of the magnetic exchange is, however, much smaller than the measurements realized by BLS on Fe<sub>60</sub>Co<sub>20</sub>B<sub>20</sub>/Pt (30 pJ/m). These findings suggest that the hybridization of the Fe<sub>1-x</sub>Co<sub>x</sub> with the Pt plays a major role in both the magnetic exchange and the DMI, as shown in Fig. 2(b). Taking into account the presence of the Pt substrate increases the exchange at 30 pJ/m, in agreement with the experimental finding. The calculations of the DMI are now under consideration. In the next period, the team will be able to study the effect of an electric field on Fe<sub>1-x</sub>Co<sub>x</sub>/Pt created by a ferroelectric layer owing to proximity effect.

## WHY BLUE WATERS

A challenge in this study was to obtain the full magnetic phase diagram of the systems with strain, temperature, and electric field. Owing to the very long period of the spin spiral in BFO (approximately 60 nm), an extremely dense k-grid is necessary to resolve the q-point corresponding to the ground-state period. Given the superior stability of the FLEUR code with a large number of cores (beyond 10,000), the study has greatly benefited from using the Blue Waters supercomputer.

## PUBLICATIONS & DATA SETS

- X. Changsong, B. Xu, B. Dupé, and L. Bellaiche, “Magnetic interactions in BiFeO<sub>3</sub>: A first-principles study,” *Phys. Rev. B*, vol. 99, p. 104420, 2019.
- B. Xu, B. Dupé, and L. Bellaiche, “Exploring the origin of non-collinear magnetism in BiFeO<sub>3</sub>,” in preparation, 2019.
- B. Dupé, “Calculating the magnetic exchange and DMI in Co<sub>x</sub>Fe<sub>(1-x)</sub>B/Pt,” in preparation, 2019.
- J. Seidel *et al.*, “Exchange splitting of a hybrid surface state and ferromagnetic order in a 2D surface alloy,” submitted, 2019.



# NUMERICAL SIMULATIONS OF A COLLAPSING CAVITATION BUBBLE NEAR AN ELASTICALLY DEFORMABLE OBJECT

**Allocation:** Innovation and Exploration/220 Knh  
**PI:** Zhen Xu<sup>1</sup>  
**Co-PIs:** Mauro Rodriguez<sup>1</sup>, Shahaboddin Alahyari Beig<sup>1</sup>

<sup>1</sup>University of Michigan

## EXECUTIVE SUMMARY

The aim of this work is to understand the structural damage to nearby objects as a result of cavitation (the formation of small vapor-filled cavities). In extreme cases, cavitation gas bubbles collapse into small volumes that produce high pressures and temperatures and generate strong shock waves that interact with the surroundings. When adjacent to a neighboring elastically deformable object, the collapse becomes asymmetric and a re-entrant liquid jet forms within the bubble. The jet hits the opposite side of the bubble and generates an outwardly propagating water-hammer shock that then interacts with neighboring object(s). To determine the effect of confinement and of an elastically deformable object, the research team conducted two high-resolution numerical simulation studies of the collapse of: (1) a single bubble near a viscoelastic object undergoing elastic deformations, and (2) a single bubble in a channel composed of rigid walls. These simulations will be used to understand damage mechanisms in elastic objects and to potentially mitigate erosion.

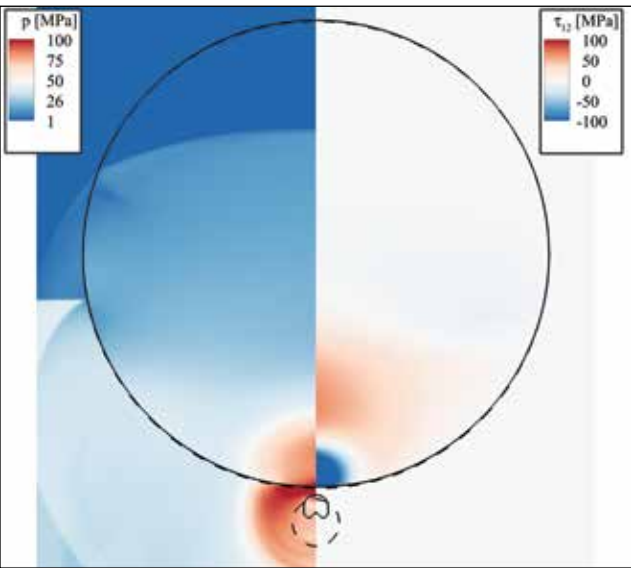


Figure 1: Pressure (left) and elastic shear stress component (right) contours of a shock-induced collapse-induced gas bubble (bottom outline) near a finite-sized elastic (calcium) stone. Isolines denote the approximate location of the bubble's and calcium stone's interface.

## RESEARCH CHALLENGE

Cavitation bubble dynamics appear in a wide range of hydraulic applications such as turbomachinery, naval structures, biomedical ultrasound, and combustion. Cavitation happens owing to the reduction in local pressure beyond the tensional threshold in the surrounding liquid, thereby incurring a mechanically driven phase change in the liquid and leading to the formation of vapor bubbles. Vapor bubbles collapsing near surfaces/objects form a re-entrant jet that impacts on the distal bubble wall, generating a water-hammer shock wave that interacts with the surroundings. The high pressures and temperatures as well as the corresponding shock waves produced by the collapse of cavitation bubbles can damage nearby objects [1–4]. This damage is recognized as one of the main consequences of cavitation and is an essential research topic in a variety of hydrodynamic and acoustic/biomedical applications. In naval applications, engineers still struggle to cope with the deleterious effects of cavitation erosion on surfaces and hydraulic machinery, which lead to degradation in performance and need for repair and/or replacement. In the context of therapeutic ultrasound, the pressure pulses from the collapse of cavitation bubbles are employed to fragment kidney stones and pathogenic tissues.

Understanding the collapse dynamics and the associated damage mechanisms will enable researchers to develop techniques to control (either by mitigation or enhancement) cavitation erosion. However, the wide range of temporal and spatial scales of these flows significantly limits the ability to obtain precise experimental measurements. As a result, numerical simulations serve as a valuable complementary tool to further our understanding of the physics, alongside analytical and experimental efforts. For this reason, the research team has developed a novel numerical framework to investigate the detailed dynamics of nonspherical bubble collapse near viscoelastic objects using high-resolution simulations [5–8]. These simulations will be used to develop a comprehensive model of the collapse dynamics and corresponding damage mechanisms of nearby objects, leading toward developing strategies to better control cavitation-induced erosion.

## METHODS & CODES

The team used its in-house petascale production code for the large-scale simulations based on MPI. This code solves the three-dimensional compressible Navier–Stokes equations with thermodynamically consistent evolution equations for the elas-

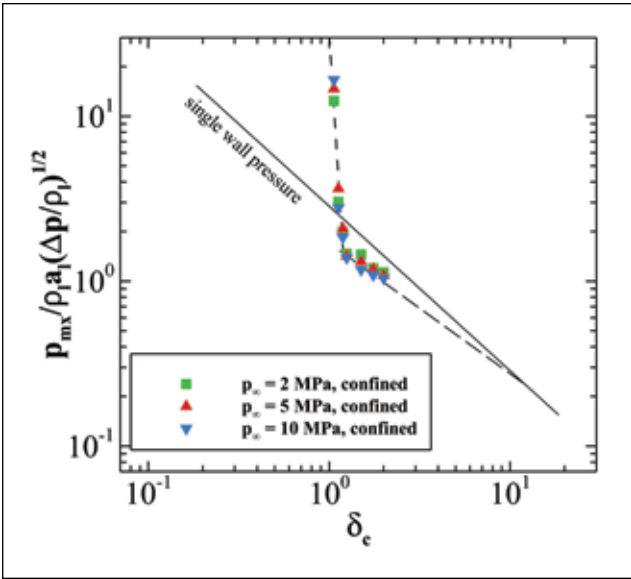


Figure 2: Scaling of maximum pressure measured along a confined channel from the collapse of a bubble centered in a channel at stand-off distance from the surface and the collapse driving pressure. Solid line: scaling for a bubble near a single wall; dashed line: scaling for a bubble centered in a channel.

tic contribution of the stress to represent multiple gases, liquids, and viscoelastic solids [5–8]. The code is based on high-order accurate (in smooth regions), nondissipative schemes with discontinuity detectors to apply high-order numerical dissipation at shock wave and material interfaces. Thus, the code can accurately solve problems involving broadband features, *e.g.*, turbulence, and discontinuous features such as material interfaces, shock, and shear waves. To perform the three-dimensional numerical simulations of the problems of interest, an in-house code was developed in C++, which was parallelized using the MPI library and implemented the parallel HDF5 library for I/O. The code has been verified and validated against a series of theoretical and experimental results.

## RESULTS & IMPACT

To better understand the detailed bubble collapse dynamics near an elastically deformable object or in confinement, two problem configurations were considered: (1) shock-induced collapse of a single bubble near a finite-sized (spherical) calcium stone, and (2) collapse of a single bubble in a confined channel. When a cavitation bubble collapses in these configurations, the collapse becomes nonspherical, leading to the formation of a high-velocity re-entrant liquid jet(s) (Fig. 1). The impact of the jet upon the opposite side of the bubble generates a water-hammer outwardly propagating shock wave, and thus can create high-pressure regions on the surface of the neighboring surfaces/objects (Fig. 2). Scaling laws of the maximum pressures (and resulting impact loads) exhibited on these surfaces based on important collapse parameters (*e.g.*, driving pressure, bubble location) were developed to predict the single-bubble dynamics.

In the first study, the synchronization of the shock wave with the bubble collapsing near the stone depends on the initial bubble size–stone size ratio for maximizing the tension inside the stone that could lead to its comminution. These results will enable fur-

ther optimization and development of ultrasound therapy tools used to fractionate elastic stones. Scaling results from the second study showed that for significant confinement, the bubble collapse jet formation dynamics change with the re-entrant jet forming in the direction parallel to the channel. This then further strengthens the collapse of the remaining bubble remnants. As a result, the channel-scaling relationship deviates from the scaling for a single bubble collapsing near a single wall.

These results enable the team to predict and quantify the damage mechanisms involving the nonspherical bubble collapse near a finite-sized elastic object in confined geometries to develop cavitation mitigation/enhancement strategies.

## WHY BLUE WATERS

The three-dimensional high-resolution simulations (requiring more than one billion grid points) that can effectively resolve the necessary small-scale features of the flow for high-fidelity results for over 36-hour simulation wall-times, as well as postprocessing and visualizations of large data files, require substantial computational power and resources. The Blue Waters petascale computing and simulation wall-time capabilities make such simulations possible and were essential for the success of this study.

## PUBLICATIONS & DATA SETS

M. Rodriguez, S. Alahyari Beig, E. Johnsen, and C. Barbier, “Inertially-driven gas bubble collapse in a channel,” in preparation, 2019.

M. Rodriguez and E. Johnsen, “Numerical investigation of a shock-induced bubble collapse near an elastic, rigid object,” in preparation, 2019.

S. A. Beig, M. Kim, and E. Johnsen, “The role of compressibility in energy budget of spherical collapse of an isolated bubble,” *Phys. Rev. Lett.*, in preparation, 2019.

S. A. Beig and E. Johnsen, “Inertial collapse of a gas bubble near a rigid boundary,” *J. Fluid Mech.*, in preparation, 2019.

S. A. Beig, “Inertial collapse of individual bubbles near a rigid surface,” presented at the 10th Int. Conf. Mult. Flows, Rio de Janeiro, Brazil, May 19–24, 2019.



NUMERICAL STUDY ON SHOCK WAVE-BOUNDARY LAYER INTERACTION AND ITS CONTROL

Allocation: Innovation and Exploration/85 Knh  
PI: Yonghua Yan<sup>1</sup>  
Collaborator: Yong Yang<sup>2</sup>

<sup>1</sup>Jackson State University  
<sup>2</sup>West Texas A&M University

EXECUTIVE SUMMARY

The overarching objective of this project is to study and improve the numerical modeling of ramp-induced shock wave-boundary layer interaction (SWBLI) control by micro-vortex generators (MVGs) in high-speed flows. The research team utilizes high-order large-eddy simulation (LES) with a large grid system to study the correlation between SWBLI low-frequency and vortex motion. SWBLI, which causes boundary layer separation and adverse pressure gradients, is a prominent problem faced by the air-breathing propulsion systems of high-speed aero vehicles. The consequences of the interaction include degraded engine performance and decreased overall propulsive efficiency of a high-speed vehicle. MVGs can alleviate or overcome the adverse effects of SWBLI and, therefore, improve the “health” of the boundary layer. An improved physical understanding of how MVGs reduce shock-induced boundary layer separation will contribute significantly to the understanding of the complex viscous-inviscid interactions that dominate high-speed aerodynamics.

RESEARCH CHALLENGE

SWBLI, which causes boundary layer separation and adverse pressure gradients, is the prominent problem faced by the air-breathing propulsion systems of high-speed aero vehicles. Particularly, SWBLI in high-speed inlets can significantly reduce the quality of the flow field by inducing large-scale separation, causing total pressure loss, flow distortions, localized heating and peak pressures, and unsteadiness. The consequences of the interaction are generally degraded engine performance and also the decreased overall propulsive efficiency of a high-speed vehicle. An improved understanding of SWBLI and flow-control technologies will directly benefit aeronautics research.

METHODS & CODES

As a result of advances in computer and code capability, high-order and high-resolution LES has become an important tool to study flow mechanisms. To capture details of instantaneous flow structure and reveal the mechanism of interaction between the vortices and the shock wave, large-scale LES was carried out. The research team used body-fitted grids for discretizing the domain. In order to obtain high-order accuracy, an elliptic grid generation with orthogonal and smoothness requirements was adopted to generate the grids [3]. The three-dimensional, time-de-

pendent, conservative Navier-Stokes equations in a curvilinear coordinate were applied as the governing system. A fully developed turbulent inflow was generated by high-order direct numerical simulation (DNS) on a flat plate with flow transition from laminar to turbulent [6]. A million inflow files were recorded as a time-dependent turbulent inflow. Adiabatic and no-slip conditions were enforced at the wall boundary on the flat plate. Periodic boundary conditions were assumed for the spanwise direction. On the inflow (the inflow could be subsonic in the lower part of the boundary layer) and the outflow boundaries, nonreflecting boundary conditions [7] were applied.

RESULTS & IMPACT

The research team collected MVG-controlled ramp flow data at two different Mach numbers (2.0 and 2.5). More data (at Mach numbers 3.0, 3.5, and higher) will be collected in the future. The study of supersonic ramp flow controlled by MVG array is ongoing and data are being collected. A series of ringlike vortices generated behind an MVG array was discovered by high-order and high-resolution LES (Fig. 1). The mechanism of the evolution and interaction of these vortices will be investigated both numerically and analytically. The numerical results obtained through this project will lead to a better understanding of the driving source of low-frequency unsteadiness of the SWBLI and thus may aid in developing methods to improve flow control in high-speed flows. The methods presented in this study as well as the resulting numerical solutions and mathematical model can pave the way for fully understanding the mechanisms of SWBLI.

WHY BLUE WATERS

Computational fluid dynamics is one of the best techniques to study complicated flows. The major scientific challenge in this study is the need to mimic the turbulent boundary layer; shock waves; the flow separation induced by the strong ramp shock; interaction between the vortices and shock waves; and the transfer of momentum, heat, and mass in the flow field. The evolution of high-performance computing (HPC) has led to significant advances in the numerical simulation of fluid flows, including turbulence, and has encouraged the direct numerical simulation of some flows. Major challenges still remain, however, and the computational requirements for turbulence-related research, driven

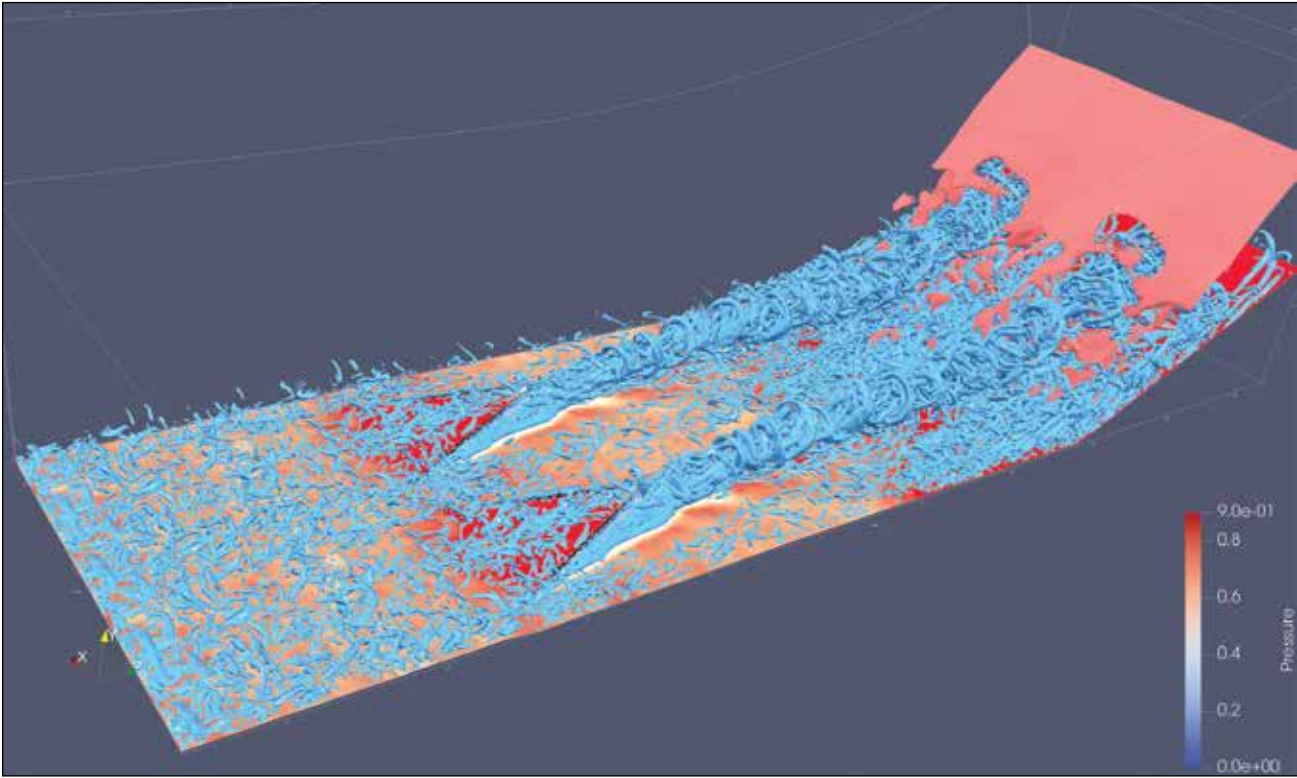


Figure 1: Vortex structures (in blue) after an MVG array (two installed). Red iso-surface represents shock wave at the ramp.

by well-established physical scaling laws, are likely to remain at the limit of the available HPC resources for some time. The main challenge in this study is to capture numerous small-scale vortices and the detailed process of the vortex-shock wave interaction in such a complex system with the existence of the turbulent boundary layer and flow separation. This requires very powerful massively parallel processing systems along with large memory and a very large amount of storage. As one of the most powerful supercomputers in the world, Blue Waters is the best choice for this project.



FREE-SURFACE FLOW MODELING OF MULTIPLE TIDAL TURBINES

Allocation: Illinois/100 Knh  
PI: Jinhui Yan<sup>1</sup>

<sup>1</sup>University of Illinois at Urbana–Champaign

EXECUTIVE SUMMARY

In this project, the PI simulated two back-to-back, full-scale, complex-geometry tidal turbines in free-surface flows using an in-house computational free-surface flow framework and the Blue Waters supercomputer. In addition, the PI investigated the wake and free-surface effects on tidal turbine performance in the downstream turbine through a case study. To quantify the free-surface effect, the researcher performed a pure hydrodynamics simulation and a free-surface simulation. He observed a significant drop in the thrust coefficient and production coefficient between the upstream and downstream turbines owing to the velocity deficit. Further, both simulations predicted almost the same thrust and production coefficients for the upstream turbine, while the pure hydrodynamics simulation predicted a noticeably higher thrust coefficient and production coefficient for the downstream turbine.

RESEARCH CHALLENGE

Although computational fluid dynamics simulations are widely used in the research and development of tidal energy, few of them consider the free-surface effect, which has been proven in experiments to have a significant influence on tidal turbine perfor-

mance. In this project, the PI developed a computational framework that is able to simulate multiple tidal turbines in turbulent free-surface flows. This framework will ultimately facilitate the research and development of tidal energy farms.

METHODS & CODES

The PI developed an in-house MPI-based parallel free-surface flow simulation framework. In the computational framework, the level-set method was adopted to track the evolution of the air–water interface. The aerodynamics and hydrodynamics were governed by a unified two-phase incompressible Navier–Stokes equation in which the fluid density and viscosity are defined by means of the level-set function. The finite-element-based Arbitrary Lagrangian Eulerian Variational Multiscale formulation enhanced with weak enforcement of essential boundary conditions was employed to discretize the free-surface flow equations. The sliding-interface formulation was used to account for the presence of the tower and nacelle, thus enabling the so-called full-machine simulation. The sliding-interface formulation was also augmented to include level-set redistancing.

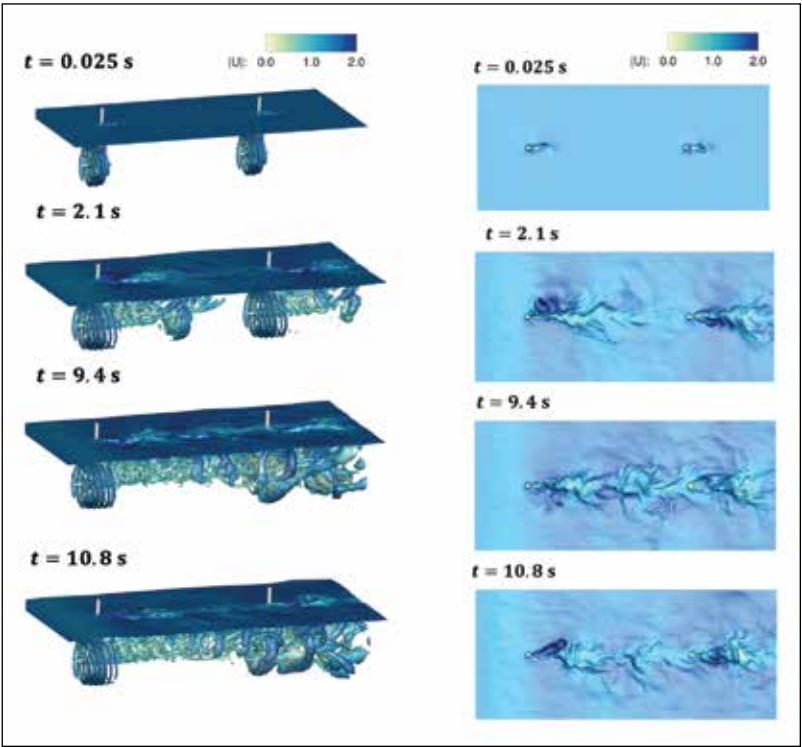


Figure 1: Free-surface deformation of two back-to-back tidal turbine simulations at four different time instances.

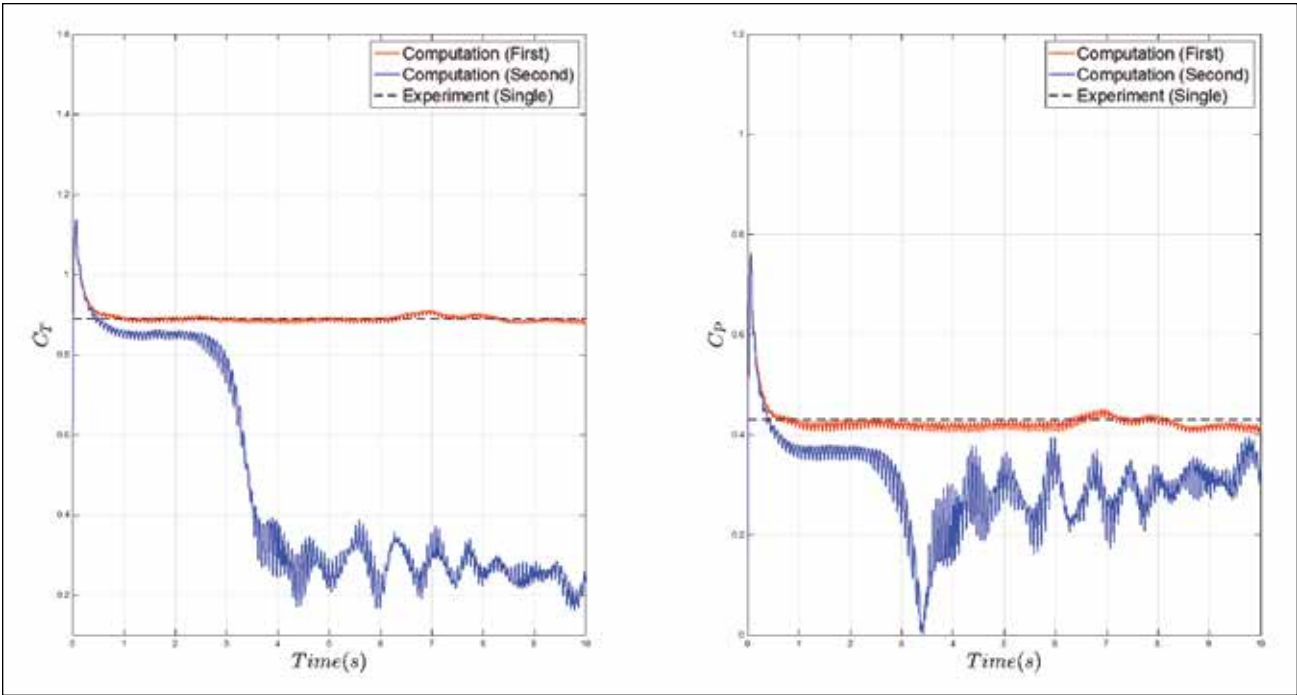


Figure 2: Time history of power and thrust coefficients of the two back-to-back tidal turbine simulations.

RESULTS & IMPACT

In this project, a computational free-surface flow framework was used to simulate two back-to-back tidal turbines in free-surface flows. The simulations were carried out at full scale and with the full complexity of tidal turbine component geometry. Without any empiricism, the simulations were able to accurately capture the effect of the free surface on the rotor hydrodynamic loading and the interaction between the upstream and downstream turbines.

For the deep-immersion operating conditions considered in this work, the free-surface simulation predicted a drop in the thrust coefficient of 70% and a drop in the production coefficient of 38.7% between the upstream and downstream turbines. By comparing the results of the pure hydrodynamics simulation and the free-surface simulation, the team found that the free surface does not affect the upstream turbine but does significantly change the performance of the downstream turbine.

This work is a first step toward using free-surface flow simulations of multiple full-scale tidal turbines with complex geometry. The proposed framework will help improve the efficiency of tidal farms by better understanding the combined wake and free-surface effects on turbines. In the future, the PI will conduct a parametric study of the distance between the upstream turbine and downstream turbine. He also plans to extend the current methodology to simulate multiple tidal turbines arranged in arrays and will consider both the fluid–structure interaction effect and cavitation.

WHY BLUE WATERS

The 3D, time-dependent, full-scale, full-geometry, multiscale, and multiphysics simulations of tidal turbines are computationally expensive, requiring supercomputers such as Blue Waters. In addition, the Blue Waters staff and technicians are knowledgeable and always willing to provide support.

PUBLICATIONS & DATA SETS

J. Yan, X. Deng, F. Xu, and S. Xu, “High fidelity numerical simulations of two back-to-back horizontal axis tidal stream turbines in free-surface flows,” *Appl. Energy*, in review, 2019.



NEW INSIGHTS ON INTERMITTENCY AND CIRCULATION STATISTICS OBTAINED FROM A MASSIVE TURBULENCE SIMULATION DATABASE

Allocation: NSF PRAC/20 Knh  
PI: Pui-Kuen Yeung<sup>1</sup>  
Co-PI: K.R. Sreenivasan<sup>2</sup>

<sup>1</sup>Georgia Institute of Technology  
<sup>2</sup>New York University

EXECUTIVE SUMMARY

Substantial advances in both domain science and computing for fluid turbulence have continued in this work on Blue Waters. A recent focus is in the study of intermittency, which is mostly characterized by multipoint descriptors of turbulence structure, whose multifractal characteristics have in turn led to significant difficulties in analyses. However, the research team has found that the circulation of velocity around a closed contour or loop (equivalent to a two-dimensional area average of a vorticity component) has much simpler (“bifractal”) properties. In particular, if all sides of a rectangular loop are within the inertial range of scale separations then the circulation depends, at least to a very close approximation, only on the size of the loop but not its shape. The study of circulation also demonstrates how a massive high-resolution turbulence database at high Reynolds numbers can provide much clearer answers than previously feasible in the literature.

RESEARCH CHALLENGE

Since turbulence is characterized by fluctuations arising over a wide range of scales, many research strategies focused on fundamental understanding have been formulated in the context of a search for scale similarity (or departures therefrom). For example, moments of the instantaneous energy dissipation rate aver-

aged over spatial regions of linear size varying over a substantial range provide a telling indicator of flow structure (Fig. 1) as well as playing an important role in corrections of classical similarity theory to account for the effects of intermittency [1]. However, researchers have also learned that [2] as a result of demanding resolution requirements [3], precise and reliable results are (whether experimentally or numerically) very difficult to obtain, which is especially the case for high-order statistics at high Reynolds numbers.

In fluid dynamics, circulation is defined as the line integral of the velocity vector around a closed loop, or, equivalently, the integral of a vorticity vector component over the area enclosed. This concept is important in the occurrence of aerodynamic lift, oceanic transport, mantle convection inside the core of the Earth, and other contexts. In principle, the statistics of circulation will depend on both the size and shape of the loop, but a theory known since the 1990s [4] suggests that only the size matters, provided that length scales on every side are within the inertial range. A wide inertial range is necessary for this theory to be tested properly. As a result, it is not surprising that several studies in the past [5,6] were not conclusive. However, since computations using machines such as Blue Waters [7] have reached resolution levels some 4,000 times larger (in terms of total number of grid points) than the state of the art of the 1990s, there is reason for new optimism.

METHODS & CODES

The research team integrated the Navier–Stokes equations over a large number of timesteps, using Fourier pseudo-spectral methods in space and finite differences in time. Circulation was computed via postprocessing of a substantial number of instantaneous snapshots of the velocity field saved during the simulations. While investigators can use either the line integral or area integral definitions as noted above, the area integration approach is more convenient. It also has the advantage of providing a more intuitive connection to the vorticity vector, which characterizes the tendency of local fluid elements to rotate on their own axes as a result of deformation by the turbulent flow. The required code development mainly resides in implementing a two-dimensional domain decomposition used to perform the simulations.

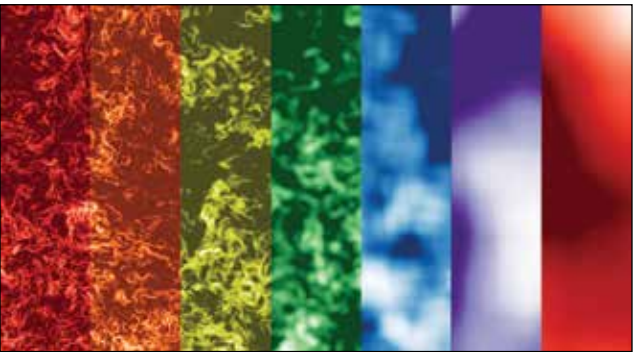


Figure 1: A mosaic of locally averaged slices of dissipation rate averaged locally over cubic subdomains of increasing scale size, from left to right, showing a change from wrinkled to smooth appearance.

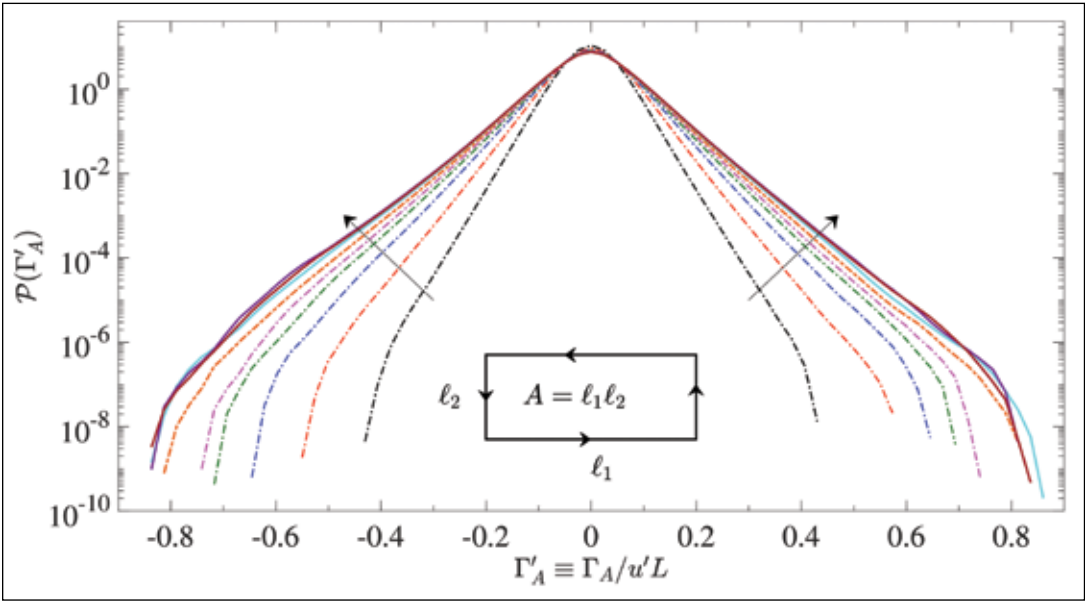


Figure 2: Probability density function of normalized circulation for loops of different aspect ratios and areas increasing in the direction of the arrows. Solid lines indicate loops with all sides in the inertial range but different shapes nearly coincide, in contrast to dashed lines for loops not wholly in the inertial range.

RESULTS & IMPACT

In [8], the research team reported on a detailed study of the statistical properties of circulation using a large database in isotropic turbulence over a range of Reynolds numbers. The highest grid resolution analyzed involved over four billion grid points (16,384 in each of the three Cartesian coordinate directions). The team primarily studied circulation computed as area integrals of the vorticity over rectangular loops. The simulation results (Fig. 2) provide strong support for the theoretical prediction (noted above) that the statistics of circulation depend only on size but not the shape of the loop, provided the entire loop is all constrained within the inertial range. The probability density function of the circulation appears to scale in a manner consistent with the classical Kolmogorov 1941 hypotheses without significant effects of intermittency. Furthermore, the circulation is shown to exhibit, to excellent accuracy, a bifractal behavior at the higher Reynolds number considered: space-filling for low-order moments, following the paradigm of Kolmogorov 1941 [9,10], and a monofractal with a dimension of about two for higher orders. This change in character, occurring roughly at the fourth moment for the highest Reynolds number considered, is reminiscent of a phase transition encountered in other branches of physics.

The evidence obtained in this study points to a reduction in complexity when considering averaged vorticity over loops, which may provide a route to circumventing the spatial complexities involving velocity differences and gradients in turbulence [11]. This conclusion suggests that a great simplification, in principle, of the intermittency problem in three-dimensional turbulence may be possible when viewed through the lens of vorticity correlators in loop spaces. This great simplification may, in turn, have the impact of invigorating the use of similar loop formulations in tackling turbulence.

WHY BLUE WATERS

The 8,192<sup>3</sup>-grid resolution of the research team’s production simulations requires access to a world-class machine such as Blue Waters. The machine capacity on Blue Waters has also proven sufficient to allow the team to obtain data at even higher resolution, *i.e.*, 12,288<sup>3</sup> and even 16,384<sup>3</sup> (although only for short periods of time), which is the highest known worldwide in the turbulence community. Indeed, 16,384<sup>3</sup> data, although short in time span, have been the focus of analysis presented in [8].

PUBLICATIONS & DATA SETS

- D. Buaria, A. Pumir, E. Bodenschatz, and P.K. Yeung, “Extreme velocity gradients in turbulent flows,” *New J. Phys.*, vol. 21, p. 043004, 2019.
- K. P. Iyer, J. Schumacher, K. R. Sreenivasan, and P. K. Yeung, “Scaling of locally averaged energy dissipation and enstrophy density in isotropic turbulence,” *New J. Phys.*, vol. 21, p. 033016, 2019.
- P. K. Yeung, “Extreme events, resolution, and onward to exascale computing,” presented at the Prospectives on Turbulence Workshop, Texas A&M University, College Station, TX, U.S.A., Aug. 2018.
- P. K. Yeung, “Massive computations in 3D turbulent flow: Science goals and an asynchronous algorithm towards the exascale,” presented at Amer. Geophys. Union 100th Fall Meeting, Washington, DC, U.S.A., Dec. 12, 2018.
- P. K. Yeung, “Extreme scale computing in turbulence: physics and algorithms,” presented at an invited seminar at the Center for Environmental and Applied Mechanics, The Johns Hopkins University, Baltimore, MD, U.S.A., Mar. 8, 2019.



EFFECTS OF SURFACE DEFECTS ON HYDROPHOBICITY AT RARE-EARTH OXIDE INTERFACES USING MOLECULAR DYNAMICS SIMULATIONS DRIVEN BY *AB INITIO*-BASED DEEP NEURAL NETWORK POTENTIALS

Allocation: Illinois/54.1 Knh  
PI: Yang Zhang<sup>1</sup>

<sup>1</sup>University of Illinois at Urbana–Champaign

EXECUTIVE SUMMARY

Hydrophobic and superhydrophobic surfaces that are robust to harsh environments have immense potential to enhance the performance of a plethora of applications. However, the successful widespread commercialization of hydrophobic surfaces has been fraught with many challenges. The biggest challenge is the lack of mechanical, chemical, and thermal robustness.

Recent studies show that rare-earth oxides (REOs) are intrinsically hydrophobic and durable owing to their unique electronic structure. However, surface defects such as adatoms (single atoms lying on surfaces) and vacancies are ubiquitous and may change the wettability of REOs. Thus, in this project, the research team investigated the influence of the defects on hydrophobicity and elucidated the mechanism governing wettability by employing first-principle neural network potentials combined with molecular dynamics (MD) simulation. This work is significant for simulating and understanding the hydrophobicity of REOs at the molecular level and providing insights to identify ideal candidates with strong hydrophobicity when considering the presence of defects.

RESEARCH CHALLENGE

Water/solid interfaces have attracted significant attention owing to their central role in many fields such as electrochemistry, corrosion, and heterogeneous catalysis. Hydrophobic materials have immense potential in enhancing the performance of materials under harsh environments, ranging from reducing ice adhesion to eliminating corrosion. Although common metals are hydrophilic while general organic materials are hydrophobic, organic materials deteriorate easily and are not durable in harsh environments.

Recently, Azimi *et al.* [1] found that REOs are intrinsically hydrophobic owing to the saturation of electrons in their outer shells and their lower tendency to form hydrogen bonds. However, surface defects such as adatoms and vacancies are pervasive, which may change the hydrophobicity of REOs because of increasing surface reactivity and attracting structural arrangements. This project aims to investigate the influence of such defects on hydrophobicity with MD simulations.

*Ab initio* MD with electronic structure theory is the gold standard for simulations of solid–liquid interfaces owing to minimal assumptions made to describe interactions between water and surface. However, the relatively high computational cost limits the system size and timescale. High-dimensional neural network

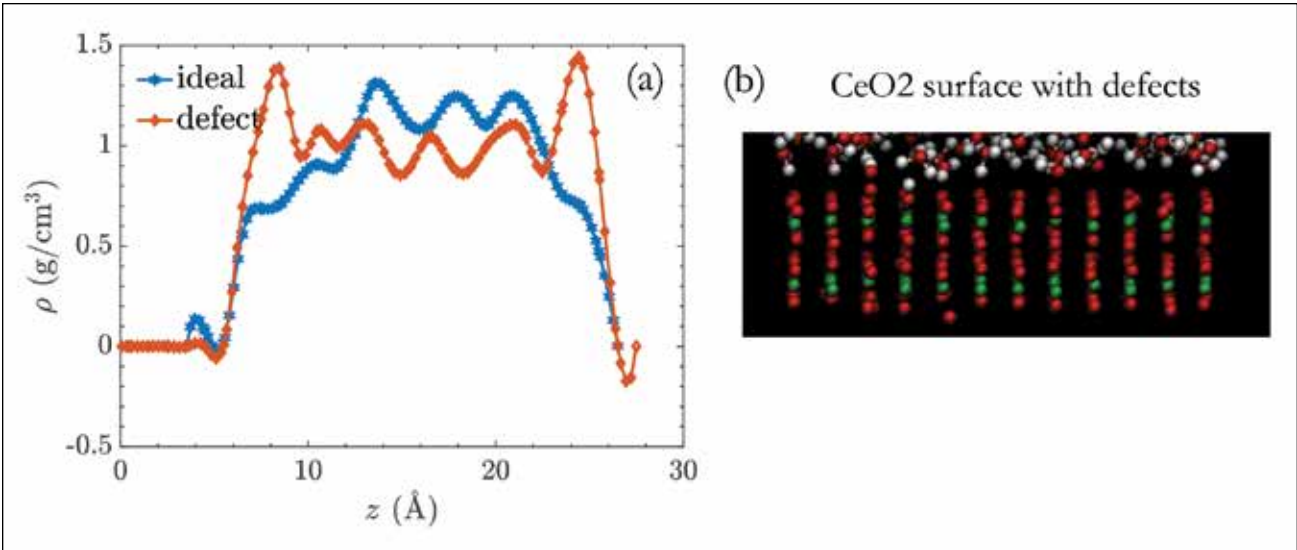


Figure 2: (a) Water density along z directions at both the ideal and defected ceria surface. (b) Snapshot of the water dissociation at the defected ceria surface.

potentials (NNPs) based on density functional theory (DFT) data have the ability to enable efficient classical MD with accuracy close to *ab initio* MD. By combining classical MD with NNPs, the research team was able to perform long-timescale simulations with high accuracy.

METHODS & CODES

Starting from bulk water, a specific bulk REO such as ceria (CeO<sub>2</sub>), and a wide range of REO–water interfaces, the team carried out DFT reference calculations using VASP [2]. Projector augmented waves with a plane wave energy cutoff of 700 eV to converge the results were used to describe the electron-core interaction in the DFT calculations. Furthermore, more uncorrelated structures for DFT calculations were generated by taking snapshots of *ab initio* MD simulation trajectories. The calculated energy and force were related with a vector of atom-centered symmetry functions defined by atomic environment. For each atom, the symmetry function vector was fed into an individual atomic neural network and the neural network was constrained to have the same architecture and parameters for each element in the system. Using these symmetry functions related to each kind of atom and energy, the reference data were split into training sets to determine the fitting parameters for each type of atom and test sets to verify the transferability of those parameters. An iterative gradient-based fitting process was used to determine the fitting parameters by minimizing the error function until a set of parameters could accurately reproduce the reference energies and forces and a converged neural network potential energy surface was obtained. Using the generated potentials, the research team performed MD simulations to collect trajectories to quantify the level of hydrophobicity.

RESULTS & IMPACT

By performing MD simulations with constructed neural network potential, both structural and dynamics properties can be calculated efficiently for ideal and defected systems. The accuracy using neural network potentials is close to DFT calculation and *ab initio* MD. Upon comparing the hydrophobicity in both the ideal and defected ceria–water interface, the researchers found that the interaction of water with the surface increases and water dissociates in the presence of the defects, suggesting the rare-earth oxides interface becomes hydrophilic when there are defects in the surface.

Using efficient and accurate *ab initio*-based neural network potentials makes it possible to explore the general structural and dynamical characterization of large-scale solid–liquid interfaces with first-principle accuracy. In the future, this work will help to identify ideal candidates with strong hydrophobicity by using a closed-loop characterization.

WHY BLUE WATERS

Collecting high-quality data requires computationally expensive DFT and *ab initio* MD simulations. To ensure the convergence of the properties, a reasonable system size of 500 atoms is essential. Tens of nodes are required to fulfill the memory and speed requirements of the calculations. Access to the Blue Waters system made the calculations possible. Furthermore, the VASP code can run efficiently on Blue Waters, and the project staff were very helpful and responded quickly to solve any computational issues.

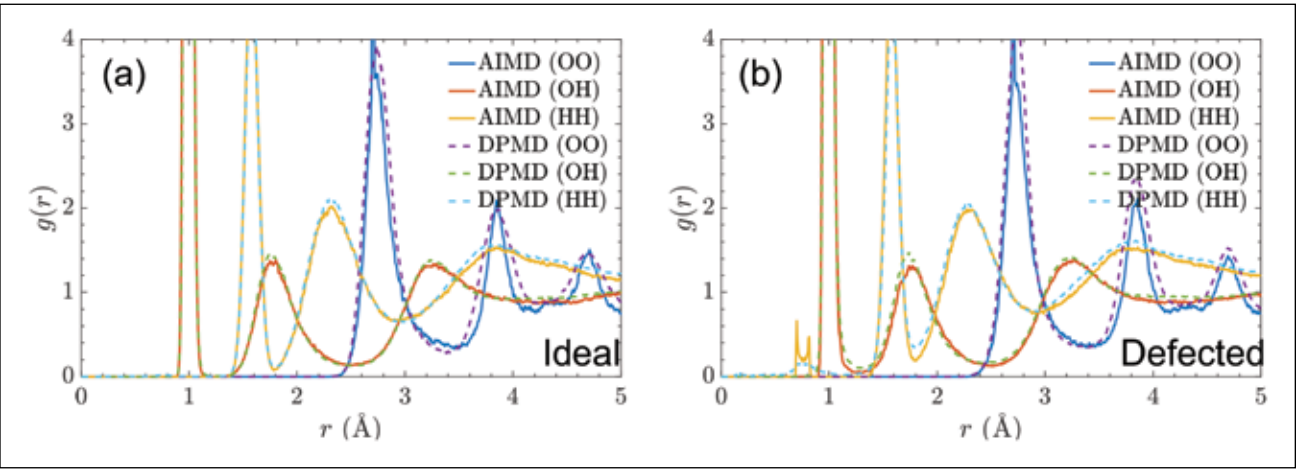


Figure 1: Comparison of the pair distribution function of O–O, O–H, H–H using *ab initio* MD (AIMD) and MD driven by neural network potentials (DPMD) at ideal ceria surface (a) and defected ceria surface (b).





# CONSTRAINING THE PROPERTIES AND INTERACTIONS OF DARK MATTER

**Allocation:** NSF PRAC/14,000 Knh  
**PI:** Phiala Shanahan<sup>1</sup>  
**Collaborators:** Balint Joo<sup>2</sup>, David Murphy<sup>1</sup>, William Detmold<sup>1</sup>, Michael Wagman<sup>1</sup>

<sup>1</sup>Massachusetts Institute of Technology  
<sup>2</sup>Thomas Jefferson National Accelerator Facility

## EXECUTIVE SUMMARY

This project addresses the Grand Challenge research problem of understanding the nature of the mysterious dark matter that permeates the universe. In particular, to interpret the results of direct experimental searches for dark matter it is necessary to quantify the interaction of potential dark matter particles with the nuclei used as targets in the detectors. This is a hugely demanding computational task that requires significant resources.

## RESEARCH CHALLENGE

Understanding the nature of dark matter is a defining challenge for contemporary particle and nuclear physics. This project will provide necessary theoretical input for interpreting dark matter searches that are currently being undertaken at laboratories around the world, and will allow optimal design of the next generation of such experiments. The interaction of a broad class of dark matter candidates with the nuclei used in direct detection experiments is governed by nuclear scalar matrix elements. In recent work, the research team determined the scalar matrix elements of light nuclei via first-principles calculations of the underlying interactions with the Standard Model, albeit with larger-than-physical values of the quark masses used in the study (allowing computationally cheaper calculations). These calculations revealed significant, and unexpected, order 10% nuclear effects in the scalar interactions in light nuclei. These significant effects in small nuclei potentially indicate even larger effects and uncertainties in the scalar matrix elements of the much larger nuclei, for example xenon, typically used in direct detection experiments. If these effects persist in a controlled study at the physical values of the quark masses, it will have significant implications for the interpretation of the results of current and future dark matter direct detection experiments around the world.

## METHODS & CODES

This study has been undertaken within the framework of lattice QCD (quantum chromodynamics), which is a first-principles method of calculating strong interaction matrix elements numerically on a discrete four-dimensional spacetime. In this approach, Monte Carlo techniques are used to create a representative set, known as an ensemble, of configurations of the background gluon fields on the links defined between points on the spacetime lattice. This ensemble is then used to perform calculations of quantities

of physical interest. As the only known direct method of studying QCD at the low energies relevant for hadronic and nuclear interactions, lattice QCD is an important source of information for tests of the Standard Model, and it provides results for various hadronic and nuclear matrix elements that are systematically improvable and model-independent. As such, it is the necessary tool for the calculations undertaken.

## RESULTS & IMPACT

This project began in mid-2019. The codebase has been successfully implemented on Blue Waters and production runs have begun. The calculations that are running are producing the time series data that will allow the determination of the nuclear quantities that govern the interactions of dark matter with light nuclei. Additionally, data structures are being generated that form the first step in calculations of nucleon–nucleon scattering and studies of light nuclear spectroscopy. Once the calculations are complete, they will enable understanding of nuclear effects in dark matter interactions with nuclei that will have far-reaching impacts on current and future terrestrial searches for dark matter. The research team anticipates that the precision they will achieve with the runs that are in progress will be sufficient to influence the design of future experiments searching for dark matter.

## WHY BLUE WATERS

Undertaking a controlled calculation of the interactions of dark matter with nuclei is a hugely demanding computational task that requires significant resources that would not otherwise be available to the project team.



# COMPUTER SCIENCE & ENGINEERING

---

## ENGINEERING & FINANCE

## NETWORK CONGESTION

## PARALLEL FUNCTIONALITY

## RESILIENCE

## SCALABILITY

## VISUALIZATION

- 212

*Cinematic Scientific Data Visualization for CADENS*
- 214

*The Structure and Statistics of Reed–Muller Codes*
- 216

*A Parallel Framework for Scaling Phylogeny Estimation Methods to Large Genomic Data Sets*
- 218

*Algorithms for Extreme-Scale Sytems*
- 220

*Kaleidoscope: Live Forensics for Large-Scale Data Center Storage Systemse*
- 222

*Extensible and Scalable Adaptive Sampling to Fold Proteins on Supercomputers*
- 224

*Improved Scalability Through Node-Aware Communicators*
- 226

*Scalable Line and Plane Solvers*
- 228

*Detection of Silent Data Corruptions Using Machine Learning*
- 230

*Pushing the Boundaries of Large-Scale Tensor Computations*
- 232

*Algorithms for Large-Scale Evolutionary Tree Construction: Improving Scalability and Accuracy through Divide-and-Conquer*
- 234

*HPC Development of Deep Learning Models in Scientific Computing and Finance*
- 235

*Optimization of a Field Data Parallel Output Library*



CINEMATIC SCIENTIFIC DATA VISUALIZATION FOR CADENS

**Allocation:** Illinois/175 Knh  
**PI:** Donna Cox<sup>1</sup>  
**Co-PIs:** Robert Patterson<sup>1</sup>, Kalina Borkiewicz<sup>1</sup>, A. J. Christensen<sup>1</sup>, Stuart Levy<sup>1</sup>, Jeffrey Carpenter<sup>1</sup>  
**Collaborators:** Thomas Lucas<sup>2</sup>, Melih Sener<sup>1</sup>, John Stone<sup>1</sup>, Barry Isralewitz<sup>1</sup>, Sarah T. Stewart<sup>3</sup>

<sup>1</sup>University of Illinois at Urbana–Champaign  
<sup>2</sup>Thomas Lucas Productions  
<sup>3</sup>University of California, Davis

EXECUTIVE SUMMARY

The Advanced Visualization Lab (AVL) at the University of Illinois at Urbana–Champaign (Illinois) has continued to work on the National Science Foundation-funded CADENS project (Centrality of Advanced Digitally Enabled Science, ACI-1445176). The AVL coproduced and rendered visualization scenes for two full-dome planetarium shows, *Birth of Planet Earth* and *Imagine the Moon*. The research team has used Blue Waters for processing data as well as rendering scenes in 4K monoscopic, 4K stereoscopic, and full-dome formats. The researchers also rendered visualizations for the Earth’s Call climate change event in Aspen, CO, U.S.A., May 17–19, 2019.

RESEARCH CHALLENGE

Drawing on data from scientists involved in high-performance computing-based research, the AVL creates visualizations in cinematic style intended for public outreach, through flat-screen science documentaries, films, and full-dome planetarium shows.

METHODS & CODES

The team’s Blue Waters visualization work depends on several externally provided packages: Houdini, commercial visual effects software from SideFX; yt, the data analysis and visualization package (yt-project.org) for ingesting and regridding some types of data; VMD, the Visual Molecular Dynamics package from John Stone of the Theoretical Biophysics group at Illinois (www.ks.uiuc.edu); and Python with numpy and scipy for many sorts of data preprocessing from the forms provided by the scientists into forms usable by Houdini. The team has also developed their own software tools, including Ytini for yt-Houdini integration, and Blurend to organize the Houdini rendering workflow for Blue Waters’ environment.

Visualizing energy harvesting in a photosynthetic purple bacterium [1] involved combining structural models from atomic, protein, organelle, and cell scales using multiple software tools [2]. The construction of structural models [3] was performed with VMD [2] and Mathematica, which was also used for the determination of relevant energy conversion steps [1]. Multiple copies of a static chromatophore model [3] were assembled in Houdini to emulate the interior of a low-light-adapted purple bacterial cell. Dynamic elements—photons, electronic excitations, protons, quinols/quinones, and ATP—were choreographed using Houdini

to illustrate energy conversion processes for a lay viewer. Since the timescales for these conversion processes span almost 12 orders of magnitude (femtoseconds to milliseconds) [1], the animations deliberately represent a simplified visual narrative rather than the results of a specific simulation at one timescale. Some visual elements were rendered using Houdini, others with VMD. In addition, the researchers relied on Blue Waters’ capacity to create a high-quality visualization of Sarah T. Stewart’s Moon-forming collision [4], as explained below.

RESULTS & IMPACT

Blue Waters enabled the research group to create and refine two data-driven cinematic animations for two full-dome planetarium shows released in 2019:

- Visualizing Energy Harvesting in a Photosynthetic Purple Bacterium[1–3]—*Birth of Planet Earth*
- Formation of the Moon [4]—*Imagine the Moon*

Full-dome shows have a lifespan of about 10 years, and though it has only been several months, *Birth of Planet Earth* has already received two awards and is being shown internationally. To suggest potential public impact for this work, it can be compared with another full-dome planetarium show in the CADENS series, *Solar Superstorms*, for which AVL also relied on Blue Waters for data visualization. Since its 2015 release, it has been booked by over 70 planetaria and science museums in 15 countries, and translated into at least 10 languages.

The research team also used Blue Waters to create data-driven graphics for the Earth’s Call climate change event in 2019, which was viewed by more than one million people globally. The group visualized a total of 12 data sets for this event, five of which [5–9] were rendered using Blue Waters owing to the length of time required to render each frame.

WHY BLUE WATERS

Access to Blue Waters has allowed the team to iterate quickly and to meet deadlines. They were able over the course of a single weekend to render a Houdini scene involving a chromatophore model [3] and its surrounding environment that was made up of six separate render layers and totaled 20,870 image frames. The capability to render a large amount of images in a short period of time has allowed the team to make several iterations of the scene

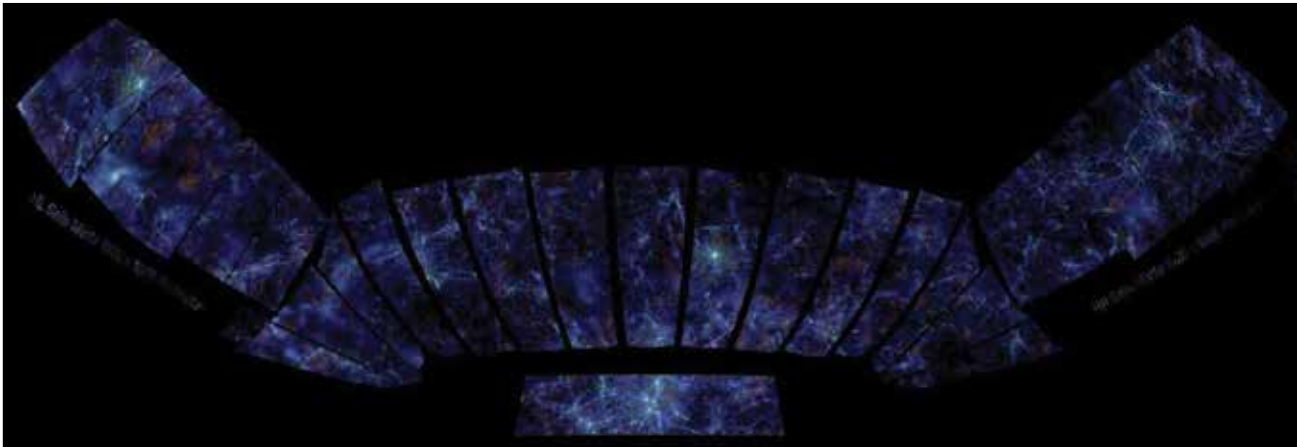
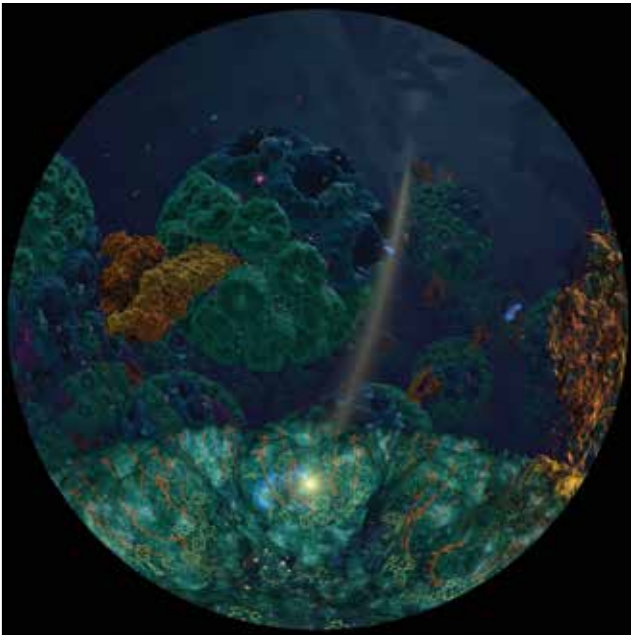


Figure 1: (top left) Photosynthetic energy harvesting in the purple bacterium *Rba. sphaeroides* [1]. Adapted from a visualization [2] of the structure [3] of chromatophores—photosynthetic vesicles—illustrating energy conversion steps from the capture of sunlight to the synthesis of ATP. (This segment is dedicated to Klaus Schulten. From the planetarium show, *Birth of Planet Earth*.) Figure 2: (top right) Visualizing the collision that formed Earth’s Moon [4] as a Mars-sized body strikes the early Earth. Appears in *Imagine the Moon*. Figure 3: (bottom) Visualization of the large-scale structure of the early universe [5], projected onto a screen as well as front and side sails around the audience at the Earth’s Call event in 2019.

before finalizing a video shown during the International Planetary Society 2018 conference. This would not have been possible on the team’s local cluster.

Furthermore, Blue Waters has made it possible to not sacrifice visual quality for render speed. One approximately 200-frame segment of the visualization of a planetary collision [4] was taking as many as 20 hours per frame to render. Blue Waters made it possible to render these images despite the long render time. Without Blue Waters, the team would have had to either change many render settings and significantly decrease the render quality, to change the camera position and lose the dramatic effect of having an arm of disk material pass closely overhead, or to spend

many days trying to come up with a different data representation of the simulation to make the render more manageable on the local cluster.

PUBLICATIONS & DATA SETS

*Birth of Planet Earth*, full-dome planetarium show, directed by Thomas Lucas, distributed by Spitz, Inc., 2019.

*Imagine the Moon*, full-dome planetarium show at Adler Planetarium, Chicago, Illinois, U.S.A., 2019.

Earth’s Call event, May 17–19, 2019, <https://earthscall.org/>.



THE STRUCTURE AND STATISTICS OF REED-MULLER CODES

Allocation: Exploratory/20 Knh  
PI: Iwan Duursma<sup>1</sup>  
Collaborator: Hsin-Po Wang

<sup>1</sup>University of Illinois at Urbana-Champaign

EXECUTIVE SUMMARY

This project focuses on computing the Tutte polynomial of the Reed-Muller (RM) codes of block length 64. RM codes are among the earliest linear block codes whose construction is easy to describe, and yet their properties have not been fully understood. They are related to polar codes, a class of codes that was introduced in the last 10 years and that is known to have good asymptotic properties. It is conjectured that RM codes also share those properties. The Tutte polynomial is an invariant of linear block codes that captures the probability of error when correct-

ing erasures in codewords. The Tutte polynomial is used more generally in graph theory, knot theory, and physics.

RESEARCH CHALLENGE

The Tutte polynomial is a two-variable generating function for the corank-nullity distribution over all subsets of columns of a given matrix. It can be computed exhaustively or recursively using a sequence of deletion-restriction operations. It is known that computing the Tutte polynomial for general block codes is sharp-P hard (the quantitative version of nonpolynomial hard-

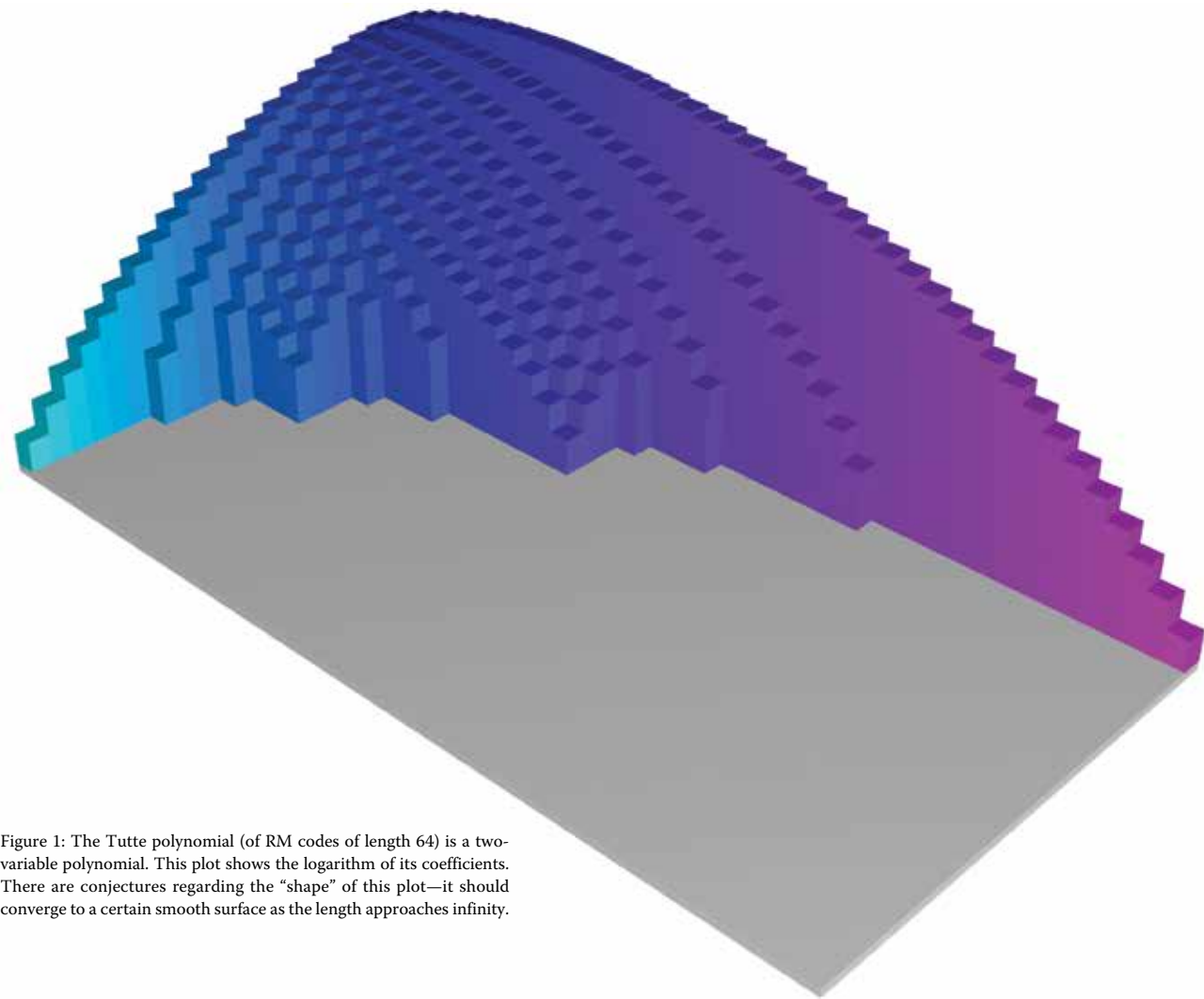


Figure 1: The Tutte polynomial (of RM codes of length 64) is a two-variable polynomial. This plot shows the logarithm of its coefficients. There are conjectures regarding the “shape” of this plot—it should converge to a certain smooth surface as the length approaches infinity.

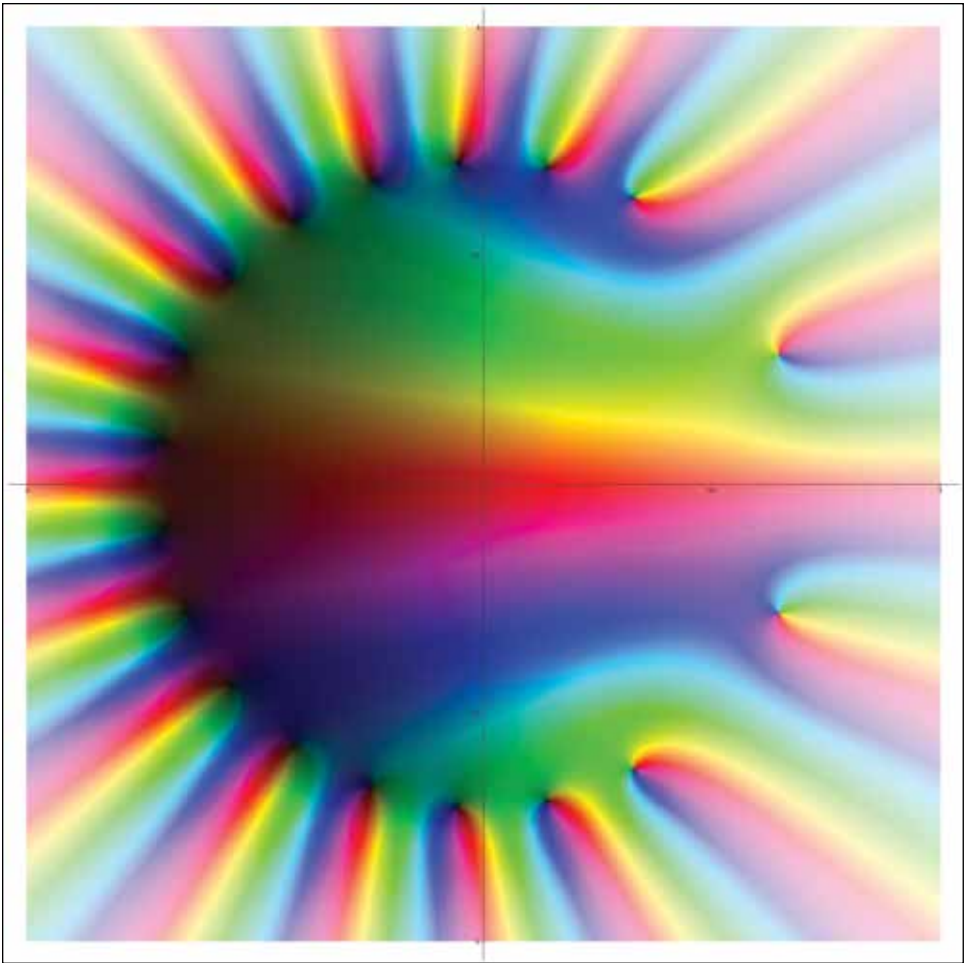


Figure 2: The zeta function of the [32,16,8] RM code using domain coloring. The zeta function in coding theory is an analog to the Riemann zeta function and the Hasse-Weil zeta function. It can be seen that the zeros all lie on a circle: the “critical line.”

ness). For Reed-Muller codes, Tutte polynomials are known up to length 32. The research team extended this to length 64 and, moreover, computed a newly defined invariant that provides more detailed information than the Tutte polynomial.

METHODS & CODES

The reduction that essentially brings this problem within a feasible scope is the Plotkin structure of RM codes. The fact that the construction of RM codes is recursive suggests that an invariant can be computed along the recursion. However, this fact only reduces the time cost by a factor of 60,000. The remaining work still costs Blue Waters 2,000 node-hours to complete. All the codes employed were produced by the team.

RESULTS & IMPACT

The researchers computed the Tutte polynomial along with extra data that allow followers to double-check whether the computation is correct. The exact polynomial may be found at <https://github.com/Symbol1/BlueWaters-RM64/blob/master/rm64tutte.txt>. The extra data with a full explanation of how they were computed and how they can be verified may be found at <https://github.com/Symbol1/BlueWaters-RM64>.

WHY BLUE WATERS

The role of Blue Waters is twofold. First, this computation is CPU-intensive, with a great deal of conditional branching and bitwise operators, and Blue Waters provides a parallel context to accelerate the computation using all 32 cores in a node. Second, the team generated an abundance of intermediate data that should be kept in RAM to avoid the hard disk’s I/O, because otherwise the I/O becomes the bottleneck. The team expected more, but the given 64 GB of memory per node barely fits. (The researchers tailored the MPI reduce function to save space.) In addition, the computation does not write any error-handling codes, but apparently Blue Waters was all green when this job was running, thanks to its sustainable design.

PUBLICATIONS & DATA SETS

*Computation of Reed-Muller codes of length 64 on Blue Waters.* [Online]. Available: <https://github.com/Symbol1/BlueWaters-RM64>



A PARALLEL FRAMEWORK FOR SCALING PHYLOGENY ESTIMATION METHODS TO LARGE GENOMIC DATA SETS

Allocation: Exploratory/50 Knh  
PI: William Gropp<sup>1</sup>  
Co-PI: Erin Molloy<sup>1</sup>  
Collaborator: Tandy Warnow<sup>1</sup>

<sup>1</sup>University of Illinois at Urbana–Champaign

EXECUTIVE SUMMARY

Phylogenetic trees are graphical models of evolution that can be used to study how species evolve and adapt to their environment. Recent advances in sequencing technology have resulted in an explosion of data and the creation of ultralarge data sets. Today, scientists need to estimate highly accurate phylogenetic trees from these data sets; however, this process is computationally challenging. Many of the best methods are not easily parallelizable and have large computational footprints (memory and running time).

The research team’s main result is a parallel framework for scaling phylogeny estimation methods to large data sets while maintaining accuracy. By running their preferred method within the team’s parallel framework, scientists will be able to estimate evolutionary trees using reduced computational resources. Thus, these computational tools will be useful for researchers attempting to build the Tree of Life, a scientific and computational Grand Challenge, as well as researchers in medicine, agriculture, and other domains.

RESEARCH CHALLENGE

Phylogenetic trees are graphical models of evolution that can be used to study evolutionary processes such as how species evolve and adapt to their environments. Such models are useful in a variety of applications, including the prediction of protein function or the classification of molecular sequences of unknown origins. Recent advances in sequencing technologies have resulted in an explosion of data, including genome-scale data for very large numbers of species. Today, scientists need to estimate highly accurate phylogenetic trees from such data sets; however, building these evolutionary trees from genome-scale data sets is challenging. For example, genome-scale data sets can be more heterogeneous owing to incomplete lineage sorting (ILS) and other biological processes that result in different regions of the genome having different evolutionary histories. In addition, many of the best methods for phylogeny estimation are heuristics for solving NP (nondeterministic polynomial-time)-hard optimization problems, and the solution space for these problems grows exponentially with the number of species. Parallelizing the existing heuristics does not typically reduce the computational effort required to search the solution space. Thus, even high-performance com-

puting implementations can require many CPU years to run on large, heterogeneous data sets.

METHODS & CODES

Divide-and-conquer is a technique used by computer scientists to run methods on large data sets. The traditional divide-and-conquer approach for estimating large evolutionary trees operates by: (1) dividing the species into overlapping subsets, (2) estimating trees on each subset, and (3) combining subset trees together by solving the supertree problem. However, supertree estimation is also computationally challenging (the best methods are heuristics for solving NP-hard optimization problems and the solution space for these problems grows exponentially with the number of species). In order to overcome this challenge, the research team introduced a new divide-and-conquer approach that divides the species into pairwise disjoint subsets and combines subset trees together by solving the disjoint tree merger (DTM) problem, which can be solved in polynomial time. TreeMerge, the DTM method developed for this Blue Waters project, is a dramatic improvement on NJMerge, requiring far less memory and far less communication.

RESULTS IMPACT

The team implemented the divide-and-conquer approach (using TreeMerge to combine subset trees) as a parallel framework and performed a benchmarking study on Blue Waters. They compared running methods within the parallel framework (in order to estimate subset trees) versus running methods to estimate a tree on the full data set. The researchers found that running methods within the parallel framework dramatically reduced running time and maintained accuracy for two leading species tree estimation methods: ASTRAL-III [1] (Fig. 1) and RAXML [2] (not shown). Novel species tree methods are continually being developed, and many of these new methods are computationally intensive (for example, new Bayesian inference methods). Because researchers can specify that the method be run on subsets, the research team’s parallel framework will be useful in the rapidly progressing field of computational biology and will further research efforts toward building the Tree a Life, a scientific and computational grand challenge.

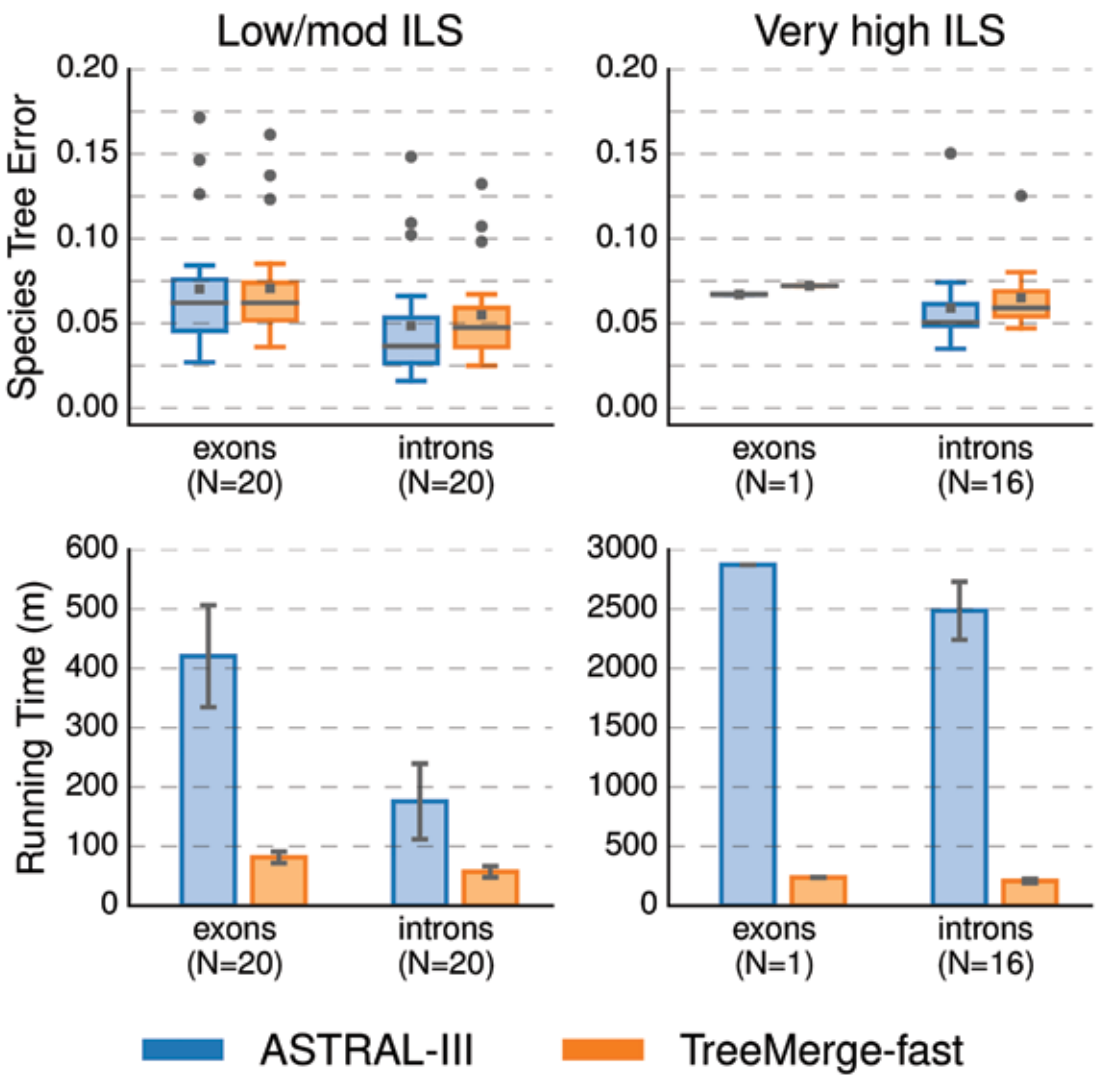


Figure 1: Running ASTRAL-III within the research team’s parallel framework (TreeMerge) maintains accuracy (top row) and reduces running time (bottom row). Left and right columns show two model conditions: low/moderate and very high levels of ILS (incomplete lineage sorting). The x-axis indicates the sequence type (exon or intron) and the number N of replicates where ASTRAL-III completed.

WHY BLUE WATERS

Method development is an iterative process that requires testing any new method on many large data sets. This process also requires that the new method be compared in terms of accuracy to the best existing methods; running the existing methods on large data sets can take months of CPU time. Without the Blue Waters system, the research team would have been unable to efficiently develop and improve upon novel methods.

PUBLICATIONS & DATA SETS

E. Molloy and T. Warnow, “Statistically consistent divide-and-conquer pipelines for phylogeny estimation using

NJMerge,” *Algo. Mol. Biol.*, vol. 14, no. 1, article 14, Jul. 2019, doi: 10.1186/s13015-019-0151-x.

*Data from statistically consistent divide-and-conquer pipelines for phylogeny estimation using NJMerge*, Illinois Data Bank, 2019. [Online]. Available: [https://doi.org/10.13012/B2IDB-0569467\\_V2](https://doi.org/10.13012/B2IDB-0569467_V2)

E. Molloy and T. Warnow, “TreeMerge: A new method for improving the scalability of species tree estimation methods,” *Bioinformatics*, vol. 35, no. 14, pp. i417-i426, July 2019, doi: 10.1093/bioinformatics/btz344.

*Data from TreeMerge: A new method for improving the scalability of species tree estimation methods*, Illinois Data Bank, 2019. [Online]. Available: [https://doi.org/10.13012/B2IDB-9570561\\_V1](https://doi.org/10.13012/B2IDB-9570561_V1)



ALGORITHMS FOR EXTREME-SCALE SYTEMS

**Allocation:** Blue Waters Professor/40 Knh  
**PI:** William Gropp<sup>1</sup>  
**Students:** Paul Eller<sup>1</sup>, Ed Karrels<sup>1</sup>

<sup>1</sup>University of Illinois at Urbana–Champaign

EXECUTIVE SUMMARY

Continued increases in the performance of large-scale systems will come from greater parallelism at all levels. At the node level, this is seen both in the increasing number of cores per processor and the use of large numbers of simpler computing elements in general-purpose computing on GPUs. The largest systems must network tens of thousands of nodes together to achieve the performance required for the most challenging computations. Successfully using these systems requires new algorithms. In the last year, the research team developed a simple yet complete implementation of the Message-Passing Interface Cartesian topology routines that significantly outperforms the available implementations, as well as further developing performance models and implementations for highly scalable Krylov methods for the solution of large linear systems.

RESEARCH CHALLENGE

This work directly targets current barriers to effective use of extreme-scale systems by applications. For example, Krylov methods such as Conjugate Gradient are used in many applications currently being run on Blue Waters and other large-scale systems. These methods depend both on high-performance matrix–vector products, which are communication intensive, and on collective all-reduce operations, which introduce synchronizations that can limit scalability. Developing and demonstrating a more scalable version of this algorithm would immediately benefit those applications. The research team’s approach begins with developing a performance model that captures the key aspects of the intra- and internode communication costs and uses that model to inform the development of new algorithms. This approach has also yield-

ed improved parallel I/O routines and a better implementation of a process placement operation that can improve the performance of applications.

METHODS & CODES

To address these challenges, the research team has developed several performance models that address limitations in off-node communication bandwidth, message matching costs, network contention, and the effect of system “noise.” Benchmarks to test these performance models have been developed, and experiments have been conducted with some applications. These performance models have led to new algorithms for providing good and efficient implementation of process mapping for regular meshes. Some of the codes are open source and available; the routines for process mapping in regular Cartesian topologies are under consideration for inclusion in the MPICH implementation of the Message-Passing Interface (MPI).

RESULTS & IMPACT

Paul Eller, working with PI William Gropp, has been using Blue Waters over the last year for an investigation into performance modeling of scalable Krylov solvers for structured grid problems. This includes developing code for measuring and processing parallel runtimes and network performance counters, developing a collection of kernels relevant to structured grid problems, exploring their use in several applications, and developing performance models with penalty terms that accurately model parallel performance at scale. He has run experiments to determine parameters for the performance models and performed scaling studies for the various parallel communication kernels and scalable conjugate gradient solvers. These results have led to a paper submitted to the 2019 Association for Computing Machinery International Conference on Supercomputing. These experiments have helped to better understand Krylov solver performance at scale, to develop more accurate performance models, and to optimize the solvers to obtain better performance. Ed Karrels, also working with William Gropp, has continued to explore parallel I/O performance, including exploring the sensitivity of several parameters for I/O operations, such as stripe count and size, and the number of processes per node performing I/O operations. The goal is to provide better guidance for users in choosing I/O parameters and to develop more automatic methods for selecting these parameters within parallel I/O libraries. The libraries that he developed in the previous year—MeshIO, zlines, and zchunk—remain available for use by applications. In addition, William Gropp has developed a new algorithm for implementing process mapping for Cartesian grids, which is needed for many applications that use structured grids. MPI provides a convenient routine for this operation, but few MPI libraries provide a good implementation of this operation. As a result, applications must either forgo the performance or use ad hoc, nonportable techniques to achieve a good mapping. Such tools do exist for Blue Waters, but these are not portable to other sys-

tems and often fail to provide good process mappings in practice. Applications should be able to rely on the features in MPI and not need to use nonstandard, nonportable methods. Finally, by using insight gained from the team’s new performance model, they developed an alternative implementation of MPI\_Cart\_create that provides a significant performance benefit compared to the vendor-supplied implementation, as shown in Fig. 1.

WHY BLUE WATERS

Scalability research relies on the ability to run experiments at large scale, requiring tens of thousands of nodes and hundreds of thousands of processes and cores. Blue Waters provides one of the few available environments where such large-scale experiments can be run. In addition, Blue Waters provides a highly capable I/O system, which the research team used in developing improved approaches to extreme-scale I/O.

PUBLICATIONS & DATA SETS

W. D. Gropp, “Using node and socket information to implement MPI Cartesian topologies,” *Parallel Comput.*, vol. 85, pp. 98–108, July 2019.

A. Bienz, W. D. Gropp, and L. N. Olson, “Node aware sparse matrix–vector multiplication,” *J. Parallel Distrib. Comput.*, vol. 130, pp. 166–178, Aug. 2019.

W. D. Gropp, “Using node information to implement MPI Cartesian topologies,” in *Proc. 25th Eur. MPI Users’ Group Meeting*, Barcelona, Spain, Sept. 23–26, 2018, pp. 1–9.

Both zchunk and zlines are available at <https://github.com/oshkosh/bioio>.

MeshIO is available at <https://github.com/oshkosh/meshio>.

The improved implementation of MPI\_Cart\_create is part of baseenv and is available from [wgropp@illinois.edu](mailto:wgropp@illinois.edu).

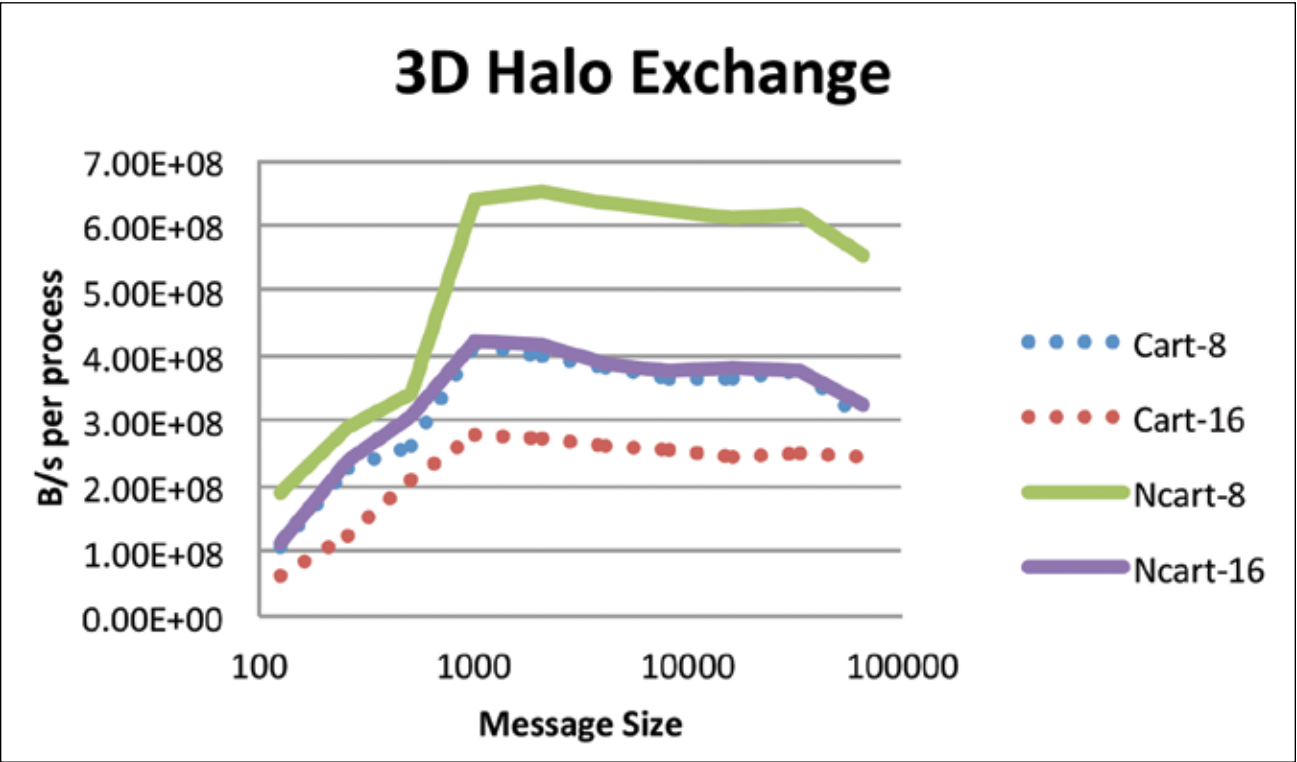


Figure 1: Communication performance for a 3D halo exchange on Blue Waters using the vendor-provided implementation of MPI\_Cart\_create (dotted lines) and the research group’s node-aware version (solid lines). The “-8” versions use an 8 x 8 x 8 process mesh (512 processes); the “-16” versions use a 16 x 16 x 16 process mesh (4,096 processes).



KALEIDOSCOPE: LIVE FORENSICS FOR LARGE-SCALE DATA CENTER STORAGE SYSTEMS

**Allocation:** Exploratory/50 Knh  
**PI:** Ravishankar Iyer<sup>1</sup>  
**Collaborators:** Saurabh Jha<sup>1</sup>, Shengkun Cui<sup>1</sup>, Tianyin Xu<sup>1</sup>, Jeremy Enos<sup>2</sup>, Mike Showerman<sup>2</sup>, Greg Bauer<sup>2</sup>, Mark Dalton<sup>3</sup>, Zbigniew Kalbarczyk<sup>1</sup>, Bill Kramer<sup>2</sup>

<sup>1</sup>University of Illinois at Urbana–Champaign  
<sup>2</sup>National Center for Supercomputing Applications  
<sup>3</sup>Cray Inc.

EXECUTIVE SUMMARY

The research team has developed Kaleidoscope, an innovative system that supports live forensics for application performance problems caused by either individual component failures or resource contention issues in large-scale distributed storage systems. The design of Kaleidoscope was driven by the team's study of I/O failures observed in a petascale storage system. Kaleidoscope is built on three key features: (1) using temporal and spatial differential observability for end-to-end performance monitoring of I/O requests, (2) modeling the health of storage components as a stochastic process using domain-guided functions that account for path redundancy and uncertainty in measurements, and (3) observing differences in reliability and per-

formance metrics among similar types of healthy and unhealthy components to attribute the most likely root causes. The research team deployed Kaleidoscope on the Cray® Sonexion®; an evaluation showed that Kaleidoscope can run live forensics at five-minute intervals and pinpoint the root causes of 95.8% of real-world performance issues, with negligible monitoring overhead.

RESEARCH CHALLENGE

Large-scale storage services are typically implemented on top of clusters of servers and disk arrays to provide high performance (e.g., load balancers and congestion control) as well as high availability (e.g., RAID, and active-active high-availability server pairs). Component failures and resource contention are chronic problems that lead to I/O timeouts and slowdown in such systems. State-of-the-art solutions focus on reliability failures and, hence, do not attempt to distinguish between resource contention and component failures in storage systems, as highlighted in Fig. 1. Knowing whether a problem is due to resource contention or component/node/subsystem failure is critical in effectively coordinating a recovery strategy.

A combination of component failures and contention issues significantly degrades application performance in production settings. This project uses a mixture of proactive monitoring and machine learning to jointly address the above issues. The team has incorporated the proposed techniques into an automated tool called Kaleidoscope. This tool has been demonstrated in live traffic on a production system to: (1) locate components such as data servers and RAID devices causing I/O bottlenecks such as I/O slowdown or timeouts, (2) differentiate between a reliability failure and a resource contention issue, and (3) quantify the negligible impact on system performance while delivering high precision and recall.

METHODS & CODES

The research team used two years of production data in excess of one terabyte from Blue Waters including system-generated storage error logs and store read/write latency logs collected intelligently using Kaleidoscope. Kaleidoscope is a generic framework

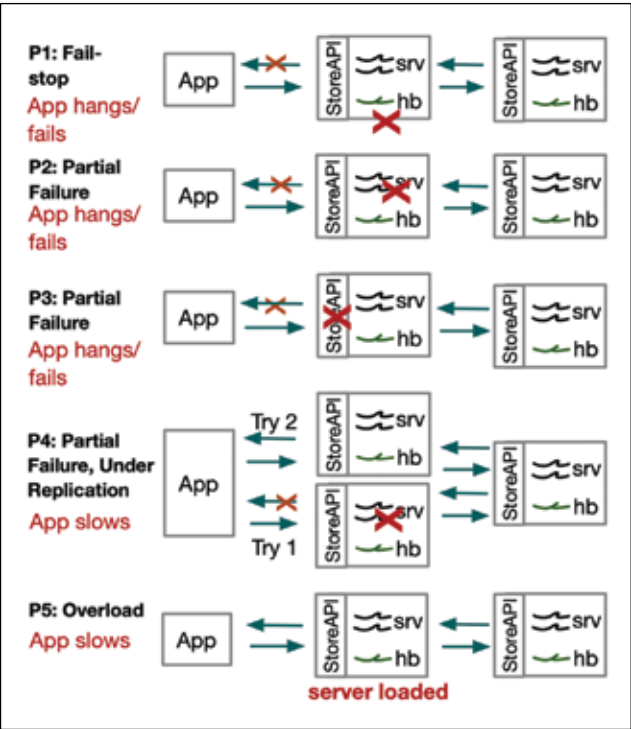


Figure 1: Common patterns of I/O failures. Notation: “hb” is heartbeat process, “srv” is service process; each box represents the storage components (e.g., data servers).

for supporting runtime detection and diagnosis of large-scale storage systems. The key components of the tools are:

- Proactive monitoring. Kaleidoscope monitors the end-to-end performance of a storage system using Store-Pings, a set of monitor primitives that covers all the storage operations involved in serving a client's I/O requests (e.g., creating, reading, writing, and deleting files). Store-Ping monitors are strategically placed to provide both spatial and temporal differential observability in real time (steps 1–3 in Fig. 2).
- Modeling and inferring component health. The health of a component in a storage system such as a metadata server or a RAID device is modeled as a stochastic process that accounts for uncertainty (owing to performance variability and asynchrony) as well as nondeterminism in distributed storage systems. The research group built a system model by using the factor graph (FG) formalization, which infers component health by ingesting the monitoring data collected by Store-Pings. The inference on the model allows Kaleidoscope to localize unhealthy components in near real time (step 4 in Fig. 2).
- Methods to determine the cause of I/O failures. A set of statistical methods (including a local outlier factor algorithm run using data on server load, disk load, and disk bandwidth utilization) and clustering of storage system error logs are used to distinguish between component failures and resource overloads. The statistical methods are based on a comparison of reliability and performance metrics (such as the number of active processes on a data server) as they are collected for healthy and unhealthy components. Note that the distinction between healthy and unhealthy components is provided by the FG-based model (Step 5 in Fig. 2).

RESULTS & IMPACT

**Deployment.** Kaleidoscope has been deployed on Blue Waters' Cray® Sonexion®, a 36-petabyte production system that employs the Lustre file system. Lustre is used by more than 70% of the top 100 supercomputers and is offered by cloud service vendors such as Amazon and Azure. Its design resembles that of many other object-based POSIX storage systems such as the IBM GPFS, Bee-GFS, Ceph, and GlusterFS. The team measured the overhead introduced by Store-Ping monitors on the production system and found the overhead to be less than 0.01% on the peak I/O throughput of the Cray® Sonexion®.

**Forensic effectiveness.** The evaluation was based on 843 production issues identified and resolved by Blue Waters system managers in a two-year period as the ground truth. Overall, Kaleidoscope correctly localized the component failures and resource overloads for 99.3% of the cases. In addition, Kaleidoscope accurately identified the likely root cause for 95.8% of the cases, i.e., disambiguation between resource contention and component failures.

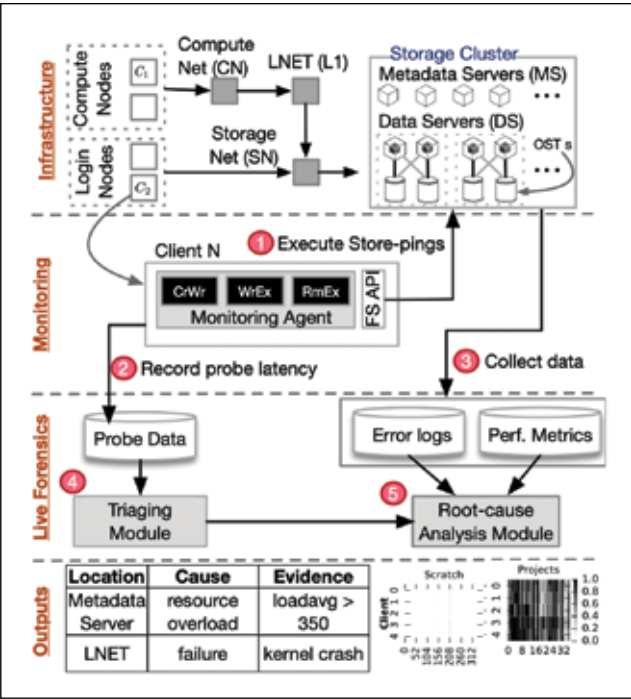


Figure 2: An overview of Kaleidoscope. Kaleidoscope consists of three components for monitoring failure localization and failure diagnosis (marked in gray).

WHY BLUE WATERS

Blue Waters is one of the few open-science capacity systems that provides a testbed for scaling computations to tens or hundreds of thousands of cores on CPUs and GPUs. It also enables the study of failures and performance degradations of applications in production petascale systems, thereby allowing researchers to understand the performance-fault-tolerance continuum in high-performance computing systems.

PUBLICATIONS & DATA SETS

- S. Jha *et al.*, “Measuring congestion in high-performance datacenter interconnects,” to be presented at *17th USENIX Symp. Networked Systems Design Implement.*, Santa Clara, CA, U.S.A. Feb. 25–27, 2020.
- S. Jha *et al.*, “A study of network congestion in two supercomputing high-speed interconnects,” presented at *26th IEEE Annual Symp. High-Performance Interconnects (HOTI)*, Santa Clara, CA, U.S.A. Aug. 14–16, 2019.
- S. Jha *et al.*, “Live forensics for distributed storage systems,” submitted, 2019, arXiv: <https://arxiv.org/abs/1907.10203v1>.



# EXTENSIBLE AND SCALABLE ADAPTIVE SAMPLING TO FOLD PROTEINS ON SUPERCOMPUTERS

**Allocation:** NSF PRAC/5,200 Knh  
**PI:** Shantenu Jha<sup>1</sup>  
**Collaborators:** Cecilia Clementi<sup>2</sup>, Eugen Hruska<sup>2</sup>

<sup>1</sup>Rutgers, The State University of New Jersey  
<sup>2</sup>Rice University

## EXECUTIVE SUMMARY

The Extensible Toolkit for Advanced Sampling and anaLYsis (ExTASY) is a software toolkit to effectively simulate protein folding on supercomputers. The use of adaptive sampling of proteins achieves a shorter time-to-solution than brute-force molecular dynamics (MD) but requires a more complex workflow. The research team has shown that ExTASY can effectively execute adaptive sampling and produce accurate simulations of protein folding and protein dynamics. The ExTASY package allows researchers to utilize and investigate different sampling strategies with great flexibility. The effective and scalable execution on supercomputers is ensured by RADICAL–Cybertools, a suite of Python modules that enables interoperability across high-performance computing machines.

## RESEARCH CHALLENGE

The previous version of ExTASY was developed to reduce the complexity of adaptive sampling and was used by the research team to show in [1] that ExTASY can scale complex workflows on supercomputers. The next step for ExTASY was to demonstrate an end-to-end execution of adaptive sampling for reference proteins. By comparison with reference results, the research team confirmed that adaptive sampling delivers accurate results for protein folding and protein dynamics and that the team could investigate the achieved speed-up. For three proteins—Chignolin, BBA, and Villin, with 10, 28, and 35 residues, respectively—the protein folding and protein dynamics are well understood and are good model reference proteins to test the performance of ExTASY.

Figure 1: Adaptive sampling requires an iterative workflow with the individual steps requiring different parallelization. Step 1 comprises long, parallel MD simulations. In contrast, steps 2–4 require only a single node. To fold a protein, these steps have to be repeated hundreds of times. A robust and effective workflow management toolkit is essential to enable more researchers to execute adaptive sampling.

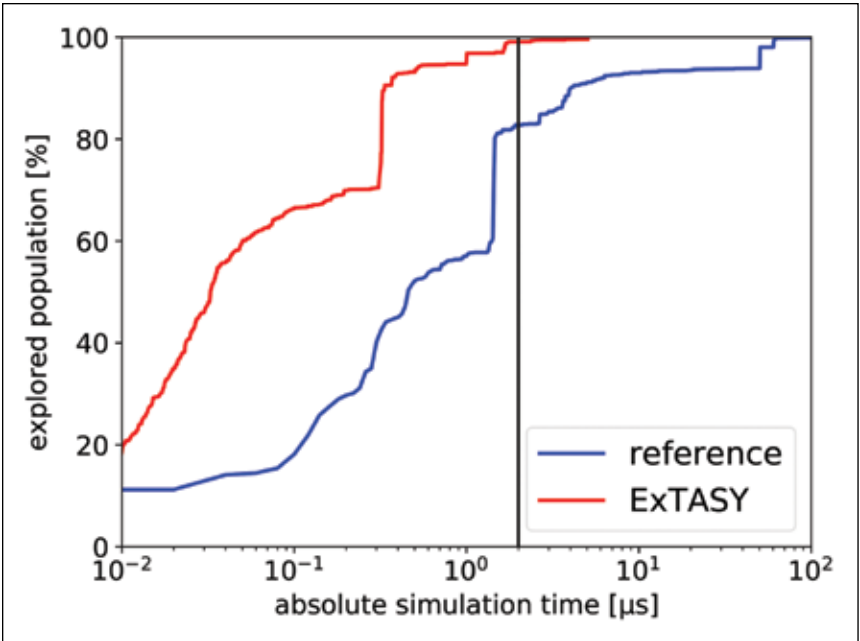
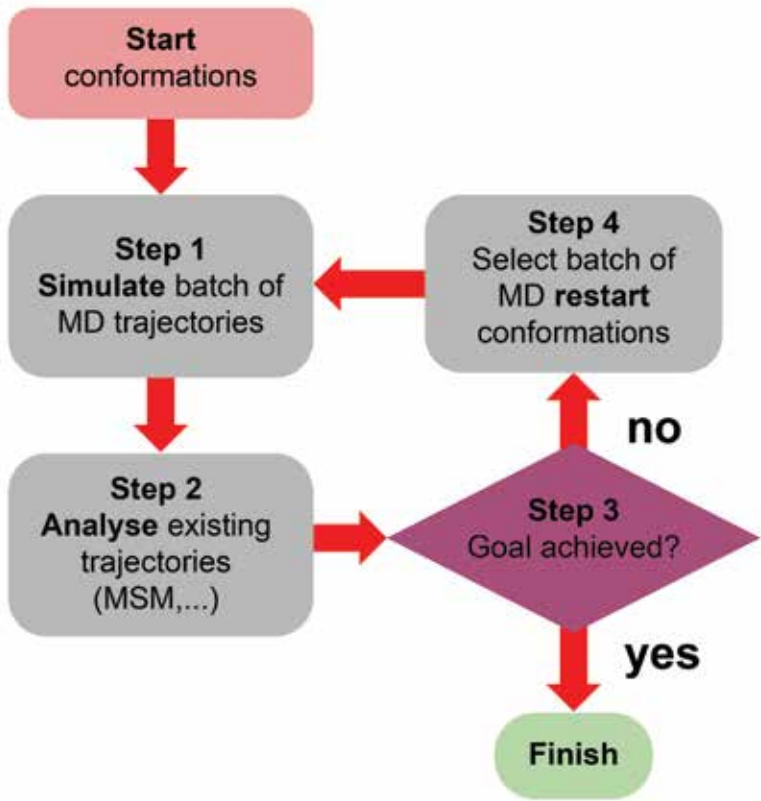


Figure 2: The brute-force MD simulation (blue) requires an order of magnitude longer time-to-solution than the adaptive sampling solution utilizing ExTASY (red). The more effective exploration strategy requires determining the restarting points adaptively during runtime. The ExTASY framework allows effective execution of these adaptive sampling strategies.

## METHODS & CODES

The ExTASY framework uses RADICAL–Cybertools or, specifically, the Ensemble Toolkit and RADICAL–Pilot, which ensures scalability, extensibility, and ease of deployment on high-performance computing platforms. The building block capabilities of RADICAL–Cybertools greatly increase the flexibility of ExTASY to utilize different sampling strategies in an easy fashion. Additionally, RADICAL–Cybertools enables execution of a complex workflow without explicit resource management, which also ensures the maintainability of the ExTASY workflow.

The adaptive sampling of proteins is an iterative process where MD and analysis steps alternate (Fig. 1). In general, adaptive sampling strategies pick optimal restarting coordinates for the next iteration of MD. This enables more effective use of computational resources in undersampled areas. In this project, the researchers utilized two different adaptive sampling strategies, *cmacro* and *cmicro*, to first effectively fold the protein and then reach accurate protein dynamics. Both strategies generate Markov state models from all generated MD trajectories in step 2. In step 4, the *cmacro* strategy picks Markov states to effectively cross transition barriers. Once the folded state is found, the *cmicro* strategy can be used to increase the accuracy of protein dynamics. Further comparison of different adaptive sampling strategies is discussed in [2]. The ExTASY code is open source and is provided at <https://github.com/ClementiGroup/ExTASY>.

## RESULTS & IMPACT

The new version of the ExTASY workflow folds proteins with a shorter time-to-solution than brute-force MD. Fig. 2 shows that adaptive sampling utilizing ExTASY is about one order of magnitude faster than brute-force MD. Additional results confirm-

ing the accuracy of protein folding and protein dynamics may be found in [3]. By utilizing different adaptive sampling strategies than in previous versions of ExTASY, the research team has shown that this workflow can be easily adapted to different exploration strategies.

## WHY BLUE WATERS

Protein folding simulations require large numbers of GPU node-hours despite the speed-up achieved by ExTASY. Blue Waters is essential to deliver these computational resources. The investigated proteins are relatively small and undergo fast folding; larger proteins would require even larger computational resources.

## PUBLICATIONS & DATA SETS

E. Hruska, V. Balasubramanian, J. R. Ossyry, S. Jha, and C. Clementi, “Extensible and scalable adaptive sampling on supercomputers,” 2019, arXiv: 1907.06954.

E. Hruska, J. R. Abella, F. Nüske, L. E. Kavraki, and C. Clementi, “Quantitative comparison of adaptive sampling methods for protein dynamics,” *J. Chem. Phys.*, vol. 149, no. 24, p. 244119, 2018, doi: 10.1063/1.5053582.

V. Balasubramanian *et al.*, “ExTASY: Scalable and flexible coupling of MD simulations and advanced sampling techniques,” in *Proc. 2016 IEEE 12th Int. Conf. e-Science*, Baltimore, MD, U.S.A., Oct. 23–27, 2016, pp. 361–370.



# IMPROVED SCALABILITY THROUGH NODE-AWARE COMMUNICATORS

**Allocation:** Exploratory/50 Knh  
**PI:** Luke Olson<sup>1</sup>  
**Co-PI:** Amanda Bienz<sup>1</sup>  
**Collaborator:** William Gropp<sup>1</sup>

<sup>1</sup>University of Illinois at Urbana–Champaign

## EXECUTIVE SUMMARY

Sparse matrix operations abound in numerical simulations and represent significant costs at scale. As researchers anticipate future machine networks and compute units, these communication-bound operations will continue to incur significant cost. The focus of this work is on reducing communication at scale, particularly in settings where machine layout and multiple compute units can be exploited. Blue Waters is an ideal setting for developing these methods since the research team can expose node-level (as well as socket-level) parallelism. The work has identified new methods for organizing communication to reduce the overall time-to-solution.

## RESEARCH CHALLENGE

Sparse matrix operations such as sparse matrix–vector multiplication and sparse matrix–matrix multiplication are key kernels in many iteration and preconditioning techniques. Yet, these operations incur a significant communication penalty in situations where the matrix patterns are highly unstructured. The goal of this work is to identify metrics for predicting communication overhead and to construct communication routines that can significantly reduce cost by utilizing certain aspects of the machine such as the node and socket layout.

## METHODS & CODES

This work has led to the development of scalable solvers within the RAPtor package [1] and has helped to shape the development of a node-aware library [2] that can be used easily by application codes from a range of scientific disciplines. Both software packages rely on node-aware MPI communication, which reroutes standard internode communication to reduce the total cost of sending data through the network.

There are two variations of this work, both aggregating data at the node level to reduce the number and size of messages being sent through the interconnect. Two-step node-aware communication consists of each process gathering all data to be sent to a single node and sending these data directly to the corresponding process on the destination node. The receiving process then distributes these data to all on-node processes that need it. Alternatively, three-step communication adds another layer of aggregation, agglomerating all data to be sent between two nodes on a single process on the node of origin. These data are then sent as a single message between nodes, after which it is redistributed locally to any process on the receiving node that needs it.

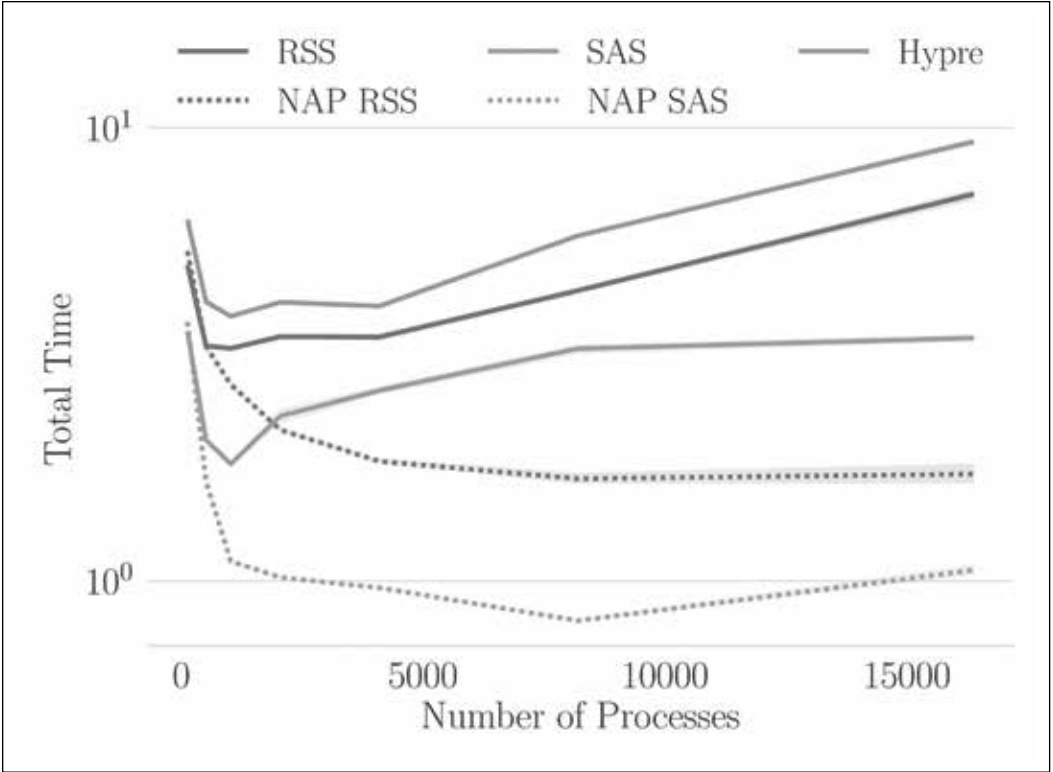


Figure 2: Total AMG times for both the Ruge–Stüben solver (RAS) and smoothed aggregation solver (SAS) for a Grad-Div problem on Blue Waters, along with the node-aware (NAP) implementation.

## RESULTS & IMPACT

The optimal method of communication varies with communication pattern as well as network topology. Therefore, optimizing communication costs throughout the algebraic multigrid (AMG) requires different strategies based on the system being solved. There are several factors that impact communication costs, including the sparsity pattern on a particular level of the AMG hierarchy, the amount of data being communicated on each level, and the number of active processes. Fig. 1 displays the cost of each strategy for matrix–matrix multiplication on each level of an AMG hierarchy, a dominating kernel in the AMG setup phase. The results show that standard communication outperforms node-aware on fine levels where communication is regular and structured, while node-aware strategies are optimal on coarse levels. Similarly, while three-step communication is often optimal, two-step node-aware communication often outperforms on the first few coarse levels when messages are large.

One goal of this work on Blue Waters is to develop a method to automatically select the communication strategy. RAPtor includes a predictive performance model that identifies the optimal communication strategy for each operation. Fig. 2 shows the cost of standard AMG versus node-aware AMG where performance models select the optimal communication strategy for each operation. Communication costs can vary widely for different computational kernels and machine settings—this work automates the communication strategy leading to significant speedups and reduced time-to-solution.

## WHY BLUE WATERS

Blue Waters was central to this work, providing a large scale of resources to test the models and routines developed by the research team. The node-level layout of Blue Waters provided the initial inspiration for the algorithms the team developed, and the consistency of the compute environment was a central component in reproducible testing at scale.

## PUBLICATIONS & DATA SETS

A. Bienz, W. D. Gropp, L. N. Olson, “Node aware sparse matrix–vector multiplication,” *J. Parallel Distrib. Comput.*, vol. 130, pp. 166–178, 2019, doi: 10.1016/j.jpdc.2019.03.016.

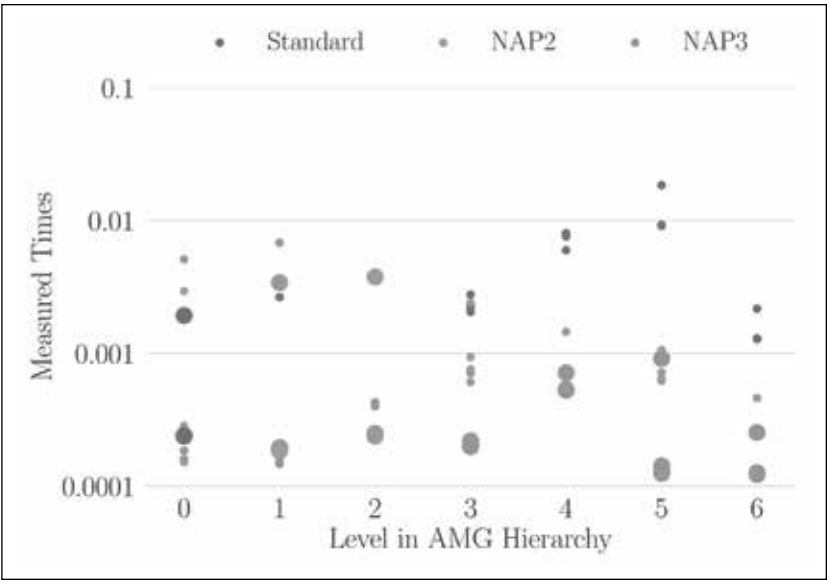


Figure 1: The matrix–matrix multiplication cost in each level of the AMG (algebraic multigrid) hierarchy using standard, two-step (NAP2), and three-step (NAP3) node-aware communication routines.



SCALABLE LINE AND PLANE SOLVERS

**Allocation:** Illinois/75 Knh  
**PI:** Luke Olson<sup>1</sup>  
**Co-PI:** Amanda Bienz<sup>1</sup>  
**Collaborator:** Andrew Reisner<sup>1</sup>

<sup>1</sup>University of Illinois at Urbana–Champaign

EXECUTIVE SUMMARY

Structured elliptic solvers play an important role in a range of applications, from plasmas to solid mechanics. However, their scalability is limited owing to the global nature of the problem. Multigrid methods have been effective at solving this class of problems; however, structured multigrid solvers rely on sweeps over lines and planes in the mesh in order to yield an effective method, which severely limits scalability. Through this work, the research team has developed scalable kernels for line and plane relaxation methods.

RESEARCH CHALLENGE

Sparse matrix problems arising from elliptic partial differential equations present a significant computational challenge in many applications at scale. Iterative approaches such as multigrid preconditioned conjugate gradient methods have proven to be cost effective. Yet, in situations with high mesh anisotropy or variable problem coefficients, multigrid solvers require the use of more robust components. Point relaxation techniques such as the Jacobi method are insufficient, thus requiring the use of line and plane forms of these methods.

METHODS & CODES

Line relaxation (and in 3D, plane relaxation) plays an important role in preconditioning structured elliptic problems. Fig. 1 highlights the importance of line relaxation in comparison to a standard weighted Jacobi (pointwise) relaxation solver in multigrid for the case of an annulus with moderate 10:1 stretching of the grid. There is a significant reduction in total iterations, yet each pass of line relaxation requires a distributed tridiagonal solver.

The key bottleneck in a tridiagonal (or banded solver) is the limited work per processor. Fig. 2 highlights the  $O(p)$  effect, with  $p$  cores increasing in one dimension. A straightforward implementation of the tridiagonal solver results in a linear relationship. This work focuses on a multilevel form of the problem where a new tridiagonal problem is formed in the communication overlap, resulting in a  $O(\log p)$  dependence and significantly reducing cost.

The current approach centers on two interrelated aspects to further advance the solver. First is the development of a scalable and accelerator-aware halo exchange library, called *Tausch* [1], which can be used in a variety of structured codes. Second is the extension of the structured solver *Cedar* [2] to accelerators, such as the XK portion of Blue Waters.

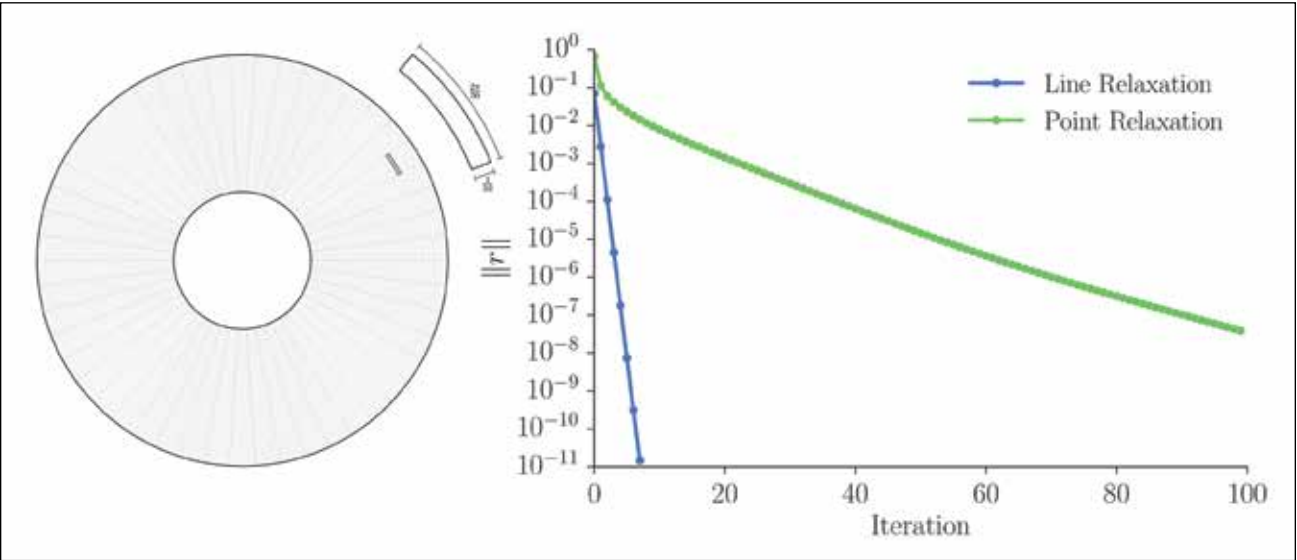


Figure 1: Point and line relaxation convergence for a Poisson problem on an annulus.

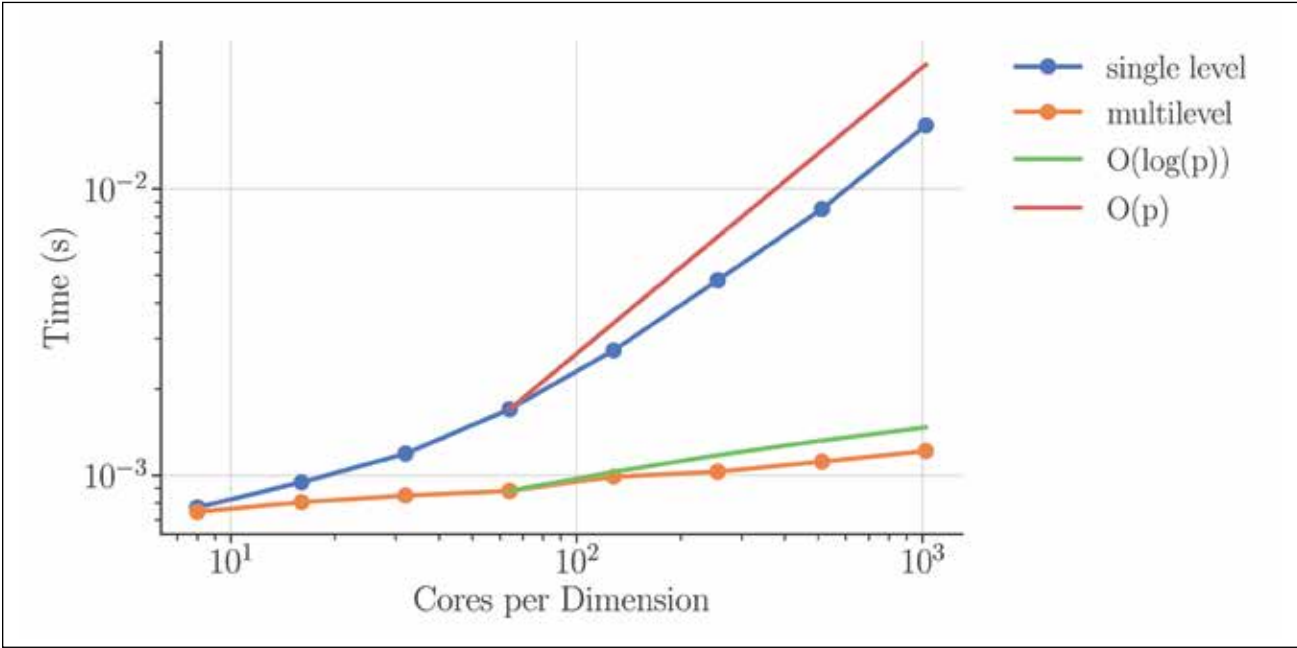


Figure 2: Scalability of sequential line relaxation versus multilevel line relaxation as the core count ( $p$ ) increases.

RESULTS & IMPACT

This work has directly contributed to the algorithms, testing, and development of the *Tausch* [1], *Cedar* [2], and *RAPtor* [3] software packages. Each of these targets general use in a wide range of scientific applications.

WHY BLUE WATERS

Blue Waters has played a key role in testing structured solvers at scale, for example in [4]. Access to large core counts and consistent testing has been instrumental in developing scalable algorithms and accurate performance models that form the core of this work.



DETECTION OF SILENT DATA CORRUPTIONS USING MACHINE LEARNING

Allocation: Exploratory/40 Knh  
PI: Marc Snir<sup>1</sup>  
Collaborator: Franck Cappello<sup>2</sup>

<sup>1</sup>University of Illinois at Urbana–Champaign  
<sup>2</sup>Argonne National Laboratory

EXECUTIVE SUMMARY

Future supercomputers are expected to encounter more frequent bit-flips as the number of devices increases and their size shrinks; furthermore, one could reduce power consumption by tolerating more frequent errors. Some errors may escape the notice of the error detection mechanisms provided in hardware. To handle those, it may be necessary to detect errors in software. One promising approach is to periodically test the state of a long-running computation and identify “anomalous” states indicating that a bit-flip occurred.

The research team is using a convolutional neural network (CNN) as a detector. The CNN is trained with multiple examples of correct and incorrect computation states, next used as a classifier. Using this method, the team achieved a high detection rate, with an acceptable overhead. The method is practical and could be used on future systems should their rate of undetected bit-flips require software detection.

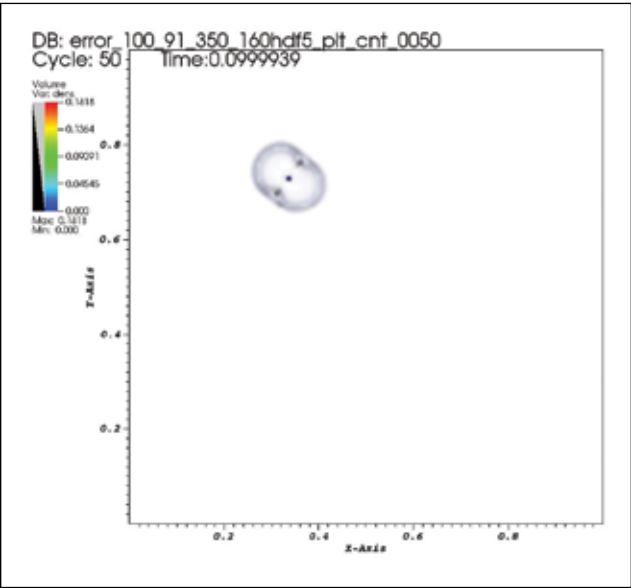


Figure 1: Error propagation after 50 iterations of the Sod application.

RESEARCH CHALLENGE

Future supercomputers are expected to encounter more frequent hardware errors, owing to the increase in the number of devices and the decrease in their size. Furthermore, one can decrease power consumption of supercomputers by tolerating more frequent errors. Some of these errors may not be detected and could corrupt the computation output. The research group’s goal is to detect silent data corruptions by periodically running tests on the state of an ongoing computation.

METHODS & CODES

A single bit-flip in one variable of a large simulation can result in an error that propagates through the entire data set over time. Figs. 1 and 2 illustrate error propagation for two different iterative simulations. The research team’s hypothesis is that machine learning can be used to detect such error patterns. The problem is that these error patterns are superimposed onto the correct computation state.

The team is treating the correct simulation as “noise” and is looking for the signal of a propagating error added to the noise. This is done by training a convolutional neural network (CNN) with multiple examples of correct and faulty simulations and then using the network as a detector. The training needs to be repeated once for each code, but the training procedure does not depend on the code.

RESULTS & IMPACT

Using this technique, the researchers have achieved a much higher detection rate than was previously achieved [1]. Furthermore, unlike previous methods, errors can be detected multiple iterations after they occurred so that the error detector need not be run at each iteration: The recall rate is around 90% and the overhead for running the detector could be as low as 1%. Follow-up work, to be published, aims at better understanding which errors do lead to a noticeable corruption of the final result, as only those need to be detected; and to extend these methods to work on 3D data sets. This research indicates that it will be possible to use future supercomputers efficiently, even if undetected bit-flips become common.

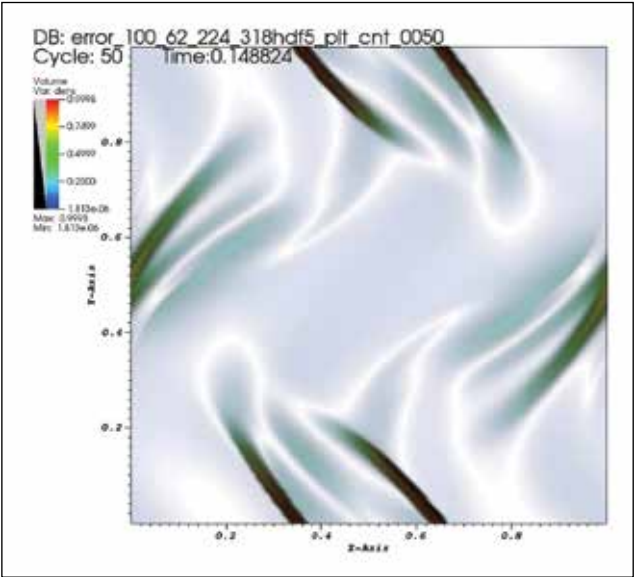


Figure 2: Error propagation after 50 iterations of the Orzag–Tang application.

WHY BLUE WATERS

GPUs are routinely used to accelerate the training of deep neural networks; such training is time consuming. A system with a large number of GPUs, such as Blue Waters, enables the research team to explore various alternatives at a much faster clip.



PUSHING THE BOUNDARIES OF LARGE-SCALE TENSOR COMPUTATIONS

Allocation: Illinois/20 Knh  
PI: Edgar Solomonik<sup>1</sup>

<sup>1</sup>University of Illinois at Urbana–Champaign

EXECUTIVE SUMMARY

This project seeks to develop new parallel algorithms and scalable productive software for matrix and tensor computations. Over the past year, the research team has made significant advances in software infrastructure of the Cyclops library for tensor computations. This library provides a productive algebraic programming interface in Python and C++ that performs numerical and combinatorial operations in a data-distributed manner. The team has developed support for hypersparse matrix representations, automatic optimization of contraction order, parallel tensor-times-tensor-product kernels, and has significantly extended capabilities at the Python level. The team currently is benchmarking tensor completion algorithms on Blue Waters using the new Python interface layer and has obtained preliminary benchmarking results for a new Cyclops application in genomic analysis. Separately, the researchers have developed a new practical communication-avoiding parallel algorithm for QR factorization, evaluating its performance via large-scale runs on Blue Waters.

RESEARCH CHALLENGE

Tensor computations are a growing area in computational science, with applications in quantum chemistry and physics, quantum circuit simulation, machine learning, optimization, and numerical PDEs (partial differential equations). They also push the boundaries of numerical linear algebra technologies, requiring

algebra on large, extremely sparse, and in some cases unstructured matrices, as well as redistribution of their data. These applications create a demand for better algorithms and software for distributed-memory tensor computations.

METHODS & CODES

The research group led the development of Cyclops, a distributed-memory library for tensor computations, which is likely the most widely used distributed tensor algebra library. Cyclops uses MPI, OpenMP, CUDA, and HPTT (High-Performance Tensor Transpose). It interoperates with ScaLAPACK and makes use of advanced sparse matrix routines in MKL. Cyclops provides a much simpler and more intuitive interface for both that is almost entirely agnostic to sparsity. It also provides a wide range of further functionality, including sparse and dense tensor contractions and generalized elementwise operations that enable graph and combinatorial algorithms. Additionally, it provides an interface to Python and a wide variety of functionality in a style similar to the NumPy library of mathematical functions.

The group has also developed standalone codes for parallel numerical linear algebra kernels. The researchers have developed and maintained suites of algorithms for tensor completion and tensor decomposition, which employ variants of alternating least squares, coordinate descent, and stochastic gradient descent methods, all parallelized using Cyclops. The team planned to benchmark the newly developed tensor decomposition and tensor completion kernels on Blue Waters in the summer of 2019.

Aside from tensor computations, the group also works on improving basic numerical linear algebra operations, such as computation of dense QR and eigenvalue decomposition of symmetric matrices. Edward Hutter, a Ph.D. student at the University of Illinois at Urbana–Champaign (Illinois), developed and maintains a new QR algorithm and library that achieves asymptotic improvements in communication efficiency by a new parallelization for the Cholesky–QR2 algorithm. One application of such large rectangular QR factorizations is in computing the Tucker decomposition of tensors.

RESULTS & IMPACT

The team’s work on Cyclops has a leading role in parallel infrastructure for numerical tensor algebra. Their library enables distributed-memory parallelism in leading quantum chemistry codes, including QChem, PySCE, and CC4S. The library has been used to execute methods for quantum chemistry at near one peta-

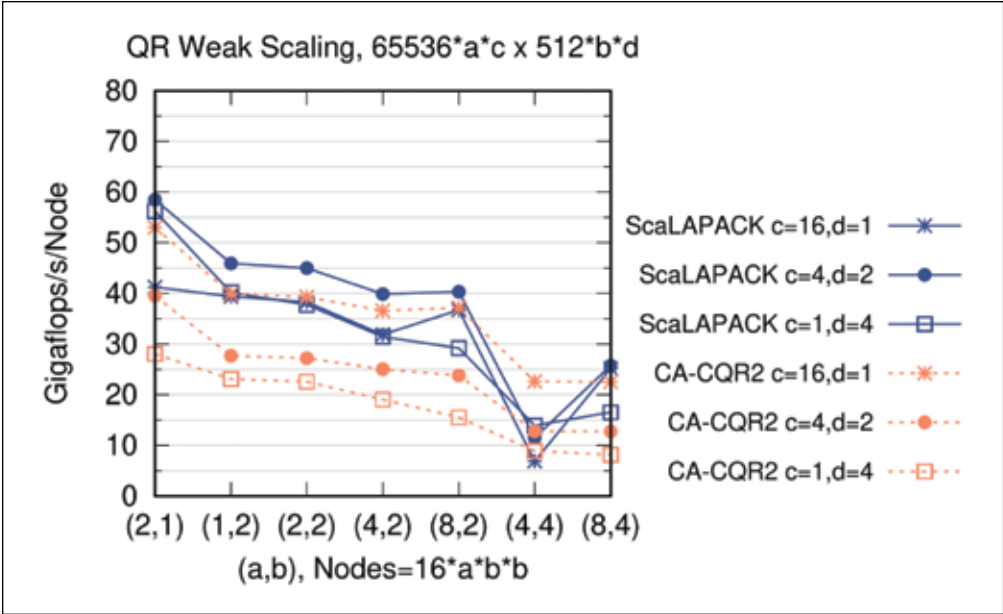


Figure 2: Weak scaling on Blue Waters of a new parallel QR factorization for various matrix sizes compared to ScaLAPACK.

flop/s and was used by a group of IBM and Lawrence Livermore National Laboratory researchers to perform a 49-qubit quantum circuit simulation, a central result in the field of quantum supremacy testing.

The research team is currently involved in collaborations with researchers in the physics department at Illinois and chemistry department at Caltech to develop the first massively parallel versions of tensor network codes. One-dimensional and two-dimensional tensor networks (namely DMRG and PEPS, two methods for computing properties such as ground state energy of quantum systems, which represent the quantum many-body state via 1-D and 2D tensor networks, respectively) provide highly accurate results for ground and excited state properties of highly correlated quantum systems. Researchers in the Illinois physics group (led by Bryan Clark) are currently prototyping initial versions of DMRG on Blue Waters.

A major impact on applications provided by Cyclops is the ability to rapidly develop massively parallel code via tensor algebra. A very recent example of this is the research group’s collaboration with a team of bioinformatics researchers who are interested in computing a Jaccard similarity matrix from the data of a large set of genomes, the first calculation of this type. Using Cyclops primitives for general sparse matrix multiplication, the team implemented the necessary kernels for logical (bitwise) operations within a week and are seeing good weak and strong scaling in small-scale initial tests on Blue Waters (Fig. 1). These results should pave the way for the bioinformatics team to apply for resources to perform a full-scale computation.

The team has obtained large-scale results for the new Cholesky–QR2 parallel algorithm, which achieves better parallel scaling trends than ScaLAPACK’s QR owing to needing less communication, but it is slightly behind in absolute performance because

it requires more FLOPS (Fig. 2). The team has also obtained results on more compute-intensive architectures, where the new Cholesky–QR2 algorithm outperforms ScaLAPACK significantly on large node-counts. The researchers are currently developing a GPU-accelerated version of the Cholesky–QR2 algorithm that they believe will be effective on the GPU nodes of Blue Waters as well as on future architectures. QR is one of the most widely used dense linear algebra primitives, so this work has the potential to impact many applications. These results will appear in the proceedings of the 2019 International Parallel and Distributed Processing Symposium.

WHY BLUE WATERS

As Illinois researchers, the team takes pride in using and showcasing results from the Blue Waters high-performance computing infrastructure. They also generally aim to test new algorithms and software on multiple supercomputing architectures, and Blue Waters is both unique and is itself diverse (providing large infrastructure for both GPU and pure-CPU runs). Using Blue Waters provided the team with a better understanding of the dependency of the performance of Cholesky–QR2 for architectures with different ratios of bandwidth and compute rate. In a number of application areas of interest, including quantum chemistry, quantum circuit simulation, tensor decomposition, and bioinformatics, the capability of calculations is often bounded by memory, making Blue Waters the architecture of choice.

PUBLICATIONS & DATA SETS

E. Hutter and E. Solomonik, “Communication-avoiding Cholesky–QR2 for rectangular matrices,” in *Proc. of the IEEE Int. Parallel and Distributed Processing Symp.*, Rio de Janeiro, Brazil, May 20–24, 2019.

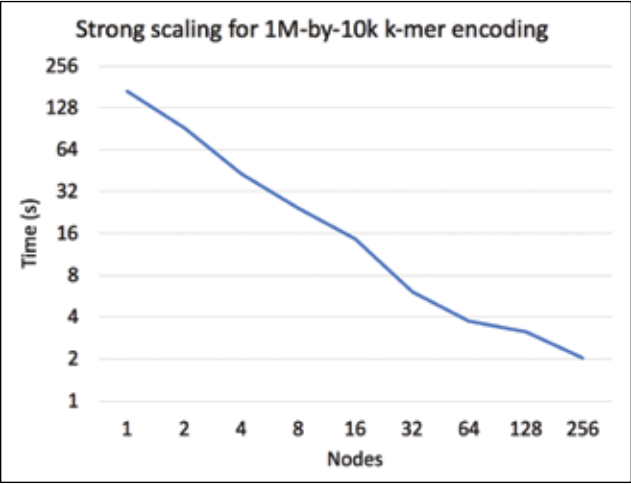


Figure 1: Strong scaling on Blue Waters of a computational biology application using specialized parallel sparse matrix multiplication with Cyclops.



# ALGORITHMS FOR LARGE-SCALE EVOLUTIONARY TREE CONSTRUCTION: IMPROVING SCALABILITY AND ACCURACY THROUGH DIVIDE-AND-CONQUER

**Allocation:** Illinois/150 Knh  
**PI:** Tandy Warnow<sup>1</sup>  
**Co-PIs:** William Gropp<sup>1</sup>, Erin Molloy<sup>1</sup>, Pranjal Vachaspati<sup>1</sup>

<sup>1</sup>University of Illinois at Urbana–Champaign

## EXECUTIVE SUMMARY

Evolutionary trees are used to advance the understanding of how life evolved on Earth, how species adapt to their environments, and to predict the structure and function of proteins. However, despite large numbers of whole genomes and increasing amounts of biomolecular sequence data available for use, the inference of a Tree of Life is beyond the reach of current methods, as even relatively small data sets can require many CPU years for analysis. This project aimed to develop new algorithms for large-scale evolutionary tree estimation, focusing on conditions with large numbers of species and/or whole genomes. Specific contributions of this work include new algorithmic strategies that greatly improve the scalability and accuracy of powerful statistical methods so they can be used to construct phylogenies on ultralarge data sets of importance to biologists.

## RESEARCH CHALLENGE

Biologists use evolutionary trees to improve the understanding of how species evolve and adapt to their environments, to predict protein structure and function, to explore the early origins of life and how humans moved across the globe, and the like. The advances in whole genome assembly have suggested that the accurate inference of a Tree of Life may be achievable. However, despite large numbers of whole genomes and increasing amounts of biomolecular sequence data available for use, the inference of a Tree of Life is much more challenging than was expected. The main issues impeding this are: (1) the inference of these evolutionary trees is computationally challenging, as the most accurate methods are based on attempts to solve hard optimization problems (such as maximum likelihood) and current optimization methods do not scale to large data sets with good accuracy;

and (2) standard approaches to phylogeny estimation, which make strong assumptions about the homogeneity of the statistical process underlying the molecular sequence data, do not have good accuracy in the presence of heterogeneity across genomes and across time, and yet substantial heterogeneity is now well established [1–3]. While some methods have been developed to enable phylogeny estimation in the presence of heterogeneity across genomes, these methods are computationally intensive, even on just moderately large data sets. Hence, current approaches to phylogeny estimation either do not provide good accuracy on large data sets or cannot even run on large data sets within reasonable timeframes.

## METHODS & CODES

The research team developed a divide-and-conquer framework that can be used with any method for constructing trees. The input is a set of species represented, for example, by sets of gene trees or DNA sequences. In the first step, the set of species is divided into disjoint subsets. Then, trees are constructed on each of the subsets using a selected method for phylogeny estimation. Finally, the trees are merged together using a distance matrix that is computed on the species. The key algorithmic innovation is the ability to merge trees together using the distance matrix. For this project, the team developed three such methods: NJMerge [4], TreeMerge [5], and INC [6]. The team used the divide-and-conquer strategies with the leading current methods for phylogeny estimation, including the maximum likelihood method RAxML [7] and ASTRAL [8,9], which estimate species trees by combining gene trees.

## RESULTS & IMPACT

The major contributions of this project are methods for large-scale evolutionary tree estimation that are capable of constructing highly accurate trees on data sets that are currently beyond the reach of existing methods. Most importantly, the divide-and-conquer strategies, TreeMerge and NJMerge, are enabling highly accurate species tree estimation using genome-scale data on tens

of thousands of species, where current methods fail to even run (owing to memory or time limitations) on these data sets. Using divide-and-conquer has reduced the running time of current methods on ultralarge data sets from multiple days to a few hours, thus dramatically advancing the capability of biologists to make biological discoveries. Specifically, this research has shown that TreeMerge and NJMerge can enable highly accurate species trees on tens of thousands of species from whole genome data sets without the need for supercomputers; this will transform research for evolutionary biologists. Furthermore, the algorithmic approach is quite general and should be useful for other problems where trees (even if not evolutionary trees) are constructed.

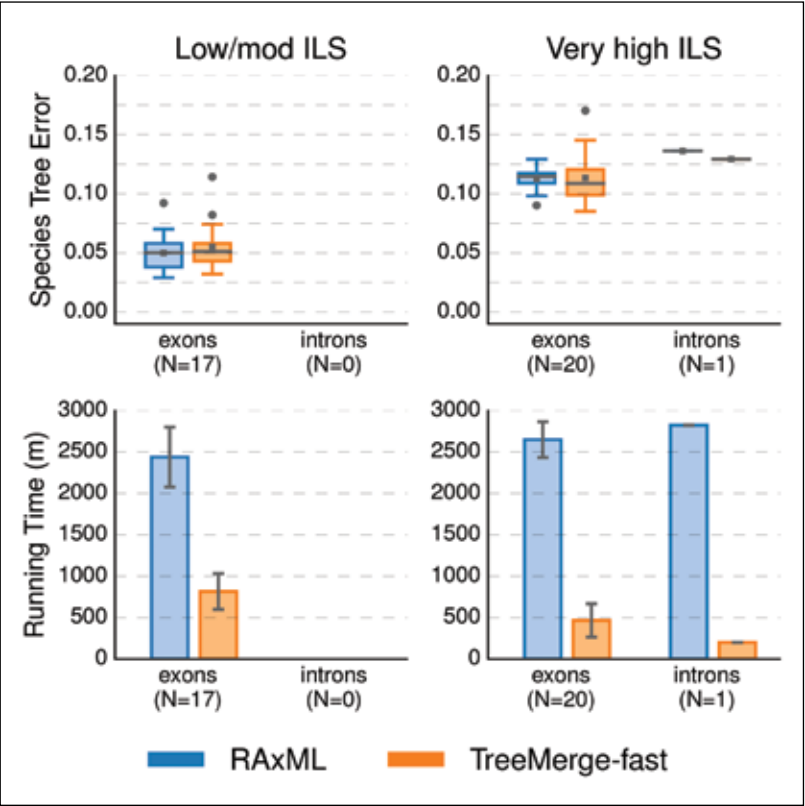
## WHY BLUE WATERS

The design of new algorithms for phylogeny estimation is an iterative process in which algorithmic strategies are explored and tested, and results are then used to improve the algorithm design. Since each analysis can be computationally intensive, this process requires large resources. The use of Blue Waters enabled this team to explore the design space effectively, and to produce new statistical and computational methods with outstanding accuracy and scalability to large and ultralarge data sets.

## PUBLICATIONS & DATA SETS

- E. K. Molloy and T. Warnow, “NJMerge: A generic technique for scaling phylogeny estimation methods and its application to species trees,” in *Comparative Genomics: RECOMB-CG 2018, Lecture Notes in Computer Science*, Magog–Orford, QC, Canada, Oct. 9–12, 2018, pp. 260–276.
- E. K. Molloy and T. Warnow, “TreeMerge: A new method for improving the scalability of species tree estimation methods,” *Bioinformatics*, vol. 35, no. 14, pp. i417–i426, Jul. 2019.
- T. Le, A. Sy, E. K. Molloy, Q. Zhang, S. Rao, and T. Warnow, “Using INC within divide-and-conquer phylogeny estimation,” in *Proc. Int. Conf. Algorithms for Comput. Biol.*, Berkeley, CA, U.S.A., May 28–30, 2019, pp. 167–178.

Figure 1: The result of using TreeMerge (a divide-and-conquer strategy) with the leading maximum likelihood method, RAxML, in species tree construction on 1,000 species with 1,000 genes when there is gene tree heterogeneity owing to incomplete lineage sorting [5]. The team explored results for two types of genes: exons and introns. The number of replicates for which RAxML returns a tree is given by N; when run by itself, RAxML cannot complete on some data sets within 48 hours on Blue Waters, but when run within the divide-and-conquer framework, it completes on all data sets. (“ILS” refers to incomplete lineage sorting.)





HPC DEVELOPMENT OF DEEP LEARNING MODELS IN SCIENTIFIC COMPUTING AND FINANCE

Allocation: Illinois/173.65 Knh  
PI: Justin Sirignano<sup>1</sup>  
Collaborators: Jonathan B. Freund<sup>1</sup>, Jonathan F. MacArt<sup>1</sup>

<sup>1</sup>University of Illinois at Urbana–Champaign

EXECUTIVE SUMMARY

This allocation was used to support several deep learning projects across a variety of applications. The three applications were: (1) scientific computing, (2) modeling high-frequency financial data, and (3) image recognition. The first application develops and evaluates deep learning closure models for large eddy simulation (LES) models of turbulent flows. The deep learning LES models are trained on direct numerical simulation (DNS) data from the Naiver–Stokes equation. The DNS data sets are generated using Blue Waters and the deep learning LES models are trained using Blue Waters’ GPU nodes. In the second application, deep learning models are trained and evaluated on massive high-frequency data sets. In the third application, which only used a small amount of the overall allocation, the asymptotics of deep learning models were analyzed on image data sets such as the Modified National Institute of Standards and Technology and Canadian Institute for Advanced Research 10 data sets.

RESEARCH CHALLENGE

Deep learning has revolutionized fields such as image, text, and speech recognition. Owing to this success, there is growing interest in applying deep learning to other fields in science, engineering, medicine, and finance. Blue Waters was used to develop and test deep learning methods and models for important applications in scientific computing and quantitative finance. The research team developed deep learning models for turbulence, which has been a longstanding challenge in computational science and engineering. The preliminary results demonstrate the ability of the machine learning-based models to predict turbulent flow energy spectrums more accurately than traditional turbulence models. In another project, the team developed a deep learning model for high-frequency financial data. Financial institutions have a strong interest in developing and using machine learning models for financial data and decisions.

METHODS & CODES

In the scientific computing project, the team developed a deep learning closure model for LES and generated DNS data sets. These DNS data sets were filtered and downsampled to yield appropriate data sets for training the deep learning LES model, which was parallelized across 24 GPU nodes. Synchronous gradient descent was used to train the deep learning model.

As part of this project, the research team also developed a new DNS/LES code, PyFlow, for training and simulation of the deep learning LES model. This code is GPU-accelerated and has potential to address modeling challenges over a wide range of flows. Similarly, the high-frequency financial data project used synchronous gradient descent (distributed across multiple compute nodes) to train deep learning models.

RESULTS & IMPACT

In the scientific computing project, the deep learning LES model is able to more accurately predict the filtered DNS data than traditional LES models. For example, in out-of-sample tests, the deep learning LES model more accurately reproduces the energy decay and energy spectrum as compared to traditional LES models such as Smagorinsky and Dynamic Smagorinsky. The use of deep learning and machine learning methods in scientific modeling has the potential to advance simulation and design in engineering.

WHY BLUE WATERS

Deep learning uses multilayer neural networks (deep neural networks) to build statistical models of data. This training of the deep learning model can be computationally intensive owing to both the large number of parameters and the large amounts of data. GPUs can be used to accelerate training of deep learning models. The researchers leveraged Blue Waters’ large amount of GPU resources to develop deep learning models for applications in engineering and finance. Blue Waters XE nodes were also extensively used to generate DNS data sets, which was extremely computationally expensive and required parallelization over thousands of cores. The Blue Waters technical staff provided invaluable help throughout the project, including solving a number of technical issues related to deep learning computational frameworks such as PyTorch.

PUBLICATIONS & DATA SETS

J. B. Freund, J. F. MacArt, and J. Sirignano, “DPM: A deep learning PDE augmentation method with application to large-eddy simulation,” submitted, 2019, arXiv: 1911.09145.  
J. Sirignano and K. Spiliopoulos, “Mean field analysis of deep neural networks,” submitted, 2019, arXiv: 1903.04440.

OPTIMIZATION OF A FIELD DATA PARALLEL OUTPUT LIBRARY

Allocation: Industry/3 Knh  
PI: David E. Taflin<sup>1</sup>

<sup>1</sup>Tecplot Inc.

EXECUTIVE SUMMARY

TecIO–MPI is Tecplot, Inc.’s parallel data output library. It is designed to output to Tecplot’s native file format the field data results of numerical simulations, such as those produced by computational fluid dynamics (CFD). Its initial implementation suffered from unacceptable parallel performance in some cases. In this work, the profiling software tools available on Blue Waters were used to identify and remedy the sources of this unacceptable performance. The researchers then ran test cases, scaling up to 912 nodes, to confirm acceptable performance and scaling. The results confirm order-of-magnitude reductions in output time and indicate that output times scale linearly with output file size.

RESEARCH CHALLENGE

Parallel simulation codes break large problems into smaller pieces and solve these smaller pieces simultaneously by using many CPU cores. At the end of the solution (and perhaps periodically during the solution), each core’s portion of the overall solution must be written to disk, perhaps all to a single file. This parallel file output must be fast enough to not hinder the overall solution progress significantly. Thus, any code that performs this output must be optimized for parallel performance. But parallel optimization requires tools that can not only examine the execution time of blocks of code but can also identify wait states caused by communication delays among the multiple processes that run together to produce the solution or output. Further, to optimize a massively parallel code requires testing that code on hardware representative of the hardware on which it is intended to run. Blue Waters provided the answer to these challenges.

METHODS & CODES

The research team ran the TecIO–MPI library with a test harness that loaded data previously serialized to disk and then called into TecIO–MPI to output the data. The test harness was a stand-in for the sort of simulation code with which TecIO–MPI is designed to run. This allowed repeated tests of the software without the overhead of an actual CFD-type solver. The researchers performed parallel profiling with Cray’s Performance and Analysis Tools, and data were written to a single file on Blue Waters’ LUSTRE file system. Hotspots that were identified were recoded to improve performance, and the software was retested to ensure the fixes were effective.

RESULTS & IMPACT

The performance of TecIO–MPI was improved to the point that CFD-solver authors should find it acceptable—roughly the cost of a single solver iteration—and measured wall-clock output times appear to scale linearly with output file size. The file format the library outputs enables Tecplot to visualize on typical desktop workstations—via contour flooding, streamline generation, slicing, etc.—much larger files than can typically be handled on such machines by loading from disk only the data required to produce the artifacts. The memory required in this postprocessing scales with the number of solution cells raised to the two-thirds power, such that its advantage grows exponentially with larger file sizes.

WHY BLUE WATERS

Blue Waters provided access to both hardware and software required for this project to be successful. Its ability to scale problems to many CPU cores, and its use of the popular LUSTRE file system, ensured that the software was being tested in a representative environment, such that performance improvements seen in this work would also be seen by the intended users of this software. Blue Waters’ Cray performance-measuring tools were indispensable in identifying performance bottlenecks and confirming their removal. And the technical support staff were very helpful with identifying the appropriate configuration of the compilers and profiling software required for this work.



# BIOLOGY, CHEMISTRY & HEALTH

- BIOINFORMATICS
- BIOPHYSICS
- HEALTH
- MEDICINE
- MOLECULAR
- NEUROSCIENCE
- PLANT SCIENCE/GENOMICS

- 238 *Resolving the Structure of Bacteriophage HK97 with Atomistic Resolution*
- 240 *A Nanopore System for Single-Molecule Protein Sequencing*
- 242 *Dynamic Interactions Between Lipid-Tethered DNA and Phospholipid Membranes*
- 244 *Influence Virulence and Transmissibility Through the Computational Microscope*
- 246 *How Blue Waters Is Aiding the Fight Against Sepsis*
- 248 *A Phylogenomic History of Protein Function and Dynamics*
- 250 *Predicting Drug-Induced Cardiac Arrhythmias Using Atomistic Simulations*
- 252 *mRNA Isoform Prediction*
- 254 *Petascale Integrative Approaches to Protein Structure Prediction*
- 256 *Discovery of Slow Kinetic Modes from Molecular Simulation Trajectories*
- 258 *Harnessing Viscous Streaming in Complex Active Systems: Minibots in Fluids*
- 260 *Molecular Dynamics Simulations of the HBV Capsid*
- 262 *Molecular Dynamics Simulations of the HBV Capsid as a Drug Target*
- 264 *Toward Predictive Computational Design of Precision Molecular Optoelectronics*
- 266 *Impact of Batch Effect and Study Design Biases on Identification of Genetic Risk Factors in Sequencing Data*
- 268 *Microscopic Identification of PIP<sub>2</sub> Binding Sites on a Ca<sup>2+</sup>-activated Cl<sup>-</sup> Channel*
- 270 *Paracellular Ion Transport*
- 272 *The TERRA Phenotyping Reference Platform: Open Data and Software for Precision Field Crop Measurement and Analysis*
- 274 *Quantum-Classical Path Integral Simulation of Proton Translocation in Biological Channels*
- 276 *A New Stabilized Fluid-Structure Interaction Method: Coupled System of Anisotropic Viscoelastic Model for Artery and Non-Newtonian Model for Blood*
- 278 *Atomic Scale Simulation of Amyloid Beta with Dismantling Peptide-Based Inhibitors*
- 280 *Transport Mechanism of POT Transporters: Employing Loosely Coupled Molecular Dynamics Simulations to Characterize Protein Structural Dynamics*
- 282 *Activation Mechanisms of the Mechanosensitive Channel of Large Conductance: Employing Loosely Coupled Molecular Dynamics Simulations to Characterize Protein Structural Dynamics*
- 284 *Molecular Mechanisms of Infection by Chlamydia*
- 286 *Calibrating the SimBioSys TumorScope for the Fight on Cancer: A Scenario Analysis Engine for Determining Optimal Therapy Choice*
- 288 *Investigating the Climate-Relevant Impacts of Chemical Complexity in Marine Aerosols*
- 290 *Molecular Dynamics Binding Free Energy Calculations Offer a Window to Understand Protein-Protein Binding Specificity*
- 292 *Molecular Basis of the Nitrate Transport Mechanism in Plants*
- 294 *Simulations Uncover the Mechanism of Serotonin Transport in the Brain*
- 296 *Elucidating the Ligand Selectivity and Activation Mechanisms of Cannabinoid Receptors*
- 298 *Full-Scale Biophysical Modeling of Hippocampal Networks During Spatial Navigation*
- 300 *Desmosomal Cadherins Beating Under Tension*
- 302 *Simulation of Viral Infection Propagation During Air Travel*
- 304 *MRI-Based Biomarkers Through High-Performance Computing*
- 306 *Mechanobiology: Using Blue Waters to Decipher the Physical Principles of Protein Mechanics*
- 308 *Atomistic Simulation of a Protocell*
- 310 *Modeling of a Zika Virus Envelope at Atomic Resolution*
- 312 *Multiscale Simulations of Complex Self-Assembling Biomolecules: Targeting HIV-1*
- 314 *Improving the Agreement of AMBER Simulation of Crystals of Nucleic Acid Bases with Experimental Data*
- 315 *Algorithms for Cancer Phylogenetics*
- 316 *Amphotericin-Driven Sterol Extraction: Probing the Mechanism*



# RESOLVING THE STRUCTURE OF BACTERIOPHAGE HK97 WITH ATOMISTIC RESOLUTION

**Allocation:** Illinois/450 Knh  
**PI:** Aleksei Aksimentiev<sup>1</sup>  
**Collaborators:** Kush Coshic<sup>1</sup>, Christopher M. Maffeo<sup>1</sup>, David Winogradoff<sup>2</sup>

<sup>1</sup>University of Illinois at Urbana–Champaign

## EXECUTIVE SUMMARY

Viruses are omnipresent, diverse, and potentially lethal biological systems that our bodies encounter every day in copious quantities without us taking notice. The essence of each virus is its genome, a biological program written using letters of the genetic alphabet. The genome is protected from the outside world by a protein shell (a viral capsid) until the conditions are met for viral invasion, at which point the genome is released from its protective shell into a host cell, initiating a new cycle of infection. The ongoing adaptation of viruses to antiviral drugs currently in use necessitates development of the next generation of drugs that target viruses from a physical standpoint. By carrying out large-scale molecular dynamics simulations on Blue Waters, the re-

search team has constructed the first atomically resolved model of a complete virus particle, including the 3D structure of its genome, an untapped resource of potential drug targets. The modeling approach demonstrated by this proof-of-principle simulation may be applied in the future to develop new antiviral drugs.

## RESEARCH CHALLENGE

Atomistic structures of protein capsids have been resolved for many viral species, but the structural organization of their genomes still remains largely unclear. One such viral species is the HK97 bacteriophage, for which experiments have characterized the packaging mechanism and resolved its protein capsid with atomistic resolution [1]. Previous cryogenic electron microsc-

py and Small-Angle X-Ray Scattering experiments [2] have not yet elucidated the precise organization of the genome in individual virus particles. Previous computational efforts have addressed the dynamic behavior of the capsid [3,4], matrix [1], or outer envelope, but the structural assignment of nucleic acids inside viral capsids has not been explored comprehensively. Using Blue Waters, the research team has met the challenge of reconstructing the 3D structure of the HK97 bacteriophage genome, a double-stranded DNA molecule containing 39,732 base pairs.

## METHODS & CODES

To obtain microscopically correct structures of DNA inside viral capsids, the researchers employed a multiscale approach whereby the results of computationally inexpensive coarse-grained (*i.e.*, with reduced representation) molecular dynamics (MD) simulations were used to set up initial conditions for fully atomistic all-atom simulations of a virus particle loaded with DNA. With four base pairs of DNA represented as one coarse-grained particle, the 39,732-base-pair-long genome was packaged into a grid-based implicit protein capsid through a narrow portal that exerted a physiological packaging force [5]. Starting from the final packaged conformation, the team obtained an atomistic model of the genome from a series of simulations gradually increasing in resolution. Subsequently, water and ions were added to mimic conditions at DNA densities typical of pressurized viruses such as HK97 [6], yielding a fully atomistic model of the HK97 genome submerged in explicit solvent. The DNA structure was then placed inside the all-atom capsid through a set of all-atom simulations carried out in the presence of a confining potential. The final atomistic system was comprised of roughly 27 million atoms. All MD simulations were carried out using NAMD [7].

## RESULTS & IMPACT

The research team validated the results of the coarse-grained simulations by comparing experimentally determined internal pressure inside a viral capsid [8] to the pressure exerted by the DNA genome on the confining potential. Small-Angle X-ray Scattering profiles generated from the atomic model of the genome were found to be in quantitative agreement with experimental data [2]. The team performed further simulations to characterize the effect of different packaging mechanisms and to determine the contribution of bending stress (attributed to confining a stiff polymer, DNA) to the internal pressure. For quantifying the geometry of the DNA inside the capsid, the scientists evaluated a toroidal order parameter that indicated a preferential organization of the packaged DNA about the axis along which the DNA was packaged into the capsid. Overall, the team found the outcome of the simulations matched experimental data extremely well, which validates the obtained structures of the HK97 genome. The simulation protocol developed through this project can now be applied to other viruses to predict their genomic structures, offering exciting avenues for designing new antiviral drugs.

## WHY BLUE WATERS

Explicit-solvent all-atom MD simulation is needed to examine the fine details of DNA–capsid interactions and to accurately characterize the surrounding ionic environment. Given the large system of approximately 27 million atoms, such MD simulations are computationally demanding. The Blue Waters petascale system is one of only a few supercomputers in the world with the computational power sufficient to carry out fully atomistic MD simulations of viral particles containing packaged genomes. The large number of GPU-accelerated XK nodes and fast Gemini interconnect of Blue Waters make it one of the best publicly available systems for performing large-scale MD simulations of virus particles.

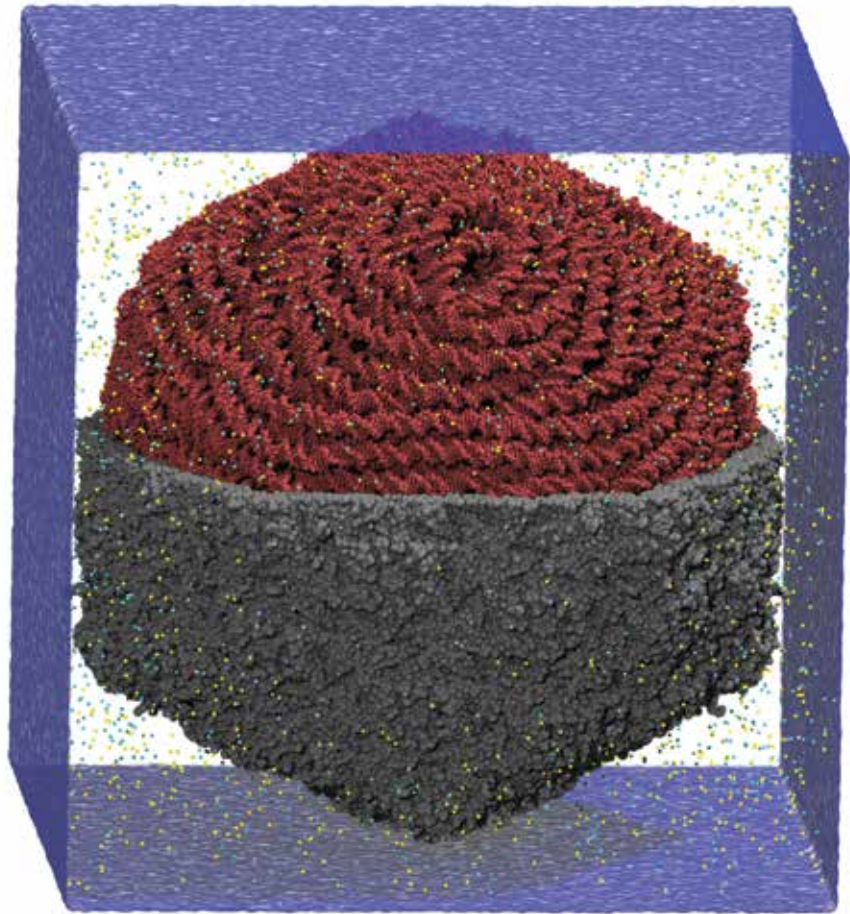


Figure 1: An all-atom model of a fully packaged HK97 bacteriophage. The DNA genome of the virus is shown in red, the icosahedral protein capsid in grey, water is shown as a blue semitransparent surface, and the ions are shown as colored spheres.



A NANOPORE SYSTEM FOR SINGLE-MOLECULE PROTEIN SEQUENCING

**Allocation:** Illinois/872.6 Knh  
**PI:** Aleksei Aksimentiev<sup>1</sup>  
**Collaborators:** Kumar Sarthak<sup>1</sup>, Hadjer Ouldali<sup>2</sup>, Fabien Piguet<sup>2</sup>, Abdelghani Oukhaled<sup>2</sup>, Tobias Ensslen<sup>3</sup>, Jan C. Behrends<sup>3</sup>

<sup>1</sup>University of Illinois at Urbana–Champaign  
<sup>2</sup>University of Cergy–Pontoise  
<sup>3</sup>University of Freiburg

EXECUTIVE SUMMARY

Single-molecule protein sequencing would provide unprecedented insights into cellular processes and the diseases that arise from protein malfunction. As compared to nucleic acid sequencing, there are no robust methods that can read the protein sequence without any challenges. In collaboration with the Oukhaled Lab at the University of Cergy–Pontoise and the Behrends Lab at the University of Freiburg, the research team is developing a nanopore system that can identify the type of amino acids that pass through the nanopore from a modulation of the nanopore ionic current—a crucial first step toward realizing nanopore protein sequencing. The researchers have already shown that a biological nanopore aerolysin can identify thirteen out of the 20 different amino acids. Using atomistic molecular dynamics (MD) simulations, the team aims to further engineer the aerolysin nanopore to enable identification of all 20 amino acids and to increase the identification fidelity by extending the time that amino acids spend within the nanopore. Successful outcome of this project will pave the way toward the first single-molecule protein sequencing method.

RESEARCH CHALLENGE

Cells express many thousands of proteins to perform a diverse set of complex tasks that are essential to the health of a living organism. A technology providing inexpensive yet sensitive and quantitative identification of the proteins in a cell will help numerous researchers elucidate biological processes, including the molecular origins of many diseases [1]. Existing protein binding and precipitation assays are relatively low-throughput, and their application is typically limited to detection of a few predetermined protein species [2,3]. However, protein sequencing methods aim to identify and quantify proteins and posttranslational modifications in a high-throughput manner without knowing *a priori* which proteins might be in a sample. The sequencing method has to be robust as even a single mutation in the amino acid sequence can lead to devastating diseases.

Sequencing a protein is not as straightforward as in nucleic acids because of the presence of 20 amino acids as compared to just four in DNA as well as the ability of protein to form complex 3D structures. Furthermore, protein function may sensitively depend on small chemical modifications to its residues, which

are difficult to characterize using existing experimental methods. Development of a method to sequence proteins as rapidly and accurately as scientists now sequence nucleic acids will revolutionize proteomics and provide the ultimate diagnostic tool.

METHODS & CODES

To determine the molecular mechanism of the blockade current modulation in aerolysin, the researchers simulated ion transport through aerolysin using the all-atom MD method. The aerolysin pore was placed in a DPhPC (1,2Diphytanoyl-sn-glycero-3-phosphocholine) lipid bilayer and solvated in a 1 M potassium chloride solution. The total number of atoms in the system was roughly 450,000.

The electrostatic potential map of the channel revealed the presence of a constant potential compartment in the middle section of the aerolysin stem, separated from the *cis* and *trans* entrances of the channel by electrostatic potential barriers. Building on this observation, the team simulated peptide transport through the aerolysin pore. In these simulations, the transport of peptides is facilitated by external forces that help the peptides to overcome local translocation barriers [4]. Averaging over all such forces at different positions in the pore yields the free energy landscape of peptide translocation through the pore [5]. All MD simulations were carried out using NAMD [6].

RESULTS & IMPACT

This ongoing project has already determined a rough free-energy landscape underlying peptide transport through an aerolysin nanopore. Repeat MD simulations have identified the location of the two major barriers to peptide transport, one each at the entrance and exit of the aerolysin nanopore. Importantly, the simulations have shown that the nanopore volume confined by the two traps acts as a single molecule trap that fully confines the translocating peptide. By combining the resulting all-atom MD trajectories with their steric exclusion model of ionic current blockades [7], the researchers have shown that the long residence of the peptides within the nanopore trap gives rise to well-defined and highly reproducible ionic current blockade values. Equipped with the knowledge of the microscopic mechanism underlying the ionic current blockades in aerolysin, the team is now in the position to suggest point mutations to aerolysin nanopores that would further increase the fidelity of amino acid identification.

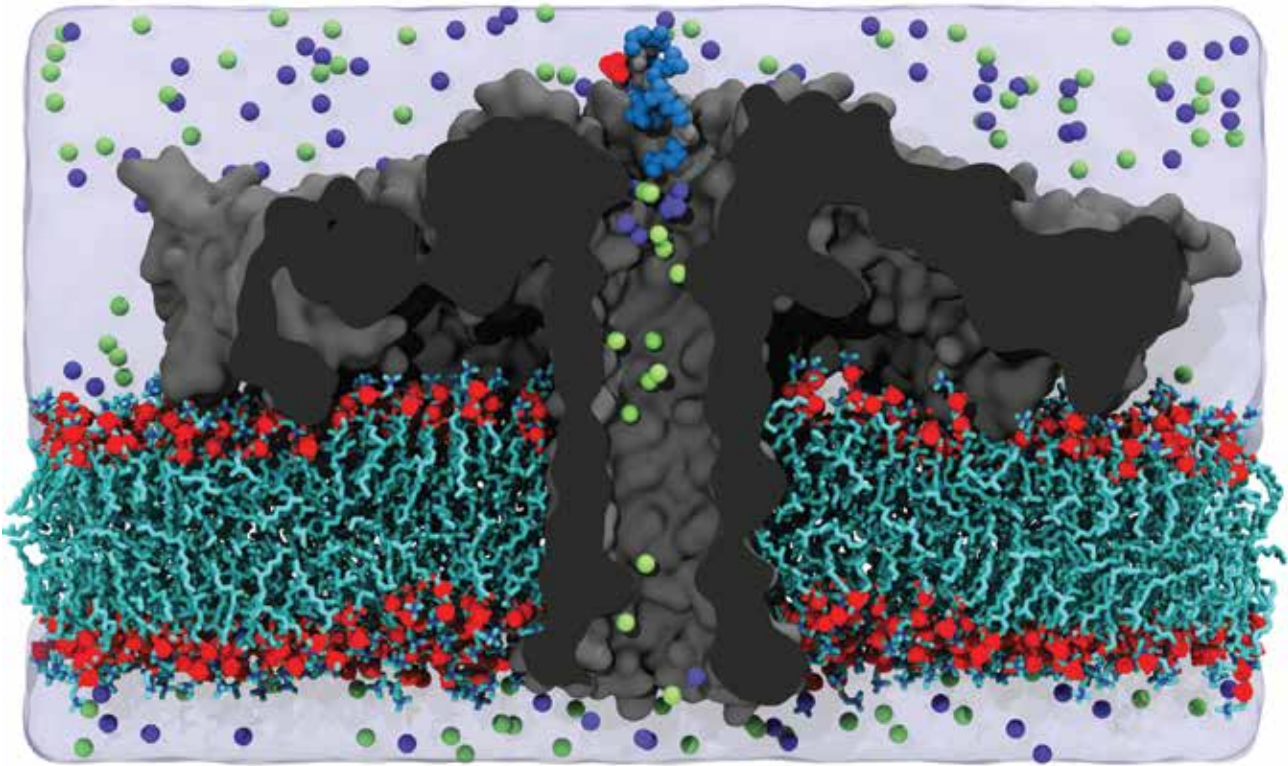


Figure 1: Nanopore sequencing system consisting of an aerolysin channel (cutaway molecular surface) embedded in a lipid membrane (cyan and red) and submerged in electrolyte (blue semitransparent surface, green and purple spheres). A single amino acid analyte (red) attached to a carrier peptide (blue) is captured by the aerolysin nanopore.

WHY BLUE WATERS

To get an accurate estimate of the ionic current and fast convergence of the free-energy landscape, the team needs explicit-solvent all-atom MD simulations that are run in parallel for exchange of structure information between different windows. Because of the long timescales and parallel MD simulations of a 0.5-million-atom system, this project is computationally demanding. The Blue Waters petascale system is one of a few supercomputers in the world with the computational power sufficient to carry out fully atomistic enhanced sampling simulations of peptide translocation through a nanopore. The large number of XK nodes on Blue Waters with graphics processing unit accelerators connected by the fast Gemini interconnect makes it one of the best publicly available systems for performing large-scale parallel MD simulations of protein systems.

PUBLICATIONS & DATA SETS

H. Ouldali *et al.*, “Electrical recognition of the twenty proteinogenic amino acids using an aerolysin nanopore,” *Nat. Biotech.*, vol. 38, pp. 176–181, 2020.



# DYNAMIC INTERACTIONS BETWEEN LIPID-TETHERED DNA AND PHOSPHOLIPID MEMBRANES

**Allocation:** Blue Waters Professor/240 Knh  
**PI:** Aleksei Aksimentiev<sup>1</sup>  
**Collaborators:** Patrick M. Arnott<sup>2</sup>, Himanshu Joshi<sup>1</sup>, Stefan Howorka<sup>2</sup>

<sup>1</sup>University of Illinois at Urbana–Champaign  
<sup>2</sup>University College London

## EXECUTIVE SUMMARY

Lipid-anchored DNA can attach functional cargo to bilayer membranes; this has applications in DNA nanotechnology, synthetic biology, and cell biology research. An understanding of DNA membrane-binding strength and structural dynamics at the nanoscale is required to optimize the DNA anchoring for these applications. Using simulations performed on the Blue Waters supercomputer, the research team elucidated how membrane binding of cholesterol-modified DNA depends on electrostatic and steric factors involving the size and charge of the lipid headgroup, duplexed or single-stranded DNA, and the buffer composition. Atomistic molecular dynamics (MD) simulations explain the experimental findings and elucidate the dynamic nature of anchored DNA, such as the mushroom-like conformation of single-stranded DNA hovering over the bilayer surface in contrast to a straight-up conformation of double-stranded DNA. The information from this study is expected to facilitate the development of biomimetic DNA versions of natural nanopores and cytoskeletons for research and nanobiotechnology.

## RESEARCH CHALLENGE

A lipid molecule attached to the end of a DNA strand can anchor the strand to a lipid bilayer membrane. This simple principle has been applied to design lipid-spanning DNA nanopores [1,2] for applications in the fields of nanobiotechnology [3], biosensing [4], and synthetic biology [5]. Controlling the interactions between anchored DNA and a bilayer membrane is critical to attaining the desired performance of such lipid-spanning DNA nanopores. There are numerous unanswered questions about the nature of such interactions; in particular, about the affinity and the conformation of the tethered DNA strands to and near the bilayer membrane and how those are affected by the charge and size of the lipid headgroup.

## METHODS & CODES

The research team performed explicit-solvent all-atom MD simulations of several lipid-conjugated DNA systems using the latest version of NAMD2 [6]; the CHARMM36 [7] forcefield to describe the bonded and nonbonded interactions among DNA, lipid bilayer membranes, water, and ions; and the team’s cus-

tom NBFIX corrections for nonbonded interactions [8]. The analysis and postprocessing of the simulation trajectories were performed using VMD and CPPTRAJ [9,10].

## RESULTS & IMPACT

Complementing the gel-shift experiments carried out by the collaborators in the Howorka group at the University College London, the researchers at the University of Illinois at Urbana–Champaign built and simulated several all-atom models of DNA–lipid systems that differed from one another by the composition of the lipid membrane, the type of DNA molecules, and the buffer conditions. The lipids were chosen on the basis of their wide use in research and to cover a representative set of headgroup charges and sizes.

This study has shown that the composition of a lipid membrane, the buffer electrolyte, and the type and length of DNA constructs can considerably affect the ability of a cholesterol-terminated DNA molecule to insert into a lipid membrane. Thus, negatively charged lipids were found to weaken the binding of DNA to the membrane. Surprisingly, the study found the type of monovalent cations in the buffer solution also affected DNA binding, which the MD simulations explained by stronger specific interactions of smaller-sized cations with the lipid headgroups. In addition to electrostatics, this study elucidated the role of steric interactions, predicting the maximum coverage of lipid membranes by various DNA constructs. It also found that, while being tethered to lipid bilayers, the DNA molecules can diffuse along the membrane surface and that such diffusion is limited by the properties of the lipid bilayer.

By synergistically combining the experiment and all-atom MD simulations, the research team examined how cholesterol-modified DNA strands interact with lipid bilayer membranes. Based on the titration results from the gel-shift assay, this study for the first time quantified the binding affinity of cholesterol-conjugated DNA to lipid membranes in terms of their equilibrium dissociation constant. The quantitative insights into the binding affinity and molecular accessibility of DNA will facilitate rational design of membrane-spanning DNA nanopores and broaden the usage of cholesterol-conjugated DNA for sculpting and assembling lipid bilayer membranes into functional biomimetic systems.

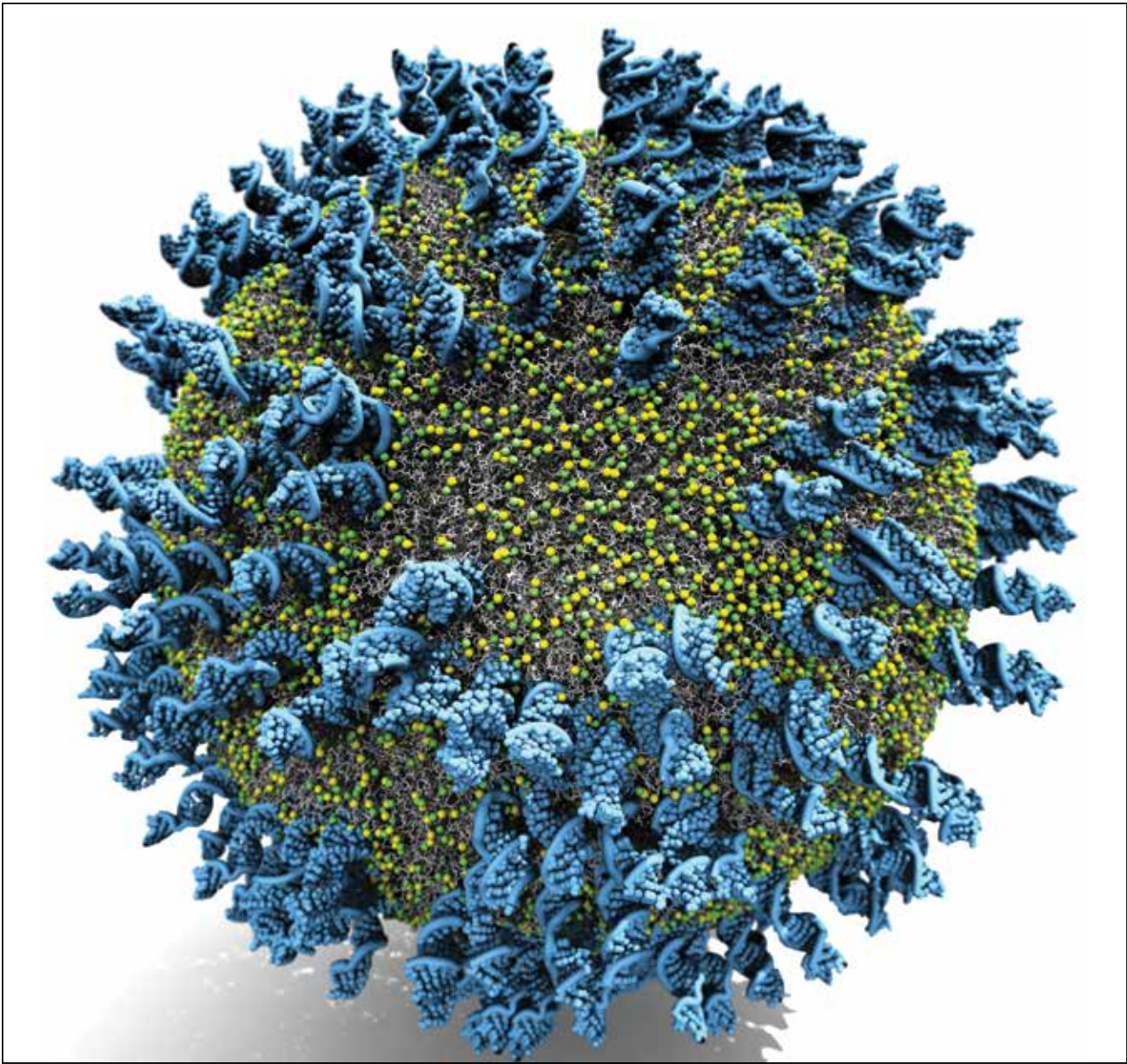


Figure 1: Molecular graphics representation of cholesterol-modified DNA fused into a phospholipid vesicle. The nitrogen and phosphorous atoms of lipid headgroups are shown in lime and yellow, lipid tails in white and gray, and DNA molecules in blue.

## WHY BLUE WATERS

Explicit-solvent all-atom MD simulations were needed to examine the fine details of lipid-tethered DNA interactions with the bilayer membrane and to accurately characterize the effect of various factors such as the charge and the size of the lipid headgroups as well as the type of the DNA constructs (single- or double-stranded DNA). Because of the long timescale needed to decipher these details, such MD simulations are computationally demanding. The large number of XK nodes on Blue Waters with graphics processing unit accelerators connected by the fast Gemini interconnect makes it one of the best publicly available systems

for performing simulations studying DNA–lipid interactions in atomistic detail. Over the past several years, the research team has used Blue Waters to carry out a set of landmark simulations in the area of nucleosome and DNA dynamics, bringing high-performance simulations to the forefront of this research field.

## PUBLICATIONS & DATA SETS

P. M. Arnott, H. Joshi, A. Aksimentiev, and S. Howorka, “Dynamic interactions between lipid-tethered DNA and phospholipid membranes,” *Langmuir*, vol. 34, no. 49, pp. 15084–15092, 2018, doi: 10.1021/acs.langmuir.8b02271.



INFLUENCE VIRULENCE AND TRANSMISSIBILITY THROUGH THE COMPUTATIONAL MICROSCOPE

Allocation: NSF PRAC/850,000 Knh  
PI: Rommie Amaro<sup>1</sup>  
Collaborator: Lorenzo Casalino<sup>1</sup>

<sup>1</sup>University of California, San Diego

EXECUTIVE SUMMARY

This work explores Influenza A Virus (IAV) biology using the “computational microscope”: a powerful, continually improving tool capable of disclosing moving pictures of the unseen atomic world of biological systems, including viruses. The team has integrated experimental structural data to push the boundaries of computer simulations toward larger scales while increasing the complexity and realism of the modeled constructs. This allowed the team to build a mesoscale IAV model of the currently circulating A/Michigan/45/2015 (H1N1) strain, and to perform all-atom molecular dynamics (MD) simulations of the massive (161-million-atom) system. MD provided unique insights on virus dynamics and glycoprotein interplay that is otherwise not accessible through individual protein simulations, which will shed light on the role played by glycans in modulating IAV virulence and transmissibility.

RESEARCH CHALLENGE

IAV causes severe illness and thousands of deaths every year. The virus can undergo random resorting of its segmented genome that may result in devastating worldwide pandemics. In this concerning scenario, the most critical mutations are usually found in two membrane glycoproteins: hemagglutinin (HA) and neuraminidase (NA), characterized by a wide population of carbohydrates located on their surface. Glycans are implicated in numerous viral processes, such as infectivity, pathogenicity, transmissibility, protein cooperativity, and small-molecule binding, thus deeply affecting IAV biology [1]. The number of glycosites exposed by HA and NA changes at regular temporal intervals and alters IAV antigenic properties [2], providing one explanation for why IAV represents a relentless health threat requiring vaccines to be updated every few years.

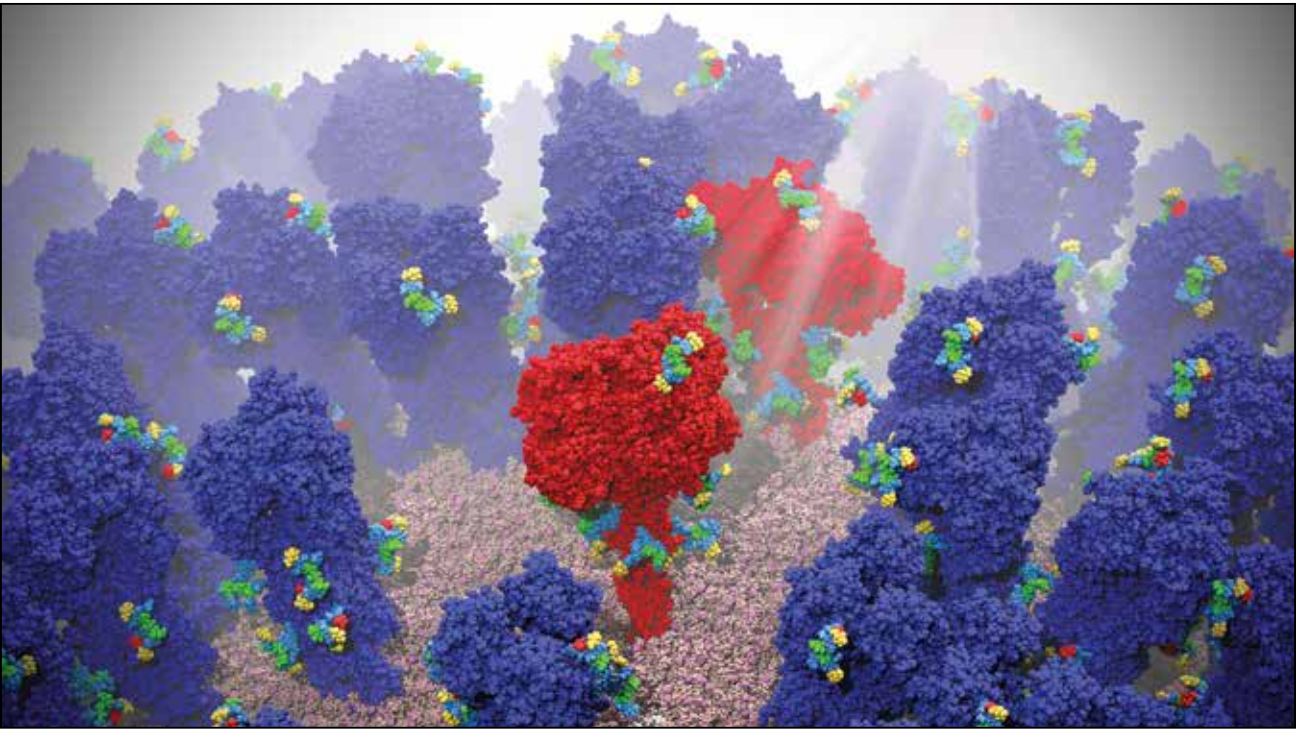


Figure 1: Snapshot after 420 nanoseconds of all-atom molecular dynamics simulation on the whole Influenza A/Michigan/45/2015 (H1N1) viral coat. Glycans (colored using the Symbol Nomenclature for Glycans standard) have been added on the two fundamental glycoproteins, neuraminidase and hemagglutinin (shown in red and blue, respectively), deeply affecting their dynamics and interplay. Credit: Dr. Lorenzo Casalino.

Experiments aimed at deciphering how minute modifications in glycoprofiles can impact IAV functions are strongly hampered by glycans’ high variability, flexibility, and small size. Computer simulations and modeling techniques, assisted by a continual growth of hardware and software technologies, constitute a viable alternative approach [3,4].

With access to Blue Waters’ high-performance resources, the research team was able to build and simulate a realistic all-atom viral coat model of the A/Michigan/45/2015 (H1N1) virus, crossing multiple spatial and temporal scales. Mutations that occurred within the antigenic site located on the HA head of this pathogen were so dramatic that in 2017 the World Health Organization decided to change the vaccine composition to target this specific strain. The research team seeks to elucidate the key roles of glycans in IAV biology and their impact on immune response escape.

METHODS & CODES

The preparation of the glycosylated A/Michigan/45/2015 (H1N1) system was based on the pandemic 2009 H1N1 viral coat built by Durrant *et al.* [4], which was in turn shaped upon cryoelectron tomography structural data [5]. This model, including NA, HA, and M2 proton channels embedded in a lipid bilayer, was overall ameliorated and upgraded with the addition of glycans on the glycoproteins’ spikes, thus notably increasing its complexity. After setting the glycosylation profiles and detecting all the N–X–S/T sequons exhibited by HA and NA, specific oligo-mannose, hybrid, and complex glycans were iteratively added on all the 236 HA trimers and 30 NA tetramers using the doGlycans tool [6] integrated with in-house scripts. After the addition of glycans and model refinement, the resulting 161-million-atom virion was ready to undergo MD simulations in explicit water and isothermal–isobaric conditions (NPT), using the memory-optimized version of NAMD 2.13 [7] and the CHARMM36 all-additive force field [8].

RESULTS & IMPACT

In this work, the research team has dramatically increased the level of accuracy of its previous IAV construct [4] by adding a total of 1,791 glycans on the HA and NA spikes. The extreme variability of the glycoprofiles exhibited by NA and HA has often discouraged investigators from modeling the glycans in computer simulations, thus neglecting their critical functional and structural contributions. By incorporating information from collaborators, bioinformatics, and experimental data into this computational approach, the research team has given rise to an even more realistic system crossing different spatial scales, from the atomic/molecular level of single glycans and proteins to the subcellular scale of the viral coat as a whole. MD simulations were conducted on Blue Waters using 4,096 XK nodes, which allowed us to collect an average of 15 nanoseconds (ns)/day, for a total of 420 ns and 18 terabytes of generated data. Although this research is still in progress, preliminary inspection and analyses have revealed an

exceptional interplay among the glycoproteins co-adjuvated by the attached glycans (Fig. 1). The observation of similar behavior with this level of accuracy and statistics, and the same biological significance, could have not been ascertained from single glycoprotein simulations. This work, embracing a multiscale computational protocol without losing the atomic detail, represents a cutting-edge attempt to bridge some gaps in the understanding of the IAV biology deriving from current experimental limitations.

WHY BLUE WATERS

Having access to a platform such as Blue Waters was absolutely crucial in order to perform MD simulations of such a massive system and to achieve a relevant amount of sampling. Carrying out the research on lesser supercomputers would have been beyond the bounds of possibility. Blue Waters’ efficient parallelism and tremendous scale provided outstanding performance in a timely fashion, enabled by using a memory-optimized version of the NAMD code specifically tweaked for XK and XE nodes and suitable for running MD simulations of multimillion-atom systems.



HOW BLUE WATERS IS AIDING THE FIGHT AGAINST SEPSIS

**Allocation:** Illinois/520 Knh  
**PI:** Rafael C. Bernardi<sup>1</sup>  
**Collaborator:** Hermann E. Gaub<sup>2</sup>

<sup>1</sup>University of Illinois at Urbana–Champaign  
<sup>2</sup>Ludwig–Maximilians–Universität

EXECUTIVE SUMMARY

Owing to their high occurrence in ever more common hos-  
pital-acquired infections, studying the mechanisms of infection  
by *Staphylococcus epidermidis* and *Staphylococcus aureus* is of  
broad interest. These pathogens can frequently form biofilms on  
implants and medical devices and are commonly involved in sep-  
sis—the human body’s often deadly response to infections.

Central to the formation of biofilms is very close interaction  
between microbial surface proteins called adhesins and compo-  
nents of the extracellular matrix of the host. The research team  
uses Blue Waters to explore how the bond between staphylococ-  
cal adhesin and its human target can withstand forces that so far  
have only been seen in covalent bonds. The team uses a syner-  
gistic combination of computational and experimental methods.  
This approach is essential to elucidating the mechanism by which  
an intricate network of hydrogen bonds makes the staphylococ-  
cal adhesion ultrastable, revealing possible routes for the devel-  
opment of antimicrobial strategies.

RESEARCH CHALLENGE

Antibiotics are increasingly powerless against a growing num-  
ber of “super bacteria” that have evolved to survive humankind’s  
pharmacological offensive. The shortage of new medicines to treat  
what the U.S. Centers for Disease Control calls “nightmare bac-  
teria” is evident. Public health agencies across the world have be-  
gun warning of the consequences of a postantibiotic era in which  
a common infection could, once again, become a death sentence.

The fight against sepsis, the human body’s often deadly re-  
sponse to bacterial infections, has become the topic of nation-  
wide campaigns in the United States. Exploring the mechanism  
by which bacteria initiate infections is, therefore, key to devel-  
oping new antimicrobial therapies. Investigations at the molec-  
ular level of the mechanism of adhesion between *Staphylococ-  
cus epidermidis* and *Staphylococcus aureus* and the extracellular  
matrix of their human hosts could lead to a detailed understand-  
ing of adhesion, one of the first steps of staph infections. This un-  
derstanding may, in turn, allow researchers to develop possible  
competitors for their interaction with humans, stopping infec-  
tion at its early stages.

METHODS & CODES

The research uses single-molecule force spectroscopy along  
with all-atom steered molecular dynamics (SMD) simulations

to investigate with exquisite precision the mechanics of interac-  
tion between SdrG and Fgβ. SdrG is an SD-repeat protein G and  
is one of the adhesin proteins of *Staphylococcus epidermidis*. Fgβ  
is the human fibrinogen β and is a short peptide that is part of  
the human extracellular matrix.

For SMD molecular dynamics simulations, the team uses Blue  
Waters’ GPU nodes (XK) and the GPU-accelerated NAMD pack-  
age. In a wide-sampling strategy, the team carried out hundreds  
of SMD runs for a total of over 50 microseconds of simulation.  
To characterize the coupling between the bacterial SdrG pro-  
tein and the Fgβ peptide, the team conducts SMD simulations  
with constant velocity stretching at multiple pulling speeds. The  
SMD procedure is inspired by experimental approaches and is  
equivalent to attaching one end of a harmonic spring to the end  
of a SdrG protein and pulling on the Fgβ peptide. To quantify  
the strength of interaction between the two molecules, the force  
applied to the harmonic spring is recorded at regular intervals.

RESULTS & IMPACT

The steered molecular dynamics simulations performed on  
Blue Waters revealed, and single-molecule force spectroscopy  
experiments confirmed, the mechanism by which this complex  
withstands forces previously only associated with the strength of  
a covalent bond. The target human peptide (Fgβ), confined in a  
screwlike manner in the binding pocket of the bacterial adhesin  
protein (SdrG), distributes forces mainly toward the peptide back-  
bone through an intricate hydrogen bond network. This behavior  
allows SdrG to attach to Fgβ with exceptionally resilient mech-  
anostability, virtually independent of the peptide’s side chains.

This unexpected mechanism expands the understanding of why  
pathogen adhesion is so resilient, which may open new ways to  
inhibit staphylococcal invasion. The development of anti-adhe-  
sion therapy could block the first steps of biofilm formation by  
staph bacteria, facilitating bacterial clearance. Understanding the  
mechanism of staph infection at the atomic level may also open  
new avenues for an intelligent design of antimicrobial therapies.

The research team’s initial findings were published in *Science*  
in 2018 [1]. Currently, the team is working on developing a new  
protocol for finding peptides with a higher affinity for staphylo-  
coccal adhesins. The initial results are promising, and a peptide  
sequence with a slightly higher affinity has already been identified.

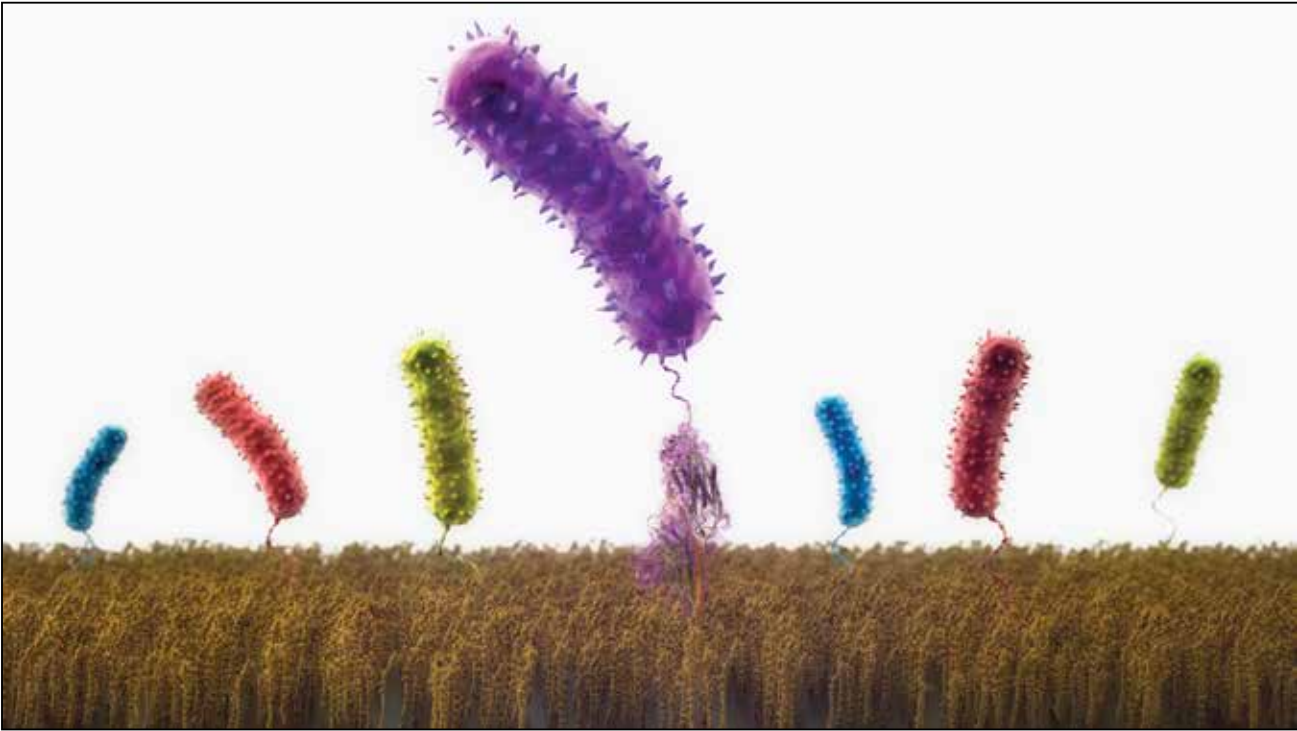


Figure 1: Staph bacteria (shown as colorful rods) adhere to their human hosts (surface at the bottom) with exceptional mechanical resilience. By combining experimental and computational approaches, the research team is deciphering the physical mechanisms that underlie the persistent stickiness of these bacterial adhesins (translucent purple structures), a major step in combating such invaders.

WHY BLUE WATERS

The research group’s work depends on obtaining multiple sim-  
ulation replicas with a fast turnaround time. This approach al-  
lows the team to quickly test any and all hypotheses both compu-  
tationally and experimentally. For this work, the team uses Blue  
Waters GPU (XK) nodes and CUDA-accelerated molecular dy-  
namics software NAMD.



# A PHYLOGENOMIC HISTORY OF PROTEIN FUNCTION AND DYNAMICS

**Allocation:** Illinois/200 Knh  
**PI:** Gustavo Caetano–Anollés<sup>1</sup>  
**Co-PI:** Frauke Gräter<sup>2</sup>  
**Collaborator:** Fizza Mughal<sup>1</sup>

<sup>1</sup>University of Illinois at Urbana–Champaign  
<sup>2</sup>Heidelberg Institute for Theoretical Studies

## EXECUTIVE SUMMARY

Studying the evolution of protein function is important for synthetic biology and translational medicine. The ability of proteins to undergo motions that perform a certain function depends on their structural flexibility. Flexibility, an evolutionarily conserved feature of protein structure, is a suitable proxy to measure protein dynamics and the underlying evolutionary drivers that sustain molecular function. In this work, the research team investigated the presence of evolutionary constraints on protein motions using molecular dynamics (MD) simulations, generating dynamics networks that capture topological features of protein structures, and constructing a three-dimensional network morphospace to trace the evolutionary emergence of protein function.

## RESEARCH CHALLENGE

Proteins perform a multitude of functions that sustain life on our planet. Understanding their evolution can impact agriculture, bioengineering, and biomedicine. Protein loops play an important role in protein function and dynamics by virtue of their structural flexibility [1]. Conservation of protein dynamics and flexibility [2] suggest the existence of signature motions corresponding to each protein function. Therefore, probing properties of loops at higher and lower levels of cellular organization may provide evolutionary clues of how protein function shapes protein dynamics. Previous work from the research group has demonstrated the

utility of networks to model evolutionary interaction at the organizational level of protein domains and loops [3].

The team is extending this methodology at the protein loop and residue level to study biophysical properties of associated functions, combining the seemingly disparate fields of physics and evolution, and leveraging nanosecond dynamics to dissect billions of years of phylogenomic history.

## METHODS & CODES

The research team extended its previous data set of 116 loops from protein domains belonging to metaconsensus enzymes [4] by including 58 additional structures. This augmented the group’s structural data set with previously underrepresented functional categories. The all-atom MD simulations were performed using an isobaric–isothermal ensemble (NPT) in TIP3P (transferable intermolecular potential with three points) water. We applied harmonic restraints of 2.1 kcal/mol Å<sup>2</sup> to the bracing secondary structure of the peptide. A sodium and chloride ionic concentration of 100 mMol was used to mimic near-physiological conditions. Depending on the number of atoms in the system, the researchers performed 50 to 70 ns (nanosecond) production runs with 1 ns of minimization using NAMD and the CHARMM36 force field. They generated networks based on the dynamic cross-correlation matrices computed from the simulations, from which they calculated network metrics for cohesion and centralities using the R

Figure 1: The maximum modularity scores (y-axis) for 170 of the 174 dynamic networks across the evolutionary timeline of protein domains (x-axis) spread across 3.8 billion years of evolution. Violin plots describe measures of central tendency with box-and-whiskers depictions of medians, quartiles, and data spread, all of them embedded within kernel density plots of the data. Modularity is high at the beginning of the timeline, and decreases with time with episodic up-down fluctuations.

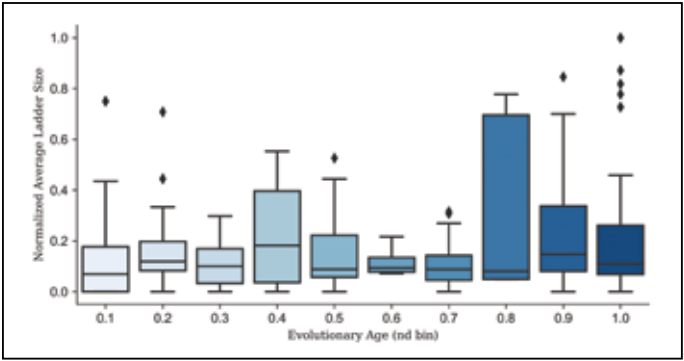
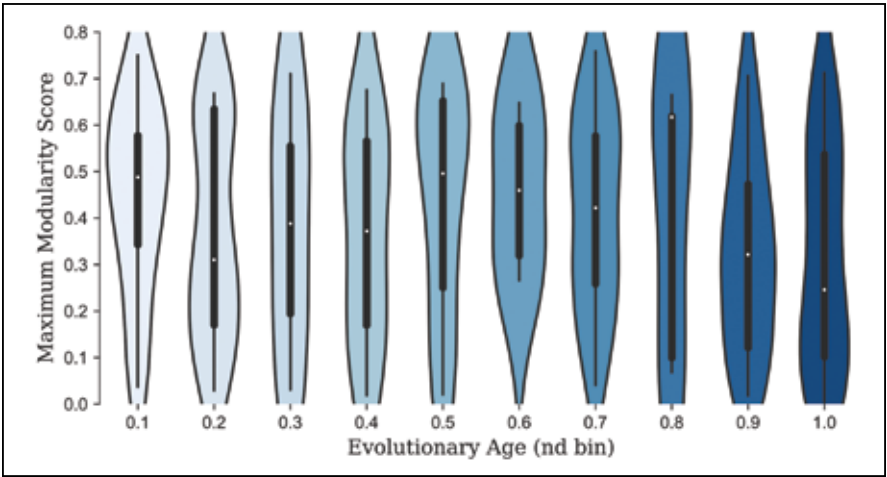


Figure 2: The average ladder size (y-axis) for trees based on community network structures for 170 of the 174 MD simulations distributed across the evolutionary timeline (x-axis). Box-and-whiskers plots describe how average ladder size stayed mostly consistent and closer to the mean for each age (nd) bin. However, a biphasic pattern was evident with average ladder sizes positively skewed late in evolution (especially at nd bin = 0.8). Skews indicate higher values than the median for the nd bin. Rhomboid symbols represent outliers.

packages bio3D and igraph, respectively [5,6]. The research team performed community structure detection on these networks that in turn generated trees for which they computed imbalance metrics using the R package phyloTop [7]. Other than network metrics, variables capturing the biophysical properties of the structure and corresponding movements were directly measured using principal component analyses, radius of gyration, and root mean square deviation. The protein structures were annotated with evolutionary age (nd) derived from phylogenomic timelines developed in the research team’s lab [8].

In addition to MD simulations, the researchers leveraged comparative genomics to assess whether protein loop structures hold significant evolutionary signal. Nearly 2,100 proteomes belonging to Archaea, Bacteria, and Eukarya along with 6,044 viral proteomes were downloaded from the RefSeq database [9]. The team included proteomes from representative and reference categories with chromosome or complete genome assembly in the study, as well as all viral proteomes listed in the National Center for Biotechnology Information viral genomes project [10]. The final set of proteomes was scanned against HMM profiles of structural domains using HMMER [11]. This genomic survey of protein domains was used to generate feature matrices to construct maximum parsimony trees of domains, following the protocol established by Kim *et al.* [8]. The resulting tree of domains with proteomes as characters are being compared against tree of domains with protein loop architectures as characters. The congruency of both the trees will help in establishing the presence (or absence) or phylogenomic signal embedded in loop architectures carried by the protein domains.

## RESULTS & IMPACT

Topological metrics of dynamic networks generated from the MD simulations provide interesting insight into molecular organization at a secondary-structure level. Network modularity decreased with time along the evolutionary timeline (Fig. 1), with episodic fluctuations embodying a biphasic model of module innovation and growth [12]. In the beginning, parts of a system have weak linkages. These weak linkages instigate diversification of parts through mutation, recruitment, and reassortment. Following diversification, competitive optimization occurs, leading

to a decrease in diversity and, subsequently, hierarchical organization of modules. The modules undergo diversification again to give rise to novel modules at yet another level of organization [12].

Phylogenetic trees are routinely used in epidemiology and evolutionary biology to study transmission patterns [13]. Quantifying the asymmetric nature of trees may help explain growth and innovation of biological features and the hierarchical structure of networks. In this study, the researchers constructed trees from communities in dynamic networks to investigate transmission of modular expansion in protein dynamics.

They measured tree asymmetry using a number of statistics, of which one is the average ladder size. A ladder is a series of nodes linked to a common ancestor, with each node having exactly one tip child (leaf) [14]. The average of all ladders in a tree was measured and normalized on a scale of 0 (balanced) to 1 (highly imbalanced). Average ladder sizes tend to be low and not as widespread for each age (nd) bin (Fig. 2). However, the average ladder sizes follow a biphasic pattern that tends to be positively skewed (above the median value) during the second phase (especially at age bin = 0.8), indicating that relatively recent dynamic networks have hierarchies that are more imbalanced than dynamic networks of protein loops from older domains.

## WHY BLUE WATERS

The computational heavy lifting of Blue Waters facilitated the completion of time-intensive research endeavors, including the scanning of the proteomes of approximately 2,000 organisms and thousands of viruses with sophisticated hidden Markov models of protein domain recognition. Other computational experiments involved MD simulations of 300 all-atom explicit water peptide systems. Access to Blue Waters helped the research team to complete these studies in a reasonable time period. Blue Waters support staff were knowledgeable in supercomputing matters and comprised domain experts. They were extremely helpful in answering both technical and field-specific queries.

## PUBLICATIONS & DATA SETS

G. Caetano–Anollés *et al.*, “Emergence of hierarchical modularity in evolving networks uncovered by phylogenomic analysis,” *Evol. Bioinform.*, vol. 15, no. 1, p. 1176934319872980, Sept. 2019. doi: 10.1177/1176934319872980



DI

GA

PREDICTING DRUG-INDUCED CARDIAC ARRHYTHMIAS USING ATOMISTIC SIMULATIONS

**Allocation:** Innovation and Exploration/196.8 Knh  
**PI:** Colleen E. Clancy<sup>1</sup>  
**Co-PIs:** Igor Vorobyov<sup>1</sup>, Kevin R. DeMarco<sup>1</sup>  
**Collaborators:** Sergei Y. Noskov<sup>2</sup>, Van A. Ngo<sup>2</sup>

<sup>1</sup>University of California, Davis  
<sup>2</sup>University of Calgary

BW

MP

EXECUTIVE SUMMARY

Induction of potentially deadly abnormal cardiac rhythms is one of the most common and dangerous risks of drugs in development and clinical use. Induction has been tightly associated with the loss of function of the cardiac ion channel protein hERG, which is responsible for transporting potassium ions out of the cell and restoring resting electric potential at the end of a heart-beat. This leads to the prolongation of the QT interval (the time of ventricular activity) on the ECG. However, not all hERG-blocking and QT-prolonging drugs cause cardiac arrhythmias resulting in withdrawal of safe and efficient pharmaceuticals. The research team has developed a computational pipeline encompassing atomic and tissue scales

that lets us estimate drug proclivity for arrhythmogenesis from its chemical structure. All-atom enhanced sampling molecular dynamics simulations of hERG–drug interactions simultaneously running on multiple Blue Waters nodes allowed the team to compute drug binding affinities and rates, which were used to predict emergent arrhythmias on a cardiac tissue scale.

RESEARCH CHALLENGE

Small-molecule pharmaceuticals form the basis of commonly used treatments for the majority of human ailments, and development of new safe and efficient drugs is a cornerstone of modern biomedical research. A challenging and yet unresolved problem plaguing these efforts is the lack of a robust and accu-

rate method for the prediction of drug cardiotoxicity in the form of deadly heart rhythm disturbances. Such cardiac arrhythmias are often caused by a drug-induced blockade of potassium channel hERG, a major cardiac membrane-embedded ion transport protein [1]. hERG block leads to an increased duration of cardiac cell membrane voltage perturbation (so-called action potential), often manifesting as a prolongation of the QT interval on the ECG. The problem, however, is that not all hERG-blocking and QT-prolonging drugs cause arrhythmias, and currently there is no methodology that can predict drug proclivity for arrhythmogenesis from its chemical structure [2]. This has led to the withdrawal from development of potentially safe pharmaceuticals. To avoid this, the research team aims to develop a multi-scale computational pipeline starting from state-dependent atomistic structural models of hERG–drug interactions all the way to functional kinetic models of cardiac cells and tissues, which would allow researchers to make such predictions. The enhanced sampling all-atom molecular dynamics (MD) simulations on Blue Waters are an integral part of this pipeline and allow the computation of drug affinities and rates, which are used as functional model parameters.

METHODS & CODES

The project’s molecular systems of approximately 128,000 atoms consisted of the hERG protein built from a cryo-electron microscopy (cryo-EM) structure (PDB ID 5VA2) and embedded in a hydrated POPC (a phosphatidylcholine) lipid bilayer. The systems were assembled using CHARMM–GUI and simulated using NAMD with CUDA support on Blue Waters’ XK nodes. After staged equilibration, umbrella sampling (US) and US–Hamiltonian-tempering replica exchange (US/H-RE) [3] MD simulations with 91 windows were used to study drug binding along the channel pore using 30 or 10 nanosecond-long production runs for each. Drug binding affinities were computed from free energy profiles, whereas drug ingress (“on”) and egress (“off”) rates were calculated based on diffusion coefficient profiles and from the ratio of “on” rates and affinities, respectively.

RESULTS & IMPACT

First, the team established that a cryo-EM hERG structure [4] likely represents an open conducting channel. The researchers also developed an inactivated hERG model by enforcing previously suggested N629 S620 intra-subunit hydrogen bonds [5] in restrained MD simulations, which also led to a distorted hERG selectivity filter conformation (Fig. 1b) thought to be important for inactivation [6]. In addition, the team developed atomistic CHARMM force field models of charged (+) and neutral (0) dofetilide (Fig. 1a). All-atom US/MD simulations revealed dofetilide binding in the channel pore, surrounded by hydrophobic F656 and Y652 residues (Fig.1b), known to be crucial for drug-induced hERG blockade [1]. Free energy profiles in Fig. 1c indicate more favorable binding for neutral dofetilide to the open hERG model compared to dofetilide(+) as well as inactivated state binding of

both drug forms (compared to  $\Delta G_{\text{bind}}$  in Fig. 1d). Both US/MD and US/H-REMD provided similar results (Fig. 1c and 1d) but at a fraction of computational cost for the latter. For open hERG, the team obtained a good agreement between experimental (3.5–11  $\mu\text{M}$  [7–9]) and the team’s computed ( $25\pm12\ \mu\text{M}$ ) drug affinities,  $K_D$ , accounting for drug form ratios at physiological pH. For inactivated hERG, experimental data suggest dofetilide binding in nanomolar range [7], but a much weaker affinity of  $320\pm140\ \mu\text{M}$  was computed. Thus, the inactivated hERG model likely is not representative of a channel state with high-affinity drug binding, and alternative models are actively being developed.

Importantly, the team computed “on” and “off” dofetilide rates for the open hERG model and directly used these values as parameters for their functional kinetic model of hERG–dofetilide interactions [10]. It was in turn integrated into the functional cardiac cell and tissue models, used to directly predict emergent arrhythmia indicators such as early afterdepolarizations in action potential profiles and beat-to-beat instabilities in computed pseudo-ECGs. The researchers’ MD simulation-informed multiscale model provided excellent agreement for a range of experimental and clinical data, including a dose-dependent high pro-arrhythmia risk of dofetilide [10]. The team is working to utilize this pipeline for other hERG blocking drugs with different proclivities for arrhythmogenesis and, more importantly, suggest drug chemical modifications, which can alter its hERG interactions to ameliorate pro-arrhythmia risks but maintain their efficacy. Thus, the atomistic MD studies on Blue Waters helped the research group to develop and test a computational transferable protocol for robust prediction of drug cardiotoxicity based on its chemical structure, which can lead to faster and cost-effective development of safe and efficient pharmaceuticals and thus save human lives.

WHY BLUE WATERS

Access to Blue Waters’ petascale architecture was indispensable for the success of these studies, since it allowed the team to efficiently conduct ~100 or more US/MD and US/H-REMD runs on GPU-equipped XK nodes at once, greatly reducing the total wall simulation time to just a few days and also permitting robust evaluation of simulation convergence.

PUBLICATIONS & DATA SETS

K. R. DeMarco *et al.*, "Atomistic modeling towards predictive cardiotoxicity," *bioRxiv*, p. 635441, 2019, doi: 10.1101/635441.

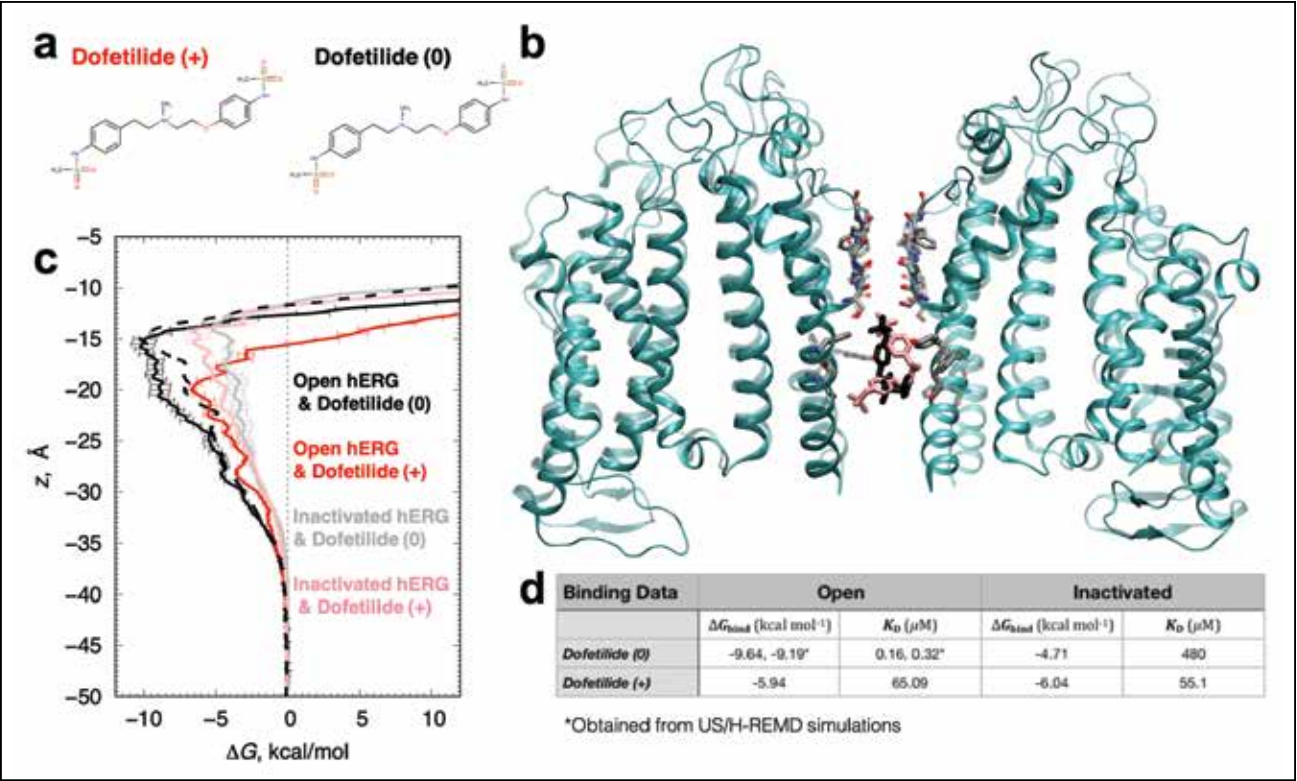


Figure 1: (a) Charged (0) and neutral (+) forms of dofetilide. (b) Open hERG (cyan ribbons) with dofetilide(0) bound (black) and inactivated hERG (semitransparent ribbons) with dofetilide(+) bound (pink). (c) Free energy ( $\Delta G$ ) profiles from the US/MD (solid) and US/H-REMD (dashed) simulations. (d) Computed binding free energies ( $\Delta G_{\text{bind}}$ ) and dissociation constants ( $K_D$ ).



MRNA ISOFORM PREDICTION

**Allocation:** Innovation and Exploration/84.9 Knh  
**PI:** Julie Dickerson<sup>1</sup>  
**Co-PI:** Gaurav Kandoi<sup>1</sup>

<sup>1</sup>Iowa State University

EXECUTIVE SUMMARY

Genes perform multiple functions, in part owing to their multiple messenger RNA isoforms. Messenger RNA (mRNA) is a large family of RNA molecules that conveys genetic information from DNA to the ribosome, where they specify the amino acid sequence of the protein products of gene expression. Alternative splicing produces multiple mRNA isoforms of genes that have important diverse roles such as regulation of gene expression, human heritable diseases, and response to environmental stresses. Alternative splicing is one mechanism that allows genes to perform multiple functions. Despite the diverse role of alternative splicing, very little has been done to assign functions at the mRNA isoform level. Therefore, differentiating the functions of mRNA isoforms is vital for understanding the underlying mechanisms of biological processes.

The goal of this study is to develop a functional network and recommendation (recommender) system to predict tissue-specific mRNA isoform function and understand implications of such alternate isoforms on metabolic pathways for the mouse and *Arabidopsis thaliana* model systems. (*Arabidopsis* is a small flowering plant that is widely used as a model organism in plant biology.)

Highlights from the research team’s outcomes include: (1) the processing of over 100 tissue-specific *Arabidopsis* RNA–Seq data sets as well as mRNA and protein sequence characterizations; (2) developing a random forest-based framework for mRNA-level functional network prediction; and (3) optimizing hyperparameters for the recommender system.

RESEARCH CHALLENGE

This research considered three major challenges in mRNA isoform function prediction. The first is the unavailability of mRNA isoform-level functional data, which is required to develop machine learning tools. However, the available data, even at the gene level, does not include all genes, further complicating the matter. The second challenge is the lack of information about tissue specificity in functional databases such as Gene Ontology, Kyoto Encyclopedia of Genes and Genomes, and UniProt. The third challenge is the lack of mRNA isoform-level “ground truth” functional annotation data. Therefore, the research team’s work includes using mRNA isoform and protein sequences, high-throughput RNA-sequencing data, and functional annotations at the gene level to develop computational methods for predicting functions for alternative spliced mRNA isoforms.

Some limitations of previous studies that have been overcome in the current work include: (1) predicting novel mRNA isoform interactions with no gene-level interaction information in current biological databases, (2) predicting tissue-specific mRNA isoform-level functional networks and mRNA isoform function, (3) limiting bias in the machine learning model by using a more biologically sound way of defining nonfunctional (negative pairs) mRNA isoform pairs, (4) formulating the task of mRNA isoform-level functional network prediction as a simple supervised learning task and formulating the task of mRNA isoform function prediction as a recommendation system, and (5) incorporating the relations between the Gene Ontology terms apart from the obvious hierarchical relations.

METHODS & CODES

The team has used STAR [1] and StringTie [2], which are codes for high-performance alignment and assembly of RNA–Seq reads and expression analysis for transcripts, without compromising mapping accuracy. These scale almost linearly with an increasing number of processing cores with a minimal increase in the memory requirement. Both tools are written entirely in C++ for higher efficiency and faster performance. All mRNA and protein sequence properties were calculated using R Bioconductor packages. These provide the means to calculate several diverse types of sequence properties. In addition, the team used TensorFlow and scikit-learn to build the machine learning systems.

RESULTS & IMPACT

One of the outcomes of this project was the evaluation and validation of the team’s supervised learning-based machine learning framework for predicting tissue-specific mRNA isoform functional networks in *Arabidopsis*. Tissue-spEcfic mrNa iSoform functional Networks (TENSION) makes use of single mRNA-producing gene annotations and gene annotations tagged with “NOT” to create high-quality mRNA isoform-level functional data. The team uses these to train random forest algorithms to develop mRNA isoform functional network prediction models. By using a leave-one-tissue-out approach and incorporating tissue-specific mRNA isoform-level predictors along with those obtained from mRNA isoform and protein sequences, the team has developed mRNA isoform-level functional networks for *Arabidopsis* tissues.

Another outcome is the evaluation of different combinations of hyperparameters of the mRNA function recommendation system (mFRecSys) for making tissue-specific function recommen-

dations for mRNA isoforms. In mFRecSys, the team considers mRNA isoforms as “users” and Gene Ontology biological process terms as “items.” By using explicit contexts for mRNA isoforms, Gene Ontology biological process terms, and tissue-specific mRNA isoform expression, mFRecSys is able to make tissue-specific mRNA isoform function recommendations.

This work emphasizes the significance of incorporating diverse biological context to develop better machine learning tools for biology. It also highlights the use of simplified supervised learning methods for biological network prediction. The machine learning models and recommendation systems developed as part of this work also draw attention to the power of simple mRNA isoform sequence-based predictors to improve mRNA isoform function prediction. The methods developed have potential practical applications, for instance, as predictive models for distinguishing the functions of different mRNA isoforms of the same gene or identifying tissue-specific functions of mRNA isoforms.

WHY BLUE WATERS

Blue Waters was essential to the two research outcomes of this project, as they both involved extensive validation and optimization. The computational resources of Blue Waters allowed the research team to perform these experiments at scale, enabling the optimization of many different experimental settings. Furthermore, conversations with the staff have been instrumental in improving job efficiency.

PUBLICATIONS & DATA SETS

G. Kandoi and J. A. Dickerson, “Tissue-spEcfic mrNa iSoform functIOnal Networks (TENSION) Collection,” 2018, doi: 10.25380/iastate.c.4275191.

G. Kandoi and J. A. Dickerson, "Tissue-specific mouse mRNA isoform networks," *BioRxiv*, 2019, doi: 10.1101/558361.

G. Kandoi and J. A. Dickerson, “Differential alternative splicing patterns with differential expression to computationally extract plant molecular pathways,” in *2017 IEEE International Conf. on Bioinformatics and Biomedicine*, doi: 10.1109/BIBM.2017.8217993.



PETASCALE INTEGRATIVE APPROACHES TO PROTEIN STRUCTURE PREDICTION

Allocation: NSF PRAC/5,600 Knh  
PI: Ken Dill<sup>1</sup>  
Co-PI: Alberto Perez<sup>2</sup>

<sup>1</sup>SUNY at Stony Brook  
<sup>2</sup>University of Florida

EXECUTIVE SUMMARY

The research team applies physics-based molecular dynamics computations to protein modeling by leveraging external information through the team’s Modeling Employing Limited Data (MELD) accelerator method. MELD runs on GPUs and is able to harness sparse, noisy, and ambiguous information using “sub-haystacking” Bayesian inference. It can lead to orders-of-magnitude speedups in protein folding and protein–protein docking over traditional molecular dynamics (MD) methods.

MELD x MD was ranked first in last summer’s Critical Assessment of Structure Prediction (CASP) protein structure prediction competition in the experimental NMR data-assisted predictions category. This event tests how well computations can utilize real-world NMR data to determine high-resolution protein structures. This is an important blind test that shows that molecular dynamics forcefields can meet the challenge and that MELD accelerates these computations sufficiently so that it can address large enough protein sizes to become an important new method for structural biology.

RESEARCH CHALLENGE

A central step in understanding how proteins perform their biological actions and how to design drugs to inhibit or activate them is to know a protein’s 3D atomistic structure. Experimental methods, such as X-ray crystallography, nuclear magnetic resonance spectroscopy (NMR), and cryo-electron microscopy (cryo-EM) provide the underlying data, but such methods require a computational means of converting that data into a meaningful structure. Different experiments have different limitations: data can be sparse, or ambiguous, or noisy and combinatoric. Researchers need computational approaches that can handle these limitations. The best would be physics-based simulations, properly sampled, to provide proper Boltzmann populations and free energies.

This is the challenge the research team is addressing with its MELD-accelerated molecular dynamics (MELD x MD) [1,2]. MELD x MD is a substantial enhancement of traditional MD in biomolecular simulations because it allows for the exploration of more limited data in determining protein structures or larger proteins [3] and larger motions in all the applications of physical molecular simulations to matters of protein structure and mechanism [4].

METHODS & CODES

The team developed a plugin (MELD) to the MD package OpenMM [5]. MELD consists of a Hamiltonian and Temperature replica exchange MD protocol in which the Hamiltonian varies according to external information coming from experiment, general knowledge, or bioinformatics. What is unique about MELD is that the information is expected to be unreliable. Hence, rather than enforcing all of it, only a fraction is enforced. The part to be enforced changes at every timestep and is chosen in a deterministic way.

RESULTS & IMPACT

High-performance computing (HPC) on Blue Waters has been essential to this work in two ways: (1) the invention and development of the MELD method, and its application to proteins, has been very costly computationally and could not have been done without these national HPC resources; and (2) the principle

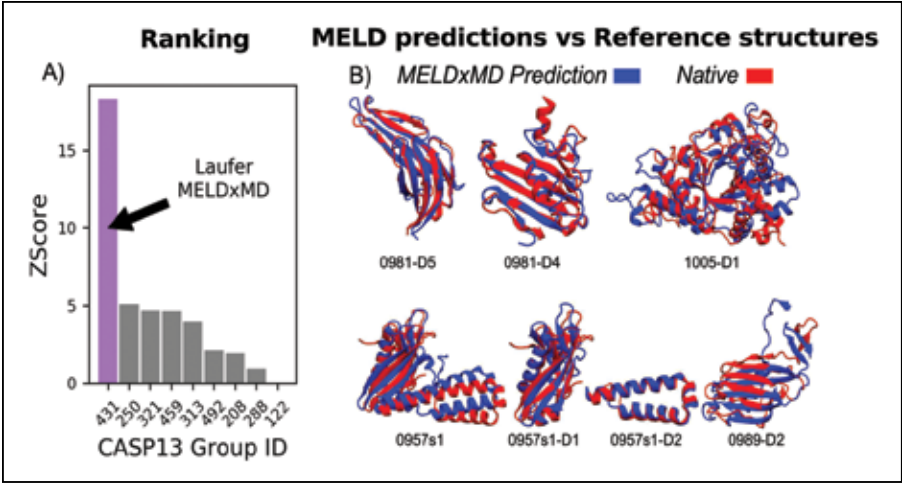


Figure 2: MELD x MD was the highest-ranked group in NMR data-assisted CASP13. a) The MELD x MD method (group code 431) had the highest Z-score of all groups that competed in NMR data-assisted CASP13. b) MELD x MD predictions are aligned to reference structures.

pal way that researchers measure success of their methods, and compare them to other methods in the field, is through blind competitions, such as the biennial CASP event. Physics-based methods have been too computationally slow to enter these many-protein time-limited events in the past. CASP13 in 2018 involved 100 protein challenges, each with a three-week computational deadline. MELD x MD now brings physical methods into the realm of these real-world comparative tests.

We believe that MELD x MD is potentially field-changing [1–4,6]. It can bring orders-of-magnitude greater speedups to MD simulations of a broad range of protein properties. From the NMR event and other such CASP tests, we are learning that the forcefields [7] and solvent models are quite good [8] and are often now better than many bioinformatics and database-driven approaches [9], with the added important power that they provide proper populations and free energies [10,11]. We are learning that MELD x MD can succeed in many venues—assisting in structure determination [3,6], *ab initio* protein folding [1,2,9], ligand binding, protein–protein docking [13], and others. Going forward, we believe the method will have additional niches important to scientists working to understand structures, dynamics, mechanisms, and to improve drug discovery. Two such applications currently in progress are assisting with cryo-EM to determine large protein structures and complexes, and determining the structures of amyloid aggregates, such as those that are problematic across the spectrum of neurodegenerative diseases.

WHY BLUE WATERS

Blue Waters is the only system in the United States that has enough GPUs for the research team to compete in CASP and allows many jobs using a relatively low number of GPUs (30 each) to run for up to 48 hours. Compilation of both Amber and OpenMM/MELD have not been trivial to optimize, and support from project staff has been invaluable, especially during the deployment of the new Python standard libraries. Furthermore, conversations with the staff have been invaluable to set up ways to run jobs efficiently during the CASP competition.

PUBLICATIONS & DATA SETS

- A. Perez, J. L. MacCallum, and K. A. Dill, “Accelerating molecular simulations of proteins using Bayesian inference on weak information,” *Proc. Natl. Acad. Sci. U.S.A.*, vol. 112, no. 38, pp. 11846–11851, 2015.
- A. Perez, J. A. Morrone, E. Brini, J. L. MacCallum, and K. A. Dill, “Blind protein structure prediction using accelerated free-energy simulations,” *Sci. Adv.*, vol. 2, no.11, p. e1601274, 2016.
- J. A. Morrone, A. Perez, J. L. MacCallum, and K. A. Dill, “Computed binding of peptides to proteins with MELD-accelerated molecular dynamics,” *J. Chem. Theory Comput.*, vol. 13, no. 2, pp. 870–876, 2017.
- J. A. Morrone *et al.*, “Molecular simulations identify binding poses and approximate affinities of stapled  $\alpha$ -helical peptides to MDM2 and MDMX,” *J. Chem. Theory Comput.*, vol. 13, no. 2, pp. 863–869, 2017.
- J. C. Robertson, A. Perez, and K. A. Dill, “MELD x MD folds nonthreadables, giving native structures and populations,” *J. Chem. Theory Comput.*, vol. 14, no.12, pp. 6734–6740, 2018.
- A. Perez, F. Sittel, G. Stock, and K. Dill, “MELD-path efficiently computes conformational transitions, including multiple and diverse paths,” *J. Chem. Theory Comput.*, vol.14, no. 4, pp. 2109–2116, 2018.
- E. Brini, D. Kozakov, and K. A. Dill, “Predicting protein dimer structures using MELD x MD,” *J. Chem. Theory Comput.*, 2019.
- J. C. Robertson *et al.*, “NMR-assisted protein structure prediction with MELD x MD,” *Proteins: Structure, Funct., and Bioinform.*, vol. 87, no. 12, pp. 1333–1340, 2019. doi: 10.1002/prot.25788

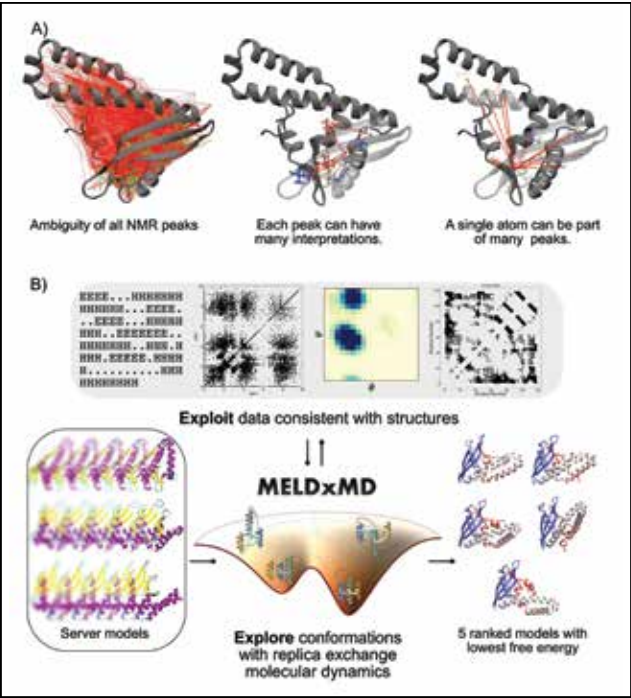


Figure 1: Noisy and ambiguous NMR data pose unique challenges during the CASP13 competition.



DISCOVERY OF SLOW KINETIC MODES FROM MOLECULAR SIMULATION TRAJECTORIES

Allocation: Exploratory/30 Knh  
PI: Andrew Ferguson<sup>1</sup>

<sup>1</sup>University of Chicago

EXECUTIVE SUMMARY

Slow modes discovery is an important topic in molecular simulation because it can help extract meaningful kinetic information for the analysis of dynamic processes and can serve as good collective variables for enhanced sampling. In this project, the researcher developed a new deep learning-based method called “state-free reversible VAMPnets” that is able to discover hierarchical nonlinear slow modes accurately with much lower computational cost than its predecessors. Moreover, this method can be well integrated with the powerful Markov model to improve its kinetic resolution significantly. This method is expected to be very helpful for understanding dominant kinetic transitions in biomolecular processes and to be useful for drug discovery.

RESEARCH CHALLENGE

Identifying the collective motions governing the longtime behaviors of biomolecules such as DNA and proteins is vital in understanding and engineering the behavior of these molecules of relevance to industrial catalysis, human health, and clean energy. This work establishes new basic science techniques combining applied mathematics and deep learning to “harness the data revolution” and perform data-driven inference of collective motions from molecular dynamics simulation trajectories. Existing methods for slow mode recovery such as time-lagged independent component analysis, or TICA, are only able to discover linear slow modes. Further, they require expert knowledge and hyperparameter tuning (e.g., kernel TICA) or fail to discover multiple hierarchical slow modes (e.g., time-lagged autoencoders or variational dynamics encoders). It is an outstanding challenge to develop a simple, robust, flexible, accurate, and efficient method to extract nonlinear hierarchical slow modes from simulation data.

METHODS & CODES

Recent advances in deep learning have made it a powerful tool to solve problems in many different fields. A variational principle developed for slow mode recovery has proved to be successful in discovering slow modes given good basis sets. The researcher’s work is founded on a combination of these two powerful ideas: Deep-learning models can naturally estimate excellent basis sets for the variational principle, whereas a variational principle defines a natural objective function with which to train deep neural networks to achieve this goal. Orthogonality and hierarchy of the modes are naturally imposed in the application of the variational principle. The new method surpasses previous methods since it

is simple, robust, accurate, efficient, and insensitive to feature selection and scaling. An open source package with extensive documentation and examples has been developed to make the method freely available to the machine learning and molecular simulation community (<https://github.com/hsidky/srv>).

RESULTS & IMPACT

The method has been tested on four different systems, including two toy models and two realistic molecular systems. Where available, the results show excellent agreement with theoretical analysis or previous calculations. This approach led to the development of kinetic models for protein folding at unprecedented time resolution and the establishment of highly efficient molecular simulators that, once trained, can perform molecular simulations at a six orders of magnitude lower cost than conventional approaches. These advances are valuable for better understanding and engineering of molecular machines for clean energy production as well as drugs and vaccines to improve human health.

WHY BLUE WATERS

This research requires running many long, large simulations; performing computationally intensive data processing; and training large numbers of machine-learning models with different training parameters. Access to Blue Waters allowed these computations to be performed at the scale and parallelism necessary to support this research. The Blue Waters staff were also invaluable in helping to improve the efficiency, performance, and workflow of the computations.

PUBLICATIONS & DATA SETS

W. Chen, H. Sidky, and A. L. Ferguson, "Nonlinear discovery of slow molecular modes using state-free reversible VAMPnets," *J. Chem. Phys.*, vol. 150, p. 214114, 2019.

State-free reversible VAMPnets, GitHub, May 2019. [Online]. Available: <https://github.com/hsidky/srv>

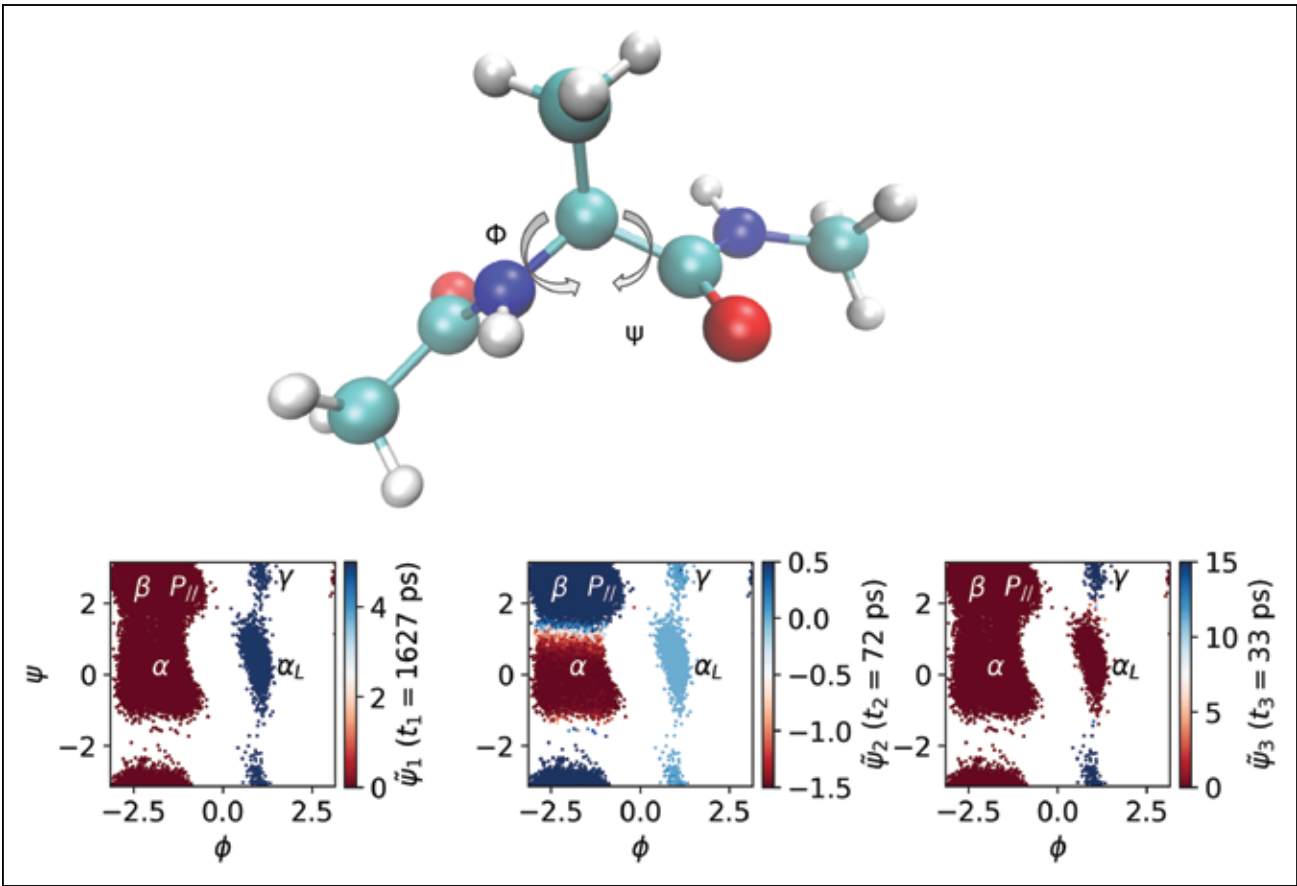


Figure 1: The molecular structure of alanine dipeptide and its three leading slowest kinetic modes: These three collective variables capture dominant transitions between the metastable basins marked in the image to providing molecular understanding of folding. The timescale associated with each transition is listed alongside each panel.



HARNESSING VISCOUS STREAMING IN COMPLEX ACTIVE SYSTEMS: MINIBOTS IN FLUIDS

Allocation: Strategic/480 Knh  
PI: Mattia Gazzola<sup>1</sup>

<sup>1</sup>University of Illinois at Urbana–Champaign

EXECUTIVE SUMMARY

Demonstrations of artificial and biohybrid (partly synthetic, partly biological) miniaturized swimming robots have underscored their potential as diagnostic and therapeutic vehicles, as well as the lack of and need for rigorous engineering methods for design and flow control. A brief survey reveals that most prototypes operate in flow regimes where viscous streaming can be leveraged. Streaming generates rectified flows in response to body-fluid oscillations and offers powerful control options for transport, mixing, drug delivery, and assembly. Thus, although streaming is not currently being exploited, it aligns well with minibots’ intended applications. One reason for this is that although streaming phenomena are well understood for simple bodies of uniform curvature, little is known in the case of complex, active geometries. Therefore, this project brings together modeling, simulations, and experiments to understand how viscous streaming relates to body morphology and aims to connect this understanding to biology and robotics to enhance the capabilities of current minibots.

RESEARCH CHALLENGE

The long-term goal of this work is to enable the rational design of miniaturized robots capable of operating in uncertain flow environments, of navigating the bloodstream, and of delivering localized treatment. The researcher is motivated by recent proof-of-concept demonstrations of artificial and living minibots in fluids, by their potential as diagnostic and therapeutic vectors, and by the lack of and need for rigorous engineering methods for design and flow control.

Toward this vision, the PI has revisited a well-known fluid dynamic effect: viscous streaming. This fluid mechanism takes place when an immersed body oscillates within specific size–frequency ratios (which happen to overlap with minibots’ operating conditions), and is responsible for the emergence of characteristic rectified flows via nonlinear fluid responses (Fig. 1). These structures offer appealing flow control options for particle manipulation, transport, and mixing. Larvae and large bacteria might also exploit and control them through body shape changes for feeding and locomotion. Thus, streaming aligns well with minibots’ intended applications, although it is not presently exploited.

Although viscous streaming is classically understood in the case of basic objects of uniform curvature such as plates, cylinders, and spheres, the PI hypothesized that morphological shape changes that introduce multiple curvatures can challenge this un-

derstanding and qualitatively affect the flow topology response, potentially providing ways to enhance flow control. The results from this project corroborate this hypothesis, and set the stage to revisit viscous streaming in a novel engineering context to improve the design and capability of minibots. Thus, the goal is to understand the relationship of body curvature to streaming and exploit it for biomedical applications.

METHODS & CODES

The characterization of propulsion and transport in fluids demands accurate, robust, fast, and flexible numerics for flow–structure interaction problems. The PI has been developing and implementing novel schemes for the direct numerical simulation of individual and multiple swimming bodies. His algorithms rely on remeshed vortex methods enhanced with a projection approach to capture the effects of the fluid on the body, and with a penalization technique, to capture the effects of the body on the fluid [1,2]. This methodology is coupled with a musculoskeletal solver developed to capture the compliant dynamics of musculoskeletal systems made of bones, tendons, and muscles or, in the case of biohybrid robots, made of muscles and artificial scaffolds [3].

RESULTS & IMPACT

This research has shown that oscillations can be utilized to improve transport robustly in an idealized two-dimensional master–slave setting across intermediate Reynolds numbers ( $1 \leq Re \leq 100$ ). The analysis of flow features identifies viscous streaming as the catalyst for this improvement. In order to leverage this information, the PI designed geometries exhibiting more favorable streaming patterns, which resulted in improved slave transport. To that extent, this project demonstrated a rational design approach by modifying the classic circular cylinder via the introduction of multiple curvatures and fore–aft symmetry breaking. Moreover, the work showed that similar concepts extend to three dimensions even though favorable streaming effects are activated differently [4].

Concurrently, the PI has developed Elastica [3], a software that captures the dynamic response of complex musculoskeletal architectures via assemblies of Cosserat rods [3,5]. In particular, the PI employed Elastica to computationally design, simulate, and optimize the structure of a biohybrid walking bot [6] and a biohybrid swimmer that combines neurons and muscles [7]. In collaboration with experimentalists, these designs were fabricated and

tested, confirming the predictive capacity. These results illustrate the biophysical accuracy of the PI’s solvers, rendering them powerful tools for the engineering design, optimization, and synthesis of microrobots operating in fluids.

WHY BLUE WATERS

Blue Waters’ sheer size and cutting-edge technology enables optimization processes that entail thousands of simulations. This allows the design of unprecedented biological architectures, bringing within reach novel high-impact applications, from soft robotics and biomedicine to precision manipulation and fabrication.

PUBLICATIONS & DATA SETS

M. Gazzola, P. Chatelain, W. van Rees, and P. Koumoutsakos, “Simulations of single and multiple swimmers with non-divergence free deforming geometries,” *J. Computat. Phys.*, vol. 230, pp. 7093–7114, 2011.

M. Gazzola, B. Hejazialhosseini, and P. Koumoutsakos, “Reinforcement learning and wavelet adapted vortex methods for simulations of self-propelled swimmers,” *SIAM J. Sci. Comput.*, vol. 36, pp. B622–B639, 2014.

M. Gazzola, L. Dudte, A. McCormick, and L. Mahadevan, “Forward and inverse problems in the mechanics of soft filaments,” *Royal Soc. Open Sci.*, vol. 5, no. 6, 2018, doi: 10.1098/rsos.171628.

T. Parthasarathy, F. Chan, and M. Gazzola, “Streaming enhanced flow mediated transport,” *J. Fluid Mech.*, accepted, 2019.

X. Zhang, F. Chan, and M. Gazzola, “Modeling and simulation of complex musculoskeletal architectures,” *Nat. Commun.*, under revision, 2019.

G. J. Pagan–Diaz *et al.*, “Simulation and fabrication of stronger, larger and faster walking biohybrid machines,” *Adv. Funct. Mat.*, vol. 28, no. 23, 2018, doi: 10.1002/adfm.201801145.

O. Aydin, X. Zhang, S. Nuethong, G. Pagan–Diaz, R. Bashir, M. Gazzola, and M. Saif, “Neuromuscular actuation of biohybrid motile bots,” *Proc. Natl. Acad. Sci. U.S.A.*, under revision, 2019.

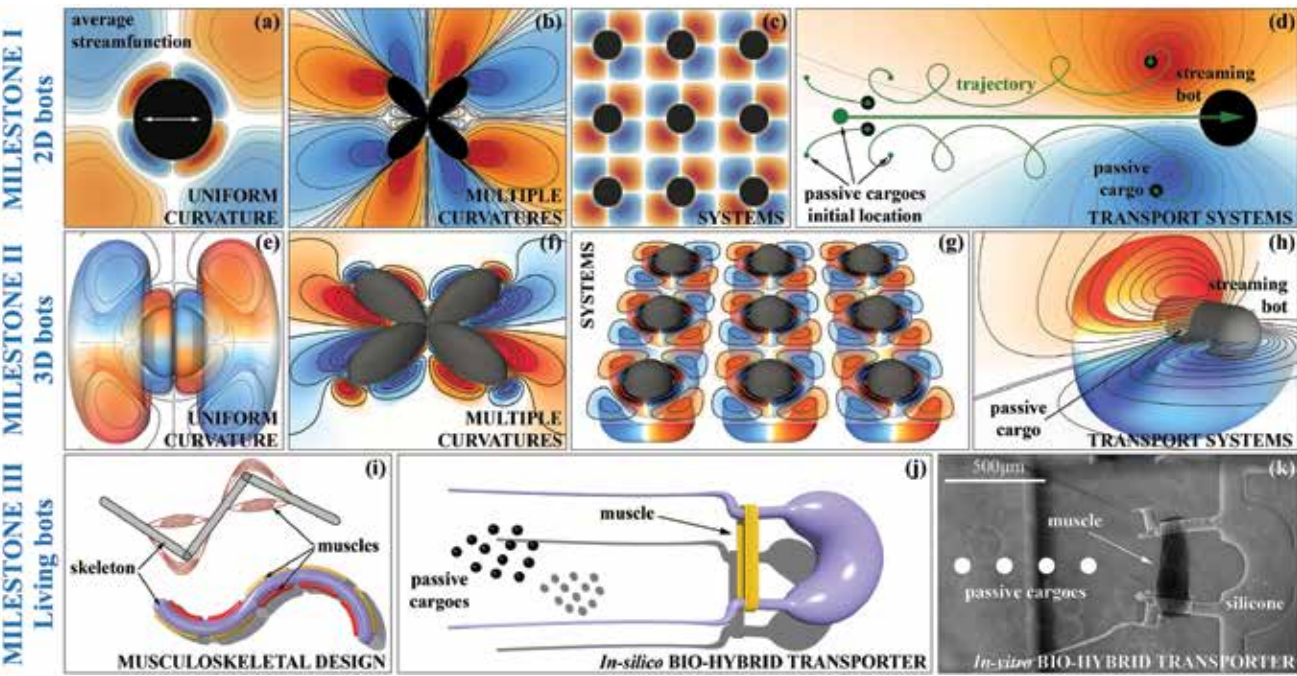


Figure 1: Goal—understand/exploit the nexus (body shape)–(viscous streaming) in 2D, 3D, and biohybrid bots for drug delivery.



MOLECULAR DYNAMICS SIMULATIONS OF THE HBV CAPSID

**Allocation:** Innovation and Exploration/199.265 Knh  
**PI:** Jodi A. Hadden–Perilla<sup>1</sup>  
**Collaborators:** Adam Zlotnick<sup>2</sup>, Juan R. Perilla<sup>1</sup>, Michael F. Hagan<sup>3</sup>, Brian Bothner<sup>4</sup>

<sup>1</sup>University of Delaware  
<sup>2</sup>Indiana University  
<sup>3</sup>Brandeis University  
<sup>4</sup>Montana State University

EXECUTIVE SUMMARY

The hepatitis B virus (HBV) is a major cause of liver disease. The World Health Organization estimates that more than 250 million persons currently suffer from chronic infection, with no cure available. Researchers aim to develop new treatments targeting HBV’s capsid, the protein shell that encloses its viral genome (Fig. 1). The research team has been leveraging Blue Waters since 2015 as a computational microscope to study the capsid, employing all-atom molecular dynamics simulations to reveal details of structure and function that are inaccessible to experiments.

Currently, the team is investigating the capsid’s response to mutations and drug compounds, probing its mechanics to discover capabilities and potential vulnerabilities. The researchers have learned that amino acids aspartate-78 and threonine-109 each play important, heretofore, unrealized roles in capsid assembly, and the team has implicated threonine-109 in drug resistance. Heteroaryldihydropyrimidine (HAP) compounds disrupt the capsid’s shape, volume, and solvent transport properties. Importantly, simulation of the multimillion-atom capsid system is only possible on a petascale supercomputer such as Blue Waters.

RESEARCH CHALLENGE

The HBV capsid is a complex molecular machine. It self-assembles from 120 capsid-protein dimers to package RNA, facilitates the maturation of the RNA to DNA, and hijacks various components of the host cell’s own machinery to transport its genomic cargo throughout the viral infection cycle. Drugs that disrupt the capsid have been identified but have not been approved for human use. The ability to produce new treatments that target the capsid depends heavily on understanding its inner workings and the mechanisms by which it carries out its function; by determining how the capsid works, researchers can also determine how best to inhibit it.

METHODS & CODES

Molecular dynamics simulations provide a powerful tool to investigate virus capsids such as that of HBV [1]. This project has demonstrated that when performed at all-atom resolution, simulations can capture remarkably subtle details of capsid structure and dynamics, including changes induced by bound drugs [2]. The simulations employed NAMD [3], a highly scalable biomolecular simulation code with a long and successful track re-

cord of deployment on Blue Waters. While all-atom simulations of the intact HBV capsid come at a great computational expense, access to NAMD on Blue Waters has enabled the research team to reveal critical new insights into its function and suggest strategies for targeting it with novel therapeutics [4,5].

RESULTS & IMPACT

*Aspartate-78.* The surface of the HBV capsid exhibits 120 spikes (Fig. 1), one for every incorporated capsid-protein dimer. The tips of the spikes contain four negatively charged amino acids: two copies each of glutamate-77 and aspartate-78 (Fig. 2a). The research team’s collaborators hypothesized that mutation of aspartate-78 to an uncharged amino acid would promote capsid assembly by decreasing electrostatic repulsion; however, experiments indicated that substitution of aspartate with serine was detrimental to the formation of capsids.

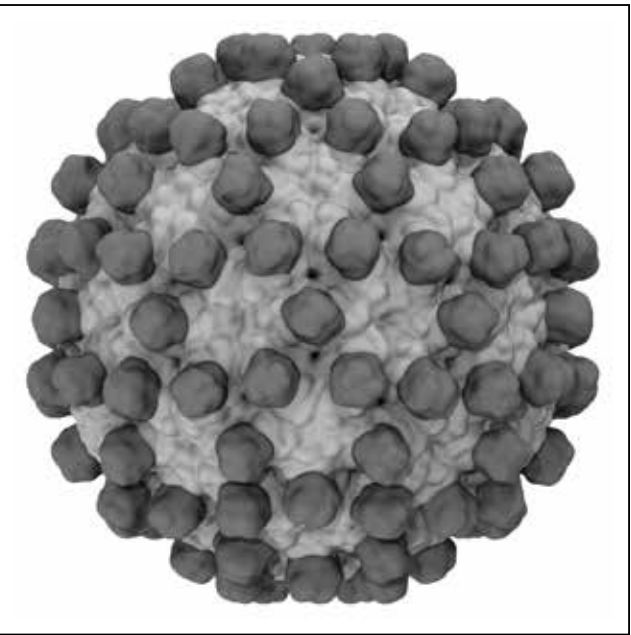


Figure 1: The HBV capsid is composed of 120 capsid-protein dimers. Each dimer contains a spike domain, shown here in dark gray.

The research team used Blue Waters to simulate multiple copies of wild type and mutant HBV capsid-protein dimers, obtaining conformational sampling totaling six microseconds. The team’s analyses showed that together, glutamate-77 and aspartate-78 induce sodium localization within their vicinity and transiently coordinate individual sodium ions within the spike tip (Fig. 2a). Substitution of aspartate-78 with serine caused the spike interface to open, dramatically reducing sodium localization and coordination. Both simulations and cryo-electron microscopy data showed that the mutation increased disorder in the spikes, demonstrating that aspartate-78 and its ion-mediated interactions are important for maintaining spike secondary structure and conformation conducive to productive capsid assembly.

*Threonine-109.* A number of drugs that target the HBV capsid are known to interfere with its assembly process. Experiments by the research team’s collaborators indicated that mutation of the amino acid threonine-109 (Fig. 2b) could increase both the rate of capsid assembly and the capsid’s resistance to assembly-disrupting drugs; however, no explanation for threonine-109’s role in either aspect was apparent from high-resolution crystal or cryo-electron microscopy structures of the capsid.

The research team used Blue Waters to simulate the intact HBV capsid on the microsecond timescale [4], producing an ensemble of 12 million samples that characterized the dynamical behavior of threonine-109. The team’s analyses showed that threonine-109 spends a significant amount of time mediating contact among neighboring dimers within the capsid (Fig. 2b), revealing why mutation of this amino acid can alter the ability of the proteins to interact and assemble. Further, the team found that the interdimer contact formed by threonine-109 can occlude a hydrophobic pocket recognized by capsid-disrupting drugs (Fig. 2b). Threonine-109’s ability to physically block the binding of such compounds confers some native drug resistance to HBV. Mutation of threonine-109 to larger, more hydrophobic amino acids enhances the rate of capsid assembly and also increases drug resistance.

*HAP compounds.* Drugs from the heteroaryldihydropyrimidine (HAP) family (Fig. 2b) can misdirect HBV capsid assembly and disrupt intact capsids. The research team previously used Blue Waters to simulate a HAP-bound capsid and observed that the compounds induced changes in the capsid’s shape [2] consistent with experimental observations. Now, the team has extended the drug-bound capsid investigation to the microsecond timescale, treating HAP as a small-molecule probe to test the capsid’s mechanics. The results indicate that saturation with HAP compounds increases the capsid’s volume and alters its solvent transport properties, revealing insight into the mechanism by which HAPs induce disruption. The team’s analyses were enabled by a novel method on which the researchers collaborated that uses ray-casting to accurately detect the interior versus exterior space of biomolecular containers, including virus capsids. The method is highly parallelizable and takes advantage of Blue Waters’ Lustre filesystem.

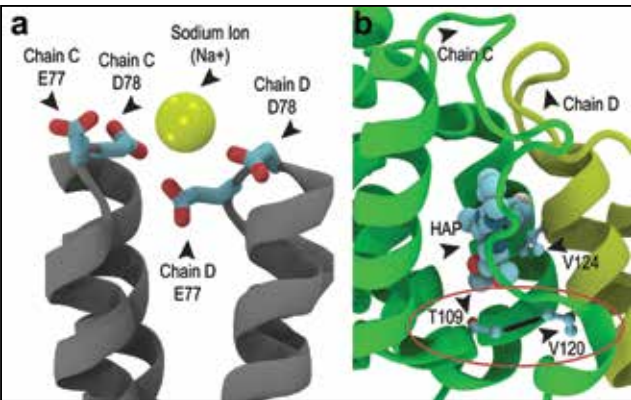


Figure 2: (a) The HBV capsid’s spikes contain glutamate-77 and aspartate-78. The team discovered that these negatively charged amino acids can transiently coordinate positively charged ions such as sodium. (b) The team also discovered that threonine-109 can form interdimer interactions that can block the binding of capsid-disrupting compounds such as HAPs to confer drug resistance.

WHY BLUE WATERS

Simulations of the intact HBV capsid are only feasible on a petascale supercomputer such as Blue Waters because of their computational expense. Investigating the capsid under physiological conditions at full chemical resolution requires calculations involving the interactions of millions of atoms. Simulations exploring the microsecond timescale can take months, even on thousands of processors. The exciting discoveries revealed by this work underscore the essential role for leadership-class computing resources in supporting basic science research toward gaining an understanding of viruses and developing novel antiviral treatments.

PUBLICATIONS & DATA SETS

L. Ruan, J. A. Hadden, and A. Zlotnick, “Assembly properties of hepatitis B virus core protein mutants correlate with their resistance to assembly-directed antivirals,” *J. Virol.*, vol. 92, no. 20, Sept. 2018, doi: 10.1128/JVI.01082-18.

J. A. Hadden and J. R. Perilla, “All-atom simulations of viruses,” *Curr. Opin. Virol.*, vol. 31, no. 8, pp. 82–91, Aug. 2018, doi: 10.1016/j.coviro.2018.08.007.

J. A. Hadden, J. R. Perilla, and J. C. Gumbart, “Biophysics on World Hepatitis Day,” *The Biophysical Society Blog*. [Online]. Available: <https://www.biophysics.org/blog/biophysics-on-world-hepatitis-day>

A. J. Bryer, J. A. Hadden–Perilla, J. E. Stone, and J. R. Perilla, “High-performance analysis of biomolecular containers to measure small-molecule transport, transbilayer lipid diffusion, and protein cavities,” *J. Chem. Inf. Model.*, vol. 59, no. 10, pp. 4328–4338, 2019.

Z. Zhao, J. A. Hadden–Perilla, and A. Zlotnick, “A pair of charged residues allosterically regulates HBV capsid assembly through secondary and tertiary structural interactions,” submitted for publication, 2019.



# MOLECULAR DYNAMICS SIMULATIONS OF THE HBV CAPSID AS A DRUG TARGET

**Allocation:** GLCPC/370 Knh  
**PI:** Jodi A. Hadden–Perilla<sup>1</sup>  
**Collaborators:** Adam Zlotnick<sup>2</sup>, Juan R. Perilla<sup>1</sup>, Michael F. Hagan<sup>3</sup>, Brian Bothner<sup>4</sup>

<sup>1</sup>University of Delaware  
<sup>2</sup>Indiana University  
<sup>3</sup>Brandeis University  
<sup>4</sup>Montana State University

## EXECUTIVE SUMMARY

Hepatitis B virus (HBV) is a major cause of liver disease. The World Health Organization estimates that more than 250 million people worldwide suffer from this chronic infection, with no cure available. Researchers aim to develop new treatments targeting HBV’s capsid, the protein shell that encloses its viral genome (Fig. 1a). The research team has been leveraging Blue Waters since 2015 as a computational microscope to study the capsid, employing all-atom molecular dynamics simulations to reveal details of structure and function that are inaccessible to experiments.

Currently, the team is endeavoring to further characterize the capsid as a drug target by investigating morphological disruption by heteroaryldihydropyrimidine (HAP) compounds, cooperativity in HAP uptake, and the structural and statistical details of HAP binding modes under physiological conditions. The team’s findings go far beyond conclusions that can be drawn from static models derived with crystallography or cryo-electron microscopy. Importantly, simulation of the multimillion-atom capsid system is only possible on a petascale supercomputer such as Blue Waters.

## RESEARCH CHALLENGE

The HBV capsid is a complex molecular machine that self-assembles from 120 capsid-protein dimers to package RNA. In addition, it facilitates the maturation of RNA to DNA and hijacks various components of the host cell’s own machinery to transport the capsid’s genomic cargo throughout the viral infection cycle. Drugs that disrupt the capsid have been identified but have not been approved for human use. Researchers’ ability to produce new treatments that target the capsid depends heavily on understanding the capsid’s inner workings and the mechanisms by which known drug compounds disrupt it; by determining how the capsid works, researchers can also determine how best to inhibit it.

## METHODS & CODES

Molecular dynamics simulations provide a powerful tool to investigate virus capsids such as HBV [1]. The research team’s work has demonstrated that when performed at all-atom resolution, simulations are capable of capturing remarkably subtle details of capsid structure and dynamics, including changes induced by bound drugs [2]. These simulations employ NAMD

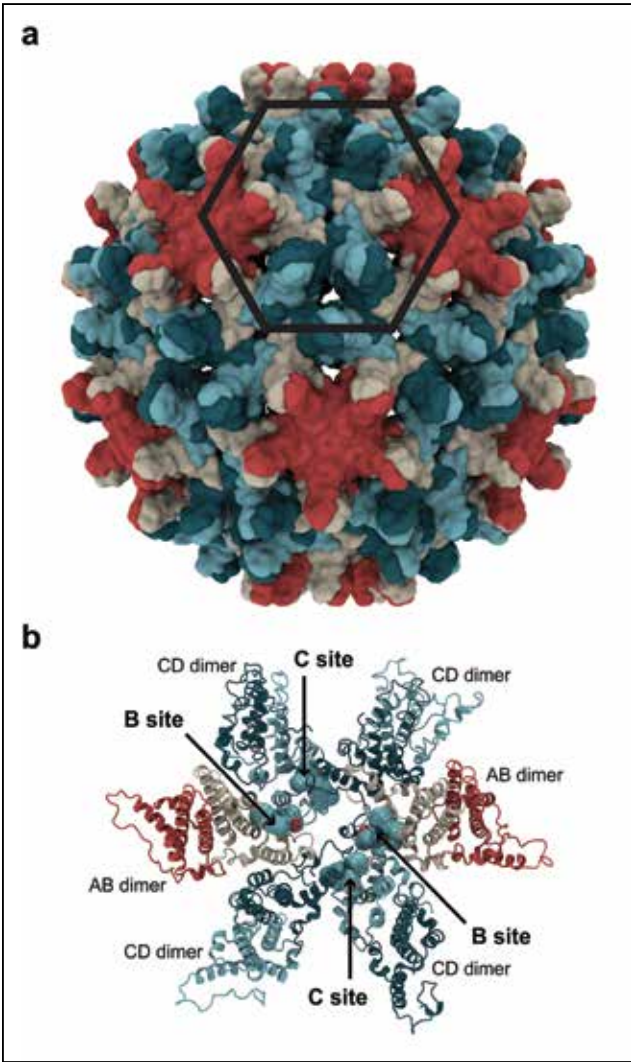


Figure 1: (a) The HBV capsid is composed of 120 capsid-protein dimers. These dimers can occupy either AB (red/beige) or CD (cyan/blue) quasi-equivalent positions. One capsid hexamer is highlighted. (b) The capsid hexamer exhibits four drug binding sites: two B sites and two C sites. Bound HAPs are shown.

[3], a highly scalable biomolecular simulation code that boasts a long and successful track record of deployment on Blue Waters. While all-atom simulations of the intact HBV capsid come at a great computational expense, access to NAMD on Blue Waters has enabled the team to reveal critical new insights into the capsid’s function and to suggest strategies for targeting it with novel therapeutics [4,5].

## RESULTS & IMPACT

The HBV capsid is composed of 120 capsid-protein dimers, arranged according to icosahedral symmetry (Fig. 1a). There are two possible orientations that dimers can occupy within the capsid structure, which are referred to as AB and CD. Drugs known to target the capsid recognize hydrophobic pockets found at the interfaces of B/C and C/D subunits, referred to as B sites and C sites (Fig. 1b). Previously, the research team used Blue Waters to simulate the apo (unbound) form of the HBV capsid on the microsecond timescale [4] and observed differences in the structure and dynamics of B sites versus C sites, despite symmetry [5]. Now, the team has used Blue Waters to study the capsid’s response to drug binding in each of these quasi-equivalent locations, furthering characterization of the capsid as a drug target. The simulations of the intact capsid with drugs from the HAP family complexed in B sites, C sites, or both B and C sites reach the microsecond timescale.

**Morphology.** Preliminary work on HAP-bound capsids demonstrated that the compounds induce changes in the capsid’s shape [2] consistent with experimental observations. The presence of drugs causes the spherical capsid to adopt a faceted morphology similar to that of a regular geometric icosahedron (Fig. 2). The latest results reveal that drug binding in C sites leads to more pronounced capsid faceting than drug binding in B sites. Saturation of the capsid with drugs in B and C sites leads to intermediate faceting and increased particle size. Importantly, these findings underscore the capsid’s resilience and ability to make structural adjustments to accommodate drug uptake.

**Cooperativity.** Experiments by the team’s collaborators suggest that drug uptake in the capsid is cooperative, meaning that the binding of one drug molecule makes the binding of a second more likely. The research team’s findings indicate that within the apo-form capsid, B sites remain mostly open, while C sites spend a significant portion of the time closed and occluded. As such, initial drug binding is most probable in B sites. Using the new simulation data, the team has learned that the presence of drugs in B sites causes C sites to become occluded less often, confirming the existence of a cooperativity mechanism. Further, the presence of drugs in C sites causes B sites to be occluded more often, reducing the likelihood of drugs binding there. Similar to the faceting effect, cooperativity between drug binding sites arises from shifts in interdimer orientations as the capsid makes structural adjustments to accommodate drug uptake.

**Binding modes.** The research team’s simulations of drug-bound capsids have provided the unique opportunity to expand charac-

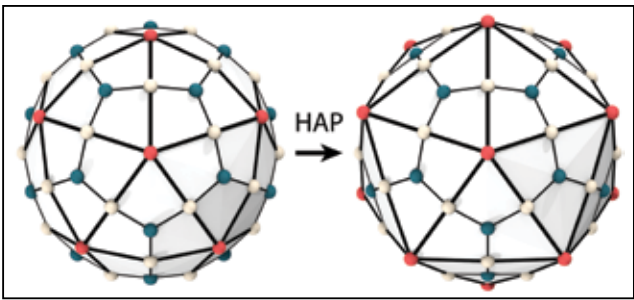


Figure 2: Bound HAPs cause the spherical capsid to adopt a faceted shape, similar to that of a regular geometric icosahedron.

terization of HAP binding modes far beyond what was previously determined from static crystal structures. Using Blue Waters, the researchers have produced ensembles of six-million samples each for B and C quasi-equivalent binding sites, which capture the dynamical interactions of HAPs with the capsid under native physiological conditions at full chemical resolution. The structural and statistical details revealed by these data sets are invaluable toward the design and optimization of new drug compounds that target the capsid and have the potential to support the discovery of new therapeutics.

## WHY BLUE WATERS

Owing to computational expense, simulations of the intact HBV capsid are only feasible on a petascale supercomputer such as Blue Waters. Investigating the capsid under physiological conditions at full chemical resolution requires calculations involving the interactions of millions of atoms. Simulations exploring the microsecond timescale can take months, even on thousands of processors. The exciting discoveries revealed by this work underscore the essential role for leadership-class computing resources in supporting basic science research toward understanding vi-ruses and developing novel antiviral treatments.

## PUBLICATIONS & DATA SETS

A. J. Bryer, J. A. Hadden–Perilla, J. E. Stone, and J. R. Perilla, “High-performance analysis of biomolecular containers to measure small-molecule transport, transbilayer lipid diffusion, and protein cavities,” *J. Chem. Inf. Model.*, vol. 59, no. 10, pp. 4328–4338, 2019.

J. A. Hadden–Perilla, B. C. Goh, C. J. Schlicksup, B. Venkatakrisnan, and A. Zlotnick, “Mechanistic insights into the drug-induced disruption of the HBV capsid revealed by all-atom molecular dynamics simulations,” submitted, 2019.



GA

TN

# TOWARD PREDICTIVE COMPUTATIONAL DESIGN OF PRECISION MOLECULAR OPTOELECTRONICS

**Allocation:** Blue Waters Professor/100 Knh  
**PI:** So Hirata<sup>1</sup>  
**Collaborator:** Alexander E. Doran<sup>1</sup>

<sup>1</sup>University of Illinois at Urbana–Champaign

## EXECUTIVE SUMMARY

The research team introduced and fully developed novel scalable algorithms and software for predictively accurate (*ab initio*) electronic-structure calculations for the key electronic parameters (electron binding energies) of large conjugated molecules and solids. These molecules and solids are potential components of precision organic optoelectronic devices such as solar cells, light-emitting diodes, field-effect transistors, smart windows,

etc. The team transformed the usual nonscalable sum-of-products expressions of many-body Green's function theories into a few high-dimensional integrals, which were then evaluated by a highly scalable Metropolis Monte Carlo algorithm. The algorithm efficiently computes energy differences (including quasi-particle energy bands) directly without a sign problem on many CPUs or GPUs.

## RESEARCH CHALLENGE

The world is entering an exciting new era of chemical technology in which synthetic chemists can now fabricate complex solid-state materials made of organic components with precise dimensions so that the fabricated materials display predicted/ designed functions and performance as optoelectronic devices. The revolutionary impact of such devices requires no explanation as they will miniaturize components for computing, light emission, memory, and the like, by many orders of magnitude. Research into such technology may be assisted powerfully by computational methods that can predict the key optoelectronic parameters of the component conjugated organic molecules, such as electron binding energies, charge mobility, exciton binding, air stability, etc. A systematically accurate series of approximations for these properties in a molecule and solid exists as many-body Green's function theory [1]. However, as compared with other *ab initio* theories, many-body Green's function theory is less well understood or developed both theoretically and algorithmically. The research group aims to address this issue with the assistance of Blue Waters.

## METHODS & CODES

The team mathematically transformed the usual sum-of-products expressions of second- and third-order many-body Green's function—GF2 [2] and GF3 [3]—theories, and their complete-basis set correction by explicitly correlated *ansätze* [4] into single high-dimensional integrals by a Laplace transform. Specifically, in the last funding cycle, 84 Feynman diagrams defining GF3 (Fig. 1) [1] were automatically generated by symbolic computing software and transformed into algebraic formulas consisting of integrals over 20-dimensional coordinates of three coupled electron pairs and over two imaginary time coordinates. These were then evaluated by a Metropolis Monte Carlo method with judiciously chosen weight functions. The resulting stochastic GF2 and GF3 methods can directly compute energy differences (electron detachment/attachment energies) without a sign problem on thousands of CPUs [5] or hundreds of GPUs [6] with an unprecedented efficiency.

## RESULTS & IMPACT

Supercomputers of the size and style of Blue Waters are changing the way chemical computing is being designed and performed. Conventional deterministic and matrix-algebra-based algorithms have given way to stochastic algorithms because of their superior parallel scalability on CPUs and GPUs; cost performance for high dimensions at the expense of having stochastic errors; naturally fault-tolerant and indefinitely restartable algorithms; and, finally and importantly, their relatively low human cost of code development. What the research team has achieved in this project is a paradigm of such changes in chemical computing.

## WHY BLUE WATERS

The stability and ease of use as well as the balanced deployment of CPUs and GPUs are all essential for the rapid coding/profiling of new scalable algorithms from scratch and their capacity testing.

## PUBLICATIONS & DATA SETS

A. E. Doran and S. Hirata, "Monte Carlo second- and third-order many-body Green's function methods with frequency-dependent, non-diagonal self-energy, *J. Chem. Theory Comput.*, Oct. 2019, doi: 10.1021/acs.jctc.9b00693.

C. M. Johnson, A. E. Doran, S. L. Ten-no, and S. Hirata, "Monte Carlo explicitly correlated second-order many-body Green's function theory," *J. Chem. Phys.*, vol. 149, p. 174112, Nov. 2018, doi: 10.1063/1.5054610.

C. M. Johnson, S. Hirata, and S. Ten-no, "Explicit correlation factors," *Chem. Phys. Lett.*, vol. 683, pp. 247–252, Sept. 2017, doi: 10.1016/j.cplett.2017.02.072.

C. M. Johnson, A. E. Doran, J. Zhang, E. F. Valeev, and S. Hirata, "Monte Carlo explicitly correlated second-order many-body perturbation theory," *J. Chem. Phys.*, vol. 145, p. 154115, Oct. 2016, doi: 10.1063/1.4964854.

A. E. Doran and S. Hirata, "Monte Carlo MP2 on many graphical processing units," *J. Chem. Theory Comput.*, vol. 12, pp. 4821–4832, Oct. 2016, doi: 10.1021/acs.jctc.6b00588.

Figure 1 (opposite page): Eighty-four Feynman diagrams of the self-energy of third-order many-body Green's function theory, computer-generated and transformed into algebraic formulas convenient for Monte Carlo integration.



# IMPACT OF BATCH EFFECT AND STUDY DESIGN BIASES ON IDENTIFICATION OF GENETIC RISK FACTORS IN SEQUENCING DATA

**Allocation:** Illinois/280 Knh  
**PI:** Matthew E. Hudson<sup>1</sup>  
**Co-PIs:** Yan W. Asmann<sup>2</sup>, Liudmila Sergeevna Mainzer<sup>1</sup>

<sup>1</sup>University of Illinois at Urbana–Champaign  
<sup>2</sup>Mayo Clinic

## EXECUTIVE SUMMARY

To explore how systematic biases within a genomic data set can impact downstream statistical analysis of genetic variants, the research team conducted stratified association analysis on Alzheimer’s disease genomic data. The researchers profiled a set of variants with highly significant, novel associations with Alzheimer’s disease that were impacted by heterogeneity in subcohort composition and exome capture. The team identified genotype quality, age, and population stratification as likely contributing factors to vastly different minor allele frequencies as well

as a batch effect across sequencing center cohorts. These findings highlight important considerations for analysis of this data set and for the design of future studies.

## RESEARCH CHALLENGE

The large sample sizes required to identify disease-associated variants in genomic sequencing data often introduce batch effects and other confounding variables or biases from study design. If not adequately addressed in the analysis, batch effects will reduce statistical power and increase false associations, and bias-

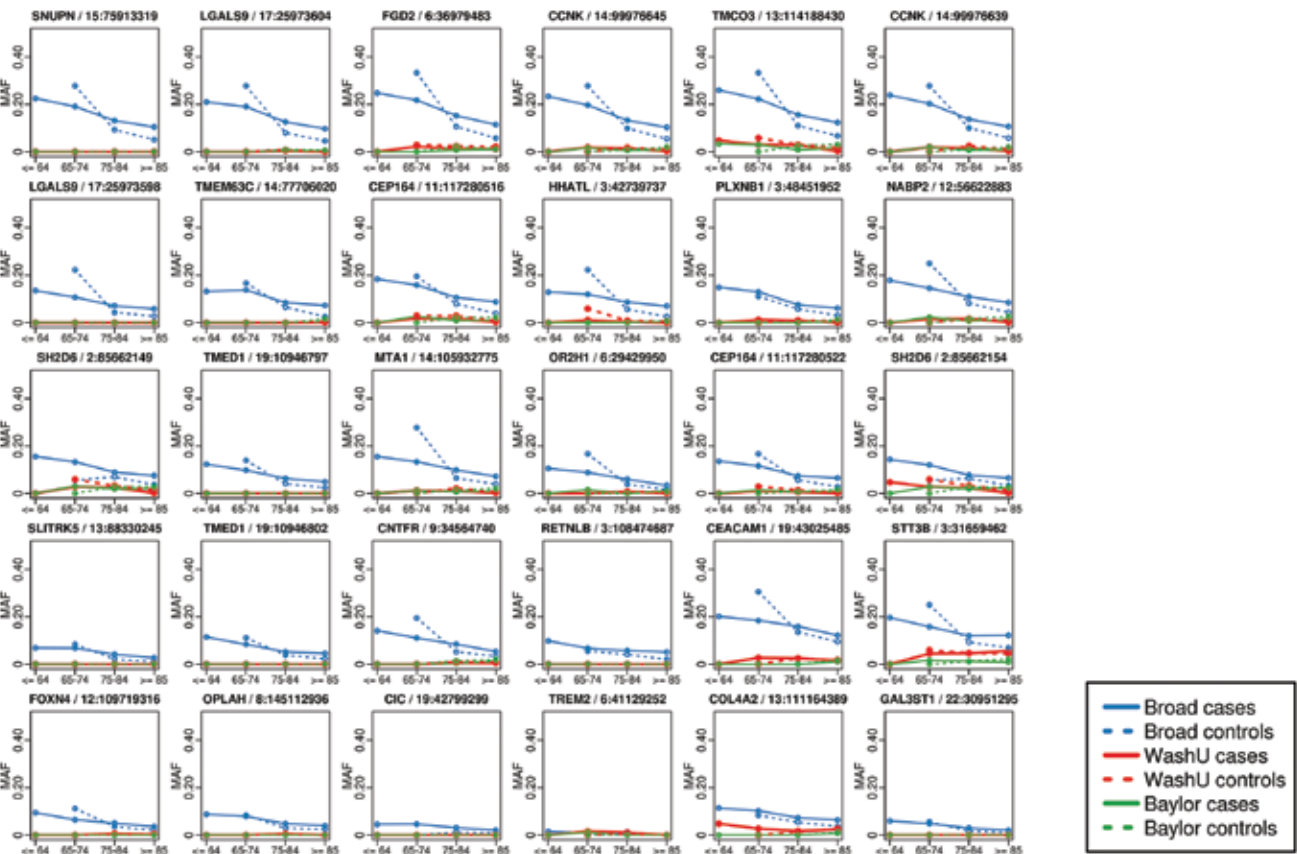


Figure 1: Minor allele frequencies in cases and controls stratified by sequencing facility and age group. Only the samples from Broad showed appreciable minor allele frequencies at the 30 variants identified as significant in the full-cohort analysis.

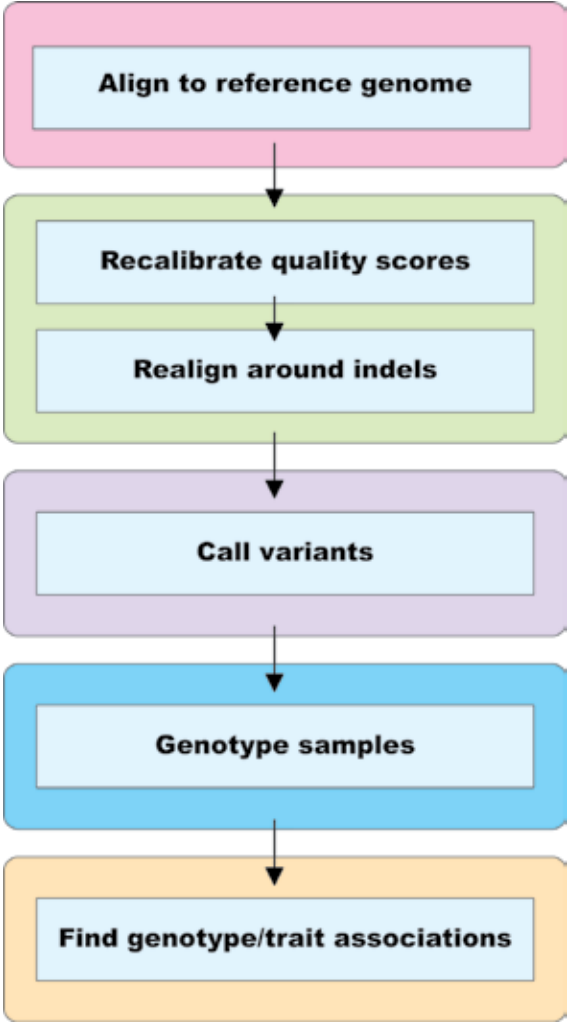


Figure 2: Workflow configuration used in this study.

es in study design, which may further hinder the ability to detect true genotype–trait associations. Common batch effects include different sequencing centers, different sample collection protocols, and different exome capture kits. For example, the Alzheimer’s Disease Sequencing Project sequenced exomes of more than 10,000 cases and controls using three sequencing centers and two exome capture kits. In addition, the controls were intentionally older than cases in an effort to increase the confidence of the Alzheimer’s variants because “true” disease-causal variants should be absent in older but cognitively normal individuals. This design introduced an age variable confounded with disease status. The research team studied both batch effects and confounding variables and demonstrated that both significantly impacted the association analysis of this data set.

## METHODS & CODES

Samples in this public data set were sequenced by three centers using two exome capture kits: Broad Institute (Illumina Rapid

Capture Exome kit), Washington University (Nimblegen VCRome v2.1 kit), and Baylor College of Medicine (Nimblegen VCRome v2.1 kit). The research team aligned paired-end reads to the hg19 and hg38 human reference genomes using Novoalign and BWA, and called variants using the Genome Analysis Toolkit. Following variant calling, the association analysis included only variants located in the common capture regions of the two exome kits. The variant associations with Alzheimer’s disease were adjusted for sequencing center, gender, the first four principal components underlying population substructure, and apolipoprotein E (APOE) genotype (a genetic risk factor for dementia).

## RESULTS & IMPACT

The research team identified 30 novel Alzheimer’s-associated genomic variants with exome-wide significance. Further examination showed that the significance of these variants originated entirely from samples processed at Broad, which used the Illumina capture kit. To investigate the cause of Broad-exclusive significance, the researchers compared multiple variant quality parameters including genotype quality, sequencing depth, and alternative allele concentration, and identified significant batch differences in genotype quality values of the variants from Broad (Illumina kit) compared to Washington University and Baylor (Nimblegen kit). The team also found several age-related differences between Broad and the other two centers. First, the Broad samples had higher minor allele frequency in both cases and controls, possibly indicating population stratification. Second, the Broad cohort consisted of a disproportionately large number of younger cases. Finally, the minor allele frequencies of the 30 variants declined with age in both Broad cases and controls, suggesting that the variants are associated with age.

## WHY BLUE WATERS

The Alzheimer’s Disease Sequencing Project data set used in this study consists of over 9,000 whole-exome sequencing samples. Processing this immense quantity of genomic data and conducting the downstream analysis required 250,000 node hours on Blue Waters. By parallelizing computational jobs across thousands of nodes, the research group was able to accomplish what would take over 100 years on a single server in months. In addition, Blue Waters is one of the few systems that allows users to keep hundreds of terabytes of data in active storage for simultaneous processing, an important step in data integration across genomic workflows.

## PUBLICATIONS & DATA SETS

Y. Ren *et al.*, “Identification of missing variants by combining multiple analytic pipelines,” *BMC Bioinform.*, vol. 19, 2018, doi: 10.1186/s12859-018-2151-0.

D. P. Wickland *et al.*, “Impact of batch effect and study design biases on identification of genetic risk factors in sequencing data,” in preparation, 2019.



DI

GA

# MICROSCOPIC IDENTIFICATION OF PIP<sub>2</sub> BINDING SITES ON A CA<sup>2+</sup>-ACTIVATED CL<sup>-</sup> CHANNEL

**Allocation:** Illinois/750 Knh  
**PI:** Tao Jiang<sup>1,2,3</sup>  
**Co-PIs:** Po-Chao Wen<sup>1,2,3</sup>, Kin Lam<sup>1,2,3</sup>, Zhiyu Zhao<sup>1,2,3</sup>, Archit Kumar Vasan<sup>1,2,3</sup>  
**Collaborators:** H. Criss Hartzell<sup>4</sup>, Kuai Yu<sup>4</sup>, YuanYuan Cui<sup>4</sup>

<sup>1</sup>University of Illinois at Urbana–Champaign  
<sup>2</sup>NIH Center for Macromolecular Modeling and Bioinformatics  
<sup>3</sup>Beckman Institute for Advanced Science and Technology  
<sup>4</sup>Emory University School of Medicine

## EXECUTIVE SUMMARY

Membrane proteins dwell in a sea of phospholipids that not only structurally stabilize the proteins by providing a hydrophobic environment but also dynamically regulate protein function. While many cation channels are known to be regulated by the negatively charged phosphatidylinositol 4,5-bisphosphate (PIP<sub>2</sub>), relatively little is known about anion channel regulation by phosphoinositides. Using atomistic molecular dynamics simulations on Blue Waters combined with experimental patch clamp electrophysiology, the research team has identified several PIP<sub>2</sub> binding sites in TMEM16A, a Cl<sup>-</sup> (chloride) channel that performs myriad physiological functions ranging from epithelial fluid secretion to regulation of electrical excitability. These PIP<sub>2</sub> binding sites form a band at the cytosolic interface of the membrane that the team proposes constitutes a network to dynamically regulate this extensively allosterically regulated protein. The microscopic description of the PIP<sub>2</sub>–TMEM16A interactions provided by this research adds a crucial layer of information for understanding the regulation mechanisms of ion channels by specific lipids.

## RESEARCH CHALLENGE

TMEM16A is a Ca<sup>2+</sup>-activated Cl<sup>-</sup> channel that regulates diverse cellular functions including fluid secretion, neuronal excitability, and smooth muscle contraction [1–3]. The channel is activated by elevation of cytosolic Ca<sup>2+</sup> and modulated by the anionic lipid PIP<sub>2</sub>. PIP<sub>2</sub> is known to be an important signaling lipid and critical for the regulation of a wide variety of ion channels [4–7]. Thus, understanding the mechanism by which PIP<sub>2</sub> binds to and affects membrane channels is of broad physiological and biophysical relevance. Although previous experimental studies have illustrated the regulatory role of PIP<sub>2</sub> on TMEM16A [8–10], it remains elusive how PIP<sub>2</sub> interacts and binds to the channel and how the binding affects molecular events underlying transport in the protein. To understand the regulatory mechanisms of TMEM16A by PIP<sub>2</sub>, it is important to identify amino acids involved in the lipid–protein interactions and characterize the conformational changes induced by these interactions. The obtained knowledge will help to unravel the atomic-level details underlying the functional mechanism of this highly allosteric protein and provide insight into the understanding of the mechanistically complicated TMEM16 superfamily.

Figure 1: The initial simulation system viewed from the membrane, in which a fraction of the acyl tails of the membrane-forming lipids is replaced by a liquid organic phase to enhance the lipid diffusion. (Bottom) Top views showing representative PIP<sub>2</sub> binding trajectories (left) and binding positions (right).

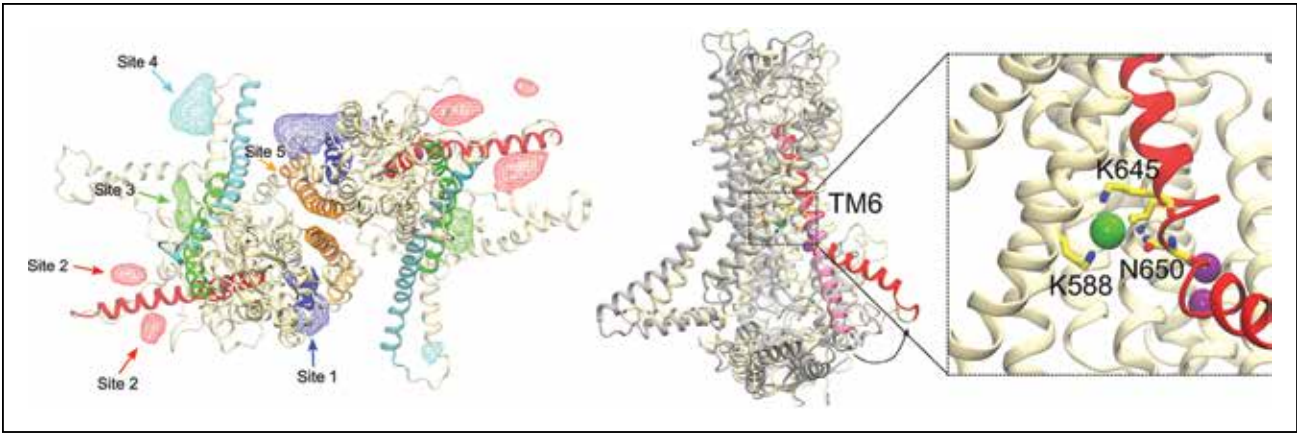
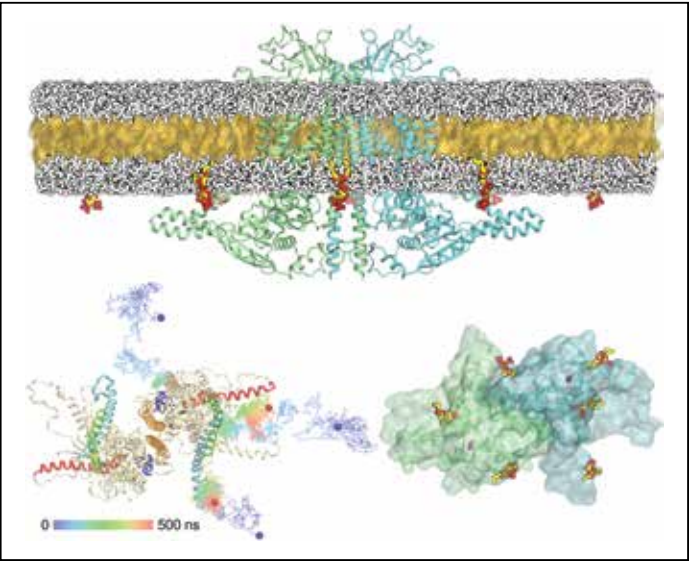


Figure 2: Maps of PIP<sub>2</sub> occupancy extracted from the simulations illustrate multiple PIP<sub>2</sub> binding sites near the functionally important transmembrane helices (individually colored). (Right) Representative snapshot showing the TM6 (transmembrane helix 6) conformational change in the presence of PIP<sub>2</sub>. The dilation of the pore allows Cl<sup>-</sup> (green sphere) to penetrate spontaneously.

## METHODS & CODES

To gain insight into the binding of PIP<sub>2</sub> to TMEM16A, extended molecular dynamics (MD) simulations were performed on the atomic model of the ion channel [11] using the highly mobile membrane mimetic model (HMMM) [12]. The HMMM model was introduced to accelerate lipid diffusion in atomistic simulations in order to obtain enhanced sampling of the interaction of lipid headgroups with proteins within simulation timescales currently achievable. This model replaces a portion of the membrane hydrophobic core by a more fluid representation using simple carbon solvent ethane (SCSE), while employing short-tailed lipids to maintain a full description of the headgroups and the initial part of the tails (Fig. 1). This model provides a more flexible and mobile environment that allows for rapid rearrangement and displacement of the lipid headgroups, thereby facilitating phenomena that might be inaccessible with conventional membrane models owing to the inherently slow dynamics of the lipids. In each of the six independent simulation systems, eight PIP<sub>2</sub> molecules were added to the inner leaflet of an otherwise phosphatidylcholine (POPC) lipid bilayer evenly surrounding the protein at the beginning of the simulations. To determine whether binding of PIP<sub>2</sub> was influenced by full-length acyl chains, after the completion of lipid-binding simulations with HMMM membranes (500 nanoseconds [ns] each), short-tailed lipid molecules were converted back to full-length lipids, and the resulting full systems were subjected to additional equilibrium simulations of 100 ns each. All MD simulations were carried out on Blue Waters using the NAnoscale Molecular Dynamics (NAMD) simulation package [13].

## RESULTS & IMPACT

The research team’s unbiased atomistic MD simulations with approximately 1.4% PIP<sub>2</sub> in POPC bilayers revealed spontaneous binding of PIP<sub>2</sub> to several potential sites on the surface of the TMEM16A channel (Fig. 1). Three of these sites captured 85%

of all PIP<sub>2</sub>–protein interactions and were validated to be critical for PIP<sub>2</sub> regulation through mutagenesis experiments by the collaborators. Simulations showed that PIP<sub>2</sub> is stabilized by hydrogen bonding between basic residues and the phosphate/hydroxyl groups on the inositol ring of the lipid headgroup. Binding of PIP<sub>2</sub> to different sites produces different conformational effects in the cytoplasmic part of transmembrane helix 6 (TM6), which forms one side of the channel pore and plays a key role in channel gating. The occupation of the major sites is especially shown to induce a dramatic rotation of the cytoplasmic end of TM6 away from the pore (Fig. 2). This pore dilation increases the accessibility of the inner vestibule of the channel to the cytosolic ions and resulted in spontaneous penetration of Cl<sup>-</sup> ions into the pore (Fig. 2). Based on this observation, the research team proposed that a network of PIP<sub>2</sub> binding sites at the cytoplasmic face of the membrane allosterically regulates channel gating. The data provided by these simulations add to a growing body of knowledge showing that TMEM16A is a highly allosteric protein that is gated by a network of interactions involving both Ca<sup>2+</sup> and PIP<sub>2</sub>.

## WHY BLUE WATERS

The state-of-the-art architecture of Blue Waters makes it an excellent computing resource for this scientific research. The GPU-optimized simulation package NAMD has been extensively tested and optimized for Blue Waters. The large number of GPUs available on the XK nodes significantly increased the overall productivity. In addition, the technical support provided by the experts and scientists of the Blue Waters team has contributed to the accomplishment of the research goals by smoothing out technical issues that have arisen during the allocation.

## PUBLICATIONS & DATA SETS

K. Yu, T. Jiang, Y. Cui, E. Tajkhorshid, and H. C. Hartzell, “A network of phosphatidylinositol 4,5-bisphosphate binding sites regulate gating of the Ca<sup>2+</sup>-activated Cl<sup>-</sup> Channel ANO1 (TMEM16A),” *bioRxiv*, May 2019, doi: 10.1101/625897.



PARACELLULAR ION TRANSPORT

**Allocation:** GLCPC/590 Knh  
**PI:** Fatemeh Khalili–Araghi<sup>1</sup>  
**Collaborator:** Christopher Weber<sup>2</sup>

<sup>1</sup>University of Illinois at Chicago  
<sup>2</sup>University of Chicago

EXECUTIVE SUMMARY

Permeation of water, ions, and small molecules through the space between neighboring cells is controlled by macromolecular structures known as tight junctions. Tight junctions seal the paracellular space and act as barriers that limit diffusion of molecules down their electrochemical gradient. Claudins are one of the major components of tight junctions and play a key role in determining paracellular permeability. Little is known about the assembly of claudins and the architecture of tight junction pores. The research team built an atomic model of claudin pores and verified its function using molecular dynamics (MD) simulations. The team then used the MD simulations to build a simple model of tight junction networks and simulate their transport properties. However, the architecture of tight junctions at the cellular level is still unknown.

RESEARCH CHALLENGE

Claudin pores are one of the major components of the tight junctions that control the transport of ions and small molecules in paracellular space between neighboring cells [1,2]. Little is known about the molecular architecture of tight junctions and the assembly of components into ion channels. The research team used all-atom MD simulations to determine: (1) the structure of claudin pores and their functional mechanism, and (2) the mechanical properties of tight junction strands at the cellular level.

The simulations carried out in this project are among the largest simulations of ion channels to date. The tight junction networks consist of a few hundred ion channels assembled into linear strands of tiny pores in two parallel lipid membranes.

METHODS & CODES

The researchers ran atomic-scale MD simulations of claudin pores in two parallel lipid membranes. The highly scalable MD program NAMD was used to build and refine the model and to simulate its ion transport function. Moreover, to investigate the macroscopic properties of tight junctions at relevant length scales (micrometers), the team ran simulations of the system using a hybrid resolution representation using the PACE force field, in which the protein was represented atomically and its environment, including lipid membranes and solvent, were coarse-grained. The size of the systems simulated in this project range from 350,000 atoms to two million particles.

RESULTS & IMPACT

In this project, the team developed the first atomic model of a new class of ion channels: claudin pores. They used MD simulations to build and refine an atomic model based on a recently proposed architectural model [3,4] and to assess its stability. Furthermore, they verified this structural model by simulating its ion transport function. The simulations verified functional characteristics of claudin pores such as their charge and size selectivity and predicted mutations that reversed the charge selectivity of the channel. These mutations were further verified in electrophysiology experiments in the collaborator’s laboratory. In addition, these simulations identified the molecular nature of ion selectivity in paracellular pores, which was again verified by experiments [5].

To investigate ion transport across tight junctions—the parallel networks of claudin pores that span the cell membrane—the team developed systems consisting of more than 180 claudins (44 pores) in two parallel membranes. Initial simulations of the system at equilibrium indicate that claudins form flexible strands with persistence lengths comparable to those obtained experimentally (approximately 200 nm). Furthermore, the simulations determined the origin of this flexibility and pairwise protein–protein interactions that are responsible for the formation of strands and their shape.

These are the first simulations of ion transport in paracellular pores ever conducted, and they have opened up new opportunities for studying the functional mechanism of these channels as well as the physical properties of tight junctions at the cellular level. Future studies could result in the development of possible inhibitors for this class of ion channel in the small intestines or kidneys, or enable the delivery of drugs across the blood–brain barrier.

WHY BLUE WATERS

Access to Blue Waters was essential in running simulations consisting of the assembly of several hundreds of proteins into functional ion channels. These simulations, which reached a few micrometers in length, were only possible through access to the large number of nodes available on Blue Waters. Equally important was the knowledge of the Blue Waters staff in compiling the codes and helping the research team to identify performance issues.

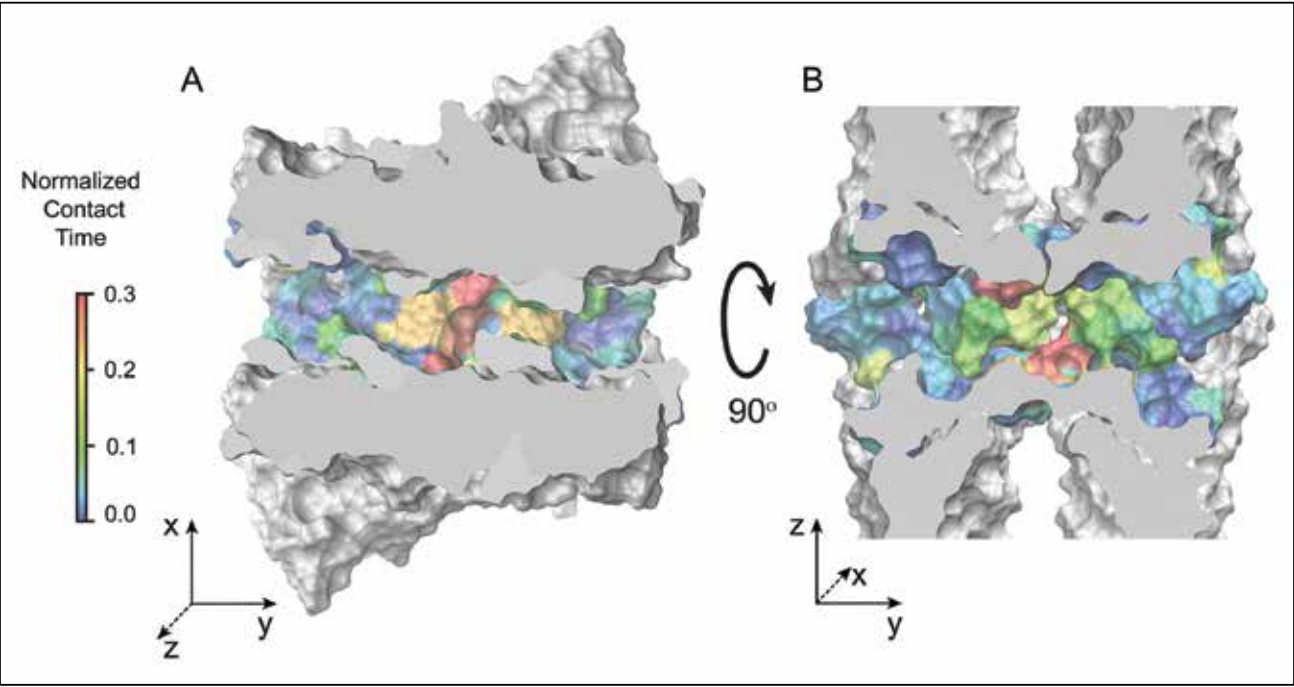


Figure 1: Ion transport pathway in claudin pore. Cross-section of a claudin pore across two parallel planes: (A) the plane parallel to the cellular membrane, and (B) the plane perpendicular to the cellular membrane. The amino acids on the pore surface are colored based on their contact time with permeating cations.

PUBLICATIONS & DATA SETS

P. Samanta *et al.*, "Molecular determination of claudin-15 organization and channel selectivity," *J. Gen. Physiol.*, vol. 150, no. 7, pp. 949–968, Jun. 2018, doi: 10.1085/jgp.201711868.



DI

THE TERRA PHENOTYPING REFERENCE PLATFORM: OPEN DATA AND SOFTWARE FOR PRECISION FIELD CROP MEASUREMENT AND ANALYSIS

**Allocation:** Director Discretionary/5 PB storage  
**PI:** David LeBauer<sup>1</sup>  
**Co-PIs:** Todd Mockler<sup>2</sup>, Robert Pless<sup>3</sup>, Roman Garnett<sup>4</sup>, Vasit Sagan<sup>5</sup>, Rob Kooper<sup>6</sup>, Duke Pauli<sup>1</sup>, Geoff Morris<sup>7</sup>

<sup>1</sup>University of Arizona  
<sup>2</sup>Donald Danforth Plant Science Center  
<sup>3</sup>George Washington University  
<sup>4</sup>Washington University of St. Louis  
<sup>5</sup>St. Louis University  
<sup>6</sup>National Center for Supercomputing Applications  
<sup>7</sup>Kansas State University

EXECUTIVE SUMMARY

The Transportation Energy Resources from Renewable Agriculture Phenotyping Reference Platform (TERRA-REF) aims to optimize breeding strategies for improving the yield and stress tolerance of sorghum. Sorghum (*Sorghum bicolor*) is a food and energy crop with high yield potential, efficient water use, and drought tolerance and has the capacity for genetic improvement using both traditional and genomic approaches. The water-use efficiency and drought tolerance of sorghum make it adaptable to future climates that are expected to be hotter and drier.

RESEARCH CHALLENGE

Improving crops is challenging, but humankind needs crops that are more productive, stress tolerant, and more efficient at using resources such as water and fertilizer. Intensive measurements of crop plants and their environment provide plant breeders with information that is required to select for these traits. Automated measurements can help breeders select the best plants more quickly and efficiently. Although new sensors and robotic platforms are becoming available, these relatively new tools are difficult for scientists to use to gather actionable information.



Figure 1: Field scanner system (with sensors in white box) operating on 200-meter steel rails in Maricopa, Arizona, U.S.A.

TERRA-REF is collecting automated high-throughput remote sensor data on diverse varieties of sorghum and wheat. Remote sensors mounted in a field scanner instrument box scan over one acre of plants, producing thousands of daily measurements with high spatiotemporal resolution. Sensor data types range from hyperspectral imaging to 3D reconstructions and thermal profiles, all at 1-mm resolution. The program is making the data available for researchers not only to study the plants but also the sensors themselves in order to learn what new information these sensors provide.

METHODS & CODES

Data from the sensors is streamed from the field scanner to the National Center for Supercomputing Applications, where it is calibrated and processed. Blue Waters' nearline tape storage is used both for backup and for staging of data for reprocessing. The data is then organized, annotated, and processed. This process uses RabbitMQ to recognize and handle each data set as it arrives or is created. Clowder extractors created using novel predictive models and algorithms process the raw data into high-resolution images, point clouds, and time series, and then extract features related to plant growth, chemistry, and physiological efficiency. Clowder stores metadata and tracks provenance for derived data sets.

Once reduced to plant and plot-level summaries, the means, trait data, and agronomic metadata are stored in the Biofuel Ecophysiological Traits and Yields database (BETYdb). Manually collected trait data and agronomic data are also imported into BETYdb for algorithm training, calibration, and validation.

RESULTS & IMPACT

All software used in the processing pipeline is available on GitHub with permissive open-source licenses. Similarly, curated data sets generated by this platform will be published and placed in the public domain to maximize the impact of these unprecedented data.

WHY BLUE WATERS

The Blue Waters nearline system provides TERRA-REF with a quickly accessible and reliable storage system for both data backup and staging data for reprocessing. Having reliable storage for backup allows the research team to focus on the data processing and delivery infrastructure.

PUBLICATIONS & DATA SETS

M. Burnette *et al.*, "TERRA-REF data processing infrastructure," in *Proc. PEARC 2018*, doi: 10.1145/3219104.3219152.  
*A Reference Phenotyping Platform for Crop Breeding and Improvement*. [Online]. Available: <https://github.com/terraref>, Accessed on: Jul. 26, 2019.



Figure 2: An algorithm to count and measure sorghum panicles (seed heads) helps agricultural researchers understand genes that increase yields. It also helps plant breeders more efficiently develop and commercialize more productive and sustainable crops that will tolerate future climates.



# QUANTUM-CLASSICAL PATH INTEGRAL SIMULATION OF PROTON TRANSLOCATION IN BIOLOGICAL CHANNELS

Allocation: Illinois/75 Knh  
PI: Nancy Makri<sup>1</sup>

<sup>1</sup>University of Illinois at Urbana-Champaign

## EXECUTIVE SUMMARY

The PI uses the quantum-classical path integral (QCPI) methodology, along with the modular decomposition of the path integral, to perform highly accurate simulations of proton translocation dynamics in water chains embedded in biological channels. These methods allow an accurate treatment of important quantum mechanical effects, accounting for the effects of the protein environment in full atomistic detail.

## RESEARCH CHALLENGE

Understanding the mechanism of the proton translocation along water chains embedded in biological channels continues to attract much theoretical and experimental effort. An interesting question is related to whether the dynamic is sequential or concerted. The accurate treatment of quantum effects, along with a faithful treatment of phase interference in the interaction of the protons with the environment, is critically important for reaching definitive conclusions in this regard. Unfortunately, quantum mechanical calculations scale exponentially with the number of coupled degrees of freedom.

## METHODS & CODES

The PI's initial calculations on a protonated water dimer are based on the QCPI [1-7] methodology she developed, with several recent advances to her code. Specifically, its ability to describe solvents has been increased with the addition of a molecular dynamics package, developed by her for QCPI, that contains an implementation of the CHARMM force field for protein environments along with accurate potentials for water clusters as well as the ability to simulate various subsystems, each with its own force field. Another development of the methods is the flexibility of augmenting the system Hamiltonian with matrix elements that depend on time through select solvent coordinates. The extension of this work to treat longer chains with multiple transferring protons will utilize the modular decomposition of the path integral [8-10] developed recently by the PI.

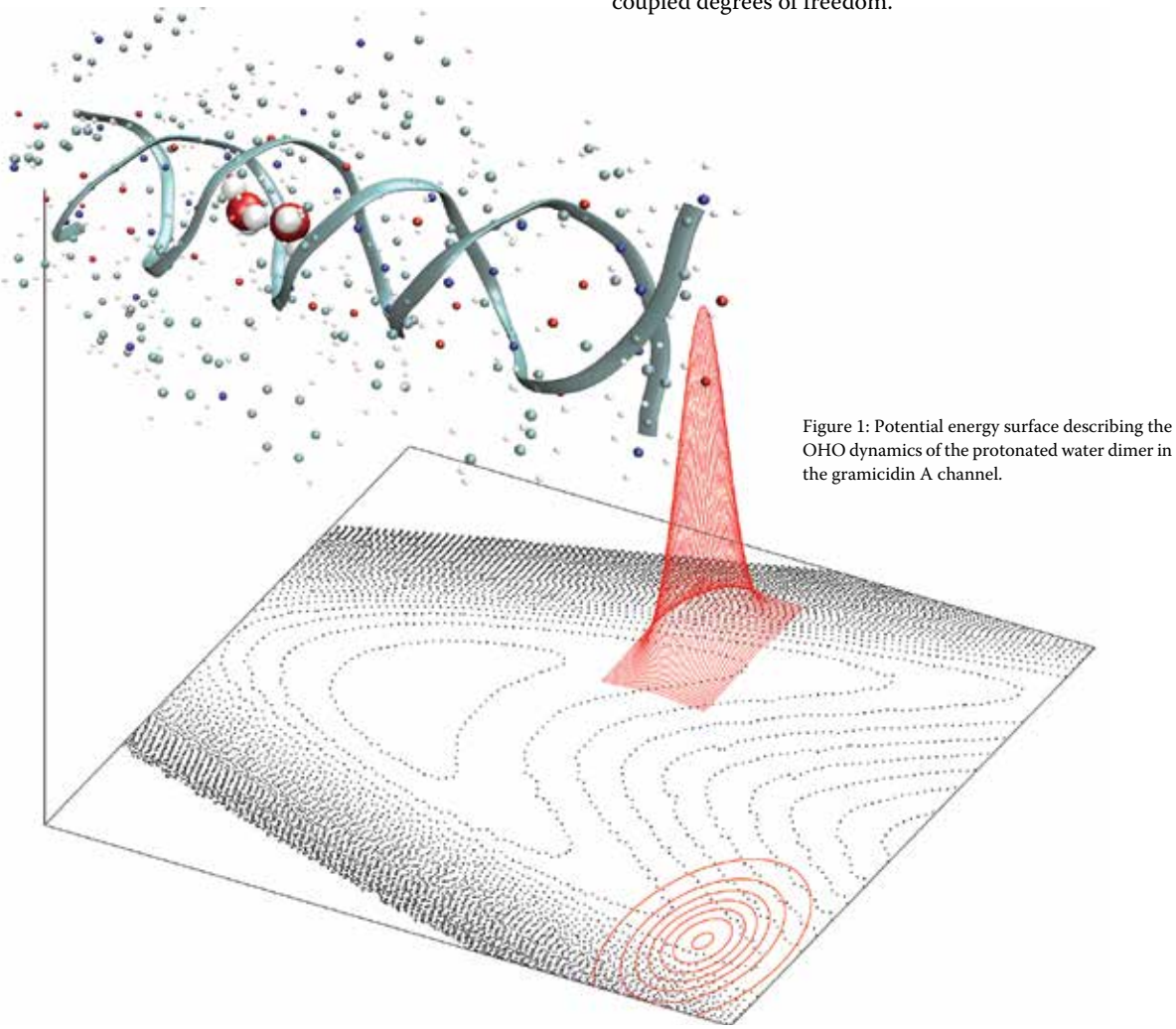
## RESULTS & IMPACT

This project involves the simulation of the dynamics of the protonated water dimer in the gramicidin A protein channel. The problem of the translocation of a proton along a water chain has fascinated scientists for decades. Since the discovery of water chains in protein channels, such a system has been considered a model for a proton pump that helps maintain osmotic pressure between a cell and its environment. The diffusivity of the proton is in fact a fast process that can be accounted for if one considers not only quantum tunneling but also its interplay with the thermal fluctuations of the oxygen atoms of the chain and the coupling to the protein that surrounds it.

Using the QCPI code, the PI has simulated the transfer dynamics of the excess proton injected in a neutral water dimer, from which she estimates a population decay on a timescale of shorter than 0.1 picosecond. This process is driven in great part by the strengthening of the hydrogen bond that pulls the oxygen atoms closer. This picture becomes more complicated in longer chains, where the motion of the oxygen atoms is correlated, and there are more reaction pathways for the proton diffusion that may correspond to a shuttling motion or a concerted displacement of multiple hydrogen atoms. The researcher plans to address these questions by simulating the dynamics of proton translocation in longer chains, taking into account the quantum mechanical character of multiple transferring protons using the modular decomposition of the path integral methodology.

## WHY BLUE WATERS

The QCPI calculations require a multilevel parallelism. Blue Waters provides the ideal platform for the implementation of the algorithm.





A NEW STABILIZED FLUID-STRUCTURE INTERACTION METHOD: COUPLED SYSTEM OF ANISOTROPIC VISCOELASTIC MODEL FOR ARTERY AND NON-NEWTONIAN MODEL FOR BLOOD

Allocation: Illinois/50 Knh  
PI: Arif Masud<sup>1</sup>  
Collaborator: Soonpil Kang<sup>1</sup>

<sup>1</sup>University of Illinois at Urbana-Champaign

EXECUTIVE SUMMARY

Fluid-structure interaction (FSI) is a class of multiphysics problems that combines fluids and solids in a single-pass simulation to capture the interactive and integrated behavior of the system. Because of the instabilities that are unique to nonlinear material models, especially in the presence of moving interfaces, a comprehensive strategy for FSI requires mathematical formulations with enhanced stability properties and coupled solution algorithms that preserve the dissipative structure of the underlying coupled continuum problem.

The research team has developed a stabilized monolithic method for coupling incompressible non-Newtonian fluids [1,3,4] with anisotropic (having a physical property that has a different value when measured in different directions) viscoelastic (materials that exhibit both viscous and elastic characteristics when undergoing deformation) models of artery walls. The method is applied to an idealized curved artery to investigate the mathematical attributes of the models as well as that of the coupled solution algorithm. The algorithm and code have been optimized on the XE nodes of Blue Waters.

RESEARCH CHALLENGE

Realistic, patient-specific models help not only to simulate pre-operative diseased configurations but also to analyze postoperative outcomes. This has evolved into the concept of computational medicine, a form of personalized medicine in which patient-specific computer modeling and engineering analysis methodologies are used to noninvasively diagnose and evaluate the efficacy of various possible treatments and to plan and design the optimal intervention based on prediction of outcomes. However, creating spatial discretizations for FSI problems that satisfy nodal

compatibility is not always an easy task and, therefore, considerable effort has been devoted to numerical methods that accommodate nonmatching interfaces [2]. Flexibility to accommodate nonmatching meshes is of great practical value in problems of industrial strength that invariably have complex geometric configurations. Relaxing the nodal continuity requirements necessitates techniques for enforcing the conditions on the continuity of the fields across the interface. The research team followed the ideas proposed in Truster and Masud [5,7] and developed an interface-stabilized method with least-squares-type terms that enforce continuity of traction at the nonmatching meshes along the fluid-solid interface.

Accurate prediction of stress and deformation in the arterial wall in patient-specific applications requires that physiologically relevant constitutive models [6] are employed for the artery wall, which is comprised of soft tissue with embedded collagen fibers. Anisotropy caused by embedded fibers in soft biological tissues is a more difficult numerical problem, and it becomes a challenge for numerical methods when large deformations are involved in an FSI simulation. In the present work, multiple layers of artery wall, intima (the innermost layer of an artery or vein), media, and adventitia (the outermost layer of the wall of a blood vessel) are modeled via a hyperelastic energy functional that accounts for finite stretching of the soft tissue as well as anisotropy induced by the directionally oriented collagen fibers. The reinforcing fibers are laid helically around the artery wall with alternate layers placed orthogonally to one another to form a network of directionally oriented layers.

W(C)=ε1(I1^ε2+I2^ε2-2)+c1(I1I2^1/3-3)+∑i=1Nαi(IiJ1^1/3-J1^1/3-2)^αi

where I1=trC, I2=detC, J1=tr[CM], and J2=tr[C^2M] are invariants of the right Cauchy-Green deformation tensor; M=a⊗a is a second-order tensor that models anisotropy of the fibrous material with a direction vector a; ε1, ε2 and c1 are the material parameters for a neo-Hookean type solid; and α1 and α2 are material parameters for reinforcing fibers.

METHODS & CODES

Numerical methods, which are not constrained by node-on-node matching, provide great advantage in developing patient-spe-

cific computational models. The researchers have employed a variational multiscale framework to develop advanced numerical techniques with enhanced stability and accuracy properties for this class of problems. The team's emphasis has been on the development of a unified mathematical framework that can have wider application both in the domain of fluids as well as in solids. The enhanced stability facilitated by the mathematical constructs results in a robust FSI algorithm that has been applied to the blood-artery interaction problem. The method is implemented in the context of a finite-element method using low-order Lagrangian elements and has been optimized on the XE nodes of Blue Waters.

RESULTS & IMPACT

Fig. 1 shows the stress-strain response for a representative material volume of the artery wall to illustrate its anisotropic response under axial stretching. A parametric study was carried out with biologically relevant values of the material coefficients, and the nonlinear stress-strain response is presented in Fig 1. The sample was first loaded in the axial direction that is aligned with the direction of the fibers and a gradual stiffening response was observed. Subsequently, the unit cube was loaded in the lateral direction and the stress-carrying capacity was reduced by five times. In addition, a gradual softening response was observed, as shown in Fig. 1b.

Fig. 2 shows the curved geometric configuration of the blood-artery model. The unidirectional collagen model was put in layers to create a network that is shown by the red and blue fibers in Fig. 2a. The mesh was comprised of hexahedral elements where the red region represents the fluid subdomain. The researchers simulated a few cardiac cycles and Fig. 2c presents an instantaneous snapshot of the deformation of the artery wall along with the streamlines of the blood flow. The team also projected the arterial wall shear stress (WSS) on the artery wall, which is one of the most significant factors affecting the progression of arterial disease. Since it is difficult to obtain spatiotemporal WSS data via *in vivo* experiments, advanced FSI simulations with appropriate constitutive models provide a virtual platform to facilitate important and insightful information for the diagnosis and treatment of arterial disease. Such information can also be critical at the planning stage for developing patient-specific strategies for surgical intervention.

WHY BLUE WATERS

The coupled solution algorithm for nonmatching FSI meshes was implemented on the Blue Waters platform and tested on XE nodes. Since element-level developments are all local to individual elements both in fluids and in solids, this part is easily and efficiently parallelized. However, the interface coupling terms need special attention as they require information from both fluid and solid subdomains across the interface. This implementation takes advantage of the local memory on the processing node, thereby expediting the calculation of element-level matrices and vec-

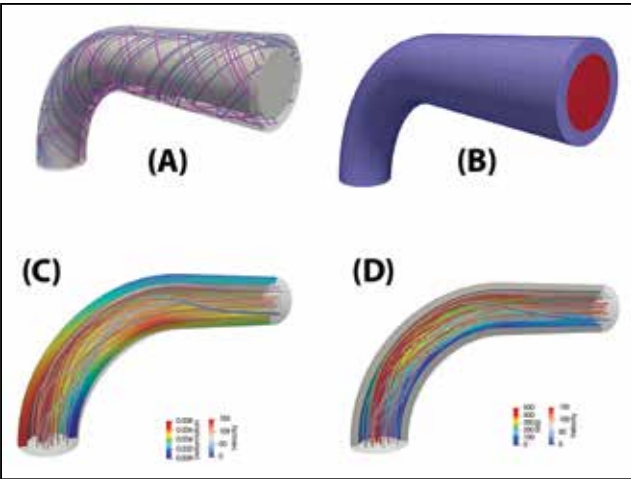


Figure 2: (a) curved artery model reinforced with helically arranged collagen fibers; (b) computational mesh for blood and artery wall; (c) deformation of artery wall with superposed streamlines; (d) wall shear stress and streamlines of blood flow.

tors. Preliminary tests confirm the robustness of the method for highly nonlinear problems. This model will now be applied to patient-specific geometry to see the scalability of the FSI method with nonmatching meshes to problems of clinical relevance.

PUBLICATIONS & DATA SETS

M. Anand, J. Kwack, and A. Masud, "A new generalized Oldroyd-B model for blood flow in complex geometries," *Int. J. Eng. Sci.*, vol. 72, pp. 78-88, 2013.

P. Chen, T. J. Truster, and A. Masud, "Interfacial stabilization at finite strains for weak and strong discontinuities in multi-constituent materials," *Comput. Methods Appl. Mech. Eng.*, vol. 328, pp. 717-751, 2018.

J. Kwack and A. Masud, "A three-field formulation for incompressible viscoelastic fluids," *Int. J. Eng. Sci.*, vol. 48, pp. 1413-1432, 2010.

J. Kwack, A. Masud, and K. R. Rajagopal, "Stabilized mixed three-field formulation for a generalized incompressible Oldroyd-B model," *Int. J. Numer. Methods Fluids*, vol. 83, no. 9, pp. 704-734, 2017.

T. J. Truster and A. Masud, "Primal interface formulation for coupling multiple PDEs: A consistent derivation via the Variational Multiscale method," *Comput. Methods Appl. Mech. Eng.*, vol. 268, pp. 194-224, 2014.

T. J. Truster and A. Masud, "A unified mixture formulation for density and volumetric growth of multi-constituent solids in tissue engineering," *Comput. Methods Appl. Mech. Eng.*, vol. 314, pp. 222-268, 2017.

T. J. Truster and A. Masud, "A discontinuous/continuous Galerkin method for modeling of interphase damage in fibrous composite systems," *Comput. Mech.*, vol. 52, no. 3, pp. 499-514, 2013.

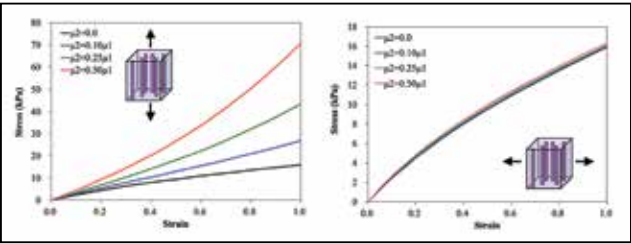


Figure 1: Anisotropic tissue/collagen-fiber model for the artery wall.



# ATOMIC SCALE SIMULATION OF AMYLOID BETA WITH DISMANTLING PEPTIDE-BASED INHIBITORS

**Allocation:** Illinois/715 Knh  
**PI:** Jeffery S. Moore<sup>1</sup>  
**Co-PIs:** Giuseppe Licari<sup>1</sup>, Xing Jiang<sup>1</sup>, Andres Arango<sup>1</sup>, Jimmy Do<sup>1</sup>

<sup>1</sup>University of Illinois at Urbana–Champaign

## EXECUTIVE SUMMARY

Aggregation of amyloid beta (AB) proteins plays a fundamental role in Alzheimer’s disease. Several inhibitors of AB aggregation have been proposed in the literature, but no effective treatment is yet available. In this work, the research team introduced novel multivalent polymer–peptide conjugates (mPPCs) that can dis-aggregate or inhibit AB formation. Moreover, the team recently evaluated new macrocyclic peptidomimetic (MP) libraries that can also hinder AB aggregation.

The researchers used all-atom molecular dynamics (MD) sim-ulations to investigate how mPPCs and MPs bind AB fibrils and alter their stability. Simulations of mPPCs alone or in the pres-ence of AB show that mPPCs self-aggregate in solution, but in the presence of AB they strongly interact with AB through both their peptide and backbone moieties. Furthermore, a docking analysis based on the simulation results reveals that inhibitors may desta-bilize AB through introducing defects in the fibril.

## RESEARCH CHALLENGE

Protein aggregation is implicated in major pathophysiological conditions including Alzheimer’s and Parkinson’s disease [1]. The occurrence of Alzheimer’s disease, in particular, has been linked to the formation of toxic aggregates of amyloid beta (AB) pep-tides. Although several inhibitors have been developed, the aggre-gation mechanism is still poorly understood. The research team has recently developed a novel class of hydrophilic, high molec-ular weight polymers, mPPCs, that bear multiple copies of pep-tides and can inhibit AB fibril formation (Fig. 1) [2,3]. This mul-tivalency strategy uses multiple simultaneous interactions of the ligand with AB to enhance the affinity. Another strategy involves the use of MP that are stable under physiological conditions and favor interactions with AB owing to their preorganized struc-ture. However, little is known about how these molecular archi-tectures interact with the target protein and what can be done to enhance their effect and increase their specificity.

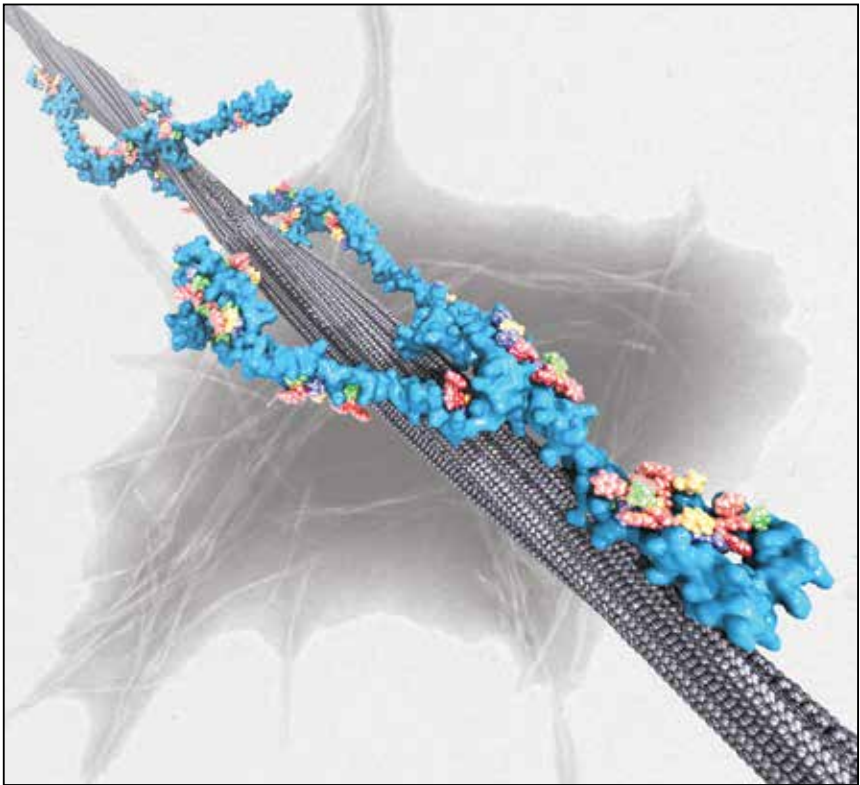


Figure 1: Molecular model of a putative inhibitory mechanism of a polymer inhibitor (backbone in cyan and active peptides in multicolor) wrapping around an amyloid beta fibril. The image in the background illustrates fibril aggregates forming an amyloid beta plaque.

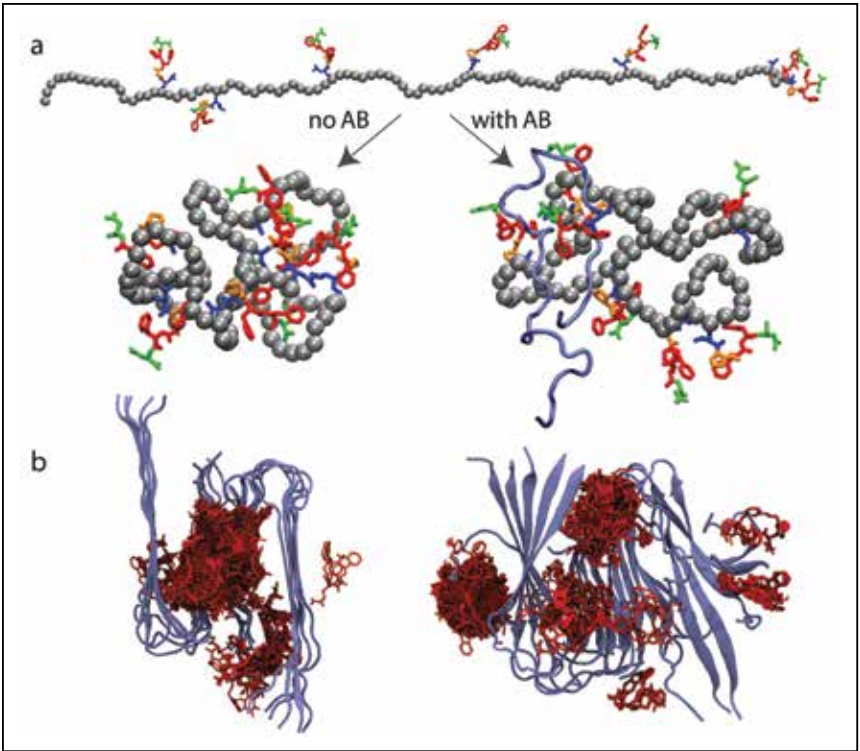


Figure 2: (a) A representative inhibitor polymer composed of 100 monomers (black beads) and multiple peptides (multicolor). In the absence of amyloid beta, the polymer forms a pseudospherical aggregate. With amyloid beta (purple strand), the polymer wraps around the proteic strand. (b) Docking poses (red) of a macrocyclic peptidomimetic for two different conformations of an amyloid beta oligomer.

## METHODS & CODES

All MD simulations were performed using NAMD [4], a GPU-accelerated code that has been optimized to run on Blue Waters. In-house scripts using VMD generated mPPC models with different compositions, molecular weights, and peptide load-ings [5]. The latter software can also be used to remotely visual-ize trajectories on Blue Waters. The AB oligomer was obtained from the Protein Data Bank (code: 2LMN).

## RESULTS & IMPACT

The simulations of mPPC polymers in solution highlighted that all polymers form pseudospherical self-aggregates within a few hundreds of nanoseconds (Fig. 2a). This self-aggregation re-sulted from hydrophobic and hydrogen bond interactions among components of the polymers. In particular, the size of mPPC ag-gregates did not increase monotonously with the peptide loading but rather decreased at high values. The size contraction of mP-PCs was ascribed to the increase of hydrophobic contacts as the number of peptides grafted onto the polymer increased, causing a more tightly packed aggregate. Interestingly, in the presence of AB, mPPCs wrap around the protein and form a stable complex. From the analysis of contacts and hydrogen bonds, the research team concluded that stabilization of this complex was significant-ly enhanced by the mPPC-backbone interaction with AB, con-firming the effectiveness of the multivalent design of the poly-mers. These analyses also revealed that proline residues of mP-PCs contribute the most to interactions with AB.

MD simulations performed on an AB oligomer provided mul-tiple conformations of the protein. Molecular docking of an MP performed on these conformations (Fig. 2b) showed that the in-hibitors can intercalate inside defects generated owing to ther-mal fluctuations, as captured during the simulation, in the AB secondary structure. Perturbed beta-sheet structures of AB can allow the insertion of inhibitors, which in turn will weaken and eventually dismantle the fibril structure.

Overall, the insights obtained from these simulations provide a molecular-level mechanism associated with the inhibition of AB aggregation and will support the design of new inhibitors.

## WHY BLUE WATERS

The resources provided by Blue Waters in terms of computa-tional power and network performance were essential to carry-ing out this project because of the need for simulation of a large data set. The large computer allocation allowed for the simula-tion of multiple systems, some of which were replicated multi-ple times with varying initial conditions to enhance sampling.

## PUBLICATIONS & DATA SETS

X. Jiang *et al.*, “Multivalent polymer–peptide conjugates—a general paradigm for inhibiting amyloid beta peptide aggrega-tion,” in preparation, 2019.

X. Jiang *et al.*, “Amyloid beta peptide aggregation inhibitors from a macrocyclic peptidomimetics library,” in preparation, 2019.



TRANSPORT MECHANISM OF POT TRANSPORTERS: EMPLOYING LOOSELY COUPLED MOLECULAR DYNAMICS SIMULATIONS TO CHARACTERIZE PROTEIN STRUCTURAL DYNAMICS

Allocation: GLCPC/576 Knh  
PI: Mahmoud Moradi<sup>1</sup>

<sup>1</sup>University of Arkansas

EXECUTIVE SUMMARY

Proton-coupled oligopeptide transporters (POTs) use the inwardly directed proton flow to uptake small peptides and peptide-like molecules. The human POT transporters PepT1 and PepT2 provide the main route through which the body absorbs and retains dietary proteins. Human POTs also recognize several important families of peptidelike drug compounds such as  $\beta$ -lactam antibiotics. POTs undergo large-scale conformational changes that are the key in understanding the transport mechanism of these proteins. Despite many experimental and computational efforts, however, the inward- (IF) to outward-facing (OF) structural transition of POTs has remained elusive owing to limitations in methodology. Therefore, the researcher has employed novel molecular dynamics (MD)-based enhanced sampling techniques to characterize the large-scale conformational changes of a bacterial POT transporter, namely GkPOT. With the help of petascale supercomputing, these MD-based techniques provide a detailed description of GkPOT's conformational landscape, which sheds light on the structure–function relationship in POT transporters.

RESEARCH CHALLENGE

Membrane transporters provide the machinery to couple active transport of materials to various forms of cellular energy. POT transporters couple the energy from proton flow to the transport of small peptides and peptidelike molecules [1]. A key feature of POTs is their substrate promiscuity [2], which is of great interest from a biomedical perspective. Human POT transporters PepT1 and PepT2, which play a key role in absorbing and retaining dietary proteins [3], recognize several important families of peptidelike drugs such as  $\beta$ -lactam antibiotics [4]. These proteins can uptake poorly absorbed/retained drugs when attached to amino acids or dipeptides as in prodrugs [5].

Recent structural studies have resulted in several crystal structures of bacterial POTs [6–9] such as GkPOT from *Geobacillus kaustophilus* [10]. These crystal structures, which are all in the IF state, provide the basis of our understanding of POTs' transport mechanism at the structural level. However, in line with the alternating access mechanism, POTs alternate between two distinct states; *i.e.*, the IF and OF states. The conformation of the OF state and the transition pathway between the two functional states have not been experimentally characterized.

Previous equilibrium MD simulations have failed to characterize large-scale conformational changes of POTs [10]. While conventional MD can provide information on local conformational changes of a protein upon binding or unbinding of a substrate, ion, or proton, the global conformational changes observed are not often statistically significant [11]. Functionally important conformational changes such as the IF–OF transition in membrane transporters typically occur on timescales beyond those accessible to conventional all-atom MD. The large-scale conformational changes, on the other hand, are usually studied using simplified modeling techniques such as coarse-graining, which could completely ignore or misrepresent the role of chemical events in the transport process. The main challenge in characterizing the large-scale conformational changes of proteins such as those associated with GkPOT is to reach the functionally relevant timescales without compromising the chemical details.

METHODS & CODES

The PI has used a novel ensemble-based simulation approach [12–15] to reconstruct the entire transport cycle of GkPOT. Bias-exchange umbrella sampling (BEUS) and string method with swarms of trajectories (SMwST) are both loosely coupled MD-based algorithms that require parallel execution of hundreds of MD simulations [14] and have recently been modified within a Riemannian geometry framework [15]. The methodology is specifically based on applying orientation-based forces on protein transmembrane helices in order to speed up the exploration of protein conformational space.

The software engine used for the simulations is NAMD, a highly scalable MD code implemented in Charm++, an object-based message-driven execution system based on C++. NAMD has been enhanced to support extremely scalable loosely coupled multiple-copy algorithms. Multiple concurrent NAMD instances are launched with internal partitions of Charm++ and located continuously within a single communication world. Messages between NAMD instances are passed by low-level point-to-point communication functions, which are accessible through NAMD's Tcl scripting interface.

RESULTS & IMPACT

The OF structure shown in Fig. 1 represents the first OF model of POT transporters and was generated using the researcher's

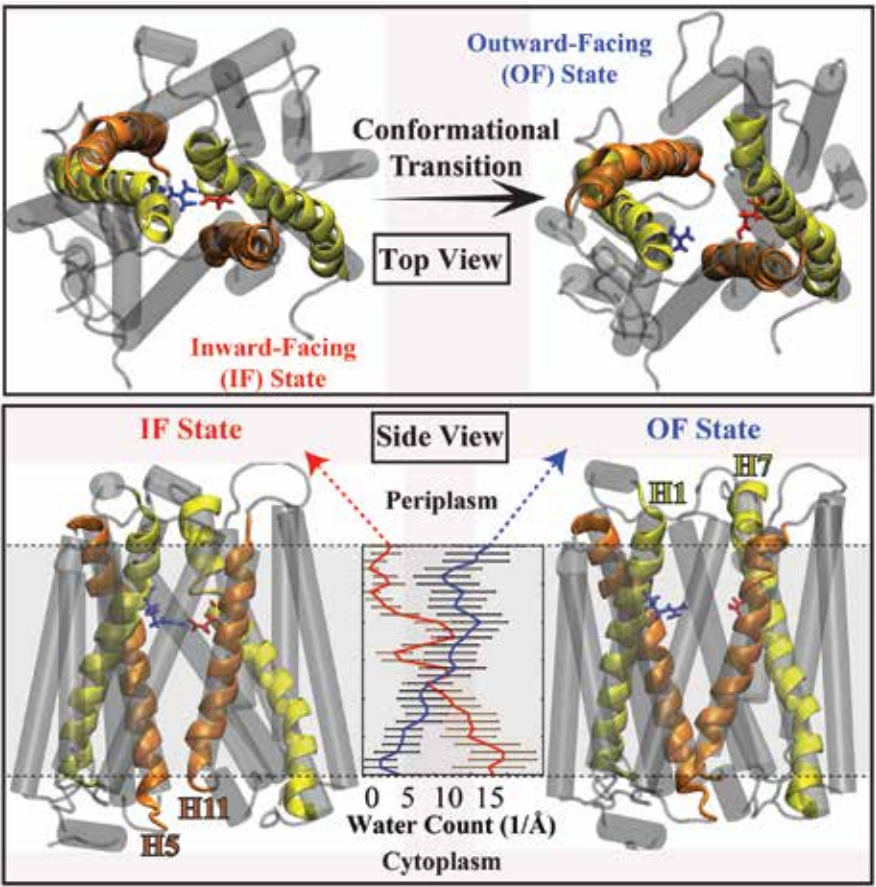


Figure 1: Top and side views of the GkPOT transporter (cartoon representation) in its inward-facing (IF) and outward-facing (OF) states, along with the water count along the pore as measured from equilibrium simulations of each model. The OF model was generated using the orientation-based enhanced sampling techniques.

all-atom MD simulations in combination with orientation-based BEUS/SMwST algorithms [16]. The model is verifiably a stable OF structure since the subsequent equilibrium simulations show a water accessibility consistent with an OF state (Fig. 1). The simulations also suggest that the full IF–OF transition requires the binding of both proton and substrate. The pathways generated reveal that the proton-bound GkPOT cannot transition to the OF state [16]. Unlike previous simulation studies, which had relied on either unbiased equilibrium simulations or simple representations (*e.g.*, coarse-graining), this new approach combines the accuracy of all-atom MD with the accessibility of long timescales provided by enhanced sampling techniques. The successful employment of these multiple-copy algorithms using Blue Waters resources opens a new window to the structural biology of membrane transporters that bypasses the limitations of computational approaches to studying structure–function relationships in these proteins.

WHY BLUE WATERS

This work has explicitly shown that the unbiased all-atom MD, which is routinely used in the field, could be quite misleading in deciphering mechanistic features of membrane transporters owing to the great gap in the timescales associated with the con-

ventional simulations and the function of these proteins [11]. On the other hand, loosely coupled multiple-copy algorithms such as BEUS and SMwST [14,15] can be used to reconstruct unknown conformational transitions of membrane transport proteins. Unlike the conventional all-atom or coarse-grained MD that can be performed on subpetascale machines, BEUS/SMwST simulations of membrane transporters are well-suited for large petascale computational resources such as Blue Waters since they require hundreds of nodes for a single job. The “weak scaling” of these algorithms makes them particularly attractive for large petascale machines, as they can utilize hundreds of compute nodes with almost perfect efficiency.

PUBLICATIONS & DATA SETS

K. Immadisetty, J. Hettige, and M. Moradi, “What can and cannot be learned from molecular dynamics simulations of bacterial proton-coupled oligopeptide transporter GkPOT?” *J. Phys. Chem. B.*, vol. 121, no. 15, pp. 3644–3656, 2017, doi: 10.1021/acs.jpcb.6b09733.

D. Ogden, K. Immadisetty, and M. Moradi, “Conformational transition pathways in major facilitator superfamily transporters,” *bioRxiv* 708289, 2019, doi: 10.1101/708289.



ACTIVATION MECHANISMS OF THE MECHANOSENSITIVE CHANNEL OF LARGE CONDUCTANCE: EMPLOYING LOOSELY COUPLED MOLECULAR DYNAMICS SIMULATIONS TO CHARACTERIZE PROTEIN STRUCTURAL DYNAMICS

Allocation: GLCPC/550 Knh  
PI: Mahmoud Moradi<sup>1</sup>

<sup>1</sup>University of Arkansas

EXECUTIVE SUMMARY

The mechanosensitive channel of large conductance (MscL) is a model system for the study of mechanosensation. Understanding MscL conformational dynamics has specific biomedical applications. The high level of conservation of MscL in bacteria and its absence from the human and animal genomes make MscL an attractive drug target for novel antibiotics. In addition, MscL has been proposed as a liposomal drug delivery nanovalve. For example, through engineering MscL can become pH-acti-

vated to release drugs when it senses the low pH of the tumor microenvironment.

The PI employed all-atom molecular dynamics (MD) simulations along with novel enhanced sampling techniques to characterize the large-scale conformational changes of MscL and its interactions with its candidate modulators. These simulations elucidate the conformational landscape of MscL at the molecular level, providing a rational design framework for designing more efficient modulators for MscL.

RESEARCH CHALLENGE

The mechanosensitive channel of large conductance (MscL) [1] is a bacterial membrane transport protein that serves as a model system for the study of mechanosensation, a process involved in hearing, touch, balance, and cardiovascular and kidney regulation [2,3]. Owing to its unique properties, MscL has also been proposed for use in various biomedical applications, both as a drug target for novel antibiotics [4] and as a stimulus-triggered nanovalve for drug delivery liposomes [5–8]). Unfortunately, MscL structure is only known in its inactive, closed state [9]. If the molecular basis of MscL activation were understood, researchers could engineer more efficient functional nanovalves and design more potent antibiotics targeting MscL.

It is vital to model the active, open state of MscL as well as the entire opening/closing process in order to characterize the MscL activation mechanism. The timescales involved in this process are beyond the currently accessible limits of traditional all-atom MD. The tilting of transmembrane helices is the main rate-limiting step required to form the open pore [10]. While many MD studies have been conducted to study MscL activation, the techniques used in these studies rely on simplifications such as coarse-graining [11]. The main challenge in characterizing the large-scale conformational changes of proteins such as those associated with MscL is to reach the functionally relevant timescales without compromising the chemical details.

METHODS & CODES

The PI used a novel ensemble-based simulation approach [12–15] to simulate the activation process of wild-type and engineered MscL. Bias-exchange umbrella sampling (BEUS) and string method with swarms of trajectories (SMwST) are both loosely coupled MD-based algorithms that require parallel execution of hundreds of MD simulations [14] and have been recently modified within a Riemannian geometry framework [15]. The methodology is specifically centered on applying orientation-based forces on protein transmembrane helices in order to speed up the exploration of protein conformational space.

The software engine used for the simulations is NAMD, a highly scalable MD code implemented in Charm++, an object-based message-driven execution system based on C++. NAMD has been enhanced to support extremely scalable loosely coupled multiple-copy algorithms. Multiple concurrent NAMD instances were launched with internal partitions of Charm++ and located continuously within a single communication world. Messages between NAMD instances were passed by low-level point-to-point communication functions, which are accessible through NAMD's Tcl scripting interface.

RESULTS & IMPACT

Using the known closed/inactive structure of MscL (Protein Data Bank: 2OAR [9]) and the enhanced sampling MD simulations conducted on an orientation-based biasing protocol, the PI

has successfully generated an open model of MscL both for the wild-type and engineered protein (Fig. 1). The engineered MscL is made by attaching positively charged MTSET ([2-(trimethylammonium)ethyl] methane thiosulfonate bromide) labels to the protein within its gate region. The opening of the wild-type MscL is triggered by an increase in the membrane tension, whereas for the engineered MscL, the opening could occur spontaneously. Microsecond-level equilibrium simulations reveal the beginning stages of the activation of the engineered MscL (Fig. 1); however, enhanced sampling techniques are required to capture the complete activation (Fig. 1). Unlike previous simulation studies that relied on either unbiased equilibrium simulations or simple representations (e.g., coarse-graining), the PI's approach combines the accuracy of all-atom MD with the accessibility of long timescales provided by enhanced sampling techniques.

The main conclusion of this project's simulations is that the wild-type and engineered MscL are associated with distinct open conformations. The knowledge of the details of the open state of MscL and its activation mechanism allows for designing more efficient modulators for the engineered MscL within a rational design framework. The successful employment of these multiple-copy algorithms using Blue Waters' resources opens a new window to the structural biology of membrane transport proteins that bypasses the limitations of computational approaches to studying structure–function relationships in these proteins.

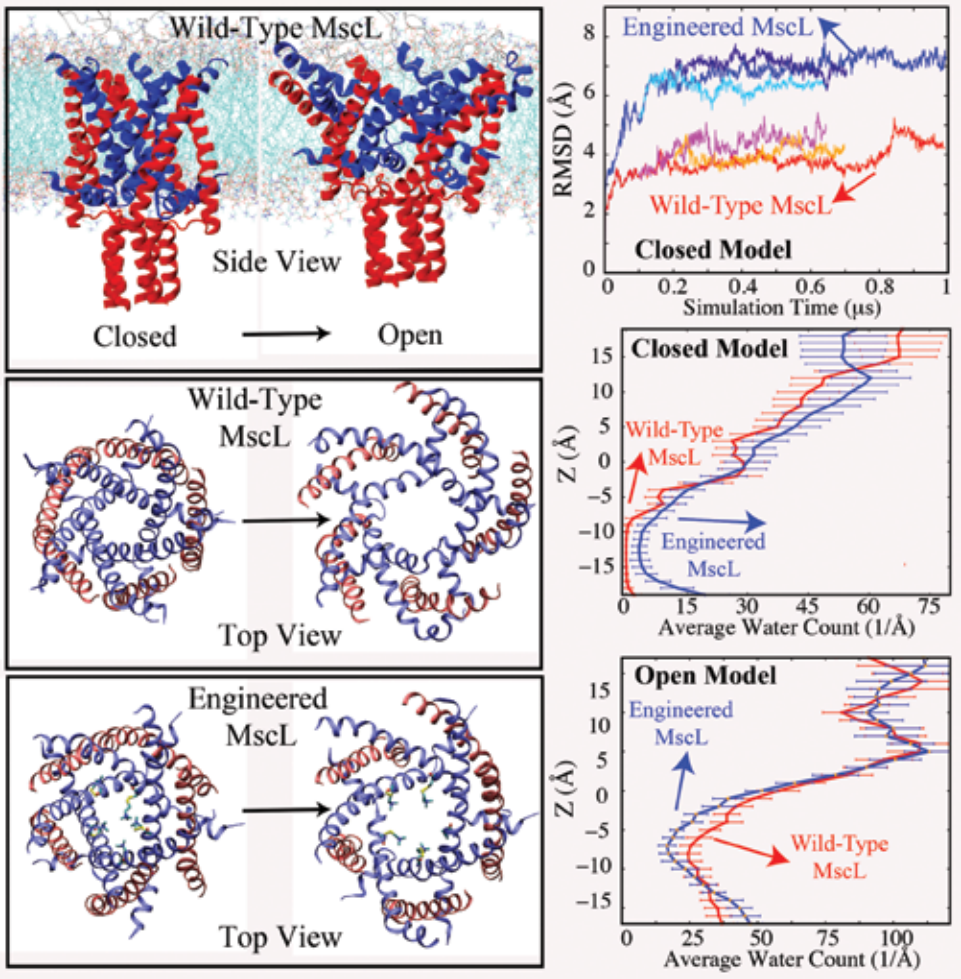
WHY BLUE WATERS

The PI has explicitly shown that unbiased all-atom MD, which is routinely used in the field, could be quite misleading in deciphering mechanistic features of membrane transporters owing to the great gap in the timescales associated with the conventional simulations and the function of these proteins. On the other hand, loosely coupled multiple-copy algorithms such as BEUS and SMwST [14,15] can be used to reconstruct unknown conformational transitions of membrane transport proteins. Unlike the conventional all-atom or coarse-grained MD that can be performed on subpetascale machines, BEUS/SMwST simulations of membrane transporters are well-suited for large petascale computational resources such as Blue Waters as they require hundreds of nodes for a single job. The “weak scaling” of these algorithms in particular makes them attractive for large petascale machines, as they can utilize hundreds of compute nodes with almost perfect efficiency.

PUBLICATIONS & DATA SETS

K. Immadisetty, A. Polasa, R. Shelton, and M. Moradi, “Elucidating the molecular basis of pH activation of an engineered mechanosensitive channel,” *bioRxiv* 707794, 2019, doi: 10.1101/707794.

Figure 1: MscL protein (cartoon representation) in its closed (inactive) and open (active) states. (Left panels) In its wild-type (top and middle left) and engineered (bottom left) forms. (Right panels) The root-mean-square deviation (RMSD) time series (top right) and water profile along the pore (middle and bottom right) from various simulations.





MOLECULAR MECHANISMS OF INFECTION BY CHLAMYDIA

Allocation: GLCPC/1,107 Knh  
PI: Juan Perilla<sup>1</sup>

<sup>1</sup>University of Delaware

EXECUTIVE SUMMARY

The most common sexually transmitted bacterium, *Chlamydia trachomatis*, is an intracellular pathogen responsible for a swath of debilitating conditions including blinding trachoma, which afflicts roughly 84 million persons worldwide and has led to total blindness in eight million people. Despite what is well understood about *C. trachomatis*, important aspects remain mysterious: particularly, how this organism accomplishes membrane fusion. Postinfection, cellular inclusions undergo this critical step whereby separate inclusions are fused together inside the host

cell before maturation and expulsion. In 2016, it was shown that the virulence of *C. trachomatis* is wholly dependent on this stage of its lifecycle, and completion thereof was traced to one protein known as IncA. With the structure of IncA provided by experimental collaborators and with access to Blue Waters, the PI performed studies that could be accomplished with no other scientific instrument: simulating the dynamics of IncA over microseconds to uncover key structural and conformational factors of this important cell machinery.

RESEARCH CHALLENGE

Using a high-resolution crystal structure of wild-type IncA provided by experimental collaborators—IncA<sub>87–246</sub>—the PI aimed to assess if the coordinates represented a thermodynamic minimum or a metastable intermediate capable of reorganizing into a fusion-competent state. For comparison, two other constructs identified experimentally were tested: a truncated mutant IncA<sub>87–237</sub> and a point-mutant IncA<sub>87–246(G144A)</sub>. Without understanding the dynamics of *C. trachomatis*’ key fusion protein, and what conformation might be assumed to initiate fusion, no conclusions leading to drug targets may be formed. In a world where classic antibiotic treatments are leading to super-resistant pathogens, the need for new drug targets cannot be overstated.

METHODS & CODES

The simulations used the NAMD molecular dynamics engine, which is optimized to take full advantage of high-performance distributed-memory architectures. Using NAMD, the PI sampled all three systems for an aggregate of twelve microseconds. These trajectories provided invaluable insights into the dynamics of IncA variants and will be used in further in-depth analyses. Blue Waters not only made collecting these data practical but enabled the researcher to efficiently gather evidence identifying key structural features of IncA that help regulate its overall structure and flexibility. Similarly, parallel I/O capabilities of Blue Waters’ Lustre file system made postprocessing of the obtained trajectories feasible.

RESULTS & IMPACT

The simulations showed that the crystal structure of IncA<sub>87–246</sub> represents a thermodynamic minimum that is not likely to unfold to initiate fusion. Analysis of obtained trajectories showed that the C-terminal domain of the fusion peptide is essential to regulating global structural flexibility, particularly in the region known as the H<sub>clamp</sub>, which was shown in simulations to have a much higher mobility than the rest of the structure (Fig. 1a). Importantly, because IncA is thought to recognize a conspecific or separate regulatory protein before fusion, the high mobility of H<sub>clamp</sub> hints at its significance in the fusion process. Truncated mutant IncA<sub>87–237</sub>—which does not have C-terminal domain—showed a wildly fluctuating H<sub>clamp</sub> region as well as a poorly maintained global structure (Fig. 1a, 1b), explaining the experimentally observed inability of this truncated construct to even initiate fusion. Addi-

tionally, when the distance between the C-terminal domain and H<sub>clamp</sub> is narrowed down, as was tested with the point-mutant IncA<sub>87–246(G144A)</sub>, overall structural stability increases. Increased stability, observed as decreased root-mean-square fluctuation (RMSF), is most apparent in the H<sub>clamp</sub> region of IncA<sub>87–246(G144A)</sub> (Fig. 1a). Consequently, this prevented the H<sub>clamp</sub> from forming any lasting hydrogen bonds with helix A, which were observed in the other constructs (Fig. 1c). This finding supports the idea that the H<sub>clamp</sub> region and its interactions with the C-terminal domain are critical determinants of structure and dynamics. Moreover, it provides insights into why the G144A construct could not form a homodimer in solution in experiments.

WHY BLUE WATERS

NAMD, the molecular dynamics engine used in this research, is well suited for distributed-memory parallel architectures and is particularly well-optimized on Blue Waters. These qualities make NAMD an ideal choice for the type of calculations the PI does in his research. Together, NAMD and Blue Waters enabled the researcher to obtain new results quickly and efficiently, whereas Blue Waters’ Lustre file system enabled the high-throughput postprocessing of the obtained results.

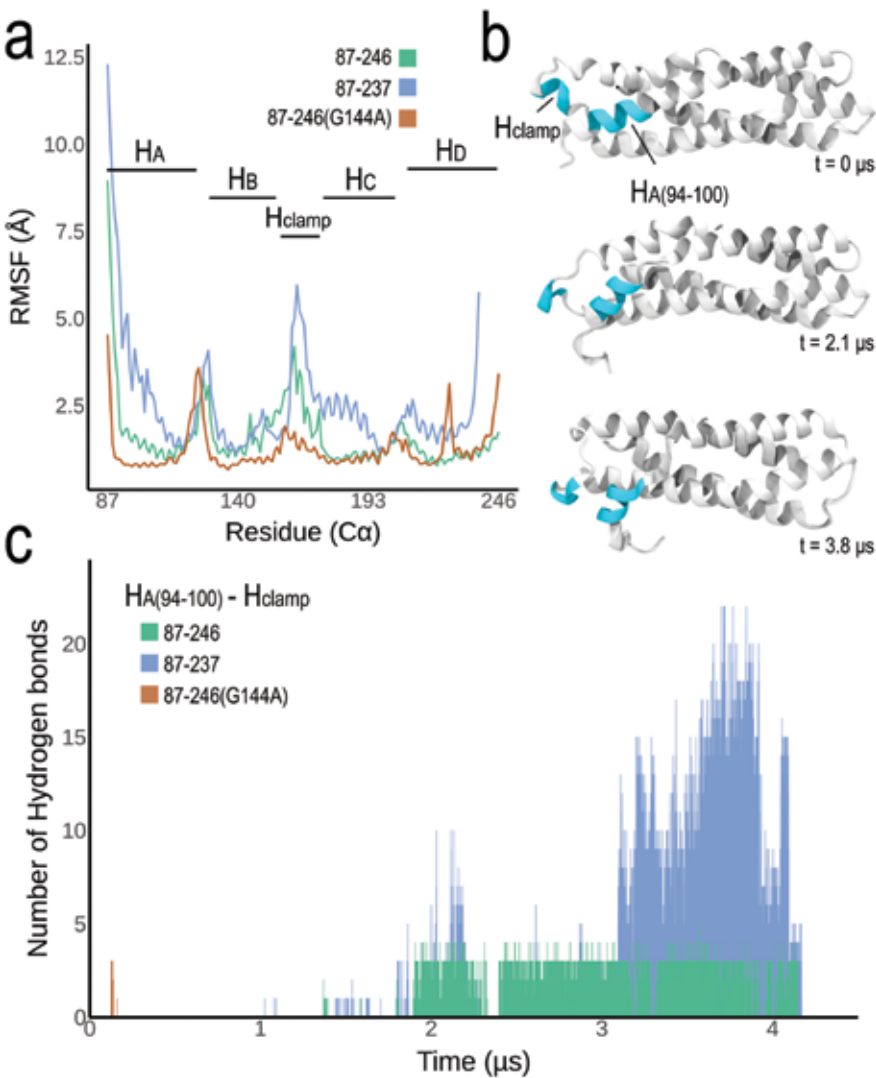
PUBLICATIONS & DATA SETS

G. Cingolani *et al.*, “Structural basis for the homotypic fusion of chlamydial inclusions by the SNARE-like protein IncA,” *Nat. Commun.*, no. 10, p. 2747, Jun. 2019.

Protein data bank entries: 6E7E, 6E6A. [Online]. Available: <https://www.rcsb.org/structure/6E7E> and <https://www.rcsb.org/structure/6E6A>

A. Bryer, J. A. Hadden–Perilla, J. E. Stone, and J. R. Perilla, “High-performance analysis of biomolecular containers to measure small-molecule transport, trans bilayer lipid diffusion, and protein cavities,” *J. Chem. Inf. Model.*, Sep. 2019, doi: 10.1021/acs.jcim.9b00324.

Figure 1: Dynamic characterization of Chlamydia’s inclusion body protein (IncA). Results of molecular dynamics study of wild-type IncA (IncA<sub>87–246</sub>) and two variants, a truncated mutant (IncA<sub>87–237</sub>) and a point-mutant (IncA<sub>87–246(G144A)</sub>). a) Root-mean-square fluctuation (RMSF) of each residue in the three IncA variants tested. b) Snapshots of IncA<sub>87–237</sub> trajectory, showing large-scale conformational change and interactions between the H<sub>clamp</sub> region and a section of helix A (H<sub>A(94–100)</sub>). c) Trace of the hydrogen bonds between H<sub>clamp</sub> and H<sub>A(94–100)</sub> for each of the three constructs over time.





CALIBRATING THE SIMBIOSYS TUMORSCOPE FOR THE FIGHT ON CANCER: A SCENARIO ANALYSIS ENGINE FOR DETERMINING OPTIMAL THERAPY CHOICE

Allocation: Industry/50 Knh  
PI: Joseph R. Peterson<sup>1</sup>  
Co-PIs: John A. Cole<sup>1</sup>, Tyler M. Earnest<sup>1</sup>, Michael J. Hallock<sup>1</sup>

<sup>1</sup>SimBioSys, Inc.

EXECUTIVE SUMMARY

Use of the most toxic drugs in a “scorched earth strategy” is common in the treatment of breast cancer, even though less toxic drugs have only single-digit lower rates of success. To address this issue, the TumorScope software was calibrated to accurately predict the response of breast cancer to the various drugs commonly used for treatment. Calibration resulted in high correlation between predicted and actual outcomes for nearly 200 patients spanning all types of breast cancer. More pertinent to treatment planning, TumorScope was able to improve the accuracy of identifying patient/drug combinations that will achieve pathological complete response (PCR) after treatment by two to three times that of the current state-of-the-art method. PCR is the strongest predictor of long-term survival for breast cancer patients and the desired outcome for all drug-based therapies. Physicians can thus use the computational analysis of different therapies produced by TumorScope to weigh likelihood of therapy success versus drug toxicity.

RESEARCH CHALLENGE

Despite significant advances in cancer treatment in both the number and efficacy of therapies, success rates of any individual therapy remain low. This is especially true in breast cancer where, while overall five-year survival rates tend to be greater than 70%, the efficacy rates of any individual therapy range from 20% to 40%. Critically missing is a way for physicians to distinguish which therapy will be most effective for each patient, or whether none of the available therapies will work. As a consequence, many patients are prescribed a therapy, often at high physical, mental, and monetary cost, that is either too extreme or completely ineffective.

Take, for example, the decision between prescribing an anthracycline-containing (ACT) or a docetaxel-containing (TC) chemotherapy. While ACT therapies, at a population level, have an approximately 3% higher efficacy, it comes at the risk of higher toxicity during treatment, an elevated risk of heart disease, and an approximately 1% likelihood of the patient developing leukemia. Yet, owing to the slightly higher efficacy rate, ACT therapies are much more commonly used, and physicians are faced with no way to argue for using the less toxic TC.

What is needed is a method of predicting the likelihood of success for different therapies that does not increase the number or cost of laboratory tests. SimBioSys set out to solve this problem by developing TumorScope, a personalized medicine approach that constructs realistic, 3D models of patients’ tumors that, when coupled with novel simulation algorithms, can be used to perform scenario analysis (Fig. 1a).

METHODS & CODES

Simulations were performed using a proprietary simulation engine known as TumorScope. Data from nearly 200 breast cancer patients who underwent neoadjuvant (presurgery) chemotherapy [1–3] were used to construct initial conditions for the simulations. The data set represents patients of all common breast cancer types as well as drugs, including chemotherapy and immunotherapy. At several times throughout the course of therapy, clinical outcomes such as tumor volume, tumor longest dimension, and overall pathological response were used to assess the accuracy and applicability of a particular parameter set. Parameter sets were varied systematically for each drug to minimize the deviation between predicted results and clinically measured results.

RESULTS & IMPACT

Simulations included a 3D representation of the local environment with the various tissues found around the breast tumor (Fig. 1a). The TumorScope software simulated the interactions among the tissues along with their biology, the drug perfusion (Fig. 1b), and resulting response of the tumor to different drug regimens (Fig. 1c). Depending on the drug regimen—that is, the combination of drugs chosen and frequency of administration—a patient’s tumor response might be drastically different. As demonstrated in Fig. 1c, a patient given the same combination of drugs every three weeks (e.g., a “standard regimen”; Fig. 1c, left) or every two weeks (e.g., a “dose-dense regimen”; Fig. 1c, right) can lead to either progression or regression of the tumor mass. Empowered with this information, the physician and patient may opt for the more toxic dose-dense regimen.

As with every predictive model, TumorScope had freely tunable parameters. To make accurate predictions, model parameters were systematically tuned to match predictions with clinical outcomes. The research team leveraged the large number of

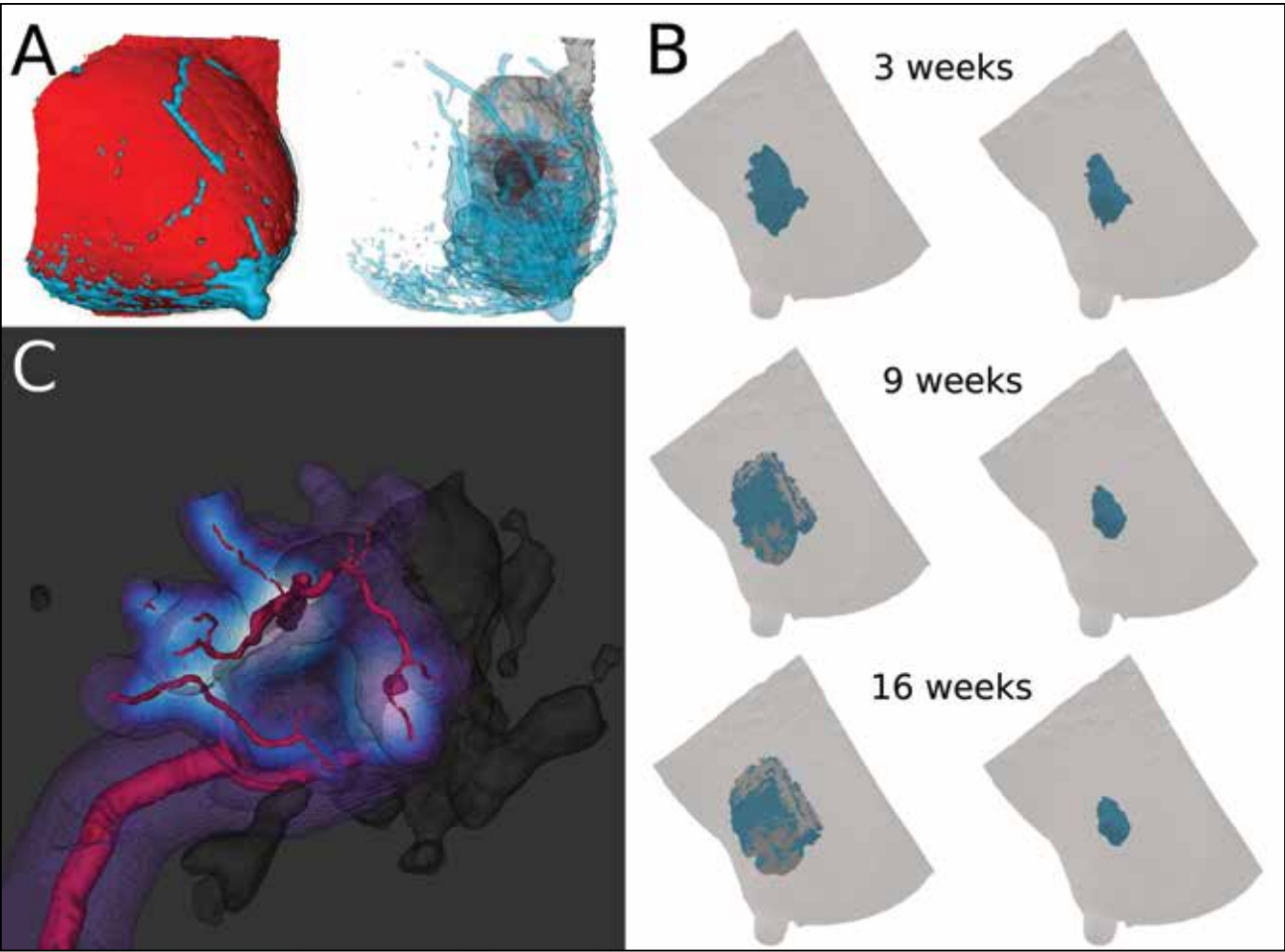


Figure 1: (A) Three-dimensional model of a patient’s breast tumor including fat (red), fibroglandular tissues (blue), and tumor tissues (black). (B) Close-up of the patient’s tumor in a realistic microenvironment including simulated drug perfusion (purple) out of the vascular network (red). (C) The same patient simulated with a dose-dense (right) and standard dose (left) chemotherapy regimen, demonstrating how choice of chemotherapy and regimen can lead to drastic differences in tumor regression (right) or progression (left) over the course of treatment. The tumor is shown in blue and the breast skin is outlined in transparent gray.

GPUs available on Blue Waters to perform thousands of simulations of the approximately 200 patients undergoing different treatments, with different parameters.

Calibration resulted in a high correlation between prediction and calibration data, indicating a good model fit. Pertinent to treatment planning, TumorScope improved the accuracy of identifying patient/drug combinations that achieve pathological complete response (PCR) by two to three times that of the current state-of-the-art method. Since PCR is the strongest predictor of long-term survival for breast cancer patients, physicians can use TumorScope to identify which therapies will achieve PCR (if any), and potentially deescalate therapy when less toxic drugs are predicted to give rise to the same extent of response as more toxic alternatives.

WHY BLUE WATERS

The Blue Waters system has been critical for the calibration of the TumorScope prediction engine. Calibration required simulating the response of hundreds of patients to various therapies while varying a number of parameters. Even with the fast time-to-solution of the predictive engine (one prediction per hour), a calibration process of this scale would have required months to years of computer time to complete. By leveraging the large number of GPUs available on Blue Waters, these calibration processes could be performed in just a few days. The computer allocation during the 2018–2019 year effectively allowed for the complete tuning of the TumorScope software for breast cancer.



DI

GA

# INVESTIGATING THE CLIMATE-RELEVANT IMPACTS OF CHEMICAL COMPLEXITY IN MARINE AEROSOLS

**Allocation:** NSF PRAC/6,400 Knh  
**PI:** Kimberly Prather<sup>1</sup>  
**Co-PI:** Rommie Amaro<sup>1</sup>

<sup>1</sup>University of California, San Diego

BW

CI

FS

BI

## EXECUTIVE SUMMARY

Marine atmospheric aerosols make up a significant portion of the planetary aerosol budget but represent one of the largest sources of uncertainty in current climate models. Aerosols impact climate through cloud seeding and multiphase heterochemistry. However, understanding of the structure and morphology of marine aerosols and the resulting impacts on atmospheric processes is extremely limited. Given the difficulty of experimental single-particle analysis techniques and the high complexity of individual particles, computational methods are used to explore how particle size and chemical composition impact climate-relevant physical and chemical properties of marine aerosols. Model sea spray aerosols are generated based on experimentally determined aerosol lipidomic and proteomic information, then sim-

ulated using Blue Waters, with the goal of understanding how chemical complexity modulates aerosol dynamics.

## RESEARCH CHALLENGE

Marine aerosols impact the climate by interacting with incoming solar radiation, serving as nuclei for cloud formation and providing interfaces for heterogeneous and multiphase atmospheric chemistry. Individual aerosols vary widely in their size, chemical profile, and internal mixing states; however, current single-particle analysis techniques in the laboratory are not able to access the nanosecond to millisecond timescale dynamics within these aerosols, which ultimately guide their climate-relevant properties. These simulations allow the researchers to understand how aerosol chemistry ultimately impacts the dynamics and interactions within aerosols at the microscale, including protein and lipid

partitioning, diffusion, and water transport across the interface, allowing the team to make predictions about macroscale aerosol behavior. The variation in lipid, protein, and saccharide (carbohydrates such as simple sugars or polymers) concentration—strongly modulated by the biochemistry at the ocean surface—allows the research team to connect ocean microbiology to aerosol properties. Integrating computational methods with experimental techniques advances the capacity to accurately model marine aerosols, inching scientists closer to unraveling the complexity of marine aerosol chemistry and how it shapes our climate.

## METHODS & CODES

While marine aerosols have yet to be fully characterized in terms of their chemical composition and structural morphology, and though they exhibit a wide variety of types and variation based on relative humidity, ocean conditions, geographic origin, etc., the team based their aerosol models on the recipe derived by Bertram *et al.* [1] for submicron particles at 70% relative humidity. The researchers’ models were constructed with PACKMOL [2] and parameterized using the CHARMM36 [3] force field and TIP3P [4] water. All simulations were carried out with GPU-accelerated NAMD [5].

To investigate the impacts of chemical complexity on nanoaerosol molecular dynamics, the team simulated three aerosol systems and increased the number of components at each step. The three aerosol systems were simulated in triplicate for up to 500 nanoseconds each. System I contains a base level of complexity, and includes four fatty acids at experimentally determined ratios, protein (*Burkholderia cepacia* lipase), and sodium chloride. In system II, the team added lipopolysaccharides from *E. coli* and increased the diversity of cation types to include magnesium, calcium, and potassium in addition to sodium. System III is the most chemically complex and includes a neutral polysaccharide (laminarin) as well as glucose monomers.

## RESULTS & IMPACT

Although work only began in early 2019, the team hopes the results will lend a more complete understanding of nascent sea spray aerosol morphology. Microscopy data show that sea spray adopt a morphology where the organics separate to the surface of the particle, leaving a salty aqueous core in the particle center. Preliminary results from this project suggest that at nanometer particle sizes, the organic and aqueous fractions within the particle do not phase separate as expected. Rather, the organic and aqueous components are distributed throughout the particle and are stabilized as an emulsion. If verified, the research team’s work will challenge the current understanding of marine aerosol morphology, and could have impacts on how scientists model multiphase atmospheric aerosol chemistry.

## WHY BLUE WATERS

The Blue Waters system has enabled the research team to explore molecular dynamics on systems of real physical and chemical relevance to the climate, which includes simulating particle sizes up to 100 nanometers in diameter, containing many millions of atoms, for microsecond timescales. The Blue Waters staff has been integral to the success of this work. Their expertise in GPU-accelerated NAMD has allowed the team to get their simulations up and running quickly, with minimal setbacks.

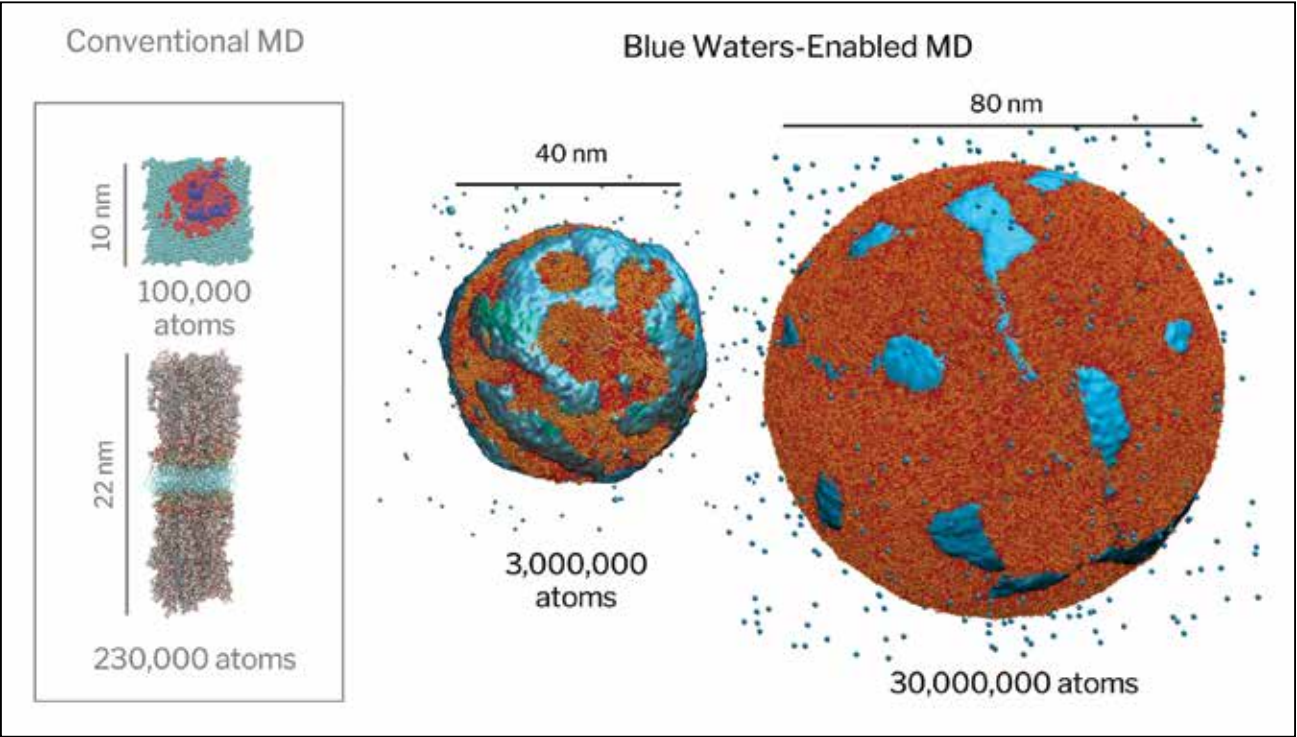


Figure 1: Blue Waters enables the simulation of much larger systems than would be feasible on conventional supercomputers or workstations. Left: Model systems of a protein embedded in a lipid monolayer (top) and a lipopolysaccharide (LPS) bilayer (bottom) with dimensions labeled. Right: 40-nanometer (nm) aerosol particle and 80-nm aerosol particle after approximately 25 nanoseconds of simulation. Lipids are colored orange and water is colored blue.



# MOLECULAR DYNAMICS BINDING FREE ENERGY CALCULATIONS OFFER A WINDOW TO UNDERSTAND PROTEIN-PROTEIN BINDING SPECIFICITY

**Allocation:** NSF PRAC/1,330 Knh  
**PI:** Benoit Roux<sup>1</sup>  
**Collaborators:** Chris Boughter<sup>1</sup>, Prithviraj Nandigrami<sup>1</sup>, Jonathan Thirman<sup>1</sup>

<sup>1</sup>University of Chicago

## EXECUTIVE SUMMARY

Extracellular domains of cell surface receptors and ligands mediate cell-cell communication, adhesion, and initiation of signaling events. In order to understand highly organized cellular systems, it is necessary to study how intricate dynamical networks arise from specific protein-protein interactions. Thus, the research team has used a free energy computational methodology to quantitatively examine the basis for protein binding specificity in two subfamilies of recently identified genes within the immunoglobulin superfamily found in the model organism *Drosophila melanogaster* (the common fruit fly).

## RESEARCH CHALLENGE

The living cell can be pictured as a collection of macromolecules that are carrying out a number of well-defined tasks. Thanks to tremendous progress made in X-ray crystallography, our knowledge of the three-dimensional structure of individual proteins present in the cell has been greatly expanded in the last decades. However, as scientists seek to comprehend the function of highly organized cellular systems, it is becoming increasingly clear that this will not be possible without understanding how intricate dynamical networks arise from specific protein-protein interactions. For example, the intercellular communications needed for the morphogenesis of the central nervous system is essentially mediated by protein-protein interactions and recognition processes involving cell surface receptors and ligands.

Some of the most intriguing of such interactions are presented by the set of recently identified Dpr-DIP complexes in the common fruit fly (*Drosophila melanogaster*), which has a total of 20 Dpr genes. The Dpr (defective proboscis extension response) and DIP (Dpr-interacting proteins) genes belong to the immunoglobulin superfamily and associate to form a molecular complex (Fig. 1). Most members of the DIP subfamily cross-react with several members of the Dprs and vice versa.

The Dpr-DIP binding specificity plays a critical role in neuronal development and synaptogenesis. The research team's objective is to use a free energy computational methodology to quantitatively explain the basis for Dpr-DIP protein binding specificity. To understand how cell surface receptors control developmental morphogenesis of the nervous system as well as its function, researchers need to answer questions such as how the distinct

structural features of cell adhesion complexes, including possible sets of highly homologous proteins, instruct the formation of synaptic networks. In other words, scientists must decipher the molecular code that governs the specific association of these proteins and understand why some of them bind together but not others despite their high structural similarities.

Predicting with quantitative accuracy when and why proteins can specifically associate and bind must begin with the statistical mechanics concept of binding free energy. For a binary protein complex, the binding affinity is determined by the equilibrium dissociation constant and the binding free energy. Based on insight gained by categorizing and observing many known protein complexes, several models have been proposed to predict the experimental binding affinities using the features discussed above. Although some have been very successful on small training sets, the published models have performed considerably less well on larger sets and their predictive value remains poor. From this point of view, a computational approach based on atomistic molecular dynamics (MD) simulation and free energy methodology is advantageous because it does not rely on any particular empirical assumptions about the binding; the approach is applicable to any protein complex and does not suffer from the limitation displayed by empirical bioinformatics/coevolution or docking/scoring methods, which often perform poorly in large-scale benchmark testing. But this raises several important issues, such as whether the atomic force fields used in the simulations are sufficiently accurate, or whether one can design an efficient computational strategy to overcome the challenges presented by conformational sampling. Another important question is whether one can truly design an effective and scalable computational strategy to tackle the processes occurring over long timescales that is well adapted for a leadership-class computer such as Blue Waters. Unbiased MD trajectories, while very valuable, can be limited. However, advanced free energy methodologies can help overcome these limitations.

## METHODS & CODES

The solution to this computational problem proposed by the research group is to break down the binding process into a large number of physically meaningful steps and express the binding free energy as a sum of free energies associated with each step.

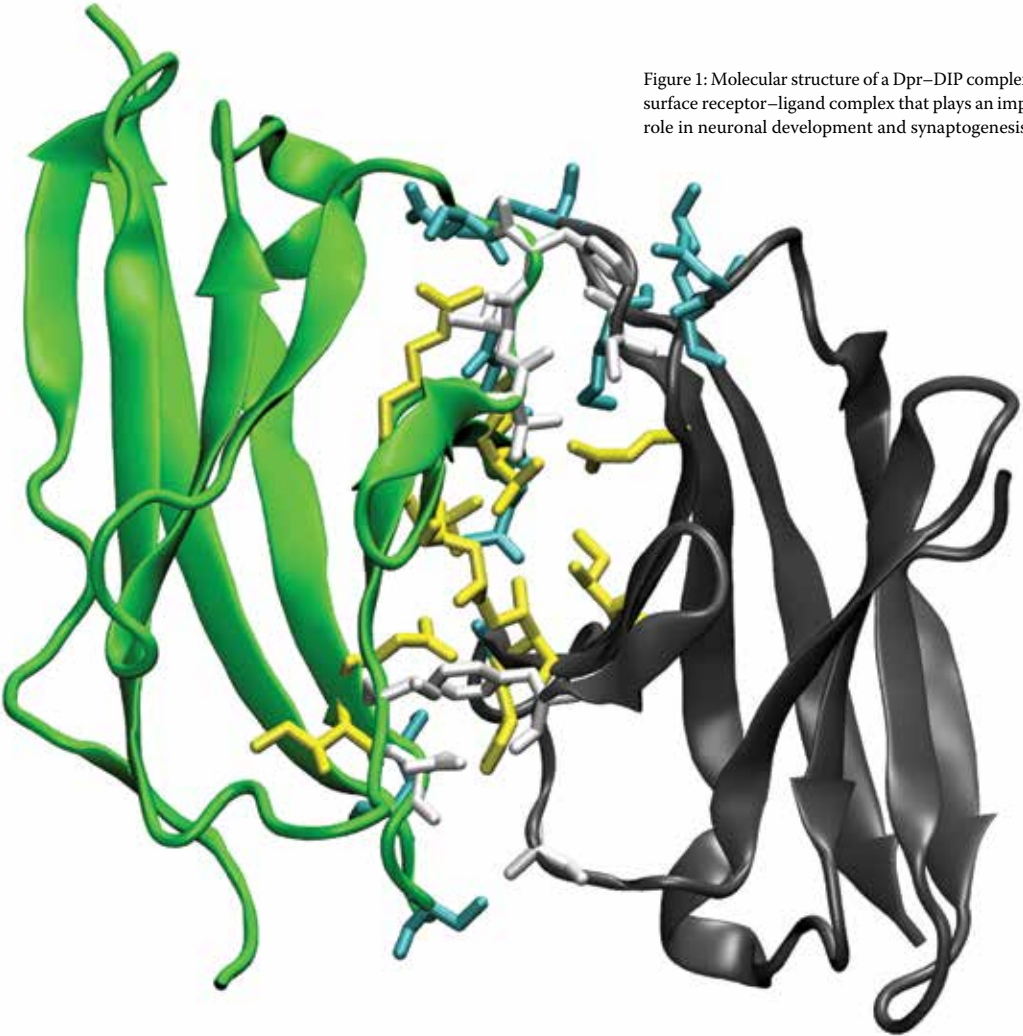


Figure 1: Molecular structure of a Dpr-DIP complex, a cell surface receptor-ligand complex that plays an important role in neuronal development and synaptogenesis.

This requires carrying out a large number of simulations with different biasing restraints acting on the system. The biased data are harvested from a large collection of copies or replicas of the molecular system via a replica-exchange algorithm. Such multiple copy algorithms (MCAs) offer a general and powerful strategy to enhance the sampling efficiency of conventional MD simulations. Further, the simulation program NAMD fully supports extremely scalable and efficient parallel active messaging interface-level simulations with MCAs on leadership computers. This makes it possible to carry out extremely scalable MD simulations on Blue Waters using the team's MD replica-exchange free energy strategy.

## RESULTS & IMPACT

The research team has developed and tested a novel theoretical framework for binding free energy calculations, leaning on the optimal curvilinear minimum free energy path determined from the string method. Fundamentally, this curvilinear pathway strategy is based on the reversible spatial separation of two binding macromolecules, which is clearly the method of choice

to quantitatively characterize the affinity of large molecular complexes in solution. Nonetheless, previous implementations based on the potential of mean force for the separation of two proteins along a predefined rectilinear path led to MD calculations that converged too slowly. The proposed methodology was validated by comparing the results obtained using both rectilinear and curvilinear pathways for a prototypical host-guest complex formed by cucurbituril binding benzene and for the barnase-barstar protein complex. The researchers found that the calculations following the traditional rectilinear pathway and the string-based curvilinear separation pathway agree quantitatively, but convergence is faster with the latter.

## WHY BLUE WATERS

The multiple-copy algorithms require a very large number of nodes to be effective. Blue Waters is a unique platform that makes it possible to fully exploit the power of this advanced sampling methodology.



# MOLECULAR BASIS OF THE NITRATE TRANSPORT MECHANISM IN PLANTS

**Allocation:** Illinois/ 400 Knh  
**PI:** Diwakar Shukla<sup>1</sup>

<sup>1</sup>University of Illinois at Urbana–Champaign

## EXECUTIVE SUMMARY

Growing evidence suggests that efficient use of nitrogen is required to improve crop productivity. The gene NRT1.1 has been identified as a nitrate transporter that enables nitrogen uptake by plants. NRT acts as a nitrate sensor and transports substrate molecules across the membrane. It belongs to the major facilitator family whose function relies on the alternate access mechanism where the substrate binding site of the transporter protein alternatively opens and closes at either side of the membrane. The crystal structure of NRT1.1 was determined in the inward-facing (IF) state as a homodimer and provided the first glimpse of structure information on this class of proteins. These proteins undergo intrinsic conformational changes; the dynamics among the functionally important intermediate states remains elusive. In this project, the research team investigated the conformational dynamics and substrate transport mechanism of NRT1.1 using molecular dynamics simulations.

## RESEARCH CHALLENGE

Nitrogen is an essential nutrient required for plant growth and development. Plants uptake nitrogen as nitrate ions through the nitrate transporter (NRT1.1), which actively mediates the transport of nitrate ions from the soil into cells. NRT1.1 exhibits a dual affinity mode [1]. At high nitrate concentration, NRT1.1 acts as a low-affinity transporter (LAT) and at high concentration, it is phosphorylated at Thr101, resulting in a high-affinity transporter [1,2]. Recently, the crystal structure of NRT1.1 bound to a nitrate ion ( $\text{NO}_3^-$ ) was obtained as a homodimer in the IF state [3,4]. NRT1.1 belongs to the major facilitator superfamily that contains

12 transmembrane helices and functions based on an alternate access mechanism to transport signaling molecules.  $\text{NO}_3^-$  was bound to the protonated His356 and Thr360 on TM7 in the binding site, and biochemical studies show that the mutation of these residues results in a loss of ion transport [3,4]. The phosphorylation of Thr101, the phosphorylation site residue, increases the affinity of the nitrate ions. However, the conformational dynamics and mechanistic basis of  $\text{NO}_3^-$  uptake remain elusive. In this project, the research team performed long microsecond simulations to understand the functional dynamics of NRT1.1 and characterize the  $\text{NO}_3^-$  transport mechanism. The results will provide molecular-level understanding of the phosphorylation-mediated increased affinity switch mechanism of NRT1.1.

## METHODS & CODES

The simulations were performed with AMBER14 [5]. The simulations produced massive amounts of data (several TB); the Python package MDTraj was used for data processing and analysis [6,7]. The MD data were featurized to biologically relevant reaction coordinates and Markov state models (MSM) were constructed using MSMBuilder [8]. MSM is a statistical model constructed to explore the kinetics of the biological events. MD simulation data will be clustered based on kinetically relevant microstates and the transition matrix that was constructed. Using the transition matrix, the rate of probability of transition to the different states were obtained. The state population are easily converted to free energies using Boltzmann distribution. The obtained free energies of individual states are called the “MSM weighted free energy” of populations.

Figure 1: Conformational dynamics of NRT1.1. The conformational free energy landscape of unphosphorylated (A) and phosphorylated NRT1.1 (B). The pore channel radius for various intermediate states is shown in 1A. The polar interactions that stabilize the key states are shown in 1B. The starting inward-facing structure is shown by the black star

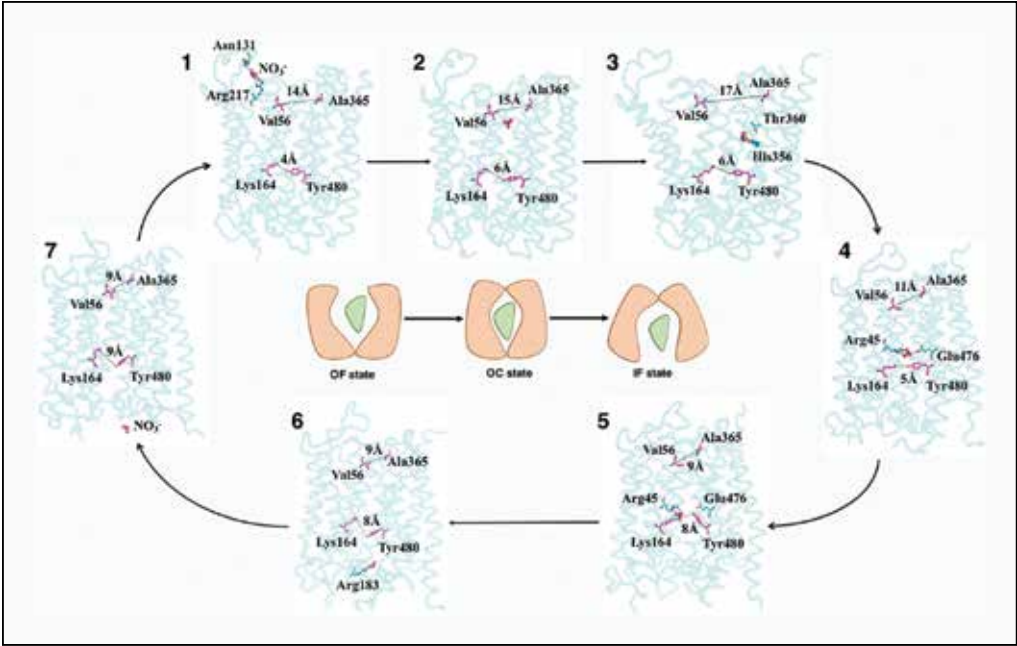
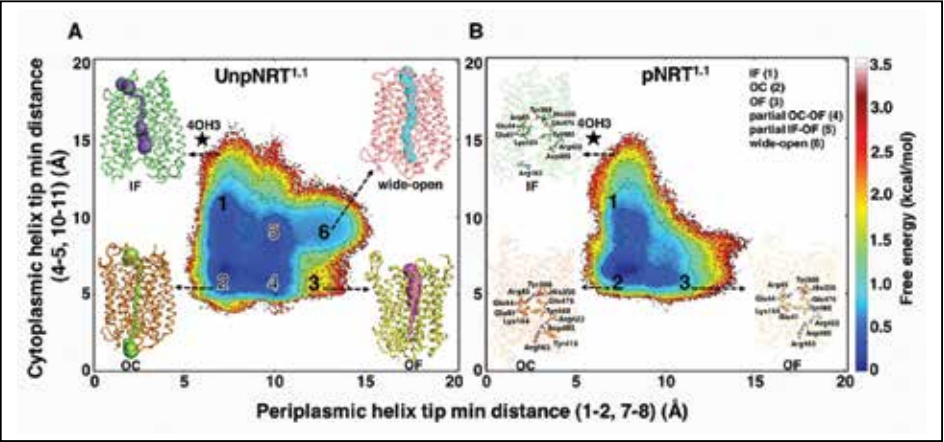


Figure 2: Nitrate transport mechanism in plants. The conformationally driven nitrate ion transport mechanism is shown in 2-1 to 2-7. The distance between the extracellular, intermediate, and intracellular gating residues is shown by black dashed lines. The nitrate ion is shown as ball and sphere.

## RESULTS & IMPACT

*Conformational dynamics of unphosphorylated (UnpNRT) and phosphorylated (pNRT) ensembles.* The IF-state crystal structure of NRT1.1 (Protein Data Bank ID: 4OH3) [5] was the starting structure of the simulations. The research team performed extensive simulations to explore the conformational dynamics and nitrate transport mechanism of UnpNRT (~110  $\mu\text{s}$ ) and pNRT (~45  $\mu\text{s}$ ). The high-dimensional simulation data were converted to the slowest process and clustered to kinetically relevant states. MSM were constructed to gain insights into the thermodynamics and kinetics of UnpNRT and pNRT conformational ensembles. An MSM-weighted free energy landscape plot was shown by projecting the MD data obtained on the minimum helical tip distance of the pore channel radius on the extracellular and intracellular region (Fig. 1). The free energy barrier for one complete conversion cycle of IF to outward-facing (OF) structures was estimated to be approximately 2.5 kcal/mol in UnpNRT and approximately 1 kcal/mol in pNRT. The phosphorylation at Thr101 results in a conformationally driven ion-coupled transport mechanism and results in a canonical L-shaped landscape. However, UnpNRT samples more intermediate states and may result in a low-affinity transporter.

*Nitrate ion transport mechanism in NRT.* The nitrate ion is recognized by the group of polar and positively charged residues at the extracellular part of the transporter in the OF state (Fig. 2); the distances between the periplasmic and cytoplasmic gating residues are approximately 12 Å and 3 Å, respectively. The ion diffuses in the pore and interacts with Thr360. Nitrate ions escape from the intermediate interaction and bind to protonated His356 on TM7; in accord with the experimental finding this

crucial interaction mediates the nitrate transport. The sidechain conformation of His356 is further favored by polar contacts of Glu476 and Tyr388. The tight binding of nitrate ions at the center of the transporter facilitates the conformational change from the OF to the occluded (OC) state. The breakage of polar contacts between Lys164 and Tyr480 leads to the opening of the intracellular gate and the release of nitrate to the cytoplasmic side.

This study reveals the molecular level detail of functionally important intermediate states and the nitrate ion transport mechanism using extensive simulation. In this work, the PI has also determined the key residues that drive the nitrate transport as well as the conformational switches. His future work will be focused on engineering NRT to improve the nitrate transport to increase the crop yield.

## WHY BLUE WATERS

Biologically important conformational transitions are slow processes that are difficult to observe by running simulations on local hardware. Powerful resources like Blue Waters are required to study such complex biological processes in full atomistic detail and over long timescales. Blue Waters provides thousands of GPUs that are used for parallel molecular dynamics simulations to perform MSM-based adaptive sampling of the conformational energy landscape of proteins. Blue Waters increases the overall compute performance by several orders of magnitude in terms of the real time required for simulation.

## PUBLICATIONS & DATA SETS

B. Selvam, J. Feng, and D. Shukla, “Atomic insights into dual affinity switch mechanism in nitrate transporter,” in preparation, 2019.



# SIMULATIONS UNCOVER THE MECHANISM OF SEROTONIN TRANSPORT IN THE BRAIN

**Allocation:** Illinois/600 Knh  
**PI:** Diwakar Shukla<sup>1</sup>  
**Collaborator:** Erik Procko<sup>1</sup>

<sup>1</sup>University of Illinois at Urbana–Champaign

## EXECUTIVE SUMMARY

The serotonin transporter (SERT) is a member of the neurotransmitter:sodium symporters (NSS) family that transports neurotransmitters in conjunction with an electrochemical gradient of ions. SERT initiates the reuptake of extracellular serotonin in the synapse to terminate neurotransmission in the nervous system. Recent cryogenic electron microscopy structures have revealed structural insights into functional conformations of SERT dynamics. However, despite being a major molecular drug target, knowledge of how serotonin is recognized, bound, and transported remains unclear. In this study, the research team performed extensive large-scale molecular dynamics (MD) simulations of the human SERT to investigate the structural transition to various states and determined the complete transport pathway in SERT. Further, the team provided a comprehensive approach for characterizing the thermodynamics of key states and critical residues involved in the substrate transport process.

## RESEARCH CHALLENGE

The neurotransmitter serotonin (5HT) regulates many physiological processes in the body with implications for cognitive function, sleep, mood, and behavior [1]. In neurons, 5HT signaling is terminated through the reuptake of 5HT from the synaptic cleft by the SERT. Imbalance of 5HT in neurons owing to malfunctions of SERT has been associated with depression, bipolar disorder, and autism. Because of its medical importance, SERT is a major drug target for the treatment of psychiatric disorders and drugs

of abuse. Selective serotonin reuptake inhibitors are commonly prescribed as antidepressants and aim to bind and inhibit SERT activity. However, these molecules can be responsible for a number of side effects. SERT and the closely related dopamine transporter (DAT) and norepinephrine transporter (NET) belong to a class of monoamine transporters. Owing to their high sequence and architecture similarities, it is difficult to design selective inhibitors to target just one transporter alone. Furthermore, these transporters undergo large conformational transitions from an outward-facing state to an inward-facing state to transport substrates across the cell membrane [2]. The lack of knowledge of the dynamics of these transporters has led to a lack of understanding of the molecular basis of selectivity for designing effective antidepressants and therapeutic molecules.

## METHODS & CODES

Obtaining sufficient sampling is a recurring challenge in simulating complex biological processes. To overcome this issue, the research team adopted a Markov state model (MSM)-based adaptive sampling methodology to efficiently explore the conformational landscape. In each round of adaptive sampling, multiple short MD simulations were conducted in parallel. The simulation data were clustered using the K-means algorithm based on a designated metric and starting structures were chosen from the least populated states to seed the subsequent rounds of simulation. The simulations were employed using the AMBER v18 suite [3]. The highly parallelized framework implemented in AM-

Figure 1: Conformational free energy landscapes of SERT obtained from MD simulations. Relative free energies from MSM-weighted simulation data plotted against the distances between extracellular and intracellular gating residues for (A) *apo*-SERT and (B) 5HT-SERT. The outward-facing SERT crystal structure (Protein Data Bank: 5I73, pink star) was used as the starting structure for MD simulations.

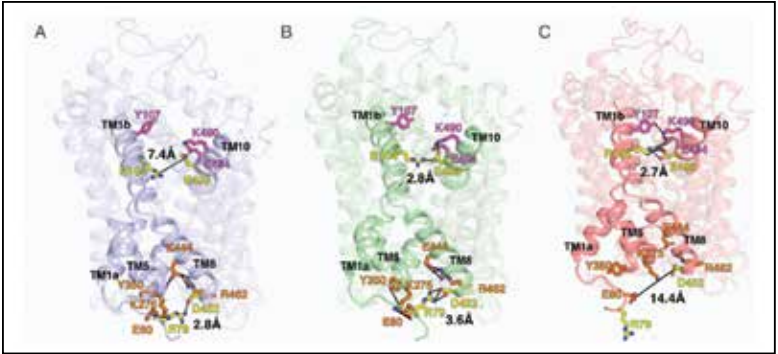
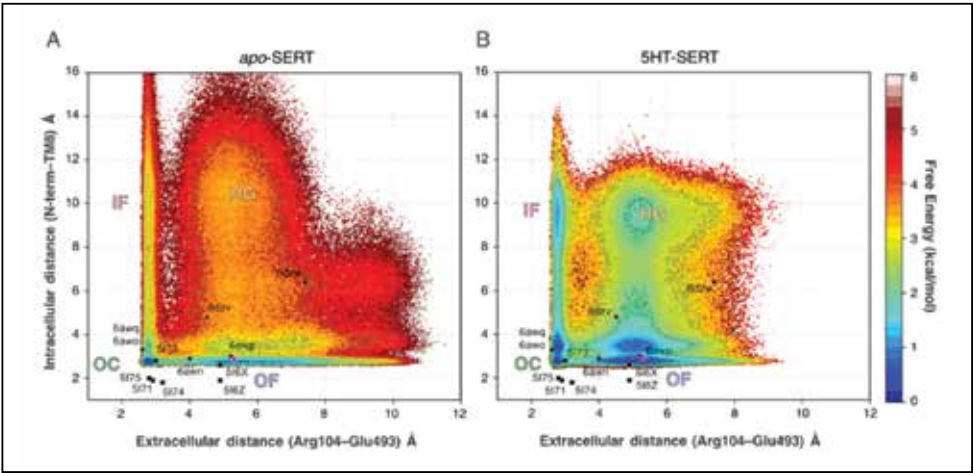


Figure 2: Switching of the hydrogen bond network drives structural transitions. MD snapshots of three conformational states in (A) outward-facing (OC), (B) occluded (OC), (C) inward-facing (IF). Residue interactions shown in sticks. Extracellular gating network colored in magenta; intracellular colored in orange. Gating residues colored in yellow.

BER offers enhanced scaling on GPU nodes to massively accelerate complex biomolecular simulations.

## RESULTS & IMPACT

**Conformational dynamics of human SERT.** To understand the effects of substrate-induced protein dynamics, the entire import process of serotonin was studied using MD simulations. Simulations were initiated from the outward-facing (OF) SERT crystal structure (Protein Data Bank: 5I73). An aggregated total of approximately 290 microseconds ( $\mu$ s) of SERT simulations were obtained and analyzed using MSM, which parses the simulation data into kinetically relevant states and calculates the transition probabilities between the states. MSM-weighted simulations data were projected onto a coordinate system defined by distances between the extracellular and intracellular gating residues (Fig. 1).

The conformational landscape plots reveal that despite the absence of serotonin binding, *apo*-SERT may undergo transitions from the OF state to the inward-facing (IF) state (Fig. 1a). Extracellular gating residues Arg104 (TM1b) and Glu493 (TM10) can extend to 10 Å, enlarging the extracellular entrance tunnel. The OF states are relatively stable, with a free energy of approximately 0.5 kcal/mol. The distance between gating residues Arg104–Glu493 decreases to 3 Å and is associated with electrostatic interactions, forming occluded (OC) conformations that are more stable than the OF state (Fig. 2). Closure of the extracellular entrance tunnel in the OF state weakens contacts on the intracellular side of the transporter, creating an energetically accessible pathway toward the IF state. The free energy barrier for transition from the OC–IF state in *apo*-SERT is estimated as approximately 3 kcal/mol, which is higher compared to the OF–OC transition (approximately 2 kcal/mol).

**Serotonin transport in SERT.** The substrate bound conformational landscape plot depicts deviations in the relative free energies of conformational states and reduced the free energy barriers in between states (Fig. 1b). Binding of serotonin in the entrance tunnel stabilizes the OF states. The gating residues interact with Gln332 (TM6) and Lys490 (TM10) and widens the extracellular vestibule (Fig. 2). The diffusion of 5HT to the S1 site via the allosteric site (S2) leads to the closure of the extracellular cavity to obtain the OC state. The OF–OC transition has a free energy

barrier of approximately 1.5 kcal/mol, which is similar to *apo*-SERT. The downward movement of 5HT facilitates the opening of the intracellular gate and leads to the IF state. The free energy barrier for the structural transition to the IF state is estimated as approximately 1.5 kcal/mol. The presence of 5HT in the intracellular pathway stabilizes SERT in the IF state, with a free energy of approximately 1 kcal/mol as compared to approximately 3 kcal/mol in *apo*-SERT.

This study reveals an atomistic-level perspective into the elusive mechanism of substrate transport in SERT. Using MSM, the research team has identified the free energy barriers associated with the transport process and key interactions that drive transport. The results provide an extensive understanding into the molecular recognition of serotonin in SERT and can serve as a model to study other closely related neurotransmitter transporters.

## WHY BLUE WATERS

Simulation of complex biological processes requires multiple parallel nodes to reach relevant timescales. The unique architecture of Blue Waters provides hybrid CPU and GPU frameworks to conduct large-scale simulations. These computations would not be achievable within a reasonable time without Blue Waters’ petascale computing capability.

## PUBLICATIONS & DATA SETS

M. C. Chan, B. Selvam, H. J. Young, E. Procko, and D. Shukla, “The substrate import mechanism of the human serotonin transporter,” in preparation, 2019.

H. J. Young, B. Selvam, M. C. Chan, D. Shukla, and E. Procko, “Deep scanning mutagenesis reveals hotspot residues that alter conformational equilibrium in SERT,” in preparation, 2019.

H. J. Young, B. Selvam, M. C. Chan, D. Shukla, and E. Procko, “Deep scanning mutagenesis reveals hotspot residues that alter conformational equilibrium in SERT,” presented at the Gordon Research Conf. Mechanisms Membrane Transport, New London, NH, USA, June 23–28, 2019.

M. C. Chan, B. Selvam, and D. Shukla, “Substrate-induced conformational transitions of the human serotonin transporter,” presented at the ACS Fall Nat. Meeting, San Diego, CA, USA., Aug. 25–29, 2019.



ELUCIDATING THE LIGAND SELECTIVITY AND ACTIVATION MECHANISMS OF CANNABINOID RECEPTORS

Allocation: Blue Waters Professor/240 Knh  
PI: Diwakar Shukla<sup>1</sup>  
Collaborator: Aditi Das<sup>1</sup>

<sup>1</sup>University of Illinois at Urbana–Champaign

EXECUTIVE SUMMARY

Cannabinoid receptor 1 (CB1) is a therapeutically relevant drug target for controlling obesity, pain, and central nervous system disorders. However, owing to the harmful side effects of full agonists (molecules that activate CB1) and antagonists (molecules that deactivate CB1), no clinical drug that targets CB1 is currently available. A deeper mechanistic understanding of CB1 selectivity and activation mechanisms with respect to homologous protein cannabinoid receptor 2 (CB2) remains elusive. To understand selectivity and partial agonism, the research team performed extensive simulations using Blue Waters to investigate the conformational dynamics of CB1 and CB2 as well as ligand binding to CB1.

RESEARCH CHALLENGE

CB1 is in a class of lipid G-protein-coupled receptors (GPCRs) that belongs to the endocannabinoid system. Besides ligands of the endocannabinoid system such as anandamide, CB1 signaling is also controlled by phytocannabinoids such as cannabidiol and synthetic cannabinoids such as rimonabant, which act as either agonists or antagonists. Despite the abundance of GPCR ligands, all available drugs are associated with serious side effects such as panic attacks, hallucinations, and addiction. For example, the potent synthetic cannabinoid fubinaca, a full agonist for CB1, caused a mass intoxication of 33 persons in one New York City neighborhood [1]. Similarly, rimonabant—an antagonist of CB1—had to be withdrawn from the market owing to side effects that included suicidal ideation. Therefore, there is a need to develop drugs that are selective and only bind to the targeted re-

ceptor as well as being partial CB agonists; in other words, drugs that do not fully activate the receptor.

METHODS & CODES

Using the hybrid CPU/GPU nodes of Blue Waters, molecular dynamics (MD) simulations were performed in parallel and were based on adaptive sampling, which is an iterative method that creates seeds for the simulations based on the clustering of the current data. Considering the trajectories of the MD simulation as a Markov chain, the research team built a Markov state model (MSM) that discretized the protein conformational space into energetically separable microstates and calculated the transition probability between these states.

RESULTS & IMPACT

Despite belonging to the same endocannabinoid system and sharing a 42% sequence identity, the selectivity of ligands for CB1 and CB2 can vary by several orders of magnitude. Understanding the selectivity of these two similar proteins has significant implications for designing new selective drugs. The research team proposes that selectivity arises owing to differences in conformational equilibrium between CB1 and CB2 that could lead to different populations of binding-competent poses of the receptor.

Compared to other class A GPCRs, CB1 exhibits a significant movement of the extracellular orthosteric binding side (binding side volume change = approximately 300 Å<sup>3</sup>) owing to the movement of helix I and II [2]. To represent the intracellular and extracellular movement of helices during activation, the research-

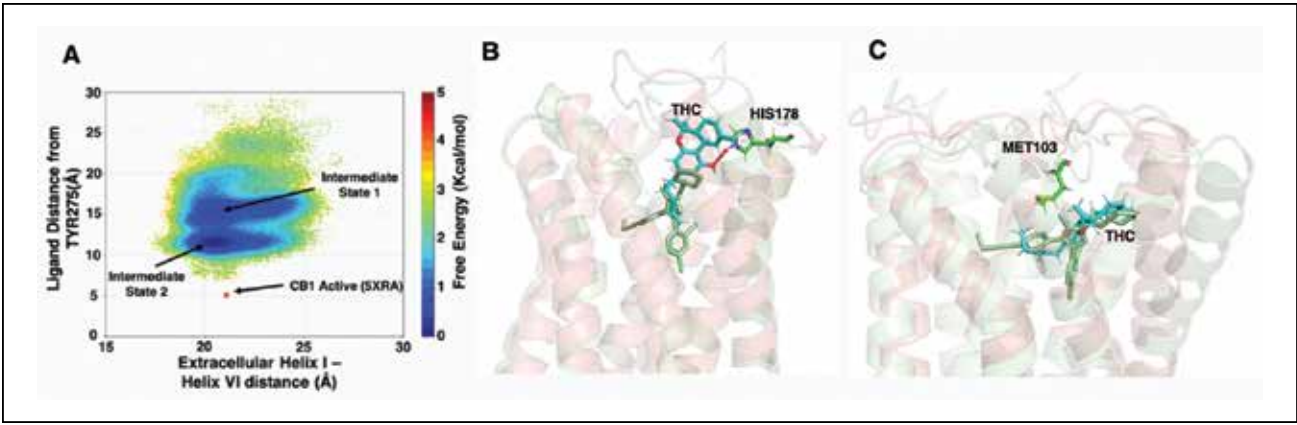


Figure 2: Agonist binding mechanism in CB1. (A) Δ<sup>9</sup>-THC binding pathway. (B) The comparison of the antagonist-bound structure with intermediate state 1 and (C) intermediate state 2, respectively. The Δ<sup>9</sup>-THC (cyan) and antagonist (green) molecules are represented as sticks.

ers considered the distance between intracellular helix III and helix VI and extracellular helix I and helix VI. MSM-weighted free energy landscapes of MD simulation data with respect to extra- and intracellular helix movements reveal the activation mechanisms of the CB1 receptor (Fig. 1a). The team also observed new stable intermediate conformations of the receptor that mediate the (de)activation process. More insights into the CB1 structural change can be revealed from the movements of the “toggle switch” residue pair (TRP 356 and PHE 200). This movement is important for intracellular helix VI to move out during the activation process. The MSM-weighted free energy landscape using root-mean-square deviation of both these residues with respect to the inactive state reveals a pathway for the flipping of the toggle switch residues (Fig. 1a).

The recent discovery of the crystal structure reveals interesting differences between inactive structures of CB1 and CB2. The extracellular part of inactive CB2 matches with the active part of CB1, while the intracellular part matches with the inactive part of CB1. A study on cardiovascular diseases reveals that inactive CB1 (or antagonist-bound) and active CB2 (or agonist-bound) protected against antipsychotic clozapine-induced carbotoxicity [3]. This opposing effect leads the research team to hypothesize that the inverse movement of the extracellular helix could be observed during CB2 activation. MD simulation shows that the extracellular helix I and the intracellular helix VI move outward (Fig. 1b). This movement could lead to the formation of the active state for CB2.

To understand the molecular mechanism of partial agonism exhibited by Δ<sup>9</sup>-THC, the researchers performed simulations of Δ<sup>9</sup>-THC binding to CB1 (Fig. 2a). They observed a gradual movement of Δ<sup>9</sup>-THC toward the binding pocket via the opening between helices I, II, and the N-terminus loop. First, the alkyl chain of Δ<sup>9</sup>-THC moves in the pocket from the extracellular water side.

The bulkier aromatic portion of the ligand faces the maximum barrier between the N-loops and helix I as the hydroxyl group of the ligand forms hydrogen bonds with polar residues, which leads to the formation of intermediate state 1 (Fig. 2b). Overcoming this resistance, the ligand moves further toward the binding pocket of the receptor and is stabilized in the antagonist bound pose of CB1 (Fig. 2c). Steric hindrance from the MET 103 residue blocks the Δ<sup>9</sup>-THC from going inside the binding pocket.

Discerning the atomistic details of CB1 and CB2 selectivity and partial agonism will aid selective drug design for CB1. The conformational space of CB1 reveals the generation of the intermediate state during activation by toggle switch pair movement (TRP 356 and PHE 200 residues moving toward the intracellular and extracellular side, respectively.) Additionally, the Δ<sup>9</sup>-THC binding simulation reveals the initial binding pathway for partial agonists and important residues that can be responsible for selectivity.

WHY BLUE WATERS

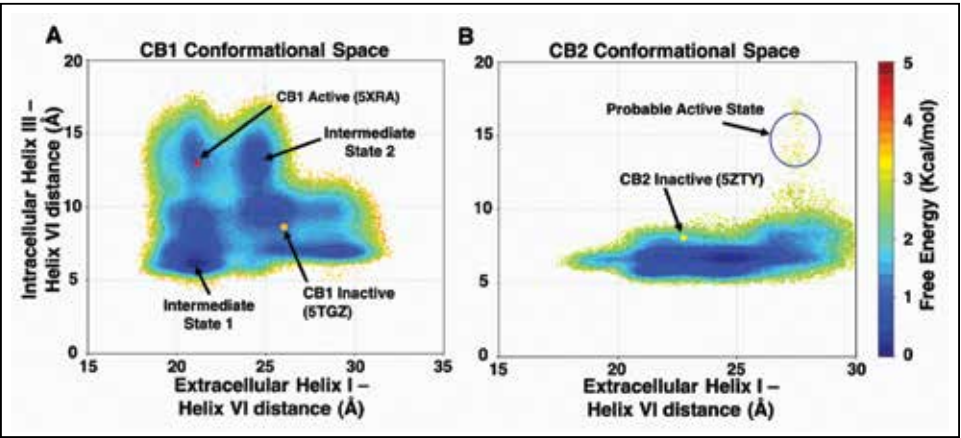
Observing the activation and ligand binding mechanism of a protein receptor is a computationally expensive process. The computer architecture of Blue Waters allowed the research team to perform hundreds of microseconds of MD simulations to understand the necessary conformational changes of these proteins. The adaptive sampling method helped the team to utilize the GPU power of Blue Waters very efficiently. The current project would not have been possible without Blue Waters’ computational facility.

PUBLICATIONS & DATA SETS

S. Dutta, B. Selvam, A. Das, and D. Shukla, “Molecular basis of Δ<sup>9</sup>-THC binding pathway in CB1,” in preparation, 2019.

S. Dutta, A. Das, and D. Shukla, “Activation mechanism of CB1 and CB2 receptors,” in preparation, 2019.

Figure 1: Activation mechanism of CB1 and CB2. (A) MSM-weighted free energy landscape of CB1 and (B) CB2.





# FULL-SCALE BIOPHYSICAL MODELING OF HIPPOCAMPAL NETWORKS DURING SPATIAL NAVIGATION

**Allocation:** NSF PRAC/8,150 Knh  
**PI:** Ivan Soltesz<sup>1</sup>  
**Collaborators:** Ivan Raikov<sup>1</sup>, Aaron Milstein<sup>1</sup>, Darian Hadjiabadi<sup>1</sup>

<sup>1</sup>Stanford University

## EXECUTIVE SUMMARY

This work is the first attempt to fully understand how memories are formed in the brain by means of a detailed computational model of each neuron in the hippocampus, a brain area important for learning and memory. The computational capacity of Blue Waters allows the research team to rapidly conduct simulations of brain function at 1:1 scale, to observe and record the behavior of millions of model neurons, and to compare the results with experimental data. The researchers’ model is capable of representing spatial location via the activation of a small number of neurons sensitive to a specific location, closely approximating the navigational system used by the brain. Further, the software infrastructure developed for this project is now allowing the team to explore several hypotheses about the formation of neural sequences as part of the process necessary for storing memory traces.

## RESEARCH CHALLENGE

This research aims to elucidate the mechanisms of sharp-wave ripples (SWRs), which are oscillatory events in the hippocampus that are required for memory consolidation and subsequent recall. To support this goal, the research group’s computational projects aim to construct the first-of-its-kind, full-scale, biophysically detailed computational models of the three major neuronal circuits within the mammalian hippocampus: the dentate gyrus, CA3, and CA1. These models will be used to provide insight into the dynamical properties of hippocampal networks that produce the SWR-specific oscillatory patterns. Furthermore, the team proposes to utilize their full-scale models to study the mechanisms of abnormal dynamics that emerge in epilepsy.

## METHODS & CODES

The principal simulation environment the researchers use is NEURON [1], <http://www.neuron.yale.edu/neuron/>, which describes neurons in terms of membrane properties and geometric structure [2] and networks in terms of connections between neurons [3]. The biophysical dynamics of the neuronal membrane are described by differential-algebraic equations solved by an implicit integrator optimized for branched structures [1]. NEURON can be fully parallelized via message passing interface with near-linear scaling [3].

The representation of the geometric structure of neurons and their connectivity requires hundreds of gigabytes for each of the models, which has necessitated a parallel computational infrastructure for data management. Thanks to the Petascale Application Improvement Discovery program, the research team has developed a parallel I/O software substrate based on the HDF5 file format that allows the rapid generation and analysis of neuronal morphology and connectivity data according to user-specified rules about neuronal structure and distribution of connectivity in a 3D volume.

The construction of biophysical models of neurons involves tuning a set of parameters in order to make the model neuron dynamics closely match experimental recordings of real neurons. Because of complex dendritic geometry and nonlinear ion chan-

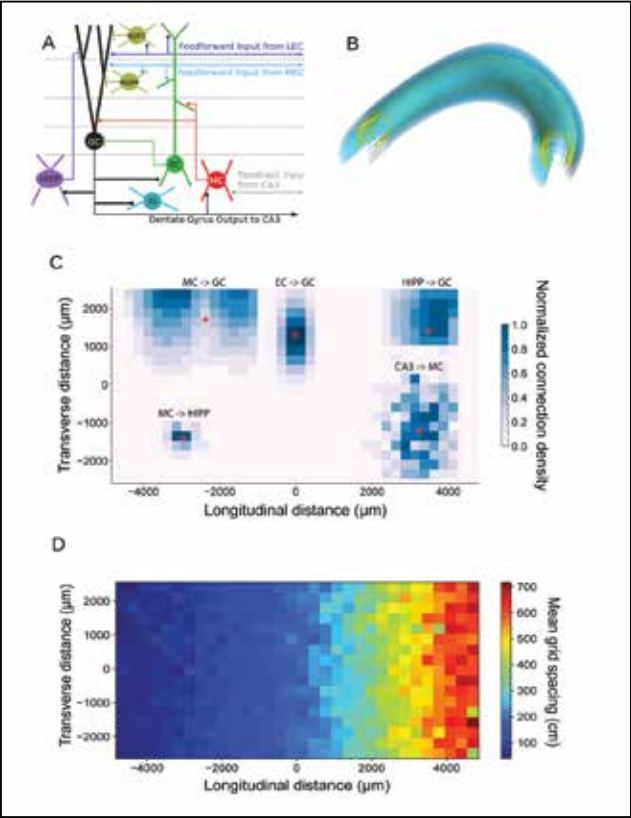


Figure 1: (A) Circuit structure of dentate gyrus model. (B) 3D anatomy and cell body distribution. (C) Distance-dependent connectivity structure between different neuron types. (D) Topological distribution of grid cell inputs with different spacing to dentate gyrus.

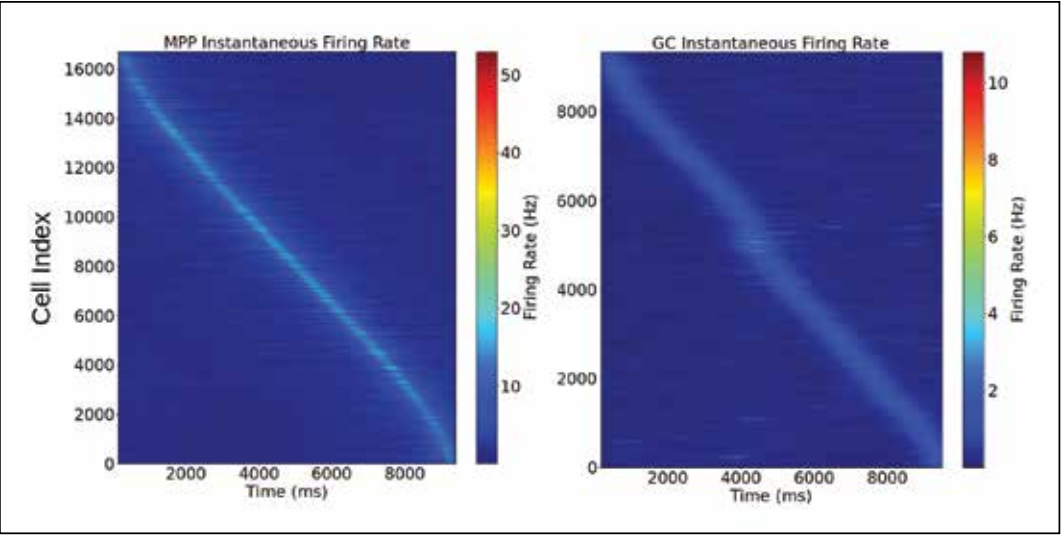


Figure 2: (A) Firing rate map of simulated spatial inputs to dentate gyrus model. (B) Firing rate map of granule cells, ordered by peak firing rate. Consistent with experimental observations, the granule cells in the model exhibit highly sparse activity, with approximately 1% active cells on a given trajectory.

nel dynamics, the model parameter space is enormous. The team has devised a multiobjective evolutionary optimization algorithm that varies ion channel distributions in order to fit the experimentally obtained electrophysiological properties of the target neuron type. This optimization method can be applied not only to models of diverse neuron types but also to a number of other problems ranging from the study of information processing in microcircuits to parameter tuning of large-scale network dynamics.

## RESULTS & IMPACT

The research team has made significant progress in developing a biophysically detailed, full-scale model of the rodent hippocampus comprised of realistically diverse cell types, cell-type-specific connectivity, and nonuniform distributions of synaptic input strengths. The researchers have constructed a full-scale model of the input layer to the hippocampus, the dentate gyrus (DG), to generate sparse, selective, and sequential population activity that matches *in vivo* experimental data [4,5]. The DG model has also served as the prototype to develop general software infrastructure to specify, simulate, optimize, and analyze large-scale biophysically detailed neuronal network models and is scalable across tens of thousands of processors. The team is extending the DG model to include CA3 and CA1 [6] and will release their modeling framework as a general open-source tool to construct models of any brain region at any level of scale and detail.

The DG model now contains over 1.2 million model cells and receives its principal input from stimulus patterns designed to mimic the spatial information content provided to the hippocampus by the entorhinal cortex. The neuronal synapses include NMDA receptors, which have a characteristic nonlinear transfer function. The researchers have found this biophysical property to be critical for a highly sparse subset of granule cells to fire at high rates within their place fields while maintaining essentially zero firing rate out of field, despite receiving a constant barrage

of excitatory inputs. The team has been able to use their model to test a prominent theory that a log-normal distribution of synaptic weights combined with lateral feedback inhibition is sufficient to generate singly peaked place fields when being driven by multiply peaked gridfield inputs from the medial entorhinal cortex [7]. The model now also contains a back-projection from CA3 pyramidal cells to DG, which is the first step toward creating an interconnected model of the hippocampus. The back-projection will enable testing of a theory of SWR-related memory replay that posits that interactions between CA3 and DG are required for proper sequence generation [8].

## WHY BLUE WATERS

The brain is one of the most complex systems studied by science. Computational modeling of the brain presents challenges in terms of both the mathematical complexity of neuronal dynamics and the large number of neurons and other cells that are part of the nervous system. This project requires the simulation of behaviorally relevant activity of approximately 2% of the total number of neurons in the rodent brain; but even then, such simulations have only been possible on Blue Waters, where producing seconds of simulated brain activity currently takes tens of hours on thousands of processors. Blue Waters has been indispensable in enabling the research team to simulate the brain at a biologically realistic anatomical scale and a behaviorally relevant timescale.

## PUBLICATIONS & DATA SETS

I. Raikov and I. Soltesz, “Data-driven modeling of normal and pathological oscillations in the hippocampus,” in *Multiscale Models of Brain Disorders*, Vassilis Cutsuridis, Ed. New York, NY, U.S.A.: Springer Nature, 2019.



DESMOSOMAL CADHERINS BEATING UNDER TENSION

Allocation: GLCPC/558 Knh  
PI: Marcos Sotomayor<sup>1</sup>

<sup>1</sup>The Ohio State University

EXECUTIVE SUMMARY

A seminal event in the history of life on Earth is the transition from single-celled organisms to multicellular life, which required the ability to form strong yet dynamic cell–cell contacts. Among the many classes of molecules that fulfill this role, the cadherin superfamily of cell adhesion proteins is one of the most prominent. Two members of the classical cadherin family, desmoglein (DSG) and desmocollin (DSC), form the robust cell–cell contacts known as the desmosomes, which provide mechanical strength to skin epithelial and cardiac tissues in the face of constant stress, normal stretching, and which protect them from cuts and abrasions. Using the molecular dynamics engine NAMD on Blue Waters, the PI was able to perform simulations on an atomistic model of the desmosome. These simulations provided insight into how this essential cellular junction may function and respond to forces, both in its wild type form and with the addition of disease-causing mutations.

RESEARCH CHALLENGE

Skin epithelial and cardiac tissues are subject to constant stress. The skin epithelium must withstand stretching and shearing forces as well as abrasions, while cardiac tissue must remain viable despite continuous contraction and expansion. To maintain the integrity of these tissues, the cells that comprise them have developed strong and mechanically robust contacts with one another, including the more apical adherens junction and the more basally situated desmosome. The desmosome is found in skin epithelial and cardiac tissues and is formed through interactions of the desmosomal proteins DSG and DSC. However, the stoichiometry and three-dimensional arrangement of these proteins in the junction remain largely unknown.

The architecture of the adherens junction is believed to be a well-ordered array of molecules, forming a zipperlike pattern in which both *trans*-interactions (those formed from opposite cells through the strand-swap mechanism) and *cis*-interactions (those between molecules from the same cell) are present in the mature lattice [1]. The molecular arrangement in three-dimensional space has been more difficult to identify in desmosomes, however. While it is known that both DSG and DSC are required for the formation of the mature desmosome [2,3], the stoichiometric relationship between DSG and DSC and the arrangement of molecules in the junction is a question that remains to be decisively answered. Models from cryo-electron tomographic imaging and those based on the structure of E-cadherin lattices have been suggested, however, and with the deposition of high-reso-

lution crystal structures of DSG and DSC, further information can be extracted from these models [4–7].

Lattices of DSG and DSC, constructed based on the crystallographic lattice of C-cadherin, offer insight into how these proteins may be arranged in the desmosome and how their interactions contribute to desmosome function. In addition, these models can offer molecular explanations for skin epithelial and cardiac diseases. Several missense mutations known to cause cardiomyopathies have been mapped to the extracellular domain of DSG and DSC, suggesting their adhesive or mechanical properties are compromised. However, the molecular mechanism underlying this often fatal disease remains unknown. Simulations on these models offer a unique ability to bridge the gap between genetics and cell biology, and this power is leveraged in these simulations to propose a model of desmosomal dysfunction in arrhythmogenic cardiomyopathy.

METHODS & CODES

Models were constructed using high-resolution crystal structures of DSG and DSC [1,7]. Lattices for the desmosome were assembled based on the crystallographic lattices present in the solution of classical cadherins and contain four DSG and four DSC molecules. This was done both in a “polarized” fashion, with all DSG on one side and all DSC on the other. To replicate the physiological conditions located in the extracellular space in which the desmosome is found, models were solvated in explicit water and ions. Models were constructed using VMD (Visual Molecular Dynamics) [8] and simulated using NAMD [9], both of which are developed at the University of Illinois at Urbana–Champaign. In addition to the wild-type systems, two separate systems were built by introducing mutations to the DSG molecules in the lattice, both of which cause the inherited and often fatal disease arrhythmogenic cardiomyopathy. All systems consist of approximately 1.8 million atoms.

RESULTS & IMPACT

The wild-type lattice was simulated in equilibrium for 20 nanoseconds (ns), and was subsequently subjected to stretching steered molecular dynamics simulations at 10, 1, and 0.1 nanometer (nm)/ns. Additionally, creep tests were performed in which a constant force was applied to one face of the lattice with the other held fixed, allowing the viscoelastic behavior of the junction to be studied. The resultant elastic response and mechanical properties of the complexes were analyzed, providing a glimpse of how these essential cellular junctions may respond to force *in vivo*. Such mod-

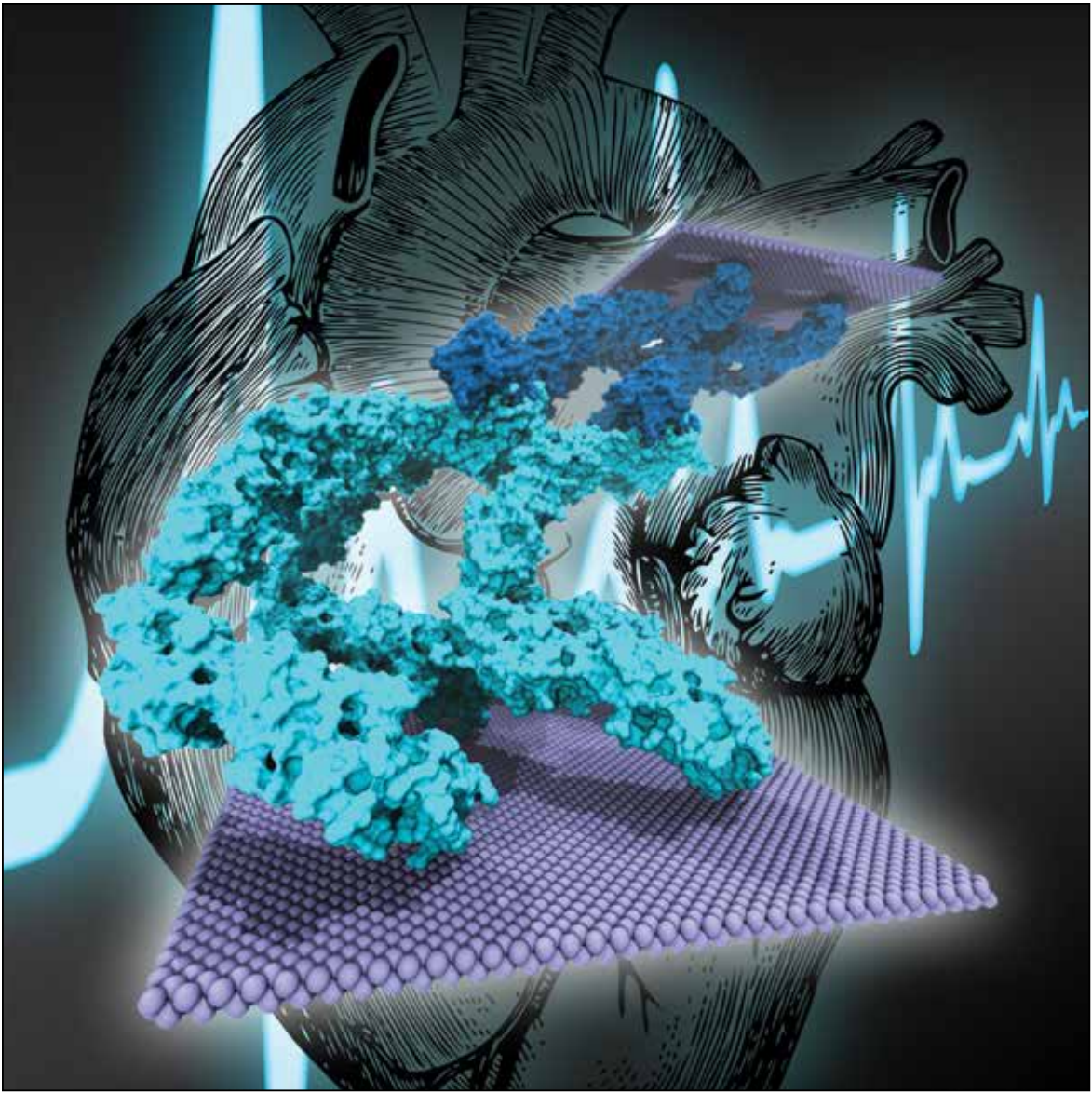


Figure 1: The desmosome represented by the extracellular domains of four molecules of DSG (cyan) and four molecules of DSC (blue; 1.8 million atom system). The system shown is of a polar arrangement of molecules. The membrane planes are shown in pale blue.

eling and simulation efforts offer unique insights into how these molecules may arrange physiologically and may guide future experimental efforts. Simulations have revealed the formation of *cis*-contacts that persist throughout equilibrium and stretching, suggesting a possible lateral contact that may have relevance in desmosome function. Creep tests revealed differing viscoelastic behavior in these proteins in the lattice as opposed to individual dimers. In particular, the damping coefficient for each individual protomer was increased when it was tested in the lattice compared to the dimer alone, suggesting the lattice acts as a shock absorber to distribute force across the junction. This property would be vital to maintain desmosomal integrity as it is subjected to very rapid and often violent perturbations.

WHY BLUE WATERS

Atomistic simulations as large as those performed on these models of the desmosome, which are composed of up to 1.8 million atoms, require significant computational resources. Such simulations need a fast networked and massively parallel system such as Blue Waters. With the slowest stretching speed, at 0.1 nm/ns, the PI was able to sample hundreds of nanoseconds, a timescale that would be unfeasible for such a large system without the capabilities provided by Blue Waters.



# SIMULATION OF VIRAL INFECTION PROPAGATION DURING AIR TRAVEL

TN

**Allocation:** NSF PRAC/25 Knh  
**PI:** Ashok Srinivasan<sup>1</sup>  
**Co-PIs:** Sirish Namilae<sup>2</sup>, Anuj Mubayi<sup>3</sup>, Matthew Scotch<sup>3</sup>, Robert Pahle<sup>4</sup>  
**Collaborator:** Sudheer Chunduri<sup>5</sup>

<sup>1</sup>University of West Florida  
<sup>2</sup>Embry–Riddle Aeronautical University  
<sup>3</sup>Arizona State University  
<sup>4</sup>New York University  
<sup>5</sup>Argonne National Laboratory

## EXECUTIVE SUMMARY

There is direct evidence of the spread of infection during commercial air travel for common infectious diseases including influenza, Severe Acute Respiratory Syndrome (SARS), tuberculosis, and measles. This has motivated calls for restrictions on air travel, for example during the 2014 Ebola outbreak. However, such restrictions carry considerable economic and human costs. Ideally, decision-makers ought to take steps to mitigate the likelihood of an epidemic without imposing restrictions. Toward that end, science-based policy analysis can yield useful insight to decision-makers.

The effectiveness of any policy depends on the human response to it. Given the inherent uncertainties in human behavior, the research team simulated a variety of scenarios and identified the vulnerability of policies under these potential scenarios. Supercomputing was used to deal with the large number of scenarios and the need for a short response time in case of national emergencies. The results identified new boarding procedures that, if implemented, can result in a substantial reduction in the risk of the spread of infectious diseases.

## RESEARCH CHALLENGE

The researchers’ goal is to develop models and a novel methodology that can provide insight to decision-makers on policies and procedures that will reduce the likelihood of infection spread during air travel. In addition, this research promises major advances in other disciplines such as pedestrian movement modeling, mathematics, epidemic modeling, computer science, and bioinformatics, with a consequent transformative effect on transportation infrastructure and management.

## METHODS & CODES

The research team modeled pedestrians during air travel as particles using the force-field approach proposed by Helbing *et al.* [1]. Both pedestrian density and the speed of the immediate neighbor in a pedestrian line determine pedestrian speed and trajectory [2,3]. The team’s modifications incorporated these aspects into the pedestrian movement model. The pedestrian trajectory information was then integrated with a discrete-time sto-

chastic susceptible-infected model for infection transmission that accounts for demographic stochasticity and variations in susceptibility of the population. This approach provides insight into the consequences of policy choices that change passenger behavior at individual levels. The team input this information into a global phylogeography model to assess the impact of these policies at a global scale.

Inherent uncertainties in human behavior and insufficient data during the initial stages of an epidemic make accurate predictions difficult. The team parameterized the sources of uncertainty and evaluated vulnerability under different possible scenarios. The researchers used Blue Waters to deal with the computational load that arises from a large parameter space, as well as a low discrepancy parameter sweep to explore the space of uncertainties efficiently.

Phylogeography (the study of the historical processes that may be responsible for the contemporary geographic distributions of individuals) uses genetic mutation information and geographic locations of viruses to model the spread of epidemics across large geographic scales. The team used Blue Waters to analyze 264 full-genome Ebola sequences from Guinea, Liberia, Sierra Leone, Italy, the United Kingdom, and the United States. The research group also used the BEAST software installed on Blue Waters to implement the phylogeography model.

## RESULTS & IMPACT

In prior work, the research team used the above approach with Ebola. They studied the impact of different procedures for boarding, disembarkation, and seat assignment on infection spread. That work showed that on a 182-passenger Boeing 757 airplane, random boarding can lead to a substantial reduction in infection transmission compared with the current zone-based boarding. The team also obtained similar results showing the potential for changes in in-plane movement, deplaning procedure, seating arrangement, and plane sizes to reduce the likelihood of infection transmission. The improvements obtained for individual flights by these policy changes can bring substantial benefits over the course of an epidemic. In fact, based on transportation data from 2013, if unrestricted air travel were to have occurred

during the 2014 Ebola epidemic, then the probability of generating 20 infectives per month from air travel could have been reduced from 67% to 40% using better pedestrian movement strategies. This could further be reduced to 13% by exclusively using smaller 50-seat airplanes.

The researchers have extended their approach to other directly transmitted diseases including SARS and influenza. This required changes to include aerosol and fomite transmission mechanisms, while pedestrian movement accounts for the proximity between infectious and susceptible individuals. The team has successfully extended the application to other high-density areas such as airport security-check areas and a generic airport gate. The research group also found that different queueing strategies can generate a decrease in infection risk of up to an order of magnitude.

The team also developed a low discrepancy parameter sweep, which reduced the number of parameter combinations that ought to be tried by one to three orders of magnitude over the conventional lattice-based sweep. Using the number-theoretic properties of a low discrepancy sequence helped balance the load on the Blue Waters machine [4].

## WHY BLUE WATERS

In an emergency, scientists usually need to model a variety of scenarios owing to a lack of data. This leads to a large parameter space of uncertainties, which requires a large computational effort. In addition, the models typically need fine-tuning, which leads to an iterative process where the model is repeatedly tuned based on results from the previous validation step. Consequently, rapid turnaround time is critical, which requires massive parallelism. Such parallelism becomes even more crucial during the course of a decision meeting, where results are typically needed in a short time span. The Blue Waters support team helped optimize parallel I/O in the code to reduce simulation time by a factor of two.

## PUBLICATIONS & DATA SETS

A. Srinivasan, C. D. Sudheer, and S. Namilae, “Optimizing massively parallel simulations of infection spread through air-travel for policy analysis,” in *Proc. 16th IEEE/ACM Int. Symp. Cluster, Cloud, Grid Computing*, 2016.

S. Namilae, A. Srinivasan, A. Mubayi, M. Scotch, and R. Pahle, “Self-propelled pedestrian dynamics model: Application to passenger movement and infection propagation in airplanes,” *Physica A Stat. Mech. Appl.*, vol. 465, pp. 248–260, 2017.

S. Namilae, A. Srinivasan, C. D. Sudheer, A. Mubayi, R. Pahle, and M. Scotch, “Self-propelled pedestrian dynamics model for studying infectious disease propagation during air-travel,” *J. Transp. Health*, vol. 3, no. 2, p. S40, 2016.

S. Namilae, P. Derjany, A. Mubayi, M. Scotch, and A. Srinivasan, “Multiscale model for infection dynamics during air travel,” *Phys. Rev. E*, vol. 95, no. 5, p. 052320, 2017.

P. Derjany, S. Namilae, A. Mubayi, M. Scotch, and A. Srinivasan, “Effect of pedestrian movement on infection transmission during air travel: A modeling study,” in *Transp. Res. Forum Proc.*, 2017.

P. Derjany, S. Namilae, A. Mubayi, M. Scotch, and A. Srinivasan, “Computational model for pedestrian movement and infectious diseases spread during air travel,” in *2018 AIAA Modeling Sim. Technol. Conf.*, p. 0419.

S. Namilae, A. Mubayi, and A. Srinivasan, “Model based policy analysis for infection spread during air transportation,” presented at the Int. Conf. Transp. Public Health, 2018.

S. Chunduri, M. Ghaffari, M. S. Lahijani, A. Srinivasan, and S. Namilae, “Parallel low discrepancy parameter sweep for public health policy,” presented at the IEEE/ACM Int. Symp. Cluster, Cloud, Grid Comput., 2018.



DI

GA

# MRI-BASED BIOMARKERS THROUGH HIGH-PERFORMANCE COMPUTING

**Allocation:** Illinois/250 Knh  
**PI:** Brad Sutton<sup>1</sup>  
**Collaborators:** Alex Cerjanic<sup>1</sup>, Aaron Anderson<sup>1</sup>, Curtis Johnson<sup>2</sup>

<sup>1</sup>University of Illinois at Urbana–Champaign  
<sup>2</sup>University of Delaware

## EXECUTIVE SUMMARY

Age-associated diseases are increasing along with the aging of the world’s population. This includes diseases that affect the brain such as Alzheimer’s disease (AD). AD annually costs \$290 billion and contributes to more than 500,000 deaths yearly in the United States [1]. Leveraging advanced Magnetic Resonance Imaging (MRI) acquisitions and high-performance GPU-enabled computing, the research team has developed a method to investigate properties of the brain blood supply that were not previously measurable. Acquiring high-resolution MRI data of the brain while correcting for small motions of the brain during the scan delivers information on the density and microstructure of the vessels during a noninvasive MRI scan. Previously, these properties of the vessels have only been observable after death but have correlated with disease progression. The current study will enable monitoring these blood vessel characteristics over the lifespan.

## RESEARCH CHALLENGE

MRI is a noninvasive, nonionizing tomographic imaging technique that enables imaging of the human brain with a variety of functional and structural contrasts, all with no more than minimal risk to the participants. For research in the basic sciences involving human subjects, MRI can measure quantities that are difficult to measure or otherwise inaccessible, such as blood flow, vascular density, and the microstructure of blood vessels in the brain. However, the MRI techniques to obtain these measures of the properties of blood vessels have lacked sufficient signal-to-noise ratio and spatial resolution to enable the monitoring of these biomarkers of health and disease. In the current work, the team leverages advanced 3D acquisitions along with a complete signal model for image reconstruction from the acquired data in order to increase the available signal and spatial resolution [2,3]. Image reconstruction can be computationally expensive for large data sets when the research team needs to incorporate compensation for physical effects in the MRI scanner along with motion of the brain during data acquisition. This can result in reconstruction times of a month or more per data set when using a single workstation, severely limiting application of these techniques.

## METHODS & CODES

The team uses a GPU-accelerated image reconstruction package called PowerGrid [4] to implement 3D, field-corrected, non-Cartesian image reconstruction with nonlinear motion correction. To speed the algorithm, the researchers use GPUs to achieve high performance via OpenACC. Specifically, the team uses an implementation of the nonuniform Fast Fourier Transform (NUFFT) [5] that has been GPU-accelerated via OpenACC.

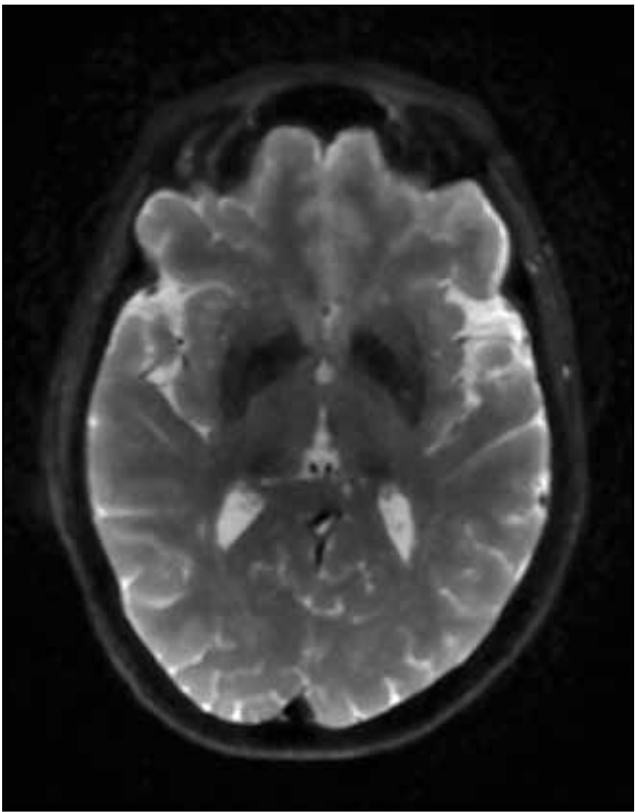


Figure 1: Example of validation image set used to test performance and accuracy between CPU and GPU parallelization via OpenACC.

## RESULTS & IMPACT

The use of OpenACC on a high-performance computing platform such as Blue Waters is not without its challenges, and the interoperability of packages and compilers needs to be addressed to more efficiently leverage the large number of GPU nodes. However, large data sets such as the one in this project are not feasible to reconstruct without utilizing massive multicore strategies. The team focused on increasing the multicore utilization for NUFFT, which is the main kernel in their code, expanding the portable parallelization to use either CPUs or GPUs depending on the environment. With only one hardware-specific library (that handles efficient Cartesian FFTs), researchers can swap CPUs or GPUs for these libraries using the latest OpenACC-capable compilers. This enables transitions to future computing platforms and more widespread adoption of this approach by enabling the code to perform either on hardware accelerators or CPU cores.

## WHY BLUE WATERS

Blue Waters provides an ideal environment for the highly parallel task of image reconstruction in MRI. Each participant’s data requires reconstruction of many images, each with a different image contrast reflecting different properties of the brain blood flow. Utilizing multiple GPUs enables the acceleration of each image’s reconstruction and the availability of a large number of GPU-enabled nodes allows the simultaneous reconstruction of all the images from several subjects’ data.



# MECHANOBIOLOGY: USING BLUE WATERS TO DECIPHER THE PHYSICAL PRINCIPLES OF PROTEIN MECHANICS

**Allocation:** NSF PRAC/9,700 Knh  
**PI:** Emad Tajkhorshid<sup>1</sup>  
**Co-PIs:** Rafael C. Bernardi<sup>1</sup>, James C. Phillips<sup>1</sup>, John Stone<sup>1</sup>  
**Collaborators:** Hermann E. Gaub<sup>2</sup>, Michael A. Nash<sup>3</sup>, Edward A. Bayer<sup>4</sup>

<sup>1</sup>University of Illinois at Urbana–Champaign  
<sup>2</sup>Ludwig-Maximilian University of Munich  
<sup>3</sup>University of Basel and ETH Zurich  
<sup>4</sup>Weizmann Institute of Science

## EXECUTIVE SUMMARY

At the heart of many cellular processes, mechanoactive proteins are responsible for converting mechanical cues into biochemical signals. Protein mechanics are also crucial in holding large macromolecular complexes together during mechanical stress, while permitting assembly and disassembly of these complexes to continue to take place at physiologically needed rates. Employing state-of-the-art computational tools and in a close collaboration with experimentalists, the research team has shown how specific protein complexes can be “activated” under shear force to withstand high mechanical loads. They have discovered that protein mechanostability is highly dependent on the direction of the force application. They have also shown that mechanical stability can be achieved by a protein architecture that directs molecular deformation along paths that run perpendicular to the pulling axis. These mechanisms of protein mechanical stabilization have potential applications in biotechnology for the development of systems exhibiting shear-enhanced adhesion or tunable mechanics.

## RESEARCH CHALLENGE

Interfacing biology, physics, and engineering, mechanobiology studies how mechanoactive proteins sense and respond to mechanical cues in different ways. Guided by advances in single-molecule force spectroscopy, researchers are acquiring a new level of understanding of mechanoactive proteins. However, protein mechanics is a challenging topic to study experimentally because molecular-level mechanical properties remain hidden to ensemble-averaging methods. They require techniques such as molecular dynamics that offer simultaneous high spatial (atomic) and temporal (femtosecond) resolutions.

The research team has employed Blue Waters to investigate how cellulosomes, large molecular machines that can efficiently degrade plant fibers, can be mechanically activated to withstand high-force loads. Understanding how cellulosomes work can have a considerable impact on rational design of more efficient enzymatic complexes for biotechnology and bioreactors. Also, cellulosome-containing bacteria were recently found in the human gut, a paradigm-shifting discovery that shows that at least some humans are able to fully digest plant fibers.

## METHODS & CODES

The research group has employed a combination of *in silico* and *in vivo* single-molecule force spectroscopy to elucidate, with atomic resolution, protein mechanics at high-force loads. Using a wide sampling approach, steered molecular dynamics simulations were performed with many replicas, allowing the team to analyze experiments and simulations in the same statistical framework. Wide sampling combined with dynamic network analysis allowed them to visualize most probable deformation pathways through the protein architecture and understand how resistance to mechanical stress arises at the level of individual protein complexes. Simulation results were validated using single-molecule force spectroscopy experiments carried out with an atomic force microscope over a range of loading rates. By recording the force required to break the protein complexes for thousands of individual interactions, the team collected sufficient statistics to analyze the interactions and unfolding pathways thoroughly. They achieved remarkable agreement between simulations and experiments, demonstrating that they probe fundamentally the same molecular process.

## RESULTS & IMPACT

The chemistry at the interface between biomolecules affected by mechanical forces is still largely unexplored. Even the best-studied mechanically stable biomolecular complex, namely the streptavidin:biotin complex, has mechanical properties that have only recently been revealed. By using Blue Waters, the research team has shown that a wide range of rupture forces previously reported for the separation of the complex are due to unspecific tethering of streptavidin [1]. In fact, different anchoring points in the same protein can lead to unfolding or unbinding of the complex [2]. In cellulosomes, which are usually found in harsh environments, mechanical stability at specific pulling geometries is crucial for their high efficiency. The team had previously shown that a cohesin:dockerin (two major components of a cellulosome) complex could withstand forces up to four times that of the streptavidin:biotin complex [3], which was previously considered to be the strongest protein complex.

In this project, the researchers discovered a cohesin:dockerin complex that is significantly stronger than the one they previ-

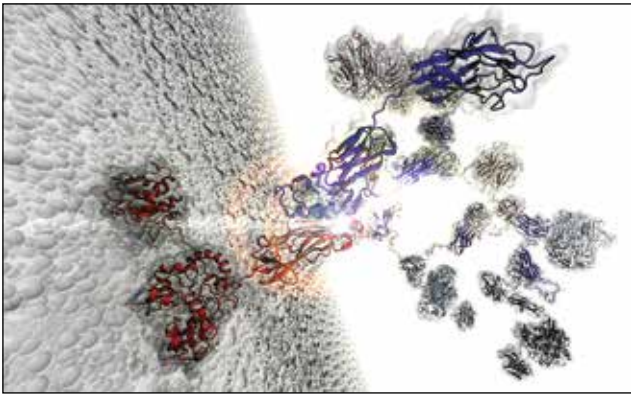


Figure 1: Illustration of a cellulosome, a multienzyme complex found in many plant-fiber degrading bacteria. Cellulosomes are usually attached to the outside of the bacterial cell wall (white surface). Like LEGO bricks, cellulosomes can be assembled in many different shapes by highly stable protein-protein (cohesin:dockerin) interactions. One such interaction is highlighted at the center.

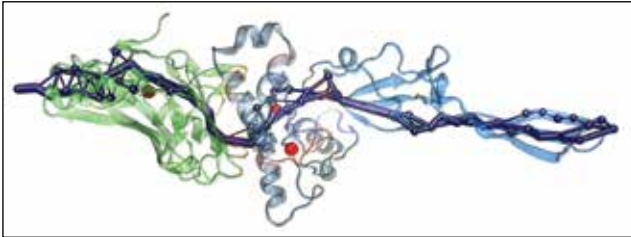


Figure 2: Illustration of an ultrastable cohesin (green):dockerin (blue) interface found in rumen bacteria's cellulosomes. Using a correlation-based network analysis, the researchers calculated the force propagation pathway at the interface. The mechanical stability is achieved by a protein architecture that directs force along paths running perpendicular to the pulling axis at the interface.

ously reported [4]. The scaffold B (ScaB) is found within the same cellulosome as the previously reported CttA complex. The role of the ScaB complex is to connect the large protein machinery of the cellulosome to the cell wall via high-affinity binding to the same cohesin of the CttA complex. This cohesin is itself covalently linked to the peptidoglycan cell wall (Fig. 1). The team's single-molecule experiments show that the ScaB complex, which is among the most mechanically stable complexes known [4], withstands forces up to a nanonewton. The team's simulation results reveal that the mechanical stability is achieved by an architecture that directs the force along pathways running perpendicular to the pulling axis (Fig. 2).

Taking advantage of Blue Waters' size and power, the team has also investigated how different cohesin domains found in a single cellulosome scaffold withstand shear forces. The results show that in the same scaffold, different cohesins can have a large (up to a factor of four) difference in mechanical stability, depending on the cohesin's position in that scaffold [5]. Remarkably, all seven cohesins in the same scaffold share high sequence identity. A striking agreement was observed between simulation and experiments, motivating the researchers to computationally explore how point mutations could affect the mechanical stability of co-

hesins. In an intelligent design strategy, simulations revealed that a single mutation in an alanine residue, replacing it by a glycine residue, could make one of the weaker cohesins much stronger. The experiments also revealed that this mutation, which replaces a methyl group by a single H, was responsible for a 2.6-fold increase in the mechanical stability of that cohesin [5]. The team's results show the vast potential offered by large supercomputers for molecular engineering and biotechnology.

## WHY BLUE WATERS

Fast turnover time is fundamental when working synergistically with experimentalists. Blue Waters' sheer size allowed the researchers to obtain many simulation replicas at about the same speed experiments can be performed. Without Blue Waters, testing new hypotheses computationally would not have been feasible.

## PUBLICATIONS & DATA SETS

- S. M. Sedlak, L. C. Schendel, H. E. Gaub, and R. C. Bernardi, “Streptavidin/biotin: tethering geometry defines unbinding mechanics,” *Sci. Adv.*, in print, 2019.
- R.C. Bernardi *et al.*, “Mechanisms of nanonewton mechanostability in a protein complex revealed by molecular dynamics simulations and single-molecule force spectroscopy,” *J. Am. Chem. Soc.*, vol. 141, no. 37, pp. 14752–14763, Sept. 2019.
- W. Yang *et al.*, “In situ conformational changes of the *Escherichia coli* serine chemoreceptor in different signaling states,” *mBio*, vol.10, no. 4, July 2019.
- S. M. Sedlak *et al.*, “Direction matters: Monovalent streptavidin/biotin complex under load,” *Nano Lett.*, vol. 19, no. 6, pp. 3415–3421, June 2019.
- J. A. Coleman *et al.*, “Serotonin transporter–ibogaine complexes illuminate mechanisms of inhibition and transport,” *Nature*, vol. 569, no. 7754, pp. 141–145, Apr. 2019.
- T. J. K. Haataja *et al.*, “Non-syndromic mitral valve dysplasia mutation changes the force resilience and interaction of human Filamin A,” *Structure*, vol. 27, no. 1, pp. 102–112, Jan. 2019.
- M. C. R. Melo *et al.*, “NAMD goes quantum: an integrative suite for hybrid simulations,” *Nature*, vol. 15, no. 5, pp. 351–354, May 2018.
- C. Sun *et al.*, “Structure of the alternative complex III in a supercomplex with cytochrome oxidase,” *Nature*, vol. 557, no. 7703, pp. 123–126, May 2018.
- W. Ma, K. D. Whitley, Y. R. Chemla, Z. Luthey-Schulten, and K. Schulten, “Free-energy simulations reveal molecular mechanism for functional switch of a DNA helicase,” *eLife*, vol. 7, no. e34186, Apr. 2018.
- L. F. Milles, K. Schulten, H. E. Gaub, and R. C. Bernardi, “Molecular mechanism of extreme mechanostability in a pathogen adhesin,” *Science*, vol. 359, no. 6383, pp. 1527–1533, Mar. 2018.
- T. Verdorfer *et al.*, “Combining *in vitro* and *in silico* single-molecule force spectroscopy to characterize and tune cellulosomal scaffoldin mechanics,” *J. Am. Chem. Soc.*, vol. 139, no. 49, pp. 17841–17852, Dec. 2017.



DI

GA

TN

# ATOMISTIC SIMULATION OF A PROTOCELL

**Allocation:** Illinois/994 Knh  
**PI:** Emad Tajkhorshid<sup>1</sup>

<sup>1</sup>University of Illinois at Urbana–Champaign

## EXECUTIVE SUMMARY

Computational biology has been critical in achieving a molecular understanding of the dynamics and function of proteins, nucleic acids, and biological membranes. Most research in this field has been limited to single-protein characterization owing to limits in computational power and analytical methods. However, to truly understand the mechanics of biological systems, proteins need to be studied in the broader context of a cell. This project aims to spearhead a new generation of computational biology research that seeks insight from molecular dynamics (MD) simulations to study cellular-scale processes in atomic detail. The immensely complex nature of a cell belies the elementary physics that govern it, and by using MD simulations a rigorous connection can be made between the underlying physical laws and the emergent phenomena they generate. By pioneering new, large molecular simulations, researchers can begin to gain insight into and to quantify mesoscale biological phenomena.

## RESEARCH CHALLENGE

The living cell represents a spatially complex and highly regulated arrangement of molecules whose coordinated motions and activities underlie all the processes that allow it to grow, reproduce, and carry out a wide spectrum of cellular functions. Recent advances in experimental techniques have targeted the identification of the positions of subcellular organelles, ribosomes, and macromolecules such as proteins, messenger RNA, and DNA at high resolution [1]. This has in turn fed a growing demand for next-generation computational tools that allow researchers to construct realistic, cellular-scale structural models at a range of resolutions; to resolve molecular functional states; and to simulate their stochastic, time-dependent behavior.

Biological membranes constitute the most fundamental component of cellular and subcellular structures, defining the active boundaries not only between the cell and its environment but also between different compartments within the cell. Diverse in form, structure, and lipid composition, they are highly heterogeneous environments containing lipids and proteins that act in concert to regulate the flow of information and materials into and out of the cell [2,3]. Modeling membranes at cellular scales is an emerging challenge that remains to be addressed and, therefore, is one of the critical steps in modeling the cell. The research team has a long tradition in modeling and simulating membranes and membrane proteins, particularly emphasizing the importance of including the natural environment in the simulations [4–8].

## METHODS & CODES

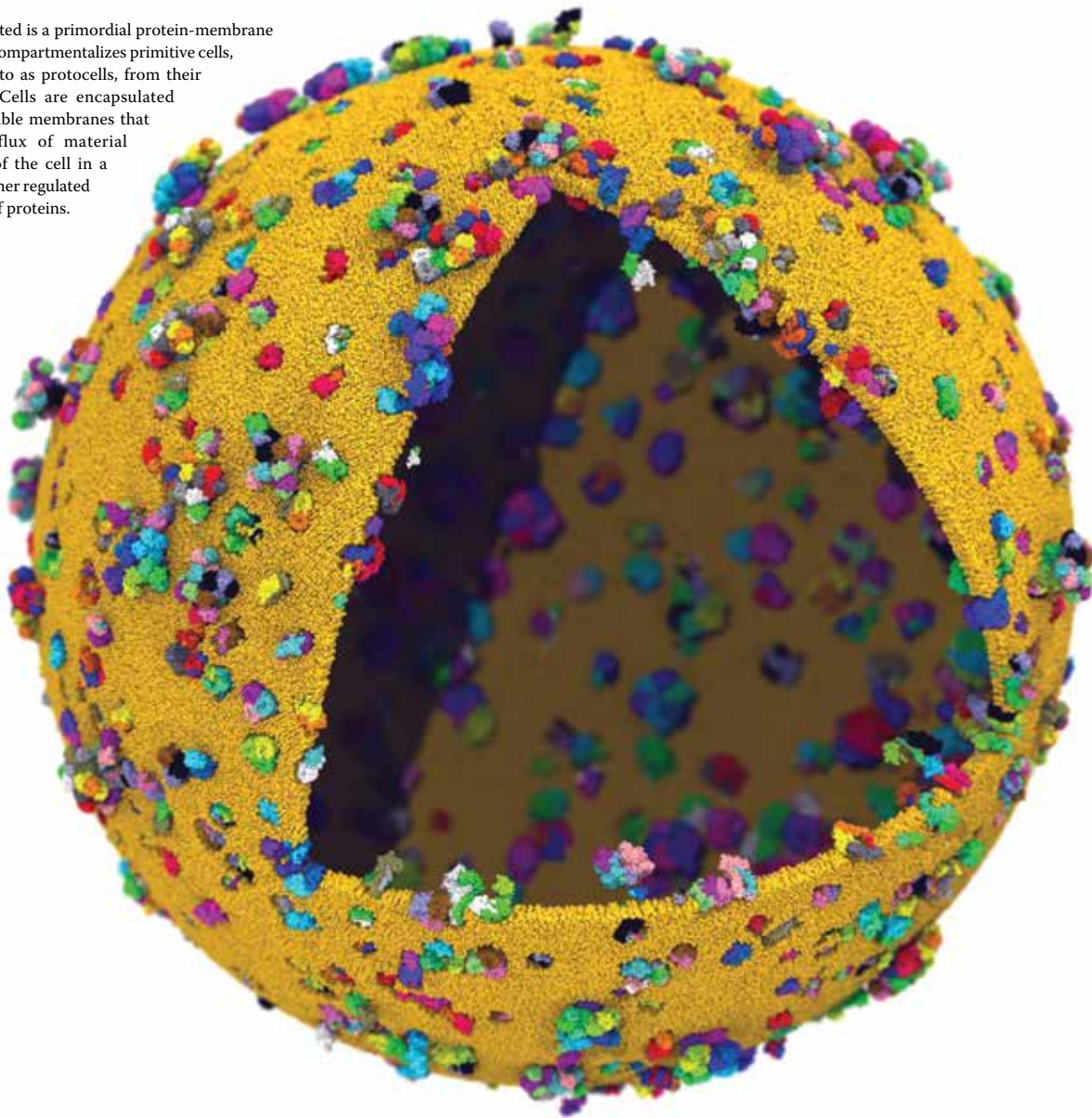
The MD simulations were performed with NAMD [9], a highly parallelized, GPU-accelerated, publicly available MD program with demonstrated scalability to hundreds of thousands of processors for both single- and multiple-replica simulations. All-atom MD simulations rely on the accurate integration of the equations of motion for all atoms of the system. The total potential energy of the system was described by the CHARMM36m force field [10,11]. Periodic boundary conditions were used to avoid surface effects at the simulated system’s boundary, allowing the efficient computation of nontruncated electrostatic interactions by the fast Fourier transform-based particle-mesh Ewald method [12].

## RESULTS & IMPACT

The research team designed, developed, and tested a protocol for the construction of spherical cellular envelopes of any lipid/protein composition, which can be divided into three stages. The first step in the designed workflow was to use experimentally derived or user-specified cell diameters to directly determine the surface area of the inner and outer leaflet of the cell envelope and to accurately calculate the number of lipids needed to construct both leaflets. The membrane shell was then generated using the Fibonacci sphere algorithm [13] to approximately solve the Tammes problem [14,15] for the thousands of lipids required to create a micrometer-scaled membranous protocell. Next, each protein was assigned a unique position on the spherical surface. In order to do this, coarse-bead approximations of each protein were generated using a hierarchical clustering algorithm. These were then simulated using a force field that properly oriented and localized them to the intended membrane surface. Lastly, the membrane and the proteins were merged, which involved removal of requisite lipids from both the inner- and outer-membrane leaflets to ensure conservation of surface area and avoid clashes between lipid and protein molecules.

The number of lipids that each protein displaced was carefully derived from conventional (smaller) protein-embedded lipid bilayer simulations and dictated how many lipids were removed from the spherical membrane. Lipids were hierarchically eliminated based on their degree of impingement into the intended protein volume and any remaining potential protein–lipid clashes were resolved via a brief grid-steered simulation of the membrane that forces any offending lipids from the protein space. This procedure ensured that the exact number of lipids matching the cross-sectional area of the protein (calculated independently for the inner and outer leaflets) were removed. Finally, the mem-

Figure 1: Depicted is a primordial protein-membrane structure that compartmentalizes primitive cells, often referred to as protocells, from their environment. Cells are encapsulated by semipermeable membranes that regulate the flux of material into and out of the cell in a concerted manner regulated by a diversity of proteins.



brane and proteins were merged together and solvated, and the final system was used for MD simulations.

The simulations performed on Blue Waters for this project informed areas of refinement for the research team’s assembly protocol, particularly for the removal of lipids during the incorporation of proteins and solvation procedures. The final simulations indicated the constructed protocell systems are stable and that the assembly protocols are adequate. The project presents the most carefully crafted and stable protocell at full atomic detail (Fig. 1).

## WHY BLUE WATERS

For this project, the researchers performed MD simulations with NAMD of systems measuring one billion atoms in size, which can

only be achieved on a petascale computing platform such as Blue Waters. The GPU accelerators on Blue Waters confer a significant boost to simulation speed (two to three times) versus the non-accelerated CPU-only nodes. NAMD has been extensively tested and optimized for Blue Waters, making use of the high-speed node interconnect and showing sustained petascale performance across hundreds, even thousands, of nodes. The team’s benchmarks demonstrated an efficient scaling performance (greater than 80%) while using up to 2,048 Blue Waters GPU (XK7) nodes for a one-billion-atom system.



GA

# MODELING OF A ZIKA VIRUS ENVELOPE AT ATOMIC RESOLUTION

**Allocation:** Illinois/855 Knh  
**PI:** Emad Tajkhorshid<sup>1</sup>  
**Co-PIs:** Soumyo Sen<sup>1</sup>, Eric Shinn<sup>1</sup>, Aaron Chan<sup>1</sup>, Hyun Park<sup>1</sup>

<sup>1</sup>University of Illinois at Urbana–Champaign

CI

## EXECUTIVE SUMMARY

In 2015, a rampant epidemic of Zika virus infection spread from Brazil to the rest of the Americas. The responsible pathogen, the Zika virus, continues to pose a major health concern. Infections have been linked to the development of Guillain–Barré syndrome in adults and microencephaly in infants. The dangers posed by the Zika virus and other flaviviruses such as the West Nile and Dengue viruses call for a better understanding of their structures and infection mechanisms.

Enhanced knowledge about such viruses can allow scientists to design effective drugs and vaccines to combat future outbreaks. The goal of this project is to provide an atomic-level description of the structure and dynamics of the Zika virus envelope—the outer shell of the virus particle made of protein and lipid—via modeling and molecular dynamics simulations. The team also

explored how the stability of the viral envelope depends on the presence of a lipid bilayer and its composition.

## RESEARCH CHALLENGE

Zika virus, a flavivirus, is a 40-nm-diameter particle consisting of an envelope and a nucleocapsid. The viral envelope of a mature Zika virus has three components: E proteins, M proteins, and a lipid bilayer (Fig. 1). Recent cryo-electron microscopy [1,2] studies have shown that 180 copies of each E and M protein are icosahedrally arranged in the viral envelope. Both E and M proteins are embedded (either fully or peripherally) into a lipid bilayer lining the inner shell of the Zika virus envelope.

The modeling of E and M proteins anchored into the lipid vesicle with proper lipid packing density is the most challenging part of this project. The number of lipid molecules needs to be as ac-

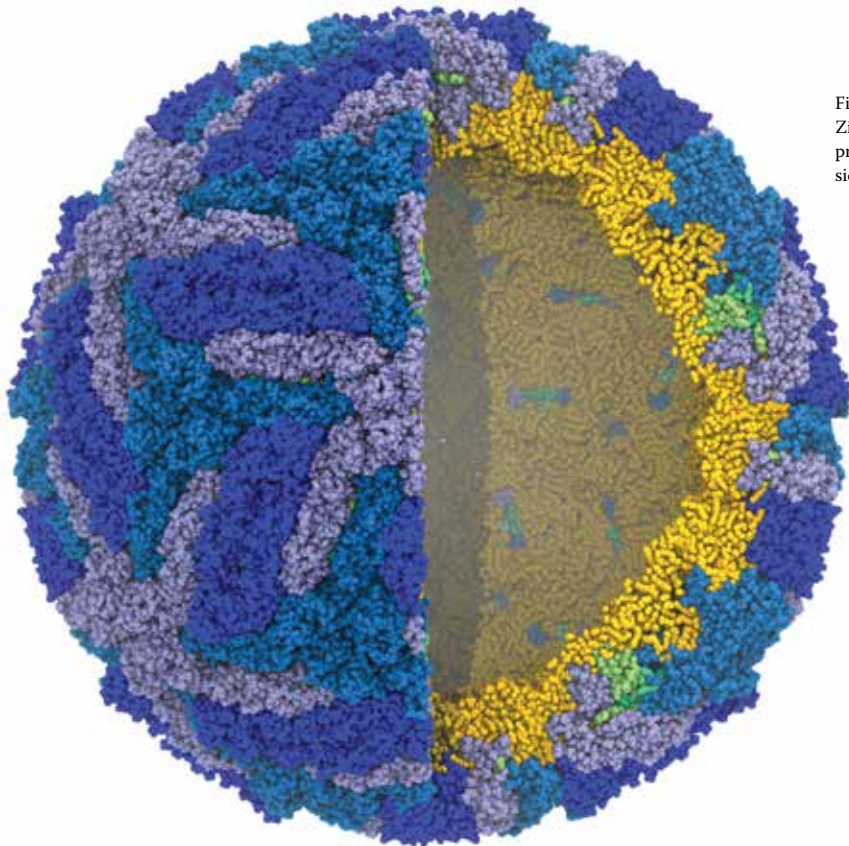


Figure 1: Cross-section of an atomistic model of a complete Zika virus envelope consisting of a protein shell (E and M proteins shown in blue and green, respectively) on the outer side and a lipid bilayer (yellow) forming the inner part.

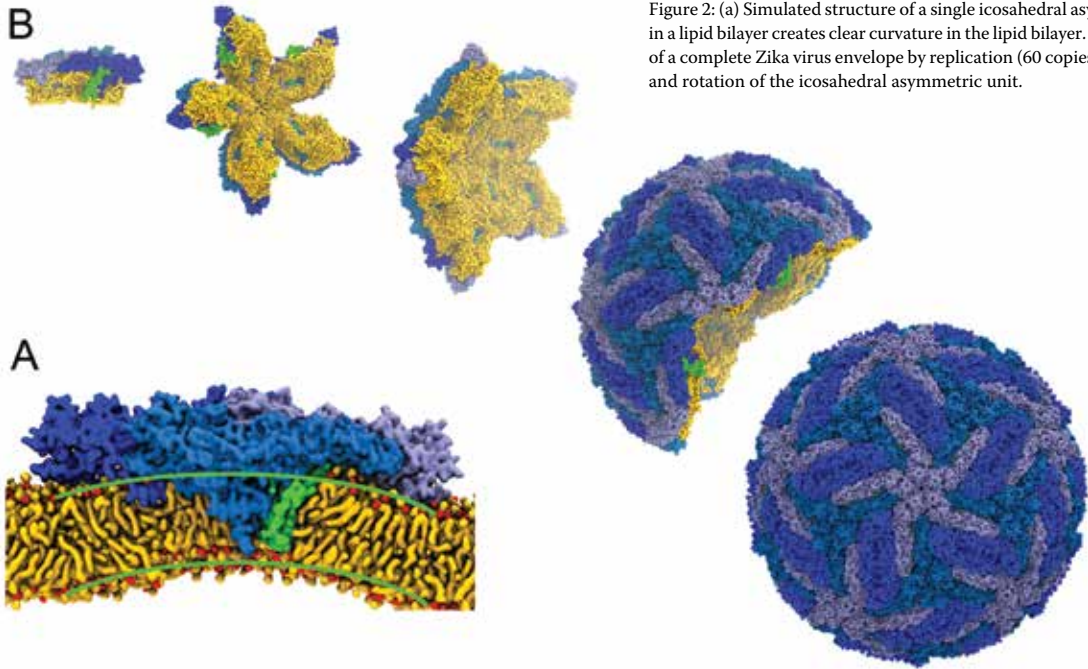


Figure 2: (a) Simulated structure of a single icosahedral asymmetric unit in a lipid bilayer creates clear curvature in the lipid bilayer. (b) Formation of a complete Zika virus envelope by replication (60 copies), translation, and rotation of the icosahedral asymmetric unit.

curate as possible to assemble a reliable and structurally stable viral envelope with correct “breathing” dynamics. The goal is to develop a model for the Zika virus envelope with full atomistic detail in explicit aqueous medium (20 million atoms) and to thereby gain dynamic information on the particle.

## METHODS & CODES

The research team developed and simulated three different systems: (1) a viral protein envelope in the absence of lipids, to serve as a control; (2) a viral protein envelope enclosing a lipid membrane composed of only neutral lipids, to study how lipids contribute to the stability of the shell; and (3) a viral protein envelope enclosing a lipid membrane with a native composition, to examine the effect of specific lipids.

The molecular dynamics (MD) simulations were performed with NAMD [3], a highly parallelized, GPU-accelerated, publicly available MD program with demonstrated scalability to hundreds of thousands of processors for both single- and multiple-replica MD simulations. All-atom MD simulations rely on the accurate integration of the equations of motion for all atoms of the system. The total potential energy of the system was described by the CHARMM36m force field [4,5]. Periodic boundary conditions were used to avoid surface effects at the simulated system’s boundary, allowing the efficient computation of nontruncated electrostatic interactions by the fast Fourier transform-based particle-mesh Ewald method [6].

## RESULTS & IMPACT

To develop a structural model for the whole Zika virus envelope, the researchers first placed a single icosahedral asymmetric unit [2] containing three E and M proteins in a lipid bilayer with

a native composition. They relied on lipidomic analysis of other flaviviruses [7,8] for the composition of the lipid bilayer. A short, 50-nanosecond simulation (Fig. 2a) showed some clear curvature generated in the lipid bilayer by the envelope proteins, which might be indicative of specific lipid–protein interactions leading to the budding process. The next step was to create a boundary defining the maximum spread of the single icosahedral asymmetric unit using a convex hull algorithm and to select a lipid patch that is covered by the proteins. This single protein–lipid patch is replicated 60 times to form the entire Zika virus envelope (Fig. 2b). Since the team’s current model is still imperfect in terms of lipid packing density, they focused on estimating the number of lipid molecules as correctly as possible by measuring the volume in the lipid layer excluded by the stem and transmembrane helices of the proteins. To overcome the lipid–protein overlaps, the team is now designing a grid-force-based simulation protocol.

This project will have a great impact in the modeling of complete virus systems, which can then be used to study the viral infection mechanism.

## WHY BLUE WATERS

The research team plans to simulate three 400-nanosecond simulations of approximately 20 million (M) atoms each by NAMD, which can only be achieved on a petascale computing platform such as Blue Waters. GPU acceleration on Blue Waters allows meaningful simulation timescales. NAMD has been extensively tested and optimized for Blue Waters and shows sustained petascale performance. The team’s benchmarks on 20M atoms show efficient performance (> 82.6%) while using up to 362 Blue Waters GPU (XK7) nodes.



DI

GA

MULTISCALE SIMULATIONS OF COMPLEX SELF-ASSEMBLING BIOMOLECULES: TARGETING HIV-1

Allocation: NSF PRAC/200 Knh  
PI: Gregory Voth<sup>1</sup>

<sup>1</sup>University of Chicago

BW

MP

CI

BI

EXECUTIVE SUMMARY

Biomolecular proteins are capable of performing extraordinary microscopic processes that are used to accomplish the useful macroscopic biological functions of the cell. From changing conformations (shape) to self-assembling into large microcompartments, these nanomachines are at the heart of the cellular machinery that underlies all living activity. The research team’s overarching research goal is to develop advances in theoretical and computational methodology that couple microscopic-scale phenomena to higher-scale descriptions. A key element of the team’s strategy is to systematically coarse-grain or reduce the representations of atomic-scale systems in a manner that is rigorously consistent within the framework of statistical mechanics. The team applies the computational tools it developed to understand biophysical phenomena that pose major public health risks. In the present research, the research team unraveled the mechanisms by which innate immune sensors assemble to block human immunodeficiency virus type-1 (HIV-1) infection and how capsid inhibitor drugs might perturb capsid assembly processes

RESEARCH CHALLENGE

HIV-1 is the causative agent of autoimmune deficiency syndrome (AIDS), which has affected millions of individuals worldwide and resulted in approximately one million AIDS-related deaths in 2017 [1]. Understanding the molecular-scale mechanisms and physical principles that govern viral processes such as the self-assembly of HIV-1 capsids is critical to the effective treatment of the disease. Retroviruses subvert normal cellular processes in order to replicate viral genetic information and assemble micrometer-sized nanostructures (viral capsids) that contain the genetic material, which is transmitted to new host cells. From a computational point of view, these large-scale viral components and the biophysical processes they are involved in consist of the concerted action of thousands of interacting proteins. Molecular simulations require not only large system sizes but also long timescale dynamics on the order of minutes to hours. Clearly, novel simulation methodologies need to be developed to tackle the intrinsic difficulties in computation at these time and length scales.

METHODS & CODES

The research group has made significant strides in developing new multiscale simulation methods to push the boundaries of phenomena that can be studied [2–5]. By combining atomic-lev-

el simulations with multiscale coarse-grained (CG) simulation algorithms, the team was able to investigate the dynamical properties of virus capsid assembly, capsid inhibitor drugs that perturb the maturation process, and intriguing innate immune proteins that wrap around the capsid to block HIV infection. The atomic-scale simulations employed NAMD, a parallel molecular dynamics code designed for high-performance simulation of large biomolecular aggregates, while the coarse-grained simulations use the research team’s custom ultra-coarse-grained software, which was designed and optimized for the Blue Waters machine.

RESULTS & IMPACT

*Viral Capsid Assembly.* During the HIV-1 lifecycle, immature viral particles are converted into mature, infectious particles in a process called maturation. Proteolytic cleavage of the Gag polypeptide releases capsid proteins (CA) that self-assemble into conical capsids that enclose the viruses’ genetic material. The team’s investigations into the self-assembly properties of the CA protein reveal that the slow kinetic growth of the capsid is carefully regulated within the viral membrane envelope. Changes to the concentration of CA, molecular crowding agents, and capsid protein interaction strengths were found to cause the growth of aberrant capsid morphologies. The results suggest that conformational changes available to structurally flexible CA domain dimers are essential for capsid assembly.

*Capsid Inhibitor Drugs.* The viral capsid is also an attractive therapeutic drug target, owing to significantly less genetic flexibility in altering the composition of the capsid as compared to other essential viral proteins such as protease, for which there already exist a wide class of inhibitors. Using CG models, the team simulated how capsid inhibitor drugs that significantly stabilize metastable intermediates in the capsid assembly process can perturb the resulting morphologies of the assembled capsid.

*Viral Capsid Restriction.* Innate immune sensors target the viral capsid and are potent restriction factors that cause the premature destruction of the capsid, thereby blocking HIV-1 infection. In particular, the tripartite-motif-containing proteins, TRIM5 $\alpha$  and TRIMCyp, are cytoplasmic proteins that confer species-specific resistance to HIV-1. The research team’s recent simulations of TRIM5 $\alpha$  revealed how they form hexagonally patterned nets on the surface of the capsid to encage the core of the virus and restrict viral activity. These results show how the coupling between two interaction sites of TRIM5 $\alpha$  can give rise to

self-assembling behavior, in which a critical balance of interaction strengths must be maintained.

These advances in multiscale simulation have opened up remarkable opportunities to study the dynamics of biomedically relevant problems that were largely intractable until now. Technological advances in computational methods that increase the accuracy, predictive power, and range of phenomena under study contribute directly to public health by elucidating the fundamental chemistry underlying infectious diseases.

WHY BLUE WATERS

Multiscale simulations of viral processes are inherently computationally expensive, requiring the evaluation of interactions among many proteins and millions of particles. An immediate advantage of CG models is that their reduced representations are very computationally efficient, allowing leadership petascale resources such as Blue Waters to probe the dynamics of molecular systems that are many orders of magnitude in both length and time scales and would otherwise be inaccessible to conventional molecular simulations. The team’s highly scalable, custom molecular dynamics software was developed with the expertise of Blue Waters staff to perform CG simulations across thousands of computer processors and has been made available to the broader scientific community. The researchers’ atomic-scale and coarse-grained simulations were made possible thanks to the massively parallel computing infrastructure of Blue Waters. These exciting advances in fundamental simulation methodology and novel discoveries in self-assembly phenomena highlight an essential role for access to state-of-the-art leadership-class computing resources.

PUBLICATIONS & DATA SETS

J. M. A. Grime and G. A. Voth, “Highly scalable and memory efficient ultra-coarse-grained molecular dynamics simulations,” *J. Chem. Theory Comput.*, vol. 10, no. 1, pp. 423–431, Jan. 2014.

J. M. A. Grime *et al.*, “Coarse-grained simulation reveals key features of HIV-1 capsid self-assembly,” *Nat. Commun.*, vol. 7, p. 11568, May 2016.

A. J. Pak, J. M. A. Grime, A. Yu and G. A. Voth, “Supercharged assembly: A broad-spectrum mechanism of action for drugs that undermine controlled HIV-1 viral capsid formation,” *JACS*, in review, 2019.

A. Yu, B. K. Ganser–Pornillos, O. Pornillos, and G. A. Voth, “TRIM5 $\alpha$  self-assembly and compartmentalization of the HIV-1 viral capsid,” in preparation, 2019.

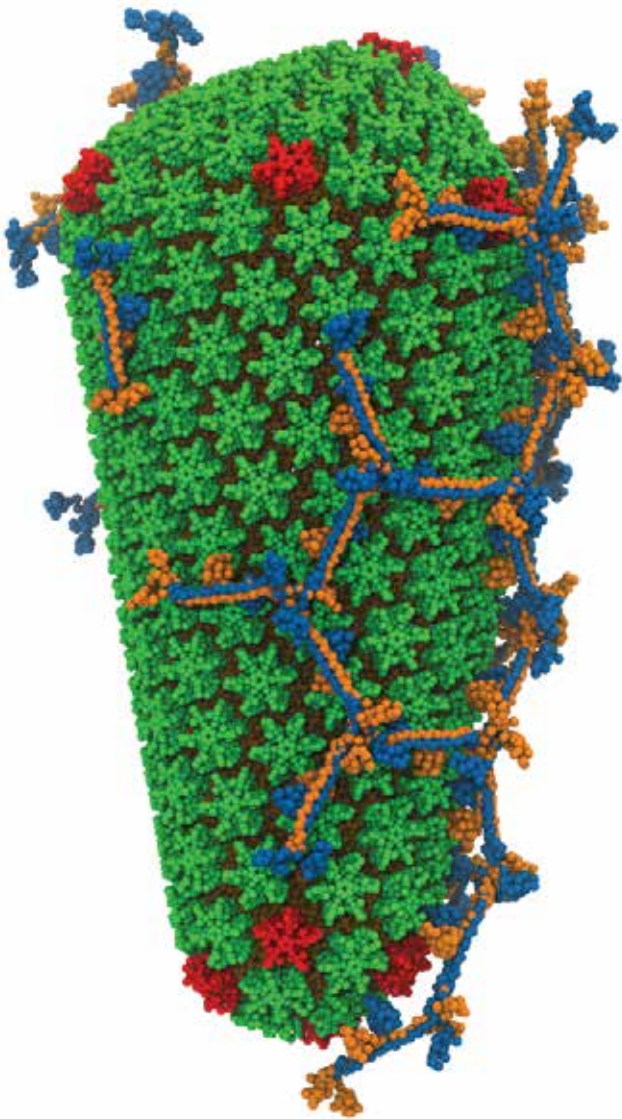


Figure 1: TRIM5 $\alpha$ , an innate immune sensor, encases the core of the HIV-1 virus to restrict viral activity.



# IMPROVING THE AGREEMENT OF AMBER SIMULATION OF CRYSTALS OF NUCLEIC ACID BASES WITH EXPERIMENTAL DATA

**Allocation:** Innovation and Exploration/100 Knh  
**PI:** Victor Anisimov<sup>1</sup>  
**Co-PIs:** Valery Poltev<sup>2</sup>, Thomas E. Cheatham III<sup>3</sup>, Jerry Bernholc<sup>4</sup>  
**Collaborator:** Rodrigo Galindo–Nurillo<sup>3</sup>

<sup>1</sup>University of Illinois at Urbana–Champaign  
<sup>2</sup>Autonomous University of Puebla  
<sup>3</sup>University of Utah  
<sup>4</sup>North Carolina State University

## EXECUTIVE SUMMARY

Classical molecular dynamics is a ubiquitous tool in the bio-pharmaceutical, chemical, and material sciences. Thousands of research teams nationwide and across the globe depend on the AMBER force field to conduct computer simulations of important practical applications. The present work extends the AMBER parameter optimization protocol to reproduce structure and thermodynamic properties of 12 crystals of five nucleobases. The target training data include enthalpies of sublimation at various temperatures, crystal volume, position of the atoms in the crystal, and key intermolecular distances. Further extending that data set, the project employs an unprecedented database of 161 base–base interaction energies to fine-tune the parameters. Based on experimental geometry of the clusters of bases, the database comprises interaction energies computed using the CCSD(T)/6-311++G\*\* level of theory. Multiconformational charge fitting to the electrostatic potential of the clusters extracted from the crystal data produces atomic point charges that closely represent the average electrostatic potential in the crystals.

## RESEARCH CHALLENGE

The utility of molecular dynamics simulations depends on the accuracy of their underlying parameters. Improving the predictive ability of molecular dynamics requires training the method against a large number of experimental data.

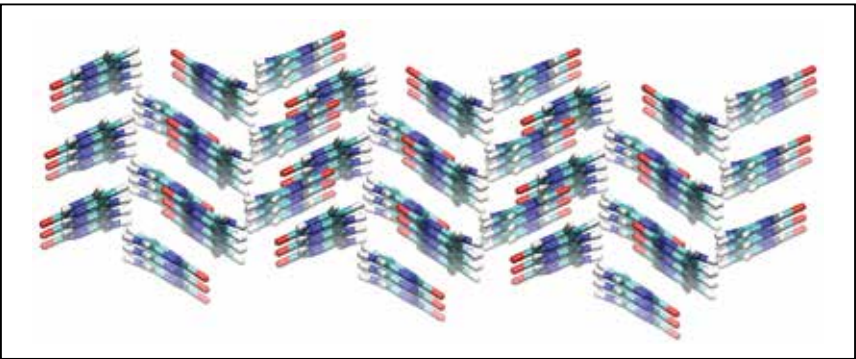


Figure 1: Packing in cytosine crystal.

## METHODS & CODES

The computational methodology involves running molecular dynamics simulations of crystals of nucleobases using the CHARMM package and performing grid searches for the optimal value of parameters to improve the agreement with experimental data.

## RESULTS & IMPACT

The optimized parameters improved the agreement with the reference data set. On that set, the derived parameters scored above the existing CHARMM and AMBER force fields for nucleic acid bases.

## WHY BLUE WATERS

The ability to run hundreds of small jobs that span to a thousand of nodes in the aggregate node allocation in backfill with high job turnaround is unique to Blue Waters.

# ALGORITHMS FOR CANCER PHYLOGENETICS

**Allocation:** Director Discretionary/50 Knh  
**PI:** Mohammed El–Kebir<sup>1</sup>

<sup>1</sup>University of Illinois at Urbana–Champaign

## EXECUTIVE SUMMARY

Cancer is a genetic disease characterized by intratumor heterogeneity, or the presence of multiple cellular populations with different sets of mutations. Cellular heterogeneity gives cancer the ability to resist treatment, and quantifying the extent of heterogeneity is key to improving our understanding of tumorigenesis (the production or formation of a tumor or tumors).

In this project, the research team developed and employed novel phylogenetic techniques (the use of information on the historical relationships of lineages to test evolutionary hypotheses) to reconstruct the evolutionary histories of individual tumors from DNA and RNA sequencing data. Typically, these data are obtained from shotgun sequencing of tumor biopsies using bulk sequencing technology. Highlights of this project’s research outcomes include a novel method to jointly infer a phylogeny and to estimate clone-specific expression profiles from matched RNA and DNA bulk sequencing samples.

## RESEARCH CHALLENGE

A tumor is a heterogeneous population composed of clones with distinct sets of genetic mutations that result from an evolutionary process. To understand and treat cancer, we must view it through the lens of evolution. In particular, we must understand how mutations that drive cancer progression achieve their function by dysregulating (disrupting normal function of a regulatory mechanism) gene expression. Distinct clones may exhibit distinct expression profiles owing to differences in somatic mutations. To elucidate cancer evolution and the functional effect of somatic mutations, we need to infer a phylogeny (evolutionary history and relationship among organisms) that describes the clonal composition of the tumor as well as to identify an expression profile for each clone.

This is a very challenging task because, in practice, the RNA and DNA of tumor biopsies are only sequenced in bulk, resulting in a mixture of sequence reads from a large population of heterogeneous cells. The tumor phylogeny estimation problem from bulk DNA sequencing data exhibits nonuniqueness of solutions, with single-sample DNA data always supporting a linear phylogeny. These challenges have prevented the inference of clone-specific expression profiles.

## METHODS & CODES

This project for the first time used matched RNA and DNA bulk sequencing samples from the same biopsy to jointly infer a phylogeny and to estimate clone-specific expression profiles. To do

so, the research team leveraged a previously described pan-cancer analysis of the *cis*- and *trans*-effects of somatic mutations in The Cancer Genome Atlas (TCGA) [1].

The team formulated an optimization problem to jointly deconvolve and estimate phylogenies from matched RNA and DNA tumor samples. They solved this problem using mixed-integer linear programming. More specifically, the researchers’ method used the RNA data to assign a likelihood to each phylogeny inferred from DNA data, prioritizing solutions that are most consistent with both types of data.

The method is available at <https://github.com/elkebir-group/PBDRJB>.

## RESULTS & IMPACT

Simulation experiments showed that this method is capable of prioritizing the true underlying phylogeny with high accuracy. The method can also accurately deconvolve the *cis*- and *trans*-effects of gene regulation at the clonal level. The research team analyzed matched single-sample DNA and RNA breast cancer data from TCGA and found that the method is able to differentiate between linear and branching phylogenies.

## WHY BLUE WATERS

Blue Waters was essential to the research outcome of this project because it involved extensive benchmarking and validation using simulated data. The computational resources of Blue Waters allowed the research team to perform these experiments at scale, enabling the study of the performance of the algorithms and the underlying problem statements in many different experimental settings. This is something that would not have been possible on other platforms.

## PUBLICATIONS & DATA SETS

Y. Luo, J. Peng, and M. El–Kebir, “Phylogenetic inference of clone-specific expression and mutation profiles from matched RNA and DNA tumor samples,” submitted, 2019.



# AMPHOTERICIN-DRIVEN STEROL EXTRACTION: PROBING THE MECHANISM

**Allocation:** Illinois/200 Knh  
**PI:** Taras V. Pogorelov<sup>1</sup>  
**Co-PIs:** Chad Rienstra<sup>1</sup>, Martin Burke<sup>1</sup>

<sup>1</sup>University of Illinois at Urbana–Champaign

## EXECUTIVE SUMMARY

The need for effective and nontoxic antifungal drugs is ever-present. By utilizing the petascale computing capabilities of Blue Waters, the research team aims to corroborate the experimental observation of the extraction of sterols from a membrane by an extramembranous amphotericin (AmB) sponge. The detail provided by all-atom molecular dynamics simulations will shed light on the nature of the AmB–sterol interactions and characterize the mechanism of extraction. Current simulations demonstrate AmB sponge–membrane surface interactions. Full characterization of the sterol extraction mechanism will open new directions for antifungal and antimicrobial drug design.

## RESEARCH CHALLENGE

Life-threatening systemic fungal infections are on the rise, and their effect is particularly severe for immunocompromised patients [1]. AmB is a potent antifungal drug with a remarkably low incidence of resistance development [2]. Unfortunately, AmB is toxic not only to fungal cells but also to human cells through cholesterol extraction [3], which limits the drug’s use. Understanding the method of sterol extraction is the first step in the modification of AmB such that it becomes less toxic to humans.

## METHODS & CODES

Molecular dynamics (MD) is a versatile and powerful technique that allows for the study of large membrane systems at atomistic detail. The research team’s simulations employed NAMD [3], a publicly available and highly scalable MD program that has demonstrated scalability on Blue Waters [4,5].

In preparation for simulations on Blue Waters, molecular systems were constructed using the highly mobile membrane mimetic model (HMMM), which reduces lipid tail lengths and fills the resulting inner leaflet space with a hydrophobic solvent [6,7]. The benefit of using this type of system in the preparation stage is that it accelerates the dynamics of the lipid headgroups by eliminating the steric bulk and hindrance caused by long lipid tails while simultaneously maintaining the atomistic detail of the headgroups. Accelerating lipid dynamics decreases the amount of simulation time necessary for sampling while maintaining the atomistic detail of the headgroups, which allows for the capture and characterization of lipid membrane interactions [8–10]. After the preparation phase, full lipid tails were grown on the lipids using the HMMM Builder [10]. The full-tail systems were then submit-

ted to replica exchange MD (REMD), which allows for the system to be simulated at different temperatures in parallel. REMD simulations are available in the NAMD software on Blue Waters.

## RESULTS & IMPACT

The researchers assembled two systems incorporating the extramembranous AmB sponge: one with a phosphatidylcholine (POPC) and cholesterol membrane and a second with a POPC and ergosterol membrane. (POPC is a lipid that is prevalent in human cells.) The team’s MD simulations of the full-tail membrane systems revealed spontaneous movement of the sponge toward the membrane and the formation of an encounter complex. The stable orientations of the sponge and membrane demonstrate the existence of favorable AmB–membrane surface interactions. The researchers expect to observe AmB–sterol-specific interactions and eventually sterol extraction from the membrane into the sponge with longer simulation timescales conducted as part of the team’s ongoing study.

The mechanism by which AmB kills fungal and human cells has not been determined. Recent studies have shed some light on that mechanism by demonstrating that the extramembranous sponge efficiently extracts ergosterol from yeast cells [11]. Using the petascale computing power of Blue Waters with MD to simulate the extraction of sterols by the AmB sponge will allow the team to more completely characterize the mechanism of sterol removal. Understanding this mechanism will enable new directions in the constant and time-demanding battle of pharmaceutical chemistry in the development of new antifungal drugs.

## WHY BLUE WATERS

To complete the proper sampling for the large membrane and extramembranous AmB sponge systems, the petascale computing power of Blue Waters is essential. These systems contain hundreds of thousands of atoms and require long simulation timescales of hundreds of connected replicas, each using hundreds of compute cores to observe AmB sponge–membrane interactions and, ultimately, sterol extraction. The goal of this project highlights the need for Blue Waters and future petascale resources for researchers to advance the development not only of less toxic antifungal medications but also the entire process of drug design.



# SOCIAL SCIENCE, ECONOMICS, & HUMANITIES

---

ECONOMICS

HUMANITIES

320 *Climate Policy in a Dynamic Stochastic Economy*

322 *Characterizing Descriptivity in Writing through  
Text Analysis of Books from the HathiTrust  
Digital Library*

324 *High-Frequency Trading in Nanoseconds: Analysis,  
Modeling, and Policy Implications*



CLIMATE POLICY IN A DYNAMIC STOCHASTIC ECONOMY

Allocation: GLCPC/130 Knh  
PI: Yongyang Cai<sup>1</sup>  
Co-PIs: Kenneth Judd<sup>2</sup>, William Brock<sup>3</sup>, Thomas Hertel<sup>4</sup>

<sup>1</sup>The Ohio State University  
<sup>2</sup>Hoover Institution  
<sup>3</sup>University of Wisconsin–Madison  
<sup>4</sup>Purdue University

EXECUTIVE SUMMARY

There are significant uncertainties in both the climate and economic systems. Integrated Assessment Models (IAMs) of the climate and economy aim to analyze the impact and efficacies of policy responses to climate change. The research team developed and solved new computational IAMs, with more than 10-dimensional continuous state space, that incorporate a spatial temperature system, climate tipping points, economic risks, carbon capture and storage, and/or regional economic activities. The team then analyzed the optimal policy under uncertainty and risks and how such a policy impacts economic activities. They found that tipping points and sea level rise significantly increase the social cost of carbon (SCC), but efficient adaptation and carbon capture and storage can significantly decrease SCC, while ignoring spatial heat transfer leads to nonnegligible bias. Moreover, they have solved dynamic stochastic cooperative and noncooperative equilibria and find that noncooperation leads to much lower carbon taxes and then to a much higher temperature in the future.

RESEARCH CHALLENGE

A major characteristic of leading IAMs is that their geophysical sector determines the mean surface temperature, which in turn determines the damage function, and then damages are related to the mean surface temperature of the planet. However, climate science shows that under global warming, the temperature at the latitudes closer to the poles will increase faster than at latitudes nearer to the Equator. This effect is called polar amplification (PA). Moreover, most countries in the tropic area are poorer and more vulnerable to climate change than rich countries in the higher-latitude regions. Furthermore, PA will accelerate the loss of Arctic sea ice, leading to a potential meltdown of the Greenland and West Antarctica ice sheets, which then could cause serious irreversible sea level rise. All of these factors call for a more realistic and regionalized IAM.

Most existing IAMs are perfect foresight forward-looking models, assuming one knows all of the future information. However, there are significant uncertainties in the climate and economic systems. For example, PA will increase the likelihood of tipping points that may significantly change the Earth system and economic productivity. On the other hand, technological progress such as carbon capture and storage as well as more efficient adaptation may reduce potential climate damage significantly. International

cooperation or noncooperation will lead to significantly different solutions. All of these uncertainties call for a richer and more dynamic stochastic IAM, which is computationally challenging.

METHODS & CODES

Cai, Brock, Xepapadeas, and Judd [1] developed a model of dynamic integration of regional economy and spatial climate under uncertainty (DIRESCU) that includes spatial heat and moisture transport from low latitudes to high latitudes, sea level rise, permafrost thaw, and tipping points. To model spatial heat and moisture transport, they disaggregated the globe into two regions: region 1 is the region north of latitude 30°N to 90°N (called the North), while region 2 is the region from latitude 90°S (the South Pole) to 30°N (called the Tropic-South). The research team adapted the computational method in DSICE [2], developed by Cai and Judd in past years using Blue Waters, to solve DIRESCU in a cooperative world. They also studied the regional climate policy under noncooperation between the North and the Tropic-South by developing an iterative method to find feedback Nash equilibrium in DIRESCU.

Cai and Judd [4] extended DSICE to study the impact of carbon capture and storage on climate policy in the face of economic risks and also with a climate target constraint: If the global average surface temperature increase is above 2° or 1.5° Celsius, then the Earth likely will incur significantly larger damages. The researchers adapted the computational method in DSICE again to solve this constrained model.

These computational methods are parallelized using the master–worker structure—the master assigns N tasks for workers to solve in parallel and then gathers the results of these tasks from workers. In principle, they follow the parallel dynamic programming method developed by Cai, Judd, and their co-authors [3]. The code shows high parallel efficiency and an almost linear speedup from 30 to 5,000 nodes.

RESULTS & IMPACT

In the past year, the DSICE paper [2] was accepted by the *Journal of Political Economy* and has been cited more than 190 times according to Google Scholar. The paper is the first in the literature to solve such a high-dimensional dynamic stochastic IAM in the face of both economic and climate risks. Its largest example would take several decades to run on a single-core machine, but

the research team solved it in less than eight hours using 3,459 nodes on Blue Waters. The paper shows that the social cost of carbon (SCC)—the present value of the marginal damage to economic output caused by carbon emissions—is substantially affected by both economic and climate risks. Moreover, SCC is itself a stochastic process with significant variation.

DIRESCU [1] and Cai and Judd [4] are two research projects continuing from last year’s Blue Waters project. This year, the team made significant revisions. In DIRESCU, they developed a computational method to solve the feedback Nash equilibrium, which is well known to be computationally challenging, particularly for high-dimensional dynamic stochastic games. The researchers studied many cases in DIRESCU [1] and Cai and Judd [4], where each case used thousands of node hours of Blue Waters.

WHY BLUE WATERS

The research team’s parallel computational package requires low-latency communications because it uses the master–worker structure and needs frequent communication between the master and workers. The problems are large. For example, DIRESCU has 10 continuous state variables and one binary state variable as well as eight continuous decision variables and a horizon of over 500 years. It corresponds to solving a Hamilton–Jacobi–Bellman (HJB) partial differential equation with 10 or 11 state variables in a cooperative world, or a system of two HJB equations in a non-cooperative world. Blue Waters allows the researchers to solve these large problems efficiently.

PUBLICATIONS & DATA SETS

Y. Cai, K. L. Judd, and T. S. Lontzek, “The social cost of carbon with economic and climate risks,” *J. Political Econ.*, 2019, doi: 10.1086/701890.



# CHARACTERIZING DESCRIPTIVITY IN WRITING THROUGH TEXT ANALYSIS OF BOOKS FROM THE HATHITRUST DIGITAL LIBRARY

**Allocation:** Exploratory/50 Knh

**PI:** J. Stephen Downie<sup>1</sup>

**Co-PIs:** Sayan Bhattacharyya<sup>2</sup>, José Eduardo González<sup>1</sup>

**Collaborators:** Boris Capitanu<sup>1</sup>, Craig Willis<sup>1</sup>, Peter Organisciak<sup>3</sup>

<sup>1</sup>University of Illinois at Urbana–Champaign<sup>2</sup>Singapore University of Technology and Design<sup>3</sup>University of Denver

## EXECUTIVE SUMMARY

This project approaches quantifying the notion of descriptivity in text. The immediate objective was to explore descriptivity in forms of writing that have been characterized as exemplifying different writing styles and genres. Digital text analysis offers an opportunity to operationalize the anecdotal notion of descriptivity by developing quantified metrics for descriptivity. This work leveraged the resource represented by the HathiTrust Digital Library repository. The requested allocation was used to create an updated data set of preprocessed, extracted features from the HathiTrust corpus with exploratory methods to support the research. These extracted “features” were quantifiable facts about the pages of the books, most usefully counts of words (unigrams) or strings of words (bigrams).

## RESEARCH CHALLENGE

There is a belief that writing has, since World War II, taken an overall turn away from "tell" toward "show"; this suggests that writers are becoming increasingly more interested in description [1]. However, estimating descriptivity is difficult. This is useful not only for literary and historical studies (the research team's immediate focus of interest) but also for researchers from beyond these areas. Characterizing books or pages in books by such a metric in the form of user-generated metadata is helpful for the use of a text corpus resource because users of the resource may want to find more descriptive or less descriptive books, depending on their needs, especially since treating text as data opens up new experiences of reading [2].

The notion of description has been the topic of a longstanding discussion in literary and historical studies. In his influential 1936 essay “Narrate or Describe?” György Lukács distinguished, in the case of fiction, between a dynamic concept of “narration,” in which verbal description of material objects is intertwined with the progression of the protagonists’ character development through action, and mere “description,” in which description is static and isolated from the flow of action [3]. Such a distinction becomes easy to operationalize in terms of the research team’s metric. Theorists from literary studies and historiography, such as Hayden White, have, likewise, associated description without narrativity with lack of meaning [4]—a notion that could similarly be operationalized to engage with ongoing debates in the humanities [5,6] and in cultural analytics [7].

In particular, bigrams are useful in several ways. In certain circumstances (such as when queries can be explicitly identified as good candidates for bigram use), greater improvements in information retrieval tasks are obtained from using bigrams as queries rather than other queries [8]. The inclusion of bigrams provides consistent gains in sentiment analysis tasks [9].

## METHODS & CODES

The research team developed code to compute a simple proxy metric for "descriptiveness" with part-of-speech-tagged unigrams (tagged with Penn Treebank part of speech categories). The script counts the total count of two "per-page co-occurring" parts of speech for any volume. When using HathiTrust Research Center unigrams, "per-page co-occurring" simply means  $\min(x, y)$ ,

Figure 1: “Data” as bags of words (developer’s view). The content of digitized text made available after crunching in the form of unigrams or bigrams is accessible to a developer as data. Each page is a separate bag of words, available for processing.

```
"pages":[
  {
    "seq":"00000007","tokenCount":26,"lineCount":17,"emptyLineCount":
    "sentenceCount":4,"languages":[{"de":{"1.00}}],"header":{"tokenCount":0,"lin
    unt":0,"emptyLineCount":0,"sentenceCount":0,"tokenPosCount":{}},body":{"to
    unt":26,"lineCount":17,"emptyLineCount":0,"sentenceCount":4,"tokenPosCount
    ILLUSTRATIONS":{"NE":1},"WITH":{"NE":1},"BROWNE":{"NE":1},"{"$.":4},"11
    CARD":1},"LONDON":{"NE":1},"K.":{"NE":1},"CHAELES":{"NE":1},"LITTLE":{"NE
    EVANS":{"NE":1},"{"$.":2},"DOORITT":{"NE":1},"1857":{"CARD":1},"H.":{"NE
    STREET":{"NE":1},"DICKENS":{"NE":1},"BOUVERIE":{"NE":1},"BY":{"NE":2},"
    {"$.":1},"BRADBURY":{"NE":1},"AND":{"NE":1}}},"footer":{"tokenCount":0,"lin
    unt":0,"emptyLineCount":0,"sentenceCount":0,"tokenPosCount":{}}},
    {
      "seq":"00000011","tokenCount":217,"lineCount":19,"emptyLineCount":
      "sentenceCount":6,"languages":[{"en":{"1.00}}],"header":{"tokenCount":0,"l
      ount":0,"emptyLineCount":0,"sentenceCount":0,"tokenPosCount":{}},body":{"
      nCount":217,"lineCount":19,"emptyLineCount":0,"sentenceCount":6,"tokenPosC
      t":{"read":{"WBW":1},"continuous":{"JJ":1},"for":{"IN":1},"Merdle":{"NP":1}
```

Figure 2: How to conceptually visualize “descriptiveness.” Descriptiveness scores of the kind the research team explores, in different genres (such as fiction, drama, etc.) and across “narrative time” (the relative position within the work, measured in percentile, of a page whose descriptiveness is being considered) can be visualized to get an intuitive comparative sense of how descriptiveness tends to wax and wane in collections of works drawn from different genres.

where  $x$  is the number of occurrences on a page of the part-of-speech-tag  $X$ , and  $y$  is the number of occurrences on a page of the part-of-speech-tag  $Y$ . This estimation using unigrams, however, typically produces an overcount, as the count generated is not restricted only to contiguous tokens among the equal number of paired tokens but also includes tokens from within the page that are not contiguously paired; this can be rectified by the use of bigrams.

## RESULTS & IMPACT

The useful impact of this work is the demonstration that as more computational power is becoming available, more complex and finer-grained analysis (such as that using bigrams) can be undertaken to approach the problem at progressively deeper levels. The research team has used this continuing work in teaching an undergraduate independent/directed study offered through the English Department of the University of Pennsylvania in 2018 and in a digital humanities course in Singapore in September 2019.

## WHY BLUE WATERS

The sheer size of the data set and the embarrassingly parallel computational nature of generating the needed features made Blue Waters an ideal environment for conducting this exploratory analysis. Using advanced features such as bigrams allowed for a better estimation of co-occurrence. This has allowed the team to use adjectives and nouns as well as adverbs and verbs in the form of actual pairs in the per-page-co-occurring parts of speech used in computing the metric to measure descriptive-

ness. For extracting these advanced features, the team created 100 partitions of approximately equal size from the 5.4-million input volumes of text and ran 100 jobs, with each job using 100 nodes and each node processing the text from twenty books (volumes) simultaneously. The average number of volumes per partition was 54,200, and the total number of nodes used was 10,000 ( $100 \times 100$ ). The total number of cores used was 200,000 ( $100 \times 2,000$ ), and the average time taken to process one partition was one hour, 45 minutes. The output produced was 6.25 terabytes (TB), consisting of 55 gigabytes of entities, 2.11 TB of bigrams, and 4.08 TB of trigrams.

## PRESENTATIONS & DATA SETS

S. Bhattacharyya, "Textual digital humanities for critique, with Lukács," presented at Lukács and the World: Rethinking Global Circuits of Cultural Production, Santa Barbara, CA, U.S.A., Apr. 20–21, 2018.



# HIGH-FREQUENCY TRADING IN NANOSECONDS: ANALYSIS, MODELING, AND POLICY IMPLICATIONS

**Allocation:** Illinois/50 Knh  
**PI:** Mao Ye<sup>1</sup>

<sup>1</sup>University of Illinois at Urbana–Champaign

## EXECUTIVE SUMMARY

This interdisciplinary project involves economics and computer science and examines how to analyze and understand financial trading and its effects on the stock market. It fosters research on the financial ecosystem by hosting a series of collaborative workshops with financial economists and data scientists. The goal is to create new metrics and data for the discipline of finance in economics and inform public policy in the era of big data.

## RESEARCH CHALLENGE

The advent of big data has reshaped not only the methodological challenges and opportunities facing financial economics but also the phenomena studied in the field. However, traditional academic research in finance emphasizes economic explanations to such an extent that, in the words of the prominent economist Susan Athey, “If you said that an economist was data mining, that was an insult.” Consequently, big data techniques have as yet gained little traction in academic research.

Existing studies in financial economics focus mostly on the economic behavior of humans. Advances in computing technology and machine-learning techniques have introduced cyber players, which use computer algorithms to make trading decisions. Existing U.S. Security and Exchange Commission (SEC) regulations, however, are designed for human traders. This project lays the groundwork for public policy for cyber traders in the era of big data. The research aims to understand the origin and impact of these cyber players in financial markets. The focus is on high-frequency traders (HFTs), which are cyber players that operate at exceptional speed, measurable in nanoseconds.

## METHODS & CODES

The NASDAQ exchange provided more than 15 TB of raw data consisting of 10 years of records consisting of all the orders submitted, executed, and canceled in the NASDAQ market during that time period. The team combined all of these messages and generated snapshots of the market at the resolution of nanoseconds using the computational power of Blue Waters. The financial data could easily be sliced to the granularity of stock-days, and the biggest stock-day (May 6, 2010, or the so-called flash crash day) is only a matter of hundreds of MB.

To balance the CPU loading, the researchers distributed larger jobs to CPUs first and then filled in smaller jobs when larger jobs were completed. Aggregation of outputs was easy and could be done with a single node.

## RESULTS & IMPACT

The research team discovered an important driver of high-frequency trading: discrete prices. The SEC’s Rule 612 imposes a minimum price variation for stock trading called a tick size. The prevailing one-cent tick size constrains price competition. The first-come, first-served rule at the same price then generates queuing, or early arrival to the market, to beat rivals.

Yao and Ye [1] found that HFTs are more active in securities with lower prices because a one-cent uniform tick size implies a larger relative tick size. Further, they showed that the current policy initiative to reduce HFTs by increasing the tick size would only encourage HFTs. Li, Wang, and Ye [2] incorporated a discrete tick size and allowed non-high-frequency traders (non-HFTs) to supply liquidity.

The findings show discrete pricing leads to higher transaction costs for non-HFTs. The Li, Ye, and Zheng [3] and Ye, Zheng, and Zhu [4] studies investigate the real effects of the secondary market. The SEC implemented a tick size pilot program from 2016 to 2018 to increase tick size from 1 cent to 5 cents for 1,200 randomly selected stocks. The Li, Ye, and Zheng [3] findings show that the increase in tick size reduces stock market liquidity and that firms that face tick-size constraints reduce their share repurchase by 45%. Further, the Ye, Zheng, and Zhu [4] findings show that after the tick size pilot, treated firms’ investment-*q* sensitivity increases, resulting from an increase in stock price informativeness and managers incorporating the information from the stock price into investment decisions.

This project has stimulated collaborations among financial economists, computer scientists, and experts on high-performance computing (HPC). The PI has organized three conferences in collaboration with the National Bureau of Economic Research, the National Center for Supercomputing Applications (NCSA), and top finance journals to jump-start big data research in finance. NCSA Director William Gropp and several experts on supercomputing from XSEDE (<https://www.xsede.org>) gave lectures during the conferences. SEC Chief Economist S. P. Kothari spoke about policy challenges and research opportunities in the era of big data (<https://bit.ly/31iKKK1>). The co-editor of *The Journal of Financial Economics*, Toni Whited, gave a speech about HPC for structural estimation. The PI spoke about big data in finance. A website has been created for the collaborative workshops, including videos of the two keynotes (<https://bit.ly/2MhL0gj>). The PI will organize five more collaborative conferences in the future.

## WHY BLUE WATERS

The research team is the first academic research group to use the power of supercomputing to analyze the financial market. Because Blue Waters and NCSA are located in Champaign–Urbana, it was easier for the researchers to access staff support. The research team also benefitted from attending the Blue Waters Symposium (<https://bluewaters.ncsa.illinois.edu/blue-waters-symposium-2019>). In addition, the file transfer is faster to/from Blue Waters than any remote sites, which is a unique advantage of using Blue Waters.



# MATHEMATICAL & STATISTICAL SCIENCES

---

## STATISTICS

**328** *A Massively Parallel Evolutionary Markov Chain Monte Carlo Algorithm for Sampling Spatial State Spaces*



# A MASSIVELY PARALLEL EVOLUTIONARY MARKOV CHAIN MONTE CARLO ALGORITHM FOR SAMPLING SPATIAL STATE SPACES

**Allocation:** Illinois/100 Knh  
**PI:** Wendy K. Tam Cho<sup>1</sup>  
**Co-PI:** Yan Liu<sup>2</sup>  
**Collaborator:** Simon Rubinstein-Salzedo<sup>3</sup>

<sup>1</sup>University of Illinois at Urbana-Champaign  
<sup>2</sup>National Center for Supercomputing Applications  
<sup>3</sup>Euler Circle

## EXECUTIVE SUMMARY

The research team has developed an Evolutionary Markov chain Monte Carlo algorithm for sampling from large, idiosyncratic, and multimodal state spaces. This algorithm combines the advantages of evolutionary algorithms (EA) as optimization heuristics for state space traversal and the theoretical convergence properties of Markov chain Monte Carlo algorithms for sampling from unknown distributions. The team encompassed these two algorithms within the framework of a multiple-try Metropolis Markov chain with a generalized Metropolis-Hastings ratio. Further, the group harnessed the computational power of massively parallel architecture by integrating a parallel EA framework that guides Markov chains running in parallel. Because the algorithm is a sampling algorithm, it is applicable to any field that samples, which is essentially all fields of science.

## RESEARCH CHALLENGE

The challenge is to create a method that randomly samples from unstructured spatial data with stringent spatial constraints.

## METHODS & CODES

The researchers' approach is two-pronged. The first prong involves developing an optimization algorithm that explores the solution landscape in a spatially aware manner. The second prong involves integrating the optimization algorithm into the sampling framework of a Markov chain Monte Carlo (MCMC) technique. The algorithm is implemented in ANSI C and can be compiled on Linux and OS X as a standard *makefile* project. It uses MPI non-blocking functions for asynchronous migration for load balancing and efficiency. In addition, the C SPRNG 2.0 library provides a unique random number sequence for each MPI process, which is necessary for running a large number of parallel MCMC chains.

## RESULTS & IMPACT

MCMC methods are used to sample from unknown distributions. While the theory ensures sampling from unknown distributions, this theoretical result is asymptotic (approaching a value or curve arbitrarily closely). For large applications, the time required before the theoretical convergence is realized may be prohibitively long. Hence, while MCMC methods are theoretic-

Figure 1: Performance of the spatial path relinking crossover operator. This operator was designed in the context of an evolutionary algorithm crossover. It has been adapted via a multiple-try framework for a Markov chain Monte Carlo algorithm. The figure shows the effectiveness of the operator in identifying increasingly optimal states.

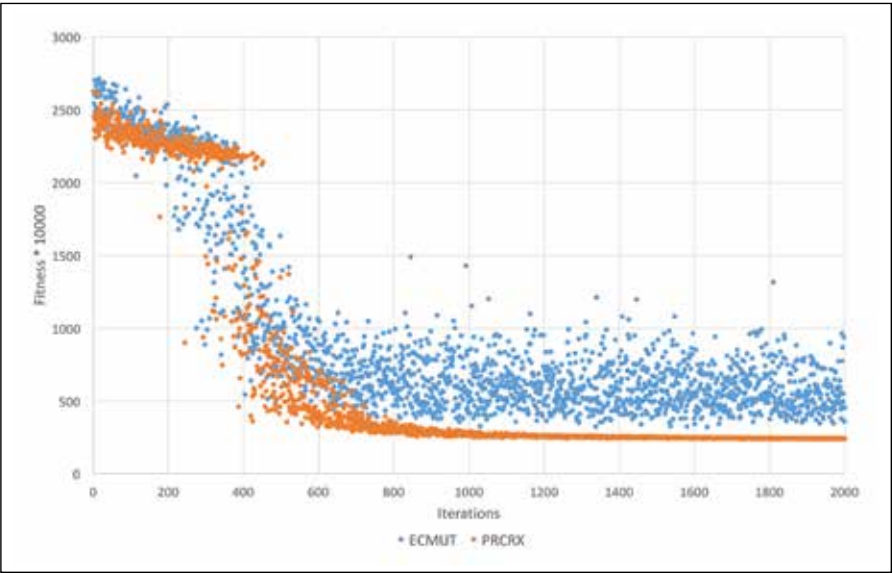


Figure 2: The algorithm produced over three million electoral maps for the State of Ohio by partitioning voter tabulation units into 16 districts. These maps were produced for the partisan gerrymandering case, APRI v. Householder. A three-judge panel found the evidence compelling and ordered a redrawing of the Ohio congressional map [1].

cally attractive, successful implementation for complex applications can be quite challenging.

A common MCMC strategy is to define a Markov transition function that amounts to a small or local random change in the current state. Small local changes are attractive for two reasons. First, they are conceptually and operationally simple. Second, the Metropolis-Hastings ratio needs to lead regularly to accepted transition proposals. Since a small change likely results in a large Metropolis-Hastings ratio, the movement of the Markov chain is then fairly fluid. At the same time, because these are small movements in a very large state space, the resulting Markov chain converges slowly and is, moreover, likely to become trapped in localized regions. Hence, for large or complex applications, it is not likely to converge rapidly enough to be practically useful.

To improve performance and hasten convergence, one strategy is to define larger steps for the Markov chain. If designed well, larger moves have the ability to more efficiently and effectively traverse a large state space, leading to faster convergence of the chain. At the same time, devising large and effective movements intelligently is not straightforward. "Large" movements often result in small Metropolis-Hastings ratios, which lead to rejected proposals, and thus to a nonfluid and ineffective Markov chain.

To devise a chain that is able to traverse the state space in both an effective and efficient manner, the team integrated movements from optimization heuristics. A central task in marrying these two techniques is to fit the mechanics of the optimization search

within the theoretical framework that enables sampling in an MCMC algorithm. One cannot use optimization operators directly, but one can adapt the operators to an MCMC framework to provide the proposal set for directional sampling according to the structure of the multiple-try Metropolis Markov chain model.

In short, the project has resulted in an evolutionary Markov chain Monte Carlo that uses evolutionary algorithm operators to guide a large number of parallel Markov chains. Statistical evidence generated by this algorithm was used in a lawsuit filed in federal court in Ohio that resulted in a three-judge panel finding that the current congressional map in that state was a partisan gerrymander [1].

## WHY BLUE WATERS:

These methods scale to all of the processor cores on Blue Waters through nonblocking MPI communication calls. The computational approach the team implemented in their solution requires generating a very large number of solutions that comprise a representative sample. Generating a large number of statistically independent draws is only feasible on a leadership-class supercomputer such as Blue Waters.

## PUBLICATIONS & DATA SETS

W. K. T. Cho, "Technology-enabled coin flips for judging partisan gerrymandering," *South. Calif. Law Rev. Postscript*, vol. 93, pp. 11-27, 2019.

W. K. T. Cho and S. Rubinstein-Salzedo, "Rejoinder to 'understanding our Markov chain significance test,'" *Stat. and Publ. Pol.*, vol. 6, no. 1, p. 54, 2019.

W. K. T. Cho and S. Rubinstein-Salzedo, "Understanding significance tests from a non-mixing Markov chain for partisan gerrymandering claims," *Stat. and Publ. Pol.*, vol. 6, no. 1, pp. 44-49, 2019.

W. K. T. Cho and Y. Y. Liu, "A massively parallel Evolutionary Markov Chain Monte Carlo algorithm for sampling complicated multimodal state spaces," extended abstract in *SC18: The International Conference for High Performance Computing, Networking, Storage, and Analysis*, Dallas, TX, USA, Nov. 11-16, 2018.

W. K. T. Cho, "Algorithms can foster a more democratic society," *Nature*, vol. 558, p. 487, June 28, 2018. Available: <https://www.nature.com/articles/d41586-018-05498-y>

W. K. T. Cho and Y. Y. Liu, "Sampling from complicated and unknown distributions: Monte Carlo and Markov Chain Monte Carlo methods for redistricting," *Physica A*, vol. 506, pp. 170-178, Sept. 2018.



# GRADUATE FELLOWS

---

- 332

*The Contributions of Root Systems to Drought Response in the Amazon Rainforest*
- 334

*Star Formation in Dwarf Galaxies: Using Simulations to Identify Key Observables to Test Models*
- 336

*Improving Convectively Induced Turbulence Forecast Parameters Through Bulk Numerical Simulations for Aviation Safety*
- 338

*Modeling Nonlinear Physical–Biological Interactions: Inertia and Sargassum in the North Atlantic*
- 340

*Predictions About the Invisible Gas in Galaxy Clusters*
- 342

*Computational Fluid Dynamics Investigation into Pulmonary Airflow Patterns in Monitor Lizards (Varanidae)*
- 344

*The Early Instability Scenario for Planet Formation in the Solar System*
- 346

*The Distribution of Shear Stress and Nutrients in a Tidally Energetic Estuary: The Role of Numerical Resolution and Vegetation*
- 348

*Exascale Astrophysics with Enzo–E: Development of Physics Modules for Galaxy-Scale and Cosmological Simulations*
- 350

*Magnetohydrodynamic Simulation: Galaxies*
- 352

*GPU-Accelerated Interstellar Chemistry with WIND: A General Ordinary Differential Equation Solver*
- 354

*Using Spectroscopic Data and Molecular Simulations to Estimate Heterogeneous Ensembles: How to Study Complicated, Flexible Proteins When Experimental Data Are Limited*
- 356

*Electron Density-Based Machine Learning for Accelerating Quantum Calculations*
- 358

*Extending the Longevity of Produced Water Disposal Wells: Evaluation Using Reactive Transport Simulation*
- 360

*Improved Trumpet Initial Lapse and Shift for Binary Black Hole Simulations*
- 361

*Understanding the Physical Processes Causing Intermediate-Depth Earthquakes*
- 362

*Escaping From an Ultracold Inferno: The Ultracold KRb Dimer Reaction*
- 364

*The Impacts of Hydrometeor Centrifuging on Tornado Dynamics: Improving the Realism of Tornado Simulations*
- 366

*Unraveling Functional Hole Hopping Pathways in the [Fe<sub>4</sub>S<sub>4</sub>]-Containing DNA Primase*
- 368

*The Transport and Dynamics of Wave-Driven Reef Jets Under the Influence of Rotation and Bottom Friction*
- 370

*Simulation of Bluff Body Stabilized Flames with PeleC: Adaptively Resolving Turbulence–Combustion Interactions in Real-World Engineering Problems*



# THE CONTRIBUTIONS OF ROOT SYSTEMS TO DROUGHT RESPONSE IN THE AMAZON RAINFOREST

Elizabeth Agee, University of Michigan  
2016–2017 Graduate Fellow

## EXECUTIVE SUMMARY

Rising global temperatures and changing patterns of precipitation are highlighting the vulnerability of tropical forests to heat- and drought-induced stress. As a critical global ecosystem, tropical forests sequester carbon, mediate weather patterns, and have impacts far beyond their regions. This research explores the role of tree root systems in mediating the impact of strong drought events. Model simulations of tree water uptake show that different combinations of tree traits will confer different degrees of resiliency to water limitation. Results from this work may be used to more effectively direct future measurement efforts and to improve predictions of tropical forest response.

## RESEARCH CHALLENGE

Tropical forests cover less than 7% of the earth’s surface but play a significant role in global water, energy, and carbon cycles. Rapid changes in forest structure owing to anthropogenic activities and changing climate may have significant imprints that extend far beyond the tropical regions. Current model projections of forest response are beset by uncertainties associated with tree water uptake under water limitation. Improvements in the representation of tree response to water limitation will help refine projections and inform conservation and policy actions.

## METHODS & CODES

The computational complexity of single-plant models has previously limited incorporation of root systems’ hydrological models at the forest plot or ecosystem scale. Over the past decade, developments in micro- and macroscale hybridization schemes have opened the door for highly scalable models of three-dimensional root water uptake and soil water physics.

In this work, root architectures that represent the structural and spatial distribution of roots were modeled using the open source RootBox model [1]. Each tree system was assigned hydraulic parameterization (*e.g.*, root hydraulic conductivity, water potential thresholds) based on statistically generated water usage strategies. These strategies may range from risky, which favor carbon assimilation over hydraulic integrity, to conservative, which will limit carbon assimilation and therefore water uptake to protect hydraulic pathways from damage. Root water uptake has been coupled with the massively parallel flow and transport model, PFLOTTRAN [2], using hybridization techniques from [3].

Using these tools, the PI explored how tree roots contributed to forest drought resilience in areas of the Amazon rainforest during the 2015–2016 El Niño drought event. To tease apart the

contributions of various ecophysiological properties, ensemble modeling approaches were employed that test a multitude of risk configurations and root distributions. Each of these approaches uses spatial distributions from the field site in the Tapajós National Forest (K67) located in the eastern Amazon River Basin in Brazil and is validated with data collected from the same region.

## RESULTS & IMPACT

High stem density and functional diversity present large challenges for representing individual water uptake processes. Different scenarios of root functional diversity, largely tested through rooting depth, displayed significant differences in the onset of water limitation. Partitioning of roots into different depth classes dependent on size helped to alleviate the impacts of water limitation by allowing individuals to tap into separate soil water reserves, reducing the overall hydraulic stress of the system as monitored by root collar potential.

## WHY BLUE WATERS

Advances in computational platforms such as Blue Waters allow for increased model fidelity, presenting an opportunity to model forests at higher degrees of detail at the scale of individual trees. In data-scarce spaces, systems like Blue Waters provide the resources needed to run many simulations. The high number of simulations allows for understanding sources of uncertainty and targeting future measurement campaigns, saving resources and increasing the impact of difficult fieldwork.

Elizabeth Agee completed a Ph.D. in environmental engineering from the University of Michigan in 2019, having worked under the direction of Valeriy Ivanov. Agee currently is a postdoctoral associate at the Oak Ridge National Laboratory.

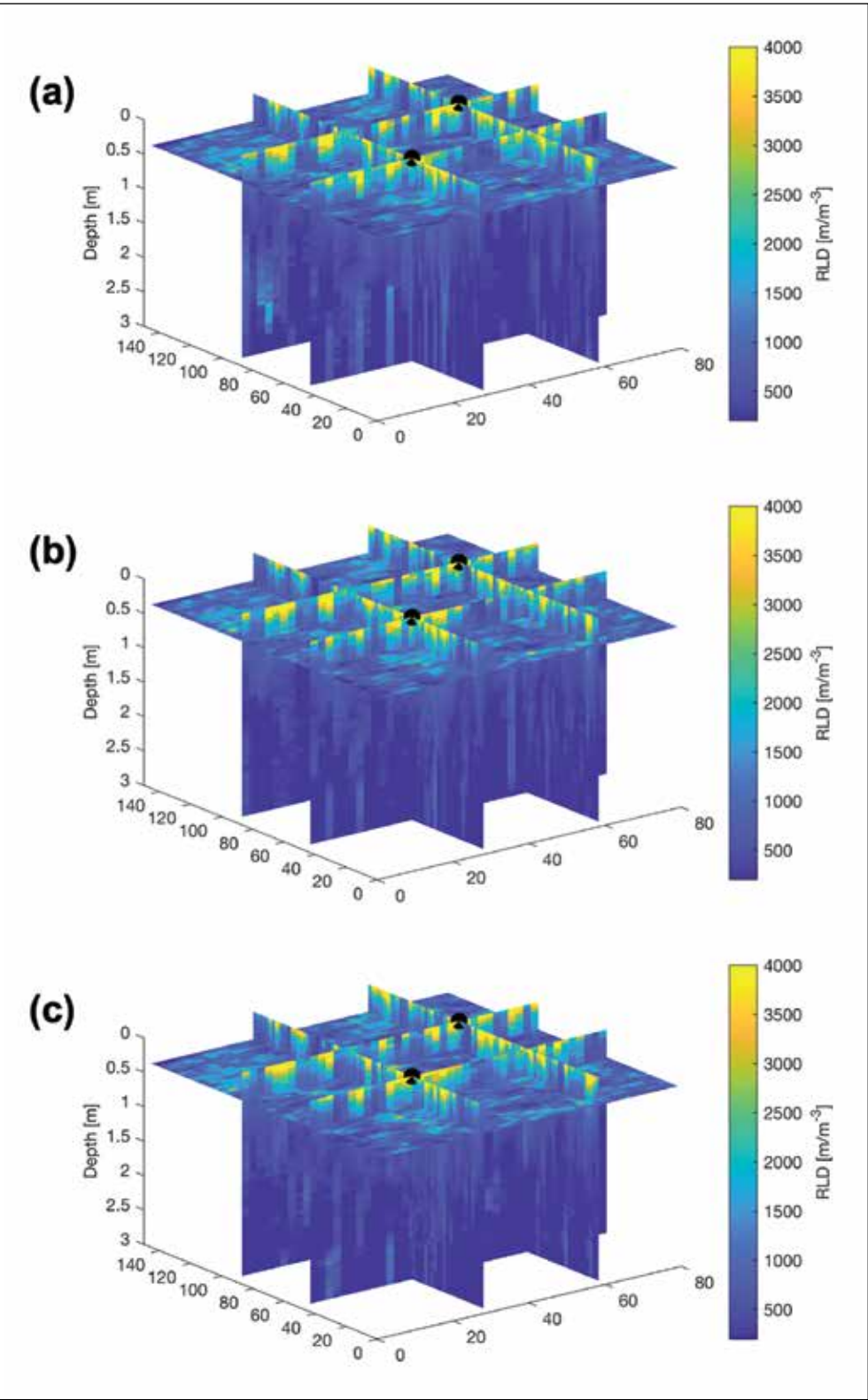


Figure 1: Three-dimensional root length density for three scenarios: (a) uniform rooting depth, (b) linear rooting depth, and (c) effective rooting depth.



STAR FORMATION IN DWARF GALAXIES: USING SIMULATIONS TO IDENTIFY KEY OBSERVABLES TO TEST MODELS

Elaad Applebaum, Rutgers, The State University of New Jersey  
2018–2019 Graduate Fellow

EXECUTIVE SUMMARY

Modern cosmological hydrodynamic simulations self-consistently model the evolution of segments of the universe under the influence of gravity and hydrodynamics. To model how gas turns into stars, these codes use analytic prescriptions taking place below the resolution of the simulation. Though implementation details differ, in large galaxies the results tend to converge. These prescriptions, however, have not been tested in the low-mass limit. In new simulations at cutting-edge resolution capable of resolving the faintest known galaxies, the PI tested how different star formation recipes affected the resulting galaxy distributions. These tests included running different models on Blue Waters and then comparing galaxies across simulations. The research found that the robustness of the models depends on the galaxy environment, with results being model-dependent in isolated regions but converged in dense environments. These results, paired with observations, can help constrain the underlying small-scale physics of how gas forms into stars. Additionally, they highlight the need for simulators to further investigate these trends to ensure accuracy in interpreting their results in faint galaxies.

RESEARCH CHALLENGE

When running cosmological hydrodynamic simulations of galaxy formation, scientists simulate the evolution of galaxies within large volumes of space. Despite strategies to maximize the already enormous range of spatial and time scales relevant to galaxy formation, it is impossible to resolve the scales in which stars form. Including realistic star formation is crucial to modeling galaxies, however, since massive stars return energy to their surroundings and modulate future star formation. To incorporate star formation, therefore, simulators include “subresolution” recipes, which are analytic prescriptions that determine under what conditions gas turns into stars. These subresolution prescriptions have led to many successes in simulating realistic galaxies. However, many galaxy properties remain unchanged even when altering these star formation recipes, as shown in [1,2]. To probe the underlying physics, researchers must look at very small galaxies, called ultrafaint dwarf galaxies, which are expected to be sensitive to changes in cosmological models.

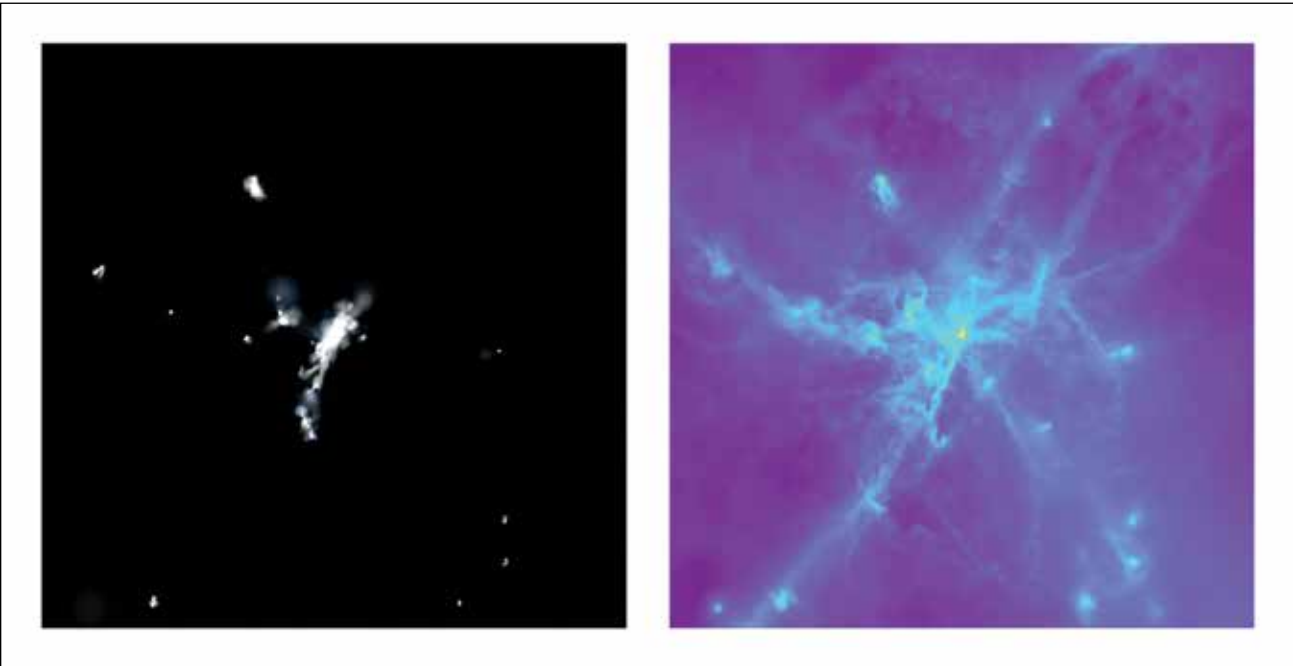


Figure 1: Mock stellar image (left) and projected gas density (right) of simulated galaxies in the very early universe. Of interest is the nontrivial relationship between local gas conditions and star formation.

METHODS & CODES

Studying such small galaxies while still capturing large volumes of space requires incredibly high resolution. This project, therefore, uses simulations at cutting-edge resolution in two different environments to compare how stars form: one is an environment such as that far from the Milky Way or any other massive galaxy and the other is an environment similar to our own Milky Way. The latter simulations are the highest resolution of their type ever run. All simulations are run with an advanced code, ChaNGa [3,4], designed to handle the unique challenges of galaxy simulations.

RESULTS & IMPACT

The results suggest that in the environment far from the Milky Way, changing the star formation model had a large impact on the resulting galaxies. In fact, changing from one model to the other led to the existence of half as many galaxies because of the greater difficulty of gas collapsing into stars, as shown in [5]. In a region such as the Milky Way, however, there is little difference because the environment is different and better able to form stars regardless of the subresolution recipe. With upcoming large surveys such as the Legacy Survey of Space and Time, researchers expect the discovery of up to hundreds of new galaxies very close to the Milky Way. It is unknown

what the properties of these galaxies will be, and theoretical work will be necessary to understand the upcoming observations. This project is a first step in constraining the uncertainty in simulation models. Furthermore, the differences among the star formation models present us with differences that are testable with future observations. Pairing the simulations with observations can therefore greatly increase the understanding of how the first star formation proceeded in the early universe.

WHY BLUE WATERS

Cosmological galaxy simulations involve enormous dynamic ranges in both space and time, and include many different computationally intensive processes relevant to galaxy formation. Additionally, modeling the faintest galaxies requires incredibly high resolution, greatly increasing the computational requirements for running these simulations. Therefore, the advanced computational capabilities of Blue Waters make it the best machine for accomplishing this work.

PUBLICATIONS & DATA SETS

F. Munshi *et al.*, “Dancing in the dark: uncertainty in ultrafaint dwarf galaxy predictions from cosmological simulations,” *Astrophys. J.*, vol. 874, no. 1, p. 40, Mar. 2019, doi: 10.3847/1538-4357/ab0085.

Elaad Applebaum, a fourth-year Ph.D. student in physics at Rutgers University, is working under the supervision of Alyson Brooks. He expects to receive his degree in 2021.



# IMPROVING CONVECTIVELY INDUCED TURBULENCE FORECAST PARAMETERS THROUGH BULK NUMERICAL SIMULATIONS FOR AVIATION SAFETY

Katelyn Barber, University of North Dakota  
2018–2019 Graduate Fellow

## EXECUTIVE SUMMARY

The number of global flight routes is projected to increase over the next five years, increasing the likelihood of airplanes being influenced by hazards associated with thunderstorms. While turbulence diagnostics are available to pilots and flight dispatchers, many of these predictors have been verified for midlatitude clear-air turbulence only. Current thunderstorm avoidance guidelines do not account for the stage of convection that influences turbulence probability.

For the first time, aviation encounters of convectively induced turbulence in both the midlatitudes and tropics were simulated at high spatial and temporal resolution to identify biases in current turbulence diagnostics and to investigate the influence of storm stage on turbulence probability. This research has found significant disagreement among the turbulence diagnostics for all cases. The probability of turbulence near developing convection was found to be greater than near mature convection, especially in tropical environments, supporting the need for storm stage and region-specific avoidance guidelines.

## RESEARCH CHALLENGE

Out-of-cloud convectively induced turbulence (CIT) is a hazard to aviation operations because it can occur vast distances away from convective sources [1] and is nearly impossible to detect using on-board radar systems. CIT is a prediction challenge owing to

the spatial (10 to 1000 meters) [2] and temporal scales (seconds to minutes) on which it occurs. Meteorological variables from forecasting systems can be used to calculate large-scale turbulence diagnostics, but generally these modeling systems are too coarse to resolve the numerous CIT generation mechanisms and turbulence propagation. While progress has been made in understanding CIT potential in the midlatitudes through field campaigns, modeling studies, and statistical examinations of pilot reports—all of which have been used to develop avoidance guidelines for aviation [3]—little is known about CIT potential in the tropics. The lack of research on tropical CIT increases the risk of turbulence encounters for tropical flight routes. As new flight routes and air traffic continue to increase, aviation will be more susceptible to convective hazards in both the midlatitudes and tropics.

High-resolution simulations of convection in the midlatitudes and tropics allow for analysis of turbulence potential for various convective regimes by capturing the majority of turbulence-generation mechanisms. These simulations also help identify the limitations of popular turbulence diagnostics and motivate the development of new diagnostics. Understanding the variation of turbulence potential with convective type, stage, and region allows for the adaptation of thunderstorm guidelines that are more specific and efficient for aviation operations, reducing aviation turbulence incidents.

## METHODS & CODES

The Weather Research and Forecasting (WRF) model [4] was used to simulate six cases of CIT on Blue Waters at 500-meter horizontal resolution and 10-minute output. Each case was a real aviation turbulence encounter that caused passenger injuries or structural damage and was associated with convection. Midlatitude turbulence cases were paired with tropical turbulence cases where the convective morphology and cause of turbulence were similar. For example, on June 29, 2018, a commercial aircraft was flying out of cloud in North Dakota in the vicinity of severe convection and encountered severe turbulence. This case was paired with a tropical case that occurred on June 20, 2017, where a commercial aircraft experienced severe turbulence while navigating out of cloud near convection [5]. Turbulence diagnostics including the Richardson number [6], eddy dissipation rate [7], and second-order structure functions [8] were computed in cloud and out of cloud ( $0.1 \text{ g kg}^{-1}$  threshold [9]). Convective stage (*i.e.*, developing and mature) was differentiated by tracking convective objects and their vertical velocities with time [5]. Environmental

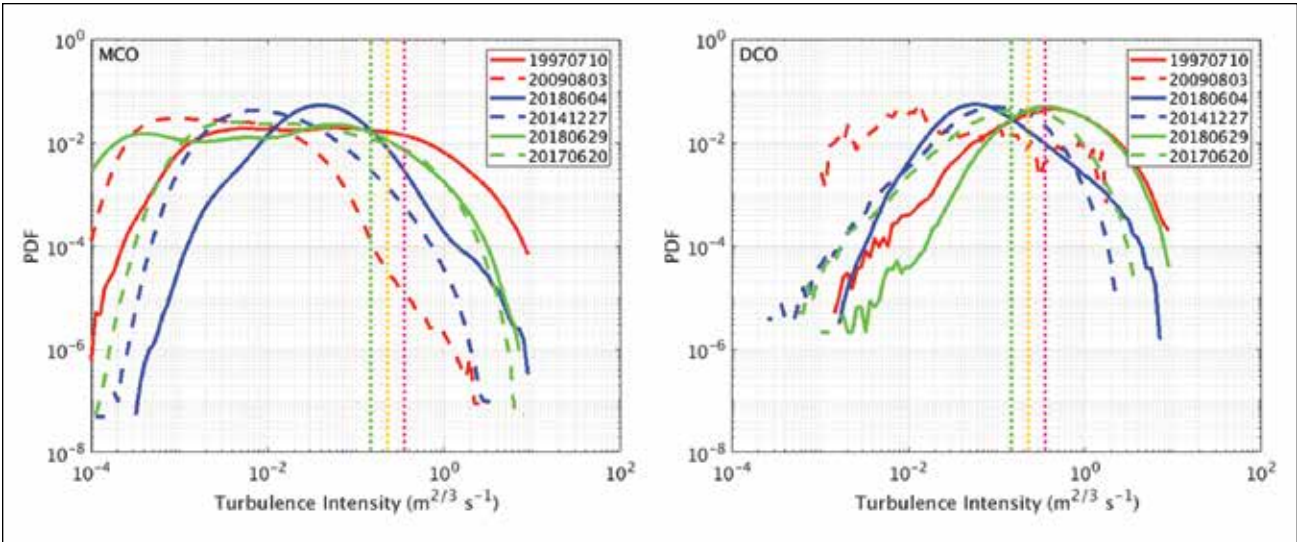


Figure 2: Distribution of out-of-cloud turbulence diagnosed by second-order structure functions between 8–12 kilometers nearest to mature convection (MCO) and developing convection (DCO) for the six simulations. The green (yellow and magenta) vertical line represents the threshold for light (moderate and severe) turbulence.

static stability and vertical wind shear were examined near developing and mature convection and related to turbulence potential.

## RESULTS & IMPACT

Six simulations of convection in the midlatitudes and tropics using high spatial and temporal resolution were performed using WRF. Simulated convective properties including morphology, strength, and location compared well against observations. The accuracy of the turbulence diagnostics for intensity and location varied drastically but had the most agreement for the midlatitude cases (Fig. 1). The eddy dissipation rate frequently underpredicted the areal coverage and intensity of out-of-cloud turbulence. Richardson number and structure functions most reproduced accurate turbulence probabilities with similar areal coverages and locations.

The examination of convective stage for the six cases illustrated the variation of turbulence probability for convective type and region (Fig. 2). The greatest probability of turbulence near mature convection occurred in the midlatitudes. Turbulence probability was found to significantly increase near developing convection in both the midlatitudes and tropics, with the greatest increases occurring in the tropics. This result highlights the increased risk for tropical aviation routes where real-time observations are limited.

The environmental static stability and vertical wind shear near developing and mature convection was found to have region-

al dependencies but not always storm-type dependencies. Static stability near tropical convection was not influenced by convective stage, but near-midlatitude convection was influenced by convective stage. Vertical wind shear was found to be influenced by region, storm type, and storm stage. Vertical wind shear increased significantly around developing convection for both regions and could be an indicator of turbulence potential. For all six cases there was a positive correlation between vertical wind shear and turbulence intensity.

This work motivates the need for more high-resolution simulations of convection to address the shortcomings of turbulence prediction and avoidance for aviation operations. More examination of turbulence diagnostics and turbulence potential in various environments is vitally needed to reduce turbulence encounters.

## WHY BLUE WATERS

Blue Waters was necessary for this project because high-resolution (spatial and temporal) simulations of convection require computational resources not available on local systems. Blue Waters allowed for thorough analysis of CIT for large domains and over long temporal periods to capture storm evolution, which required thousands of computing cores. Storage of high-resolution simulations of convection is an additional challenge that was adequately addressed through Blue Waters' resources.

A fourth-year Ph.D. student in atmospheric sciences at the University of North Dakota, Katelyn Barber is working under the direction of Gretchen Mullendore. Wiebke Deierling of the National Center for Atmospheric Research served as a collaborator on this study.

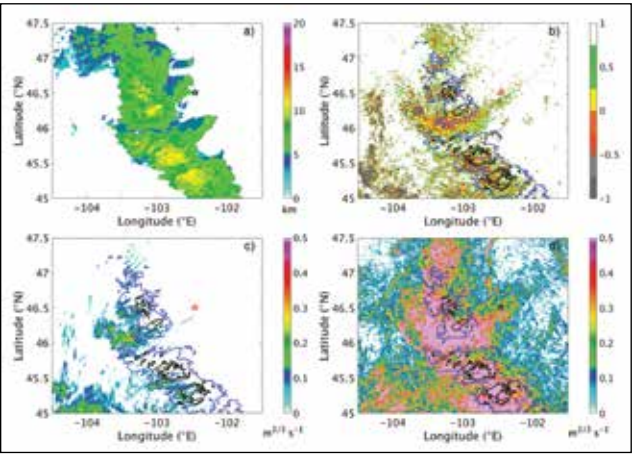


Figure 1: (a) Echo top heights, (b) Richardson number, (c) eddy dissipation rate, and (d) structure functions on June 29, 2018, between 8–12 kilometers (km). Echo tops greater than 8, 10, and 12 km are shown in b–d as blue, black, and magenta contours, and the star represents the location of the aircraft.



MODELING NONLINEAR PHYSICAL-BIOLOGICAL INTERACTIONS: INERTIA AND SARGASSUM IN THE NORTH ATLANTIC

Maureen T. Brooks, University of Maryland Center for Environmental Science  
2015–2016 Graduate Fellow

EXECUTIVE SUMMARY

The floating seaweed of the genus *Sargassum* serves as critical habitat in the open Atlantic but causes economic harm to coastal communities when it washes ashore in large aggregations. This study links *Sargassum* dispersal and growth to underlying ocean circulation features. A model framework and satellite observations were used to determine how *Sargassum* responds to inertial forces, and the implications for its basinwide distribution. The resources of Blue Waters facilitated model development and allowed implementation at high resolution over the entire *Sargassum* habitat, covering over  $4 \times 10^7$  km<sup>2</sup>. This enabled the calculation of *Sargassum*'s inertial parameters. Accounting for inertia leads to an increase in *Sargassum* export from the Sargasso Sea, providing a return pathway to the tropics. It also leads to increased retention in the Gulf of Mexico and Caribbean Sea, where the retention can cause management challenges. Including inertial effects in models of *Sargassum* could improve forecasting of coastal inundation events.

RESEARCH CHALLENGE

Floating *Sargassum* supports a diverse ecosystem in an otherwise nutrient-poor region of the ocean, supporting invertebrates, fish, and even sea turtles [1]. However, changes in *Sargassum* abundance and distribution over the past decade have resulted in millions of dollars in economic harm when it washes ashore [2]. Accurate predictions of these beaching events require an understanding of both the ocean currents that transport *Sargassum* and how much it grows along the way. Understanding the effects of inertia is particularly important because that can alter trajectories and potentially change the rate of entrainment in eddies, where growth conditions can differ from the surrounding water. Cyclonic eddies tend to propagate westward and northward in the North Atlantic, which could potentially drive more *Sargassum* to vulnerable coastal areas, while anticyclonic eddies would tend to carry *Sargassum* south toward the equator. The strength of inertial effects determines which of these two scenarios is more likely. Modeling inertial effects on *Sargassum* is difficult because it

requires estimates of density and radius, yet *Sargassum* rafts are highly nonspherical. While density can be determined directly from field samples, estimating radius requires a novel approach.

METHODS & CODES

This research uses a system of four coupled models to simulate *Sargassum* growth and transport. A Hybrid Coordinate Ocean Model (HYCOM) [3] domain was implemented at  $1/12^\circ$  ( $< 10$  m) resolution with 28 hybrid vertical layers, encompassing the known *Sargassum* distribution from  $15^\circ\text{S}$  to  $64^\circ\text{N}$  and  $100^\circ\text{W}$  to  $15^\circ\text{E}$ . Coupled to this is a biogeochemical model adapted from the work of Fennel [4], which includes nitrogen and phosphorus, phytoplankton, zooplankton, and detritus to effectively capture the dynamics of biologically mediated nutrient cycling in the upper ocean. *Sargassum* rafts are modeled using an individual-based physiology model embedded within a Lagrangian particle model. The particle model is modified from the HYCOM Lagrangian particle package to allow for *Sargassum* buoyancy, inertial effects, reproduction (particle splitting), and sampling of the underlying nutrient availability to allow for growth.

The effective radius of a modeled *Sargassum* raft was determined via an inverse method. Lines of visible *Sargassum* from satellite remote sensing [5] were compared with the finite-size Lyapunov exponent field to determine the deflection angle. A total of 91 *Sargassum* lines were measured from four dates in 2018 when there was high abundance. These angles were compared with angles calculated from model simulations with varying particle radius. An Anderson–Darling  $k$ -sample test was used to compare model and observed probability density functions and determine the *Sargassum* effective radius. This parameter was then applied to the *Sargassum* particles in the coupled model system to examine the effects of inertia on growth and distribution.

RESULTS & IMPACT

This multiscale modeling project provides the first estimates of *Sargassum* parameters for implementing inertial effects. Although the size of *Sargassum* rafts can vary from centimeters up to aggregations spanning kilometers, they respond to inertia comparably to a sphere with a radius of 0.95 m and a density of 92% of ambient sea water (Fig. 1). Accounting for these inertial properties changes how *Sargassum* moves and grows. Inertial *Sargassum* is entrained in eddies much more frequently,

with 61% entrainment compared to 12% entrainment of noninertial particles. *Sargassum* is 48% more likely to be retained in the Western Gulf of Mexico and 36% more likely to be retained in the Caribbean Sea than noninertial particles. Finally, there is a seasonal increase in export of up to 20% out of the Sargasso Sea, which helps explain how the seasonal pattern of *Sargassum* distribution can restart every year.

These inertial effects and changes in trajectories also have implications for *Sargassum* growth. Although simulations did not show significant differences in growth for *Sargassum* inside versus outside of eddies, there were differences in overall growth between inertial and noninertial *Sargassum* particles. The annual mean biomass is 8% higher when inertia is accounted for, growth rates more frequently approach their theoretical maximum, and survival time is increased. This is owing to higher transport into and retention within regions with optimal growing conditions. Accounting for these physical and biological consequences of inertia can help improve predictions of *Sargassum* beaching events and allow coastal communities to better mitigate their harmful effects.

WHY BLUE WATERS

The resources of Blue Waters have made the scale and scope of this project possible. High-resolution ocean circulation modeling alone has a high computational cost. By utilizing Blue Waters, this was accomplished along with coupling it with ocean biogeochemistry, Lagrangian particles, and individual organism physiology at temporal and spatial scales that span orders of magnitude. The NCSA staff has also been key to the success of this project. Their responsiveness and expertise was critical to implementing and running this code on Blue Waters.

PUBLICATIONS & DATA SETS

M. T. Brooks, V. J. Coles, R. R. Hood, and J. F. R. Gower, “Factors controlling the seasonal distribution of *Sargassum*,” *Mar. Ecol. Prog. Ser.*, vol. 599, pp. 1–18, Jul. 2018, doi: 10.3354/meps12646.  
M. T. Brooks, V. J. Coles, and W. C. Coles, “Inertia influences pelagic *Sargassum* advection and distribution,” *Geophys. Res. Lett.*, vol. 46, pp. 2610–2618, Mar. 2019, doi: 10.1029/2018GL081489.

Maureen T. Brooks received a Ph.D. in marine–estuarine–environmental sciences in 2019, working under the direction of Victoria Coles at the University of Maryland Center for Environmental Science Horn Point Laboratory.

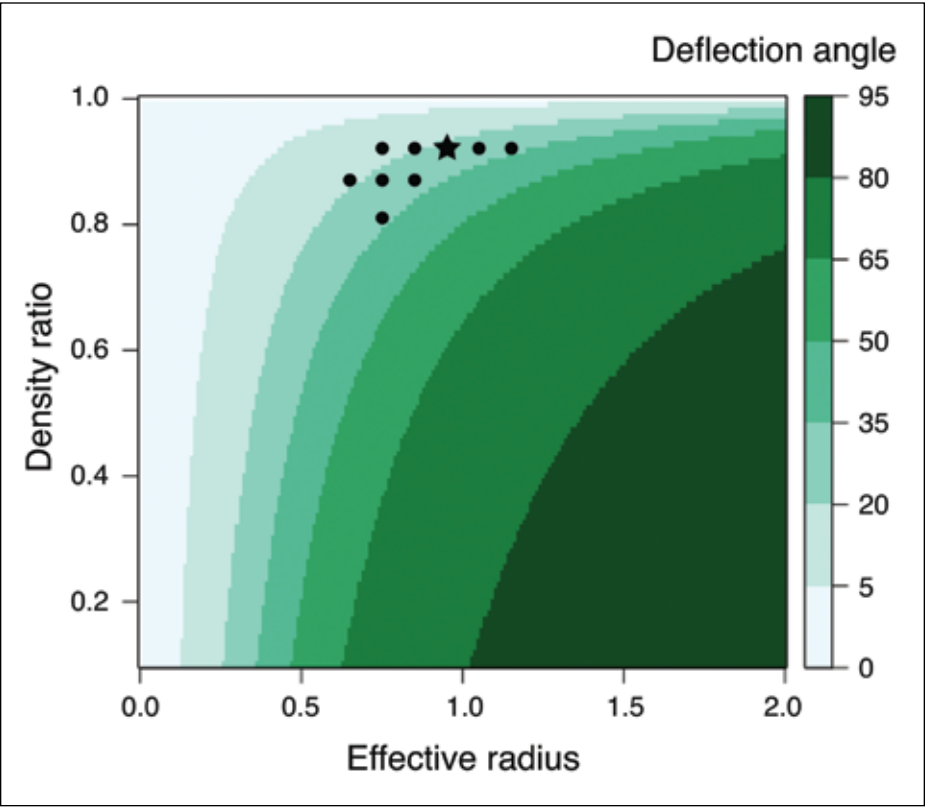


Figure 1: Deflection angle derived from the inertial equations (shading). Density ratio is relative to ambient sea water. Plotted symbols indicate where model and observed deflection angle distributions were most similar. The star indicates the effective radius and density ratio of *Sargassum* determined in this study.



DI

PREDICTIONS ABOUT THE INVISIBLE GAS IN GALAXY CLUSTERS

Iryna Butsky, University of Washington  
2016–2017 Graduate Fellow

EXECUTIVE SUMMARY

Quasar absorption-line studies in the ultraviolet (UV) can uniquely probe the nature of the multiphase cool–warm gas in and around galaxy clusters, promising to provide unprecedented insights into the interactions between infalling galaxies and the hot X-ray-emitting intracluster medium (ICM). This work resulted in a high-resolution simulation of a galaxy cluster used to study the physical properties and observable signatures of the cool ICM gas. The PI extracted synthetic spectra to demonstrate the feasibility of detecting and characterizing the thermal, kinematic, and chemical composition of this cool gas and demonstrated the feasibility of observations with the existing Cosmic Origins Spectrograph aboard the Hubble Space Telescope.

RESEARCH CHALLENGE

Galaxy clusters are the largest structures in the universe, composed of thousands of galaxies that are gravitationally bound to a massive central galaxy. Compared to isolated galaxies, cluster galaxies are more likely to have stopped forming stars—becoming “quenched.” The cluster environment plays a critical role in governing the gas cycle that fuels star formation in galaxies. There are several ways in which the cluster environment can deprive galaxies of gas, causing them to quench. For example, galaxies in a crowded cluster environment can lose gas during gal-

axy mergers or owing to tidal forces from close encounters. Additionally, the space between galaxies in a galaxy cluster is composed of a hot, dense gas known as the intracluster medium (ICM). Galaxies simply moving through the cluster experience a headwind force from the ICM that can be strong enough to remove all of the galaxy’s gas. The gas that has been stripped from galaxies enriches the ICM with heavy elements and contributes to its unique density, temperature, and chemical structure. Therefore, studying the structure of the ICM in detail places stringent constraints on the mechanisms responsible for quenching galaxies. Traditionally, cluster galaxies have been observed directly through X-ray emission. Although this method is ideal for studying the hot ( $T > 106\text{K}$ ) and dense inner ICM, it cannot detect the diffuse, cool ( $T = 104\text{--}6\text{K}$ ) gas that makes up a large fraction of the ICM at the cluster outskirts. Although the diffuse, cool ICM gas is virtually invisible to X-rays, it can be observed indirectly through UV absorption. This research uses an extremely high-resolution simulation of a galaxy cluster, RomulusC (Fig. 1; [1]) to perform a detailed characterization of the properties of the cool ICM gas and to make predictions for its observational signature using existing and future UV telescopes.

METHODS & CODES

The RomulusC cluster simulation was run on Blue Waters using the modern astrophysical hydrodynamics simulation code, ChaNGa [2]. ChaNGa is parallelized through the Charm++ parallel infrastructure framework [3] and scales up to 500,000 cores on Blue Waters.

Postprocessing of the simulation was done using the Python analysis tool, yt [4]. The PI generated realistic, instrument-specific synthetic absorption spectra from the simulation using Trident [5]. The synthetic spectra were then analyzed in an analysis pipeline developed for observers, using the Veeper software [6]. The unique combination of state-of-the-art simulations with observational analysis techniques is the key strength of this work.

RESULTS & IMPACT

Using the extremely high-resolution RomulusC simulation, the PI found that the cool ICM phase comprised a significant fraction of the total gas mass at the cluster outskirts. This cool gas (observable through UV absorption) traces a highly complementary structure to that traced by the hot ICM gas that is traditionally studied through X-ray emission. Furthermore, the results found that although the hot phase of the ICM has a uniform distribution of heavy elements, the cool phase of the ICM has a more clustered distribution of “metals” that traces gas stripping from cluster galaxies.

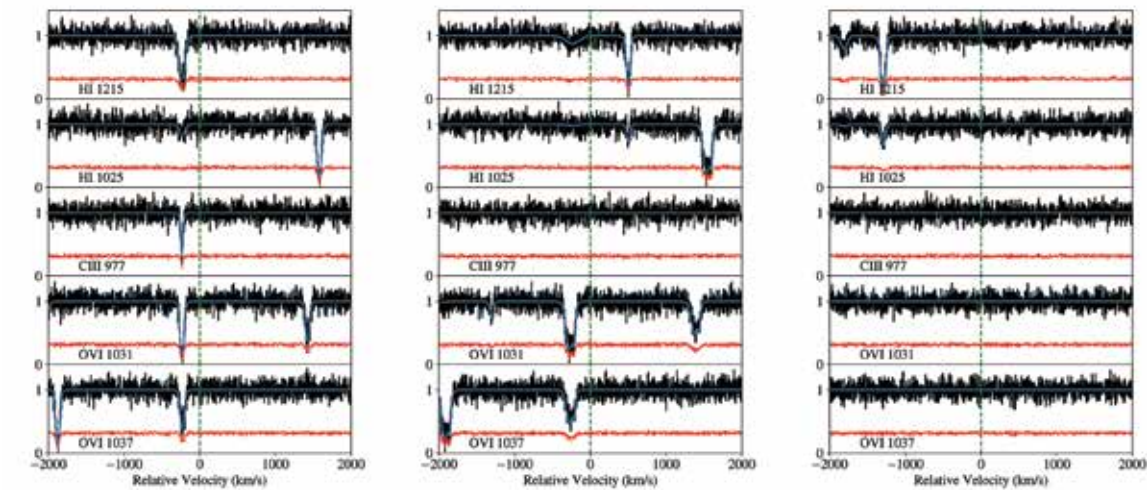
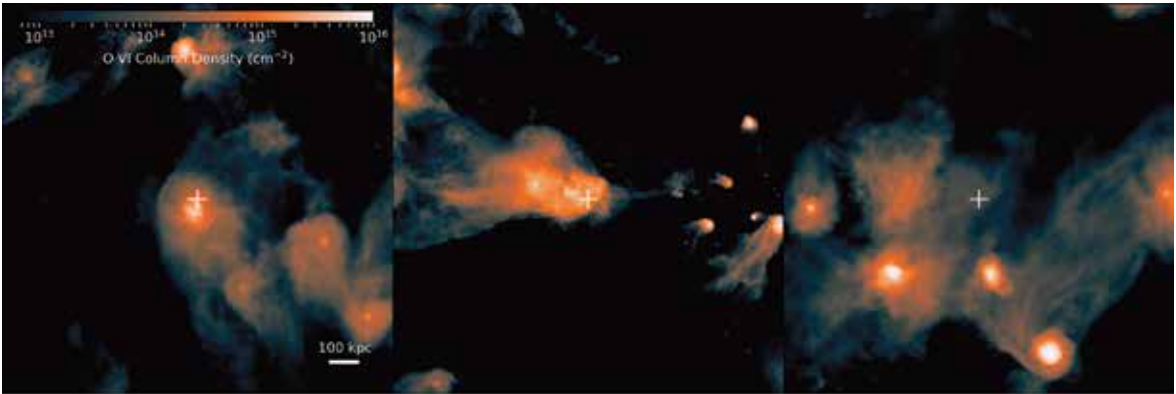


Figure 2: (top) The oxygen VI (O VI) column density (tracing cool gas abundance) at several zoomed-in regions of the RomulusC cluster, and (bottom) the corresponding predicted spectra obtainable by the Cosmic Origins Spectrograph on board the Hubble Space Telescope. Although the O VI signal is strongest around cluster galaxies, there is sufficient signal to observe the ICM between galaxies as well.

The use of synthetic spectroscopy allowed for detailed predictions for the observational signature of this cool ICM gas in clusters using the currently available Cosmic Origins Spectrograph on board the Hubble Space Telescope (Fig. 2). The results showed that random sightlines throughout a galaxy cluster should have a 40% chance of detecting cool gas (traced by hydrogen I absorption) and a 15% chance of detecting warm gas (traced by oxygen VI absorption). Future UV space telescopes such as LUVOIR will increase both the probability of detecting absorption features and the number of possible sightlines probing each galaxy cluster to revolutionize our understanding of galaxy evolution.

WHY BLUE WATERS

While galaxy clusters span several million parsecs in size, their structure is dictated by star formation and feedback processes

that happens on subparsec scales. Capturing this large span of physical scales requires detailed subgrid models and significant computational resources. Therefore, these simulations require the use of massively parallel, high-performance supercomputers such as Blue Waters.

PUBLICATIONS & DATA SETS

I. Butsky, J. Burchett, D. Nagai, M. Tremmel, T. Quinn, and J. Werk, “Ultraviolet signatures of the multiphase intracluster and circumgalactic media in the RomulusC simulation,” MNRAS, in press, 2019, doi: 10.1093/mnras/stz2859.

A fourth-year Ph.D. candidate in astronomy at the University of Washington, Iryna Butsky works under the direction of Tom Quinn and Jessica Werk, and plans to graduate in June 2021.

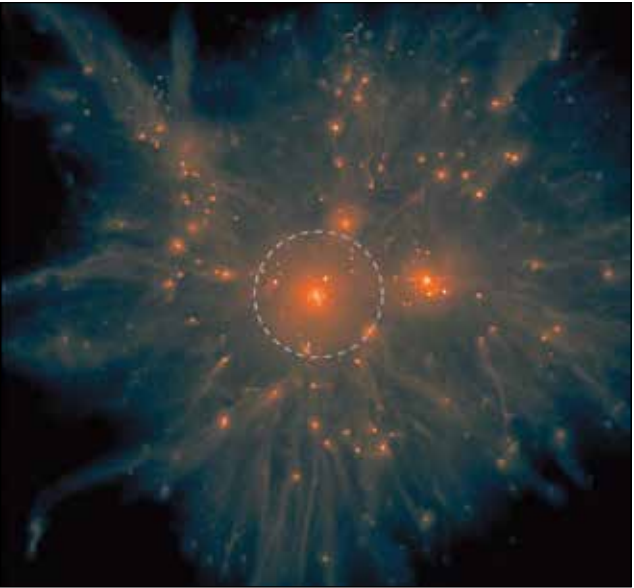


Figure 1: The projected gas density in the RomulusC simulation at a redshift of 0.3. This density map shows the distribution of galaxies within the cluster. X-ray observations are most sensitive to gas within the dashed circle. UV absorption is necessary to study the structure of the cluster outskirts.



# COMPUTATIONAL FLUID DYNAMICS INVESTIGATION INTO PULMONARY AIRFLOW PATTERNS IN MONITOR LIZARDS (VARANIDAE)

Robert Cieri, University of Utah  
2018–2019 Graduate Fellow

## EXECUTIVE SUMMARY

This research used computational fluid dynamics modeling to simulate how air flows through the lungs of monitor lizards in order to investigate why there are different types of lung airflow patterns among various types of animals (e.g., birds, mammals, reptiles). The PI found that monitor lizard lungs have a unique net-unidirectional airflow pattern where, although air moves in multiple directions through each part of the lung during breathing, each part transports more air in a certain direction over the whole breath cycle. This fascinating airflow pattern has features in common with both bird lungs (fully unidirectional) and mam-mal lungs (fully tidal).

## RESEARCH CHALLENGE

An important open question in comparative physiology con-cerns a functional explanation for the variety of lung designs in vertebrates: Why are lungs so diverse? Unidirectionally pul-monary airflow, a condition where certain lung gases travel in a consistent direction during both inspiration and expiration, was long thought to be found only in birds and to have been a re-quirement for the high metabolic demands of powered flight [3]. Recent work, however, discovered unidirectional flow patterns in alligators, monitor lizards, and iguanas, which are animals that do not fly and have relatively low metabolic rates [4–6]. We

thus need research to explain why pulmonary airflow patterns and lung designs vary in vertebrates [7]. Unfortunately, lungs are very small, delicate, and complex, so determining how air flows through them directly is difficult. Computational fluid dynam-ics (CFD) modeling offers a new approach: Anatomically accu-rate computational domains can be constructed from computed tomography scans, and simulations of how air flows through the lungs can be point-validated to reconstruct the diversity of pul-monary flow patterns in vertebrates. This project sought to in-vestigate these patterns in monitor lizards, an extremely diverse group of lizards [8] that nonetheless have a conserved body plan and unidirectional pulmonary airflow [6].

## METHODS & CODES

CFD simulations were run in OpenFOAM, an open source continuum mechanics library using a custom moving-boundary code based on the SIMPLE algorithm (transientSimpleDyMfoam). Model geometry was segmented from computed tomography scans of a live, anesthetized lizard. Simulation data were visual-ized in ParaView. Simulations were validated by means of visu-alization of aerosolized lipids through a microendoscope placed at various points in the lung while the lung was ventilated with a 25 mL syringe. In almost all cases, there was good correspon-dence between the smoke measurements and the CFD model.

Figure 1: Different types of airflow patterns in animal lungs. Mammals (top left) have a tidal airflow pattern where air moves to and fro through a ramifying network of bronchi. Birds (top right) have looping bronchi and air moves in the same direction during both inspiration and expiration through many of their bronchi. Monitor lizards (below) have a net-unidirectional airflow pattern where air moves to and fro but more air moves forward or backward over the whole ventilatory cycle in each part of the lung. (Top images from [1]; bottom image from [2].)

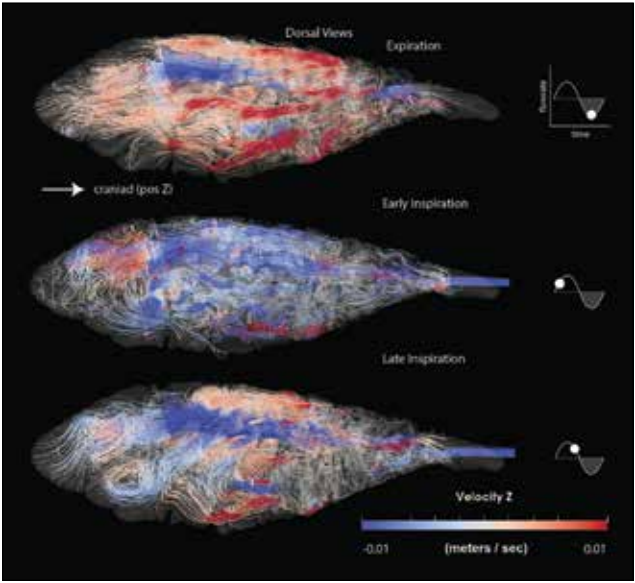
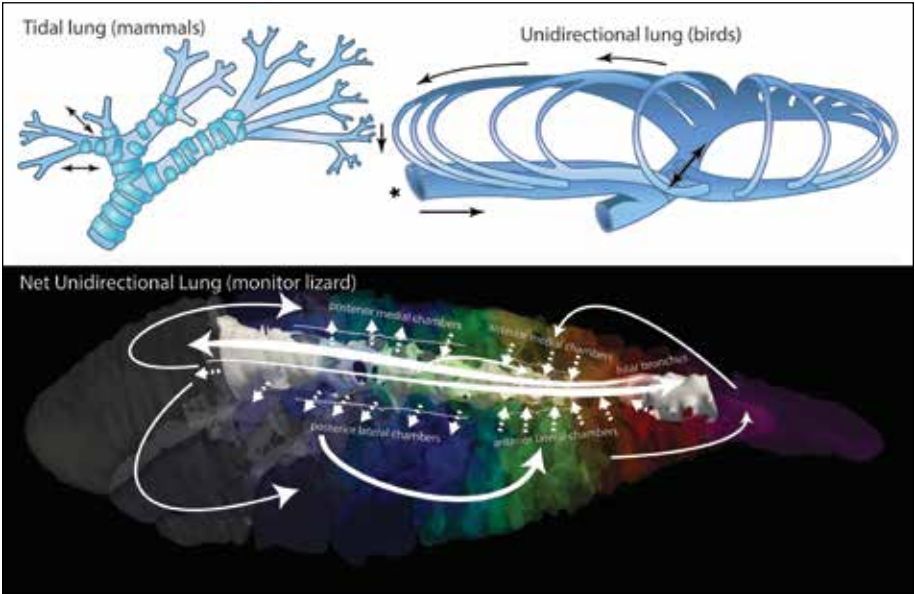


Figure 2: Dorsal views of streamlines from the CFD simulation of pulmonary airflow in monitor lizards run on the Blue Waters machine. Red streamlines indicate flow toward the front of the lung and blue streamlines indicate flow toward the back of the lung. Expiration and late inspiration contain overall similar flow patterns, but early inspiration is largely reversed. (Image from [2].)

## RESULTS & IMPACT

The lung of the savannah monitor lizard (*Varanus exanthematicus*) consists of a central intrapulmonary bronchus (IPB) and multiple ostia (openings) into lateral, medial, and dorsal second-ary bronchi. The main finding of this investigation is that the pat-tern of flow in the lung of monitor lizards is uniquely net-unidi-rectional (Fig. 2). Air flows primarily toward the back of the lung through the distal portion of the IPB and toward the front of the lung through intercameral perforations in the walls of the lat-eral, medial, and dorsal bronchi. The filigree that separates the bronchi allows the passage of air through a route other than the main IPB and thus has a function for airflow that is similar to the parabronchi of birds. The net-unidirectional flow throughout the lung arises from the collective action of multiple unidirection-al bronchi, where the direction of flow often reverses during the ventilatory cycle, such that early inspiratory flow is in a different direction from late inspiratory flow and expiratory flow.

The researcher is not suggesting that the lung of *V. exanthematicus* represents a transitional state; it has a unique airflow pattern that is an interesting blend of features similar to the bird and the iguana lung. Similar to the bird lung, inspired air travels net-caudad (toward the tail as opposed toward the head) through the center of the lung and net-cranial (toward the head) through the walls connecting bronchi within the lungs. Also like the bird lung, the cranial bronchi ostia are net-incurrent and the caudal ostia off the IPB are net-excurrent openings with respect to the IPB. The intercameral perforations in the varanid lung are simi-lar to the parabronchi of birds and even more like the intercam-eral (within a chamber) perforations of crocodilians [9], in that they move air net-cranially, but in monitors they are simple holes and not elaborate gas-exchange channels. The dorsal bronchi and proximal aspects of the secondary bronchi, on the other hand, are similar to the avian parabronchi in that they are the sites of gas exchange. On the other hand, both the varanid and the igua-na lung feature a strong jet of air that bends flow cranially at the caudal aspect of the lung.

## WHY BLUE WATERS

With Blue Waters, it was possible to run many models at once to test hypotheses about the importance of different anatomical lung traits on the resulting pulmonary airflow pattern. For ex-ample, several models were run where the frequency and size of the intercameral perforations between the lung bronchi were ei-ther increased or decreased, the first secondary bronchus was re-moved, and the type of lung motion was changed. Blue Waters also made possible the timely simulation of multiple breaths, as many simulations calling for hundreds of computational nodes could be computed simultaneously. Finally, a current investiga-tion into how airflow patterns vary between different species of monitors is only possible through Blue Waters because of the sheer number of node hours necessary!

## PUBLICATIONS & DATA SETS

R. L. Cieri and C. G. Farmer, "Computational fluid dynamics reveals a unique net-unidirectional pulmonary airflow pattern in the savannah monitor (*Varanus exanthematicus*), in review, 2019.

In August 2019, Robert Cieri received a Ph.D. in biology from The University of Utah, where he worked under the direction of C. G. Farmer.



THE EARLY INSTABILITY SCENARIO FOR PLANET FORMATION IN THE SOLAR SYSTEM

Matthew Clement, University of Oklahoma  
2017–2018 Graduate Fellow  
Collaborators: Nathan A. Kaib<sup>1</sup>, Sean N. Raymond<sup>2</sup>, Kevin J. Walsh<sup>3</sup>, John E. Chambers<sup>4</sup>, Alessandro Morbidelli<sup>5</sup>, David C. Rubie<sup>6</sup>

<sup>1</sup>University of Oklahoma  
<sup>2</sup>Université Bordeaux  
<sup>3</sup>Southwest Research Institute  
<sup>4</sup>Carnegie Institution for Science  
<sup>5</sup>Observatoire de la Côte d’Azur  
<sup>6</sup>University of Bayreuth

EXECUTIVE SUMMARY

The solar system’s outer planets (Jupiter, Saturn, Uranus, and Neptune) formed rapidly, while gas was still present in the infant solar system [1]. While the outer planets’ evolutionary history is rather well understood by theorists [2,3], the leading models seem to be incompatible with the solar system’s terrestrial system (Mercury, Venus, Earth, and Mars) [4]. The research team has used large suites of N-body simulations of the solar system’s earliest epochs of formation and growth to develop a new, robust model for the solar system that explains both its inner and outer regimes. Additionally, the team conducted the largest-ev-

er suite of planet formation simulations using a realistic code that accounts for the effects of fragmentation as bodies collide [5]. The team also used GPU acceleration [6] to accurately model dynamics down to realistic mass resolutions during the solar system’s earliest epoch, and in the young asteroid belt. Finally, the team performed a detailed investigation into the origin of the solar system’s most peculiar planet, Mercury.

RESEARCH CHALLENGE

Accurately modeling the late stages of planet accretion is subject to numerical limitations and simplifications. In particular, to keep the calculation tractable, most authors [7,8] employ integration schemes that neglect collisional fragmentation. The initial planet-forming disk, which in reality contained millions of solid objects with a range of masses, must be approximated with just over a thousand bodies (the majority of which are assumed not to interact gravitationally with one another). Nevertheless, such studies have proved successful at replicating the general orbits of the inner planets.

However, explaining Mars’ small mass (just 10% that of Earth) and rapid formation (about 10 times faster than Earth, as inferred from isotopic dating [9]) requires substantial modification to the standard theory of planet formation [8,10]. Furthermore, the asteroid belt’s low total mass (only a few percent that of the Moon) and unique dynamical structure are still largely unexplained [7,8,10]. Earth and Mars are both in the Sun’s “potentially habitable” zone, yet Mars is small, barren, and unable to support a robust atmosphere. Understanding the dynamical mechanisms that prevented Mars from growing into an Earth-like planet will give us insight into how special our own world really is.

METHODS & CODES

For the fragmentation simulations, the team used a modified version of the Mercury6 hybrid integrator, written in FORTRAN [11,13]. The simulations of terrestrial accretion begin with the simplest initial conditions, consistent with observations of proto-stellar disks [1,7,8]. To systematically test the effects of a giant planet instability, the team performed several batches of integrations and triggered the instability during different epochs of ter-

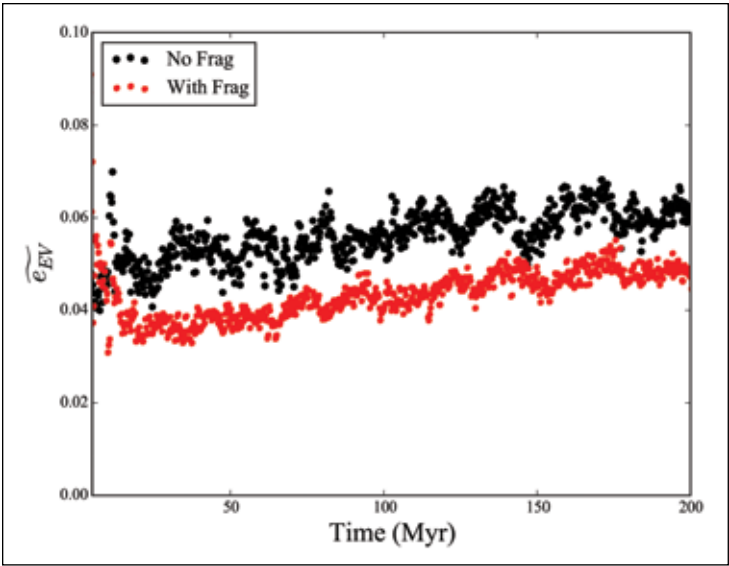


Figure 2: Eccentricity evolution of Earth and Venus analogs with (red) and without (black) collisional fragmentation. Since the total particle number stays higher for longer when fragmentation is included, the orbits of the growing analogs are damped more strongly by dynamical friction.

restrial growth. To investigate the effect on the asteroid belt, the researchers used a GPU code written in CUDA C (GENGA, [6]) to reperform successful integrations with a larger number of objects in the belt region. Because gas–disk interactions are complex, the exact mass and planetesimal size distribution profiles that emerge from the primordial gas (and go on to form the inner planets) are not well known. Therefore, “simple” initial conditions might not be representative of physical reality. The team investigated this problem using GPU acceleration as well. Further, they employed a forcing function to mimic the effects of gas drag, and utilized a multi-annulus approach to track the accretion of millions of small objects in the infant terrestrial disk.

RESULTS & IMPACT

This work offers a simple and elegant explanation for Mars’ small size and rapid growth (Fig. 1). The instability simulations consistently outperform the control runs when measured against a variety of success criteria. In successful simulations, Mars undergoes no further accretion events after the instability, while Earth and Venus continue to grow (thus matching their relative geological formation times [9]). Additionally, the team found that accounting for collisional fragmentation results in fully grown systems of terrestrial planets that are better matches to the actual solar system in terms of their orbital excitation (eccentricities and inclinations; Fig. 2) and planet spacing (particularly that of Earth and Venus). Furthermore, the instability proves successful at depleting a primordially massive asteroid belt (consistent with disk models [8]) at the 99.9% level. Thus, an early dynamical instabili-

ty among the giant planets can simultaneously explain the structure of both the inner and outer solar system.

Of further note, recent work performed on Blue Waters has uncovered a potential solution to another long-standing problem in planetary science. All previous evolutionary models of the solar system generate populations of highly inclined asteroids inconsistent with the observed solar system. This work has found that these asteroids are naturally removed during Saturn’s phase of orbital migration. It also found a fossilized record of this phase of depletion in the distribution of asteroidal precession rates.

WHY BLUE WATERS

Blue Waters boasts state-of-the-art resources that were invaluable to the success of this project. The research relied on GPU accelerators on XK nodes almost exclusively. Having the ability to efficiently run large suites of GPU-accelerated jobs led the PI to seek out a Blue Waters allocation.

PUBLICATIONS & DATA SETS

M. Clement *et al.*, “Excitation and depletion of the asteroid belt in the early instability scenario,” *Astron. J.*, vol. 157, 2019, doi: 10.3847/1538-3881/aaf21e.  
M. Clement *et al.*, “The early instability scenario: Terrestrial planet formation during the giant planet instability and the effect of collisional fragmentation,” *Icarus*, vol. 321, 2019, doi: 10.1016/j.icarus.2018.12.033.  
M. Clement *et al.*, “Dynamical constraints on Mercury’s collisional origin,” *Astron. J.*, vol. 157, 2019, doi: 110.3847/1538-3881/ab164f.

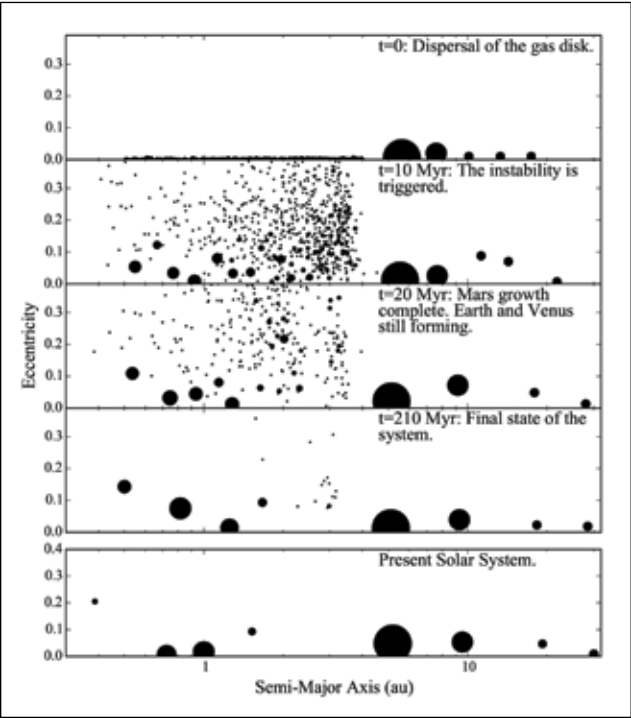


Figure 1: Semi-major axis vs. orbital eccentricity plot depicting the evolution of a successful system. The size of each point corresponds to the mass of the particle. (Because Jupiter and Saturn are hundreds of times more massive than the terrestrial planets, the team used separate mass scales for the inner and outer planets.)

Matthew Clement received a Ph.D. in astrophysics in May 2019 from the University of Oklahoma, There, he worked under the direction of Nathan A. Kaib.



# THE DISTRIBUTION OF SHEAR STRESS AND NUTRIENTS IN A TIDALLY ENERGETIC ESTUARY: THE ROLE OF NUMERICAL RESOLUTION AND VEGETATION

Salme Cook, University of New Hampshire  
2017–2018 Graduate Fellow

## EXECUTIVE SUMMARY

Ocean tides represent a major forcing mechanism in many coastal environments, responsible for the transport of salt, temperature, sediment, nutrients, and pollutants. Increases in population density and associated anthropogenic impacts have altered the productivity of estuarine environments, resulting in increased nutrient loading and amplified suspended sediment that reduces water quality. To accurately predict sediment transport, a good understanding of the bed shear stress that drives the sediment erosion, suspension, and deposition is essential. In this work, a high-resolution three-dimensional coupled hydrodynamic–wave–sediment transport numerical model (COAWST) was implemented and verified in a tidally dominated estuary located in the Gulf of Maine. The model was used in conjunction with available observational data sets to predict the shear stress distribution from the tidal channels across the mudflat.

## RESEARCH CHALLENGE

The coastal ocean includes diverse ecosystems encompassing both terrestrial and marine habitats that support approximately a third of the world’s population [1]. We are only beginning to understand the economic and environmental value of these resources and how to protect them in the face of climate change, sea level rise, extreme storm events, and increased human impact and pollution associated with population growth. These highly nonlinear systems are difficult to observe; however, the advent of numerical models and increased computational resources has made predicting the dynamics of these systems more accessible.

Coastal managers and decision-makers rely on data to understand how to protect them for future generations. Coastal currents and waves and their interactions with vegetation are complex and until recently, researchers have relied on sparse observations and theory alone. With the increase in numerical modeling formulations and computing power, the predictive capability of managers to help create more resilient coastal communities has increased. Better management of human impact and better understanding of the complex biogeochemical interactions and physical properties of these systems can help us sustainably interact and rely on them.

An open question this research addressed is the resolution required to capture the shallow water dynamics and how to couple these higher-resolution models to coarser regional and global models. In this work, the PI employed a validated high-reso-

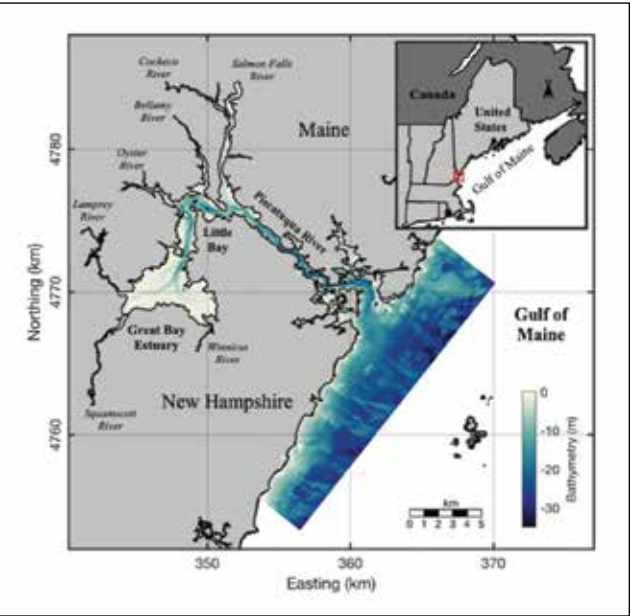


Figure 1: The Great Bay estuary and model bathymetry within the Gulf of Maine.

lution numerical model of a New Hampshire estuary (Fig. 1) to estimate the nutrient loading from sediments in the estuary to the coastal ocean.

## METHODS & CODES

This research was unique in that it relied on both collecting observational data in the estuary and the use of numerical model data sets. The observations were used to validate the model for currents and estimated shear stress in several locations. This provided fidelity in the overall distribution of shear stress in currents temporally and spatially within the estuary. The numerical model used was the COAWST modeling system. The 10-meter and 30-meter grid models were forced on the ocean side with five tidal constituents, three semidiurnal (M2, S2, N2) and two diurnal (O1, K1), using the Oregon State University Tidal Prediction Software package with the United States East Coast Regional Tidal Solution. River flow is generally low in the summer months, and since this modeling effort was representing those conditions, river flow was not considered. Shear stress was determined using a logarithmic bottom boundary layer assumption that is common in modeling shallow coastal environments.

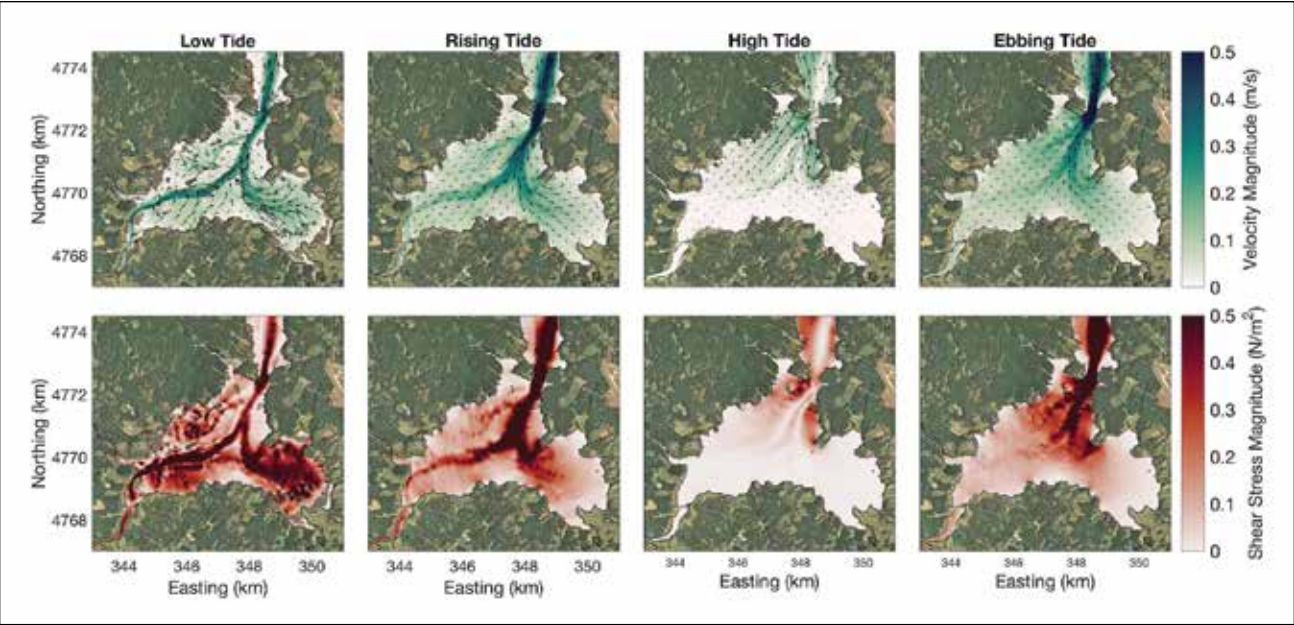


Figure 2: Spatial variability in velocity and shear stress for low (a), rising (b), high (c), and ebbing (d) tide. Velocity arrow resolution is decimated by a factor of 3.

## RESULTS & IMPACT

The results of this study present an estimate of spatial distribution of shear stress in a tidally dominant estuary using a verified numerical model and compared with stress estimates from observed currents at several locations. The spatial distribution of depth averaged velocities and shear stress are presented in Fig. 2 for four stages of the tide. The model runs (not shown) included the effects of vegetation and resolution and found that incorporating vegetation was an important improvement to the model, whereas the higher-resolution 10-meter model vs. the coarser-resolution 30-meter model had little effect on estimates. This provides a significant savings in computational resources when considering this issue. Future work looking at the distribution of waves or sediment types might require the higher-resolution model.

The distributions of shear stress were then used to estimate internal nutrient loading from sediments for a typical tidal cycle, spring, and neap cycle, and averaged over a month. When compared with rivers, model results suggest that internal sources of nutrient loads from sediment are on the same order as rivers for at least half of the year. These results indicate that when eelgrass populations are healthy and abundant, they lower the availability of sediment for resuspension and subsequent release of nutrients. This study demonstrates that a coupled hydrodynamic–vegetation model is capable of estimating the distribution of shear stress for a tidally dominant estuary.

Although nutrient loading from sediments is considered an internal load to the system, estuarine managers do not incorporate it into current nutrient loading estimates. This is not a process that can be mitigated; therefore, greater attention must be placed on those processes that can be controlled in terms of impervious surface cover, fertilizer use in residential and agricultural lands, industrial outputs of nitrogen and phosphorus, and wastewater treatment plants.

## WHY BLUE WATERS

The Blue Waters system and associated project staff were incredibly reliable and provided a powerful tool that was integral for this research, whereas the other machines the PI has used were incapable. The professionalism and efficient nature of the project staff were and are simply unparalleled, and have created a new standard in high-performance computing scientific support.

## PUBLICATIONS & DATA SETS

S. Cook, T. C. Lippmann, and J. D. Irish, “Modeling nonlinear tidal evolution in an energetic estuary,” *Ocean Model.*, vol. 136; pp. 13–27, Apr. 2019.

Salme Cook received a Ph.D. in oceanography from the University of New Hampshire in May 2019, having worked under the direction of Tom Lippmann.



EXASCALE ASTROPHYSICS WITH ENZO-E: DEVELOPMENT OF PHYSICS MODULES FOR GALAXY-SCALE AND COSMOLOGICAL SIMULATIONS

Andrew Emerick, Columbia University  
2018–2019 Graduate Fellow

EXECUTIVE SUMMARY

The Enzo–E/Cello project involves a new astrophysical hydrodynamics code written to take advantage of the next generation of exascale computing systems currently under construction. This research has been under development for several years, during which time the code has demonstrated impressive scaling capabilities on test problems utilizing the entire Blue Waters system. In addition, the PI worked to implement key physics modules into Enzo–E/Cello that are required for its use on production-level science runs studying the evolution of individual galaxies and large, cosmological volumes of the universe. This code development is necessary to enable cutting-edge astrophysics research on upcoming next-generation systems.

RESEARCH CHALLENGE

Advancements in our understanding of the universe over the past several decades have largely been due to an investment in ever-improving high-performance computational resources and the development of computational codes to model astrophysical phenomena on these machines. However, much of the current generation of astrophysical hydrodynamics codes—used to study everything from planet formation, star formation, stellar evolution, galaxies, and cosmology—were first written in the 1990s, when scaling to hundreds or thousands of concurrent tasks was the newest technology. Many of these codes are not well suited to take full advantage of the next generation of exascale computing systems that will allow for millions (or more) of concurrent tasks. In particular, Enzo, a grid-based adaptive mesh refinement (AMR) code and one of the most widely used astrophysical hydrodynamics codes, has multiple design shortcomings that prevent it from making efficient use of next-generation computing systems. While continued optimization has improved Enzo’s performance over the years, a complete refactoring is required. For this reason, the Enzo–E/Cello project began with the aim of completely redesigning the AMR hierarchy control system (Cello) to scale well to large systems while layering on modern physics modules and algorithms (Enzo–E). Like its predecessor, Enzo,

this is an open-source code project that will be available to the astrophysical community for use in a broad range of contexts to tackle new challenging open questions in astronomy. In this project, the PI worked to develop and port the physics modules necessary to conduct the first set of science/production simulations with Enzo–E/Cello, including models for star formation, chemistry, radiative cooling and heating, stellar feedback, analytic gravitational potentials, and isolated galaxy initial conditions.

METHODS & CODES

This project utilized both the well-established AMR, cosmological, hydrodynamics code Enzo and the newly developed code project Enzo–E/Cello. Both codes are written predominantly in C++ with some underlying routines written in Fortran. While Enzo uses MPI for communication across processors/nodes, Enzo–E/Cello uses Charm++, a parallel programming library developed at the University of Illinois at Urbana–Champaign. The change means that Enzo–E/Cello utilizes task-based parallelism rather than domain decomposition or data parallelism.

RESULTS & IMPACT

The Enzo–E/Cello project has made significant strides over the past few years in developing a powerful, scalable AMR hydrodynamics code by demonstrating nearly ideal scaling using the entire Blue Waters computing system on a handful of simple test problems. In this project, the PI successfully implemented physics modules for star formation, stellar feedback, chemistry, radiative cooling and heating, analytic gravitational potentials, and isolated galaxy initial conditions. In addition, he helped improve the code with additional documentation, bug fixes, and error testing.

This work has enabled researchers to begin planning in detail the types of simulations possible with the new code that would not be possible with the current version of Enzo and other similar hydrodynamics codes. These projects will immediately target open questions in galactic evolution and the chemical enrichment of the gas and stars in and around galaxies, the diffuse gas—the circumgalactic and intergalactic media—that fills much

of the universe, and the formation of the first- and second-generation of stars in the early universe. Finally, the code is entirely open source and well documented, allowing for its use by the broader astrophysical community.

WHY BLUE WATERS

Blue Waters has provided the computing environment and technical support needed to develop the new code and to conduct a variety of scaling and performance tests. This petascale machine and its support of large simulation runs with large node counts has enabled the PI to explore the limits of Enzo–E/Cello to better target future development and optimization.

As a sixth-year Ph.D. student in astronomy at Columbia University, Andrew Emerick successfully defended his dissertation in May 2019, having worked under the direction of Greg Bryan and Mordecai–Mark Mac Low.



MAGNETOHYDRODYNAMIC SIMULATION: GALAXIES

Forrest Glines, Michigan State University  
2019–2020 Graduate Fellow

EXECUTIVE SUMMARY

Magnetized plasmas contribute to many phenomena in the universe, from stars and black hole jets to galaxy clusters. However, computational models of these plasmas require efficient use of vast computational resources. Implementing these models is further complicated by the growing number of architectures of upcoming machines. Most of these machines use different accelerators from different manufacturers, each with a different programming environment that normally requires a rewrite of the simulation code. In order to prepare for future supercomputers, the PI developed K-Athena, a conversion of the magnetohydrodynamics code Athena++ using Kokkos, a performance portability library, which allows one code base that runs efficiently on many computer architectures. In this work, the researcher used K-Athena on the GPUs on Blue Waters to study the transference between kinetic and magnetic energies in the magnetic turbulence by modeling the Taylor–Green vortex. The next step will be to develop physics-rich galaxy cluster simulations with magnetic fields, cosmic ray physics, and active galactic nuclei feedback.

RESEARCH CHALLENGE

Plasmas dominated by magnetic fields are ubiquitous in the universe. Scientists know from observations of synchrotron radiation that galaxy clusters, the largest gravitationally bound structures in the universe, host large-scale magnetic fields. Although the precise coupling of these fields and larger clusters is not yet understood, researchers know they influence cluster evolution. Magnetic fields generated within the active galactic nuclei (AGN), the supermassive black holes at the center of galaxy clusters, drive jets that carry energy out into the cluster, playing a key role in the quasistability of many clusters [1]. Cosmic rays, charged particles with relativistic velocities, stream along the magnetic fields and drive winds, carrying metals from stars and providing pressure support in the cluster [2]. Magnetic fields also couple to turbulence within the cluster gas through the small-scale dynamo effect, where small turbulent eddies wind up and grow magnetic fields, transporting small-scale kinetic energy into magnetic energy [3].

Although these effects have been observed, it is unclear how important they are in the evolution of galaxy clusters. Researchers can explore this, however, through simulation. Galaxy cluster simulations modeling the dark matter, gas dynamics, magnetic fields, and aforementioned effects would advance our understanding of galaxy clusters. However, such accurate models require the computational resources of next-generation supercomputers.

Most upcoming supercomputers, though, are moving to new hardware such as the GPUs on Blue Waters and other accelerators instead of the traditional CPUs. These accelerators use unique application programming interfaces (APIs), requiring codes to be rewritten for each API. To circumvent writing multiple codes, new tools such as RAJA, Kokkos, and additions to OpenMP allow for writing performance-portable code that executes efficiently across many hardware platforms. Taking advantage of these tools, the PI investigated creating a magnetohydrodynamics code that will be able to run on these upcoming accelerators. This led to the development of K-Athena, a conversion of the astrophysical magnetohydrodynamics code Athena++ [4] using Kokkos [5], a performance portability library, which attains high performance on CPUs and GPUs using thousands of nodes.

For this fellowship, the PI is using K-Athena on Blue Waters to study magnetic turbulence in the magnetized Taylor–Green vortex as the first application of K-Athena. The magnetized Taylor–Green vortex is a periodic initial field that decays into a turbulent flow. By modeling the vortex at high resolution, its energy spectrum can be measured to create simplified magnetic turbulence models. These models can be inserted into other simulations to account for turbulence below the simulation resolution. The next effort will be to add additional physics to K-Athena such as cosmic rays and AGN to do state-of-the-art galaxy cluster simulations on Blue Waters.

METHODS & CODES

Magnetized plasmas are expensive to evolve, requiring high resolution to accurately model many phenomena. Efficient usage of hardware is required to achieve these resolutions. To meet this challenge, the PI converted the existing Athena++ code using the Kokkos library [4] to enable high-performance runs on both CPUs and accelerators. Kokkos allows a single kernel to be compiled with OpenMP for CPUs, CUDA for NVIDIA GPUs, and other APIs for future machines.

The coding began using the plasma code Athena++ owing to its well-written structure and extremely efficient performance on CPUs [5]. The simple kernel design and robust data structures in Athena++ took minimal effort to incorporate Kokkos. The resulting K-Athena code runs near peak performance on state-of-the-art CPU and GPU machines using thousands of nodes.

RESULTS & IMPACT

This research using Blue Waters is just beginning. Results have already shown K-Athena performs on a variety of supercomputers, but simulating the Taylor–Green vortex on Blue Waters will be its first scientific application. By simulating magnetized turbulent flow with high resolution, this research will capture enough of the energy spectra to be able to extrapolate the effects of turbulent flows far below the simulated resolution in magnetized plasma simulations. Accounting for the turbulent cascade and small-scale dynamo effect will be crucial for modeling accurate galaxies and galaxy clusters with magnetic fields. These turbulence models will be used in next-generation astrophysical and cosmological simulations, consequently helping constrain the properties of dark matter and dark energy.

WHY BLUE WATERS

These magnetic turbulence simulations and idealized galaxy cluster simulations using the GPU-accelerated K-Athena code are well served by Blue Waters (BW). The resolution and scale of the simulations require large computational resources that are only available on a few supercomputers, including BW. The large number of GPU nodes on the BW system allows the simulations to use less energy and fewer resources. Additionally, the BW staff were very helpful in providing suggestions for compiling code to run at the best performance.

As a fourth-year Ph.D. student in astrophysics, Forest Glines works under the direction of Brian O’Shea at Michigan State University. He expects to receive his degree in April 2021.



# GPU-ACCELERATED INTERSTELLAR CHEMISTRY WITH WIND: A GENERAL ORDINARY DIFFERENTIAL EQUATION SOLVER

**Alexander Gurvich, Northwestern University**  
2018–2019 Graduate Fellow

## EXECUTIVE SUMMARY

Radiative cooling owing to interstellar chemistry is an important component of modern cosmological simulations, but the fully time-dependent calculation is often too computationally expensive to include for large-volume and high-resolution simulations. To address this limitation, the PI created WIND, a general GPU-accelerated ODE (ordinary differential equation) solver that supports systems that are coupled and stiff. After a naïve first implementation, we have seen a speedup by a factor of three for our problem. In addition, we have identified a number of bottlenecks whose mitigation can improve the code substantially, with the goal of making WIND a public code that is applicable to a wide variety of problems.

## RESEARCH CHALLENGE

Many simulations of radiative cooling approximate the processes of interstellar chemistry by assuming that the gas in the simulation is in chemical equilibrium; however, comparisons with time-dependent interstellar chemistry in the FIRE simulations

show that there is a significant difference. Including time-dependent chemistry results in more accurate cooling processes during the course of the simulation also enhances the predictive power and applicability of the simulation to interpreting real-life observations. Nevertheless, it is computationally expensive—taking up to 98% of the computational cost of the simulation when enabled, which has restricted the volume and resolution of the simulations we have been able to run.

## METHODS & CODES

We developed WIND, a brand-new GPU-accelerated code, to address the challenge of simulating radiative cooling so that we can apply our time-dependent chemistry model in a new regime of simulation. WIND is a general ODE solver and can be applied to any system of ODEs, either coupled or independent. WIND also includes two numerical methods for integrating ODEs, one of which is an “implicit method” that allows WIND to efficiently solve stiff systems of ODEs (ones that involve many very different timescales). We have implemented these algorithms for both

GPUs in CUDA and for CPUs in C, allowing users to take advantage of GPUs when they are available but switch to the CPU version if they are not.

## RESULTS & IMPACT

We have preliminarily found that our new ODE solver is three times faster on GPUs than when run on CPUs, with increasing speed as the system of equations is made larger. Additionally, there are at least three concrete targets for improving this result where naïve first-attempt implementations were used. Future work will be focused on these hotspots to improve the speed of WIND so that it can be made public and applicable to a wide variety of problems.

## WHY BLUE WATERS

Blue Waters gave me access to GPUs when I had none; the XK nodes on Blue Waters were absolutely critical to the development and testing of WIND.

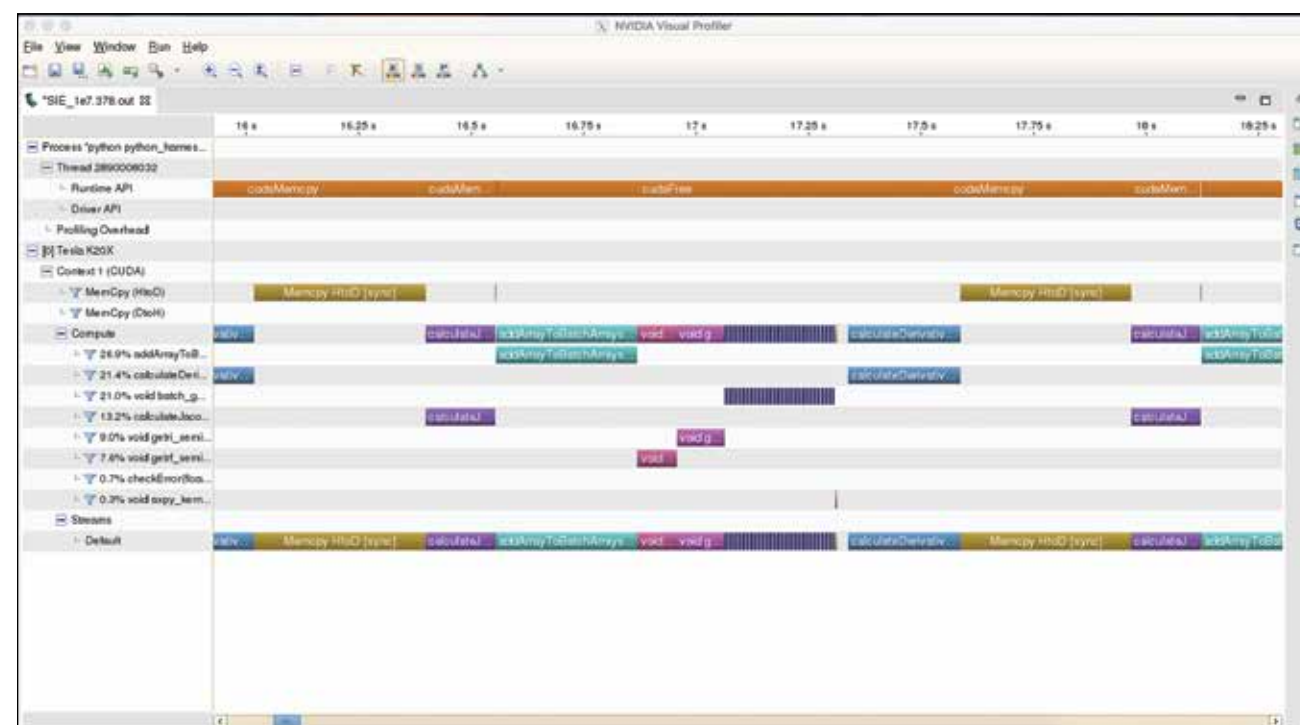


Figure 1: An example profiling session in NVidia’s visual profiling program NVVP. Time is along the horizontal axis and different CUDA kernels called in WIND are represented in rows. Colored blocks represent time spent by the GPU accelerator within each CUDA kernel. Visual profiling can quickly identify severe bottlenecks and algorithmic inefficiencies. In this session, memory communication (top row) is shown to be a major limiting factor.

Alexander Gurvich is a third-year Ph.D. candidate in astronomy and astrophysics at Northwestern University. There, he works under the direction of Claude-André Faucher-Giguère. He hopes to graduate in 2023.



# USING SPECTROSCOPIC DATA AND MOLECULAR SIMULATIONS TO ESTIMATE HETEROGENEOUS ENSEMBLES: HOW TO STUDY COMPLICATED, FLEXIBLE PROTEINS WHEN EXPERIMENTAL DATA ARE LIMITED

Jennifer M. Hays, University of Virginia  
2017–2018 Graduate Fellow

## EXECUTIVE SUMMARY

Flexible proteins play critical roles in cellular transport but are extremely challenging to model at high resolution. Experimental techniques such as double electron–electron resonance (DEER) report on conformational heterogeneity but are sparse over atom coordinates. The PI developed a methodology, Bias-Resampling Ensemble Refinement (BRER), to incorporate multimodal DEER data into molecular dynamics (MD) simulations to obtain high-resolution, experimentally validated models of flexible proteins. The results from ensemble simulations of unbound syntaxin-1a, a protein involved in the formation of SNARE complexes, which drive neuronal vesicle fusion, show that the PI’s method better reproduces experimental data than current state-of-the-art methods. Specifically, the methodology promotes sampling of significant backbone conformational change, unlike any other existing methods. In addition, BRER simulations of the soluble domain of syntaxin-1a revealed a previously unresolved open conformation of syntaxin-1a.

## RESEARCH CHALLENGE

It is difficult to study flexible proteins that play critical roles in infectious disease and cellular transport because so many states contribute to their conformational ensembles. High-resolution models of these systems are important for innovation in drug development and for answering fundamental questions in biophysics. It is challenging to develop atomic-resolution models using experiments alone because experimental techniques that report on heterogeneity often do so for only a few atomic degrees of freedom. Therefore, new hybrid methods are needed that include experimental and computational approaches to understand these systems.

## METHODS & CODES

The PI developed a new method for incorporating distributional data into MD simulations (Hays, Cafiso, and Kasson, 2019; see Publications & Data Sets below). This was done using the software package gmxapi, a Python interface for the GROMACS MD engine. The Python package for BRER simulations is freely available at [https://github.com/jmhays/run\\_brer](https://github.com/jmhays/run_brer) and the gmxapi code is at <https://github.com/kassonlab/gmxapi>.

## RESULTS & IMPACT

The PI developed both a method and an open source software package to integrate sparse experimental data using MD simulation to better understand flexible proteins. This will enable scientists, specifically spectroscopists, to study systems that would be too heterogeneous and complicated for standard refinement methods.

## WHY BLUE WATERS

The researcher has run multiple sets of ensemble simulations on Blue Waters to test the novel method and accompanying software package. A petascale, multi-GPU resource like Blue Waters was absolutely essential for completing both testing and production of all-atom ensemble simulations, which demanded over 200K node-hours over the course of the Graduate Fellowship.

The Blue Waters staff were also critical both in terms of the software development and in professional development. The PI learned to compile and run complicated software packages on CRAY systems. Further, she learned a great deal about general high-performance computing through the online and NCSA Symposium workshops, including how to utilize singularity containers, which have become an essential part of the researcher’s laboratory workflow development.

## PUBLICATIONS & DATA SETS

J. M. Hays, D. S. Cafiso, and P. M. Kasson, “Hybrid refinement of heterogeneous conformational ensembles using spectroscopic data,” *J. Phys. Chem. Lett.*, vol. 10, no. 12, pp. 3410–3414, 2019, doi: 10.1021/acs.jpcllett.9b01407.

Jennifer M. Hays obtained her Ph.D. in biomedical engineering from the University of Virginia in November 2019 and worked under the direction of Peter M. Kasson.

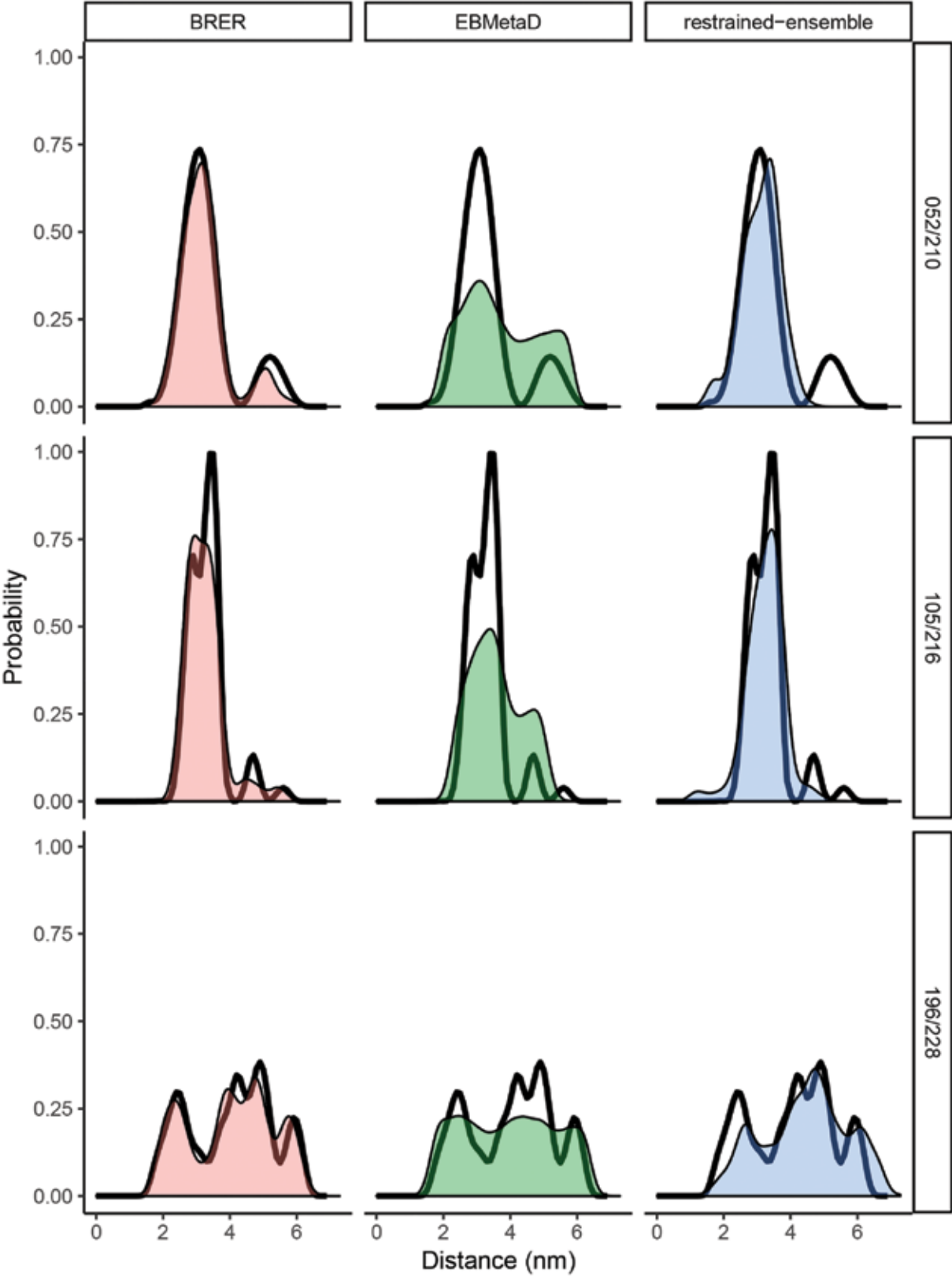


Figure 1: Ensemble-MD simulations refined using BRER better reproduce three experimental distributions (52/210, 105/216, 196/228) than two other preexisting refinement methods: EBMetaD and restrained-ensemble MD.



ELECTRON DENSITY-BASED MACHINE LEARNING FOR ACCELERATING QUANTUM CALCULATIONS

Joshua Lansford, University of Delaware  
2019–2020 Graduate Fellow

EXECUTIVE SUMMARY

This project aims to develop a machine learning (ML) algorithm that can be used to generate catalytic reaction mechanisms and kinetic models at reduced computational cost. The ML-based model will be trained to electron density to accelerate the three most computationally intensive components of the reaction energy profile: transition states, global minima, and entropy. The work will for the first time combine structural and density data to enhance ML convergence. The representation of the molecular systems will be invariant to rotations, translations, and re-ordering of atoms. Because it will use distances, partial charges, and bond orders, it should be generalizable to any size system. All generated surface data will be made available to the public upon publication of the work. On Blue Waters, the PI has computed electron density for local minima of carbon monoxide platinum nanoparticles; these data will also serve as the input for the data generation necessary for entropy predictions.

RESEARCH CHALLENGE

A vast array of materials properties can be computed using quantum chemical calculations that cannot be identified with even the best experiments available today. Machine learning (ML) algorithms can emulate these calculations at speeds that are orders of magnitude faster. For bulk materials properties, training ML algorithms to either structural or one-dimensional electron density data can result in very accurate predictions [1]. However, such techniques are currently lacking for much more demanding calculations involving catalytic reaction mechanisms and predictions of better materials.

Construction of a reaction mechanism and building a kinetic model require mapping of the reaction coordinate along the minimum energy path of the potential energy surface. Identifying global minima and their transition states on this path is expensive because of the large number of states that must be sampled and calculated. Vibrational contributions to entropy require separate energy calculations for every degree of freedom of every atom at a high convergence [2].

METHODS & CODES

The electronic density distribution completely specifies the energy of a chemical system's state and can be calculated using Density Functional Theory (DFT) based on the Kohn–Sham equation [3]. With ML, descriptors of input data are mapped to desired quantitative output, often using complex nonlinear fitting or distances between input data and training data [4]. ML algo-

rithms based on atomic positions have already sped up energy and transition state calculations by an order of magnitude [5–7]. Without an infinitely large training set, however, relying only on atomic positions does not allow extrapolation. Combining geometric and electronic density information will help alleviate this issue and is an innovation of this project.

The method for accelerating quantum-based calculations used in this work includes electronic density as input for ML and should significantly accelerate transition state, energy minima, and entropic calculations. This representation will also enable the extrapolation of the algorithms to systems that include atom types not found in the original training set. This method will combine structural and three-dimensional electron density data to accelerate calculations of molecular reactions on catalyst surfaces. Partial charges, atomic dipoles, and possibly effective bond orders can represent the electron density, as partial charges represent the electron density localized to specific atoms, atomic dipoles represent the asymmetry of the charge, and bond orders represent the density in the space between atoms. This representation of electron density is easily combined with geometric information and can be made invariant to translations, rotations, and reindexing of atoms.

The PI uses the Vienna Ab initio Simulation Package (VASP) [8] for generating electron densities of the ground state and CHARGEMOL [9] for integrating those densities. In addition, he uses his own code for calculating frequencies and spectral intensities used in the physics- and data-based surrogate models.

RESULTS & IMPACT

This project has generated thousands of DFT calculations already and the PI has implemented parallel versions of VASP and CHARGEMOL on Blue Waters, which are generating data. These calculations serve as the basis of the modeling. Previous work, which the PI has submitted for publication, provides evidence that the calculated electronic densities are accurate and can be used in surrogate modeling. That work used the electron density and physics-based surrogate models to generate complex infrared spectra that were then used to learn the mapping from spectra to local structure via multinomial regression. The PI plans to implement this electron density-based machine learning algorithm to speed up the calculations of spectra for this model and to aid automatic structure generation. Further plans are to learn the forces and electron asymmetry directly to compute vibrational frequencies and spectral intensities. This approach should also drastically reduce the number of calculations needed to re-

lax a structure to its local optimum. A training set of local optima will be used to identify global optima and identify transition states. Upon completion of this work, a database of all DFT calculations will be freely available to research groups studying surface science and catalysis.

WHY BLUE WATERS

Blue Waters provides the necessary computational power to generate massive data. In the first week of the fellowship, the PI had already used several thousand node hours. The highly parallel architecture is necessary to train neural networks on large amounts of data with many features.

Joshua Lansford is in the fourth year of a Ph.D. program in chemical engineering at the University of Delaware and expects to graduate in 2020. His advisor is Dionisios G. Vlachos.



# EXTENDING THE LONGEVITY OF PRODUCED WATER DISPOSAL WELLS: EVALUATION USING REACTIVE TRANSPORT SIMULATION

Kara Marsac, Colorado School of Mines  
2018–2019 Graduate Fellow

## EXECUTIVE SUMMARY

For every barrel of oil an average of seven barrels of water is produced [1]; disposal of this produced water adds significantly to the cost of domestic oil and gas development and is one of the largest problems facing the industry today. Reactive transport simulations investigate treating produced water before deep well disposal in the Permian Basin, Texas, U.S.A. This study found that a reduction of  $\text{HCO}_3^-$  and  $\text{Ca}^{2+}$  concentrations led to increases in well lifetime of more than 13 years. Increasing the lifetime of a produced water disposal well near a producing well lowers the cost of transporting each barrel of produced water. These cost savings have the potential to create an increase in the number of economically viable oil and gas plays in the United States.

## RESEARCH CHALLENGE

Produced water management is one of the biggest challenges associated with oil and gas development [2]. It is estimated that onshore wells in the United States generate 14 to 21 billion barrels of produced water every year [1,3]. Of that total, 92% is managed using injection [1]. Disposal of produced water can be expensive, with the use of disposal wells costing between \$0.05 and \$2.65 per barrel [4]. With each barrel of extracted oil creating multiple barrels of produced water, disposal costs can add significantly to the cost of oil.

The interaction between produced water and the water in the disposal formation can create mineral precipitation, which impacts the continued disposal of water in that formation and eventually leads to the sealing and abandonment of the well. Reduction of mineral precipitation through produced water treatment

can increase the lifetime of disposal wells, thereby saving transportation costs.

## METHODS & CODES

PFLOTTRAN, a massively parallel, multiphase, multicomponent reactive transport code [5] is used to create simulations of a control produced water injection in the Permian Basin, TX, U.S.A. This control water simulation was generated from the Permian Basin Brine Database [6] as a synthetic water that is representative of an amalgamation of brines present in the basin. Dolomite, anhydrite, and calcite minerals precipitate in the Permian Basin from four ions:  $\text{Ca}^{2+}$ ,  $\text{Mg}^{2+}$ ,  $\text{HCO}_3^-$ ,  $\text{SO}_4^{2-}$ . In this work, we created simulations that reduced each of these ions in the control water by 25%, 50% and 75%, for a total of 12 simulations. In each simulation, produced water is injected at 20,000 gallons/day from a center point into a 100 m x 100 m x 100 m observation area. The grid is variable, with higher-discretization 1-m cells concentrated around the injection zone. The injected produced water mixes with the formation water already present in the matrix, a median of the Permian Basin Brine Database. Each simulation runs until porosity reaches zero, which we consider the end of the well lifetime.

## RESULTS & IMPACT

The results show that reducing calcium, magnesium, or bicarbonate increases injection time: our measure of increased well lifetime. The control simulation mineral volume results show calcite and dolomite as the dominant minerals controlling precipitation volume, while only a small amount of anhydrite precip-

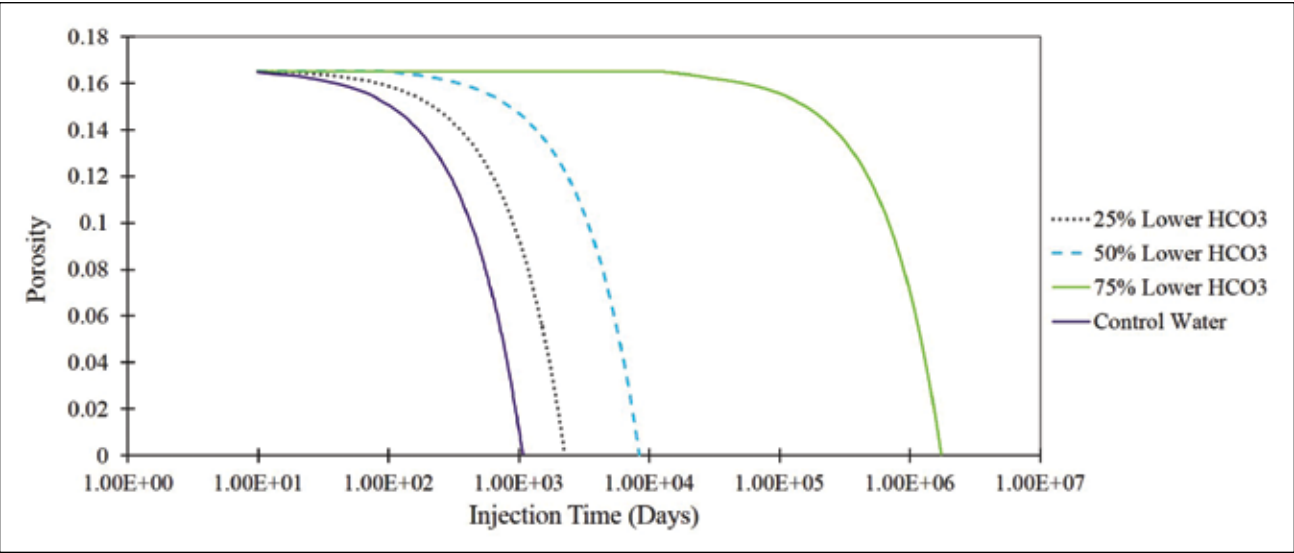


Figure 2: Results from the bicarbonate concentration reduction simulations compared to the control water.

itates. A calcium concentration reduction of 75% increases the well lifetime by over 5,000 days owing to a decrease in anhydrite, calcite, and dolomite. Reducing bicarbonate concentration by 75% increases the well lifetime by over 4,500 years by decreasing the volume of calcite and dolomite precipitated. Reducing magnesium concentration by 75% leads to an increase in well lifetime of over 300 days owing to the reduction of dolomite precipitation. Reducing sulfate concentration by 75% actually decreases injection time by 50 days. Reducing the concentration of sulfate only affects anhydrite, the mineral with the lowest volume. When anhydrite is not formed, there is more  $\text{Ca}^{2+}$  available in the water. This calcium forms calcite, leading to the precipitation of more calcite more quickly than the control simulation, and thus shows a decrease in well lifetime.

These results indicate that in the Permian Basin, oil and gas companies should pursue the treatment of calcium or bicarbonate in produced water to extend the lifetime of produced water disposal wells. This pursuit of treatment should come in the form of funding research of calcium and bicarbonate treatment for Permian Basin produced waters. By increasing the lifetime

of a disposal well near the producing well, companies save money on transporting produced water. These cost savings can lead to cheaper oil and gas and create more economically viable hydrocarbon plays in the United States, assisting in the development of domestic assets.

## WHY BLUE WATERS:

Blue Waters is essential for this research owing to the timescales of the simulations and the amount of output data. These simulations represent thousands of days of time; without the power of parallelization of Blue Waters, running each simulation could take months. Over the course of each simulation more than 6,000 output files can be printed, totaling over 600 GB of data. Blue Waters allows us to process such large data volumes.

## PUBLICATIONS & DATA SETS

K. Marsac and A. K. Navarre–Sitchler, “Extending the longevity of produced water disposal wells: evaluation using reactive transport modeling,” in review, 2019.

Kara Marsac graduated in May 2019 with a Ph.D. in hydrology from the Colorado School of Mines, having worked under the direction of Alexis Navarre–Sitchler.

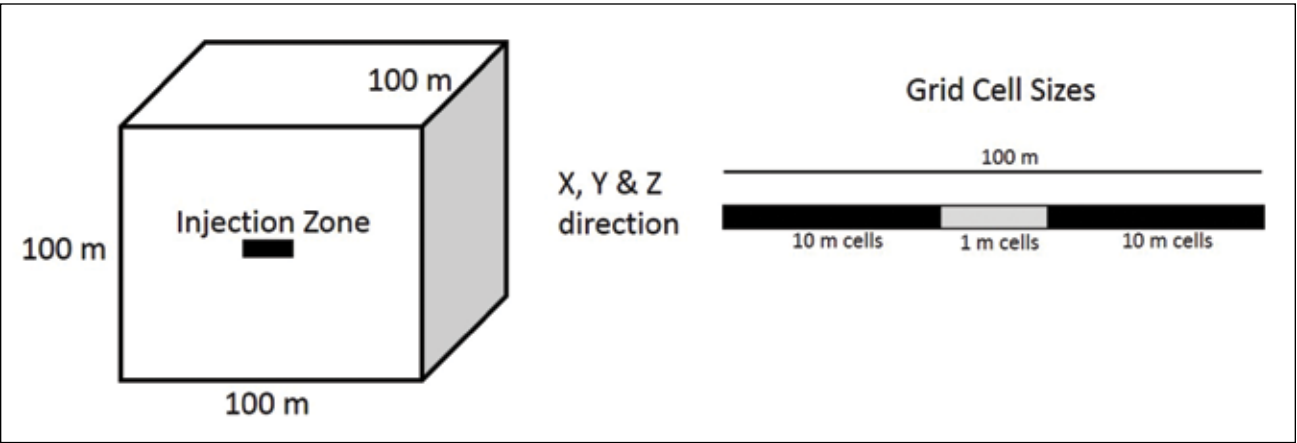


Figure 1: Diagram of grid size and discretization of all simulations.



DI

IMPROVED TRUMPET INITIAL LAPSE AND SHIFT FOR BINARY BLACK HOLE SIMULATIONS

Nicole Rosato, Rochester Institute of Technology  
2019–2020 Graduate Fellow

EXECUTIVE SUMMARY

This research project focuses on reducing error in simulations of merging pairs of black holes in order to better study the gravitational wave emission at merger. It currently takes months of supercomputer time to perform difficult simulations (for example, where the black holes are spinning quickly or are very differently sized), and increasing the accuracy of these simulations translates to an increase in required computational resources. We would like to change the initial values given to the gauge equations to be closer to their settled shape, allowing the gauge to settle more quickly and reducing error in the simulations.

RESEARCH CHALLENGE

The challenge of this research is to gain accuracy in the most challenging binary black hole simulations without increasing computational cost. The low mass-ratio and high-spin areas of parameter space are very sparsely covered; we hope this project will fill out that parameter space so that researchers have waveforms to

compare to potential detections from the Laser Interferometer Gravitational-Wave Observatory.

METHODS & CODES

To perform numerical relativity simulations, we use the Einstein Toolkit and specifically modify the Rochester Institute of Technology TwoPunctures initial data thorn. We are constructing new initial data values for the gauge based on their expected settled shape.

RESULTS & IMPACT

We are seeing a reduction in error using these new initial data, and therefore are gaining accuracy without actually having to use more computational resources.

WHY BLUE WATERS

These simulations require the use of large-scale computing resources due to computational intensity. The staff itself is knowledgeable about both the system and about the software I use.

Nicole Rosato is in the second year of a doctoral program in mathematical modeling at the Rochester Institute of Technology. She is working under the direction of Carlos Lousto and hopes to graduate in 2021.

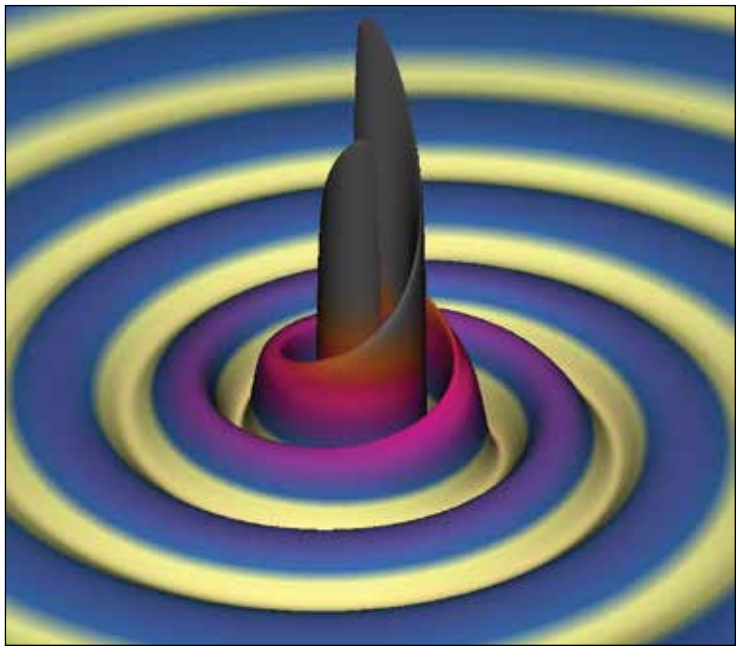


Figure 1: A visualization of gravitational radiation emitted 7 milliseconds after the merger of binary black hole system GW150914, detected by the Laser Interferometer Gravitational-Wave Observatory on September 14, 2015.

UNDERSTANDING THE PHYSICAL PROCESSES CAUSING INTERMEDIATE-DEPTH EARTHQUAKES

Shanna Chu, Stanford University  
2018–2019 Graduate Fellow

EXECUTIVE SUMMARY

This project aims to constrain the mechanisms of intermediate-depth earthquakes, which are rare but can be incredibly damaging, especially at large magnitude. (The recent magnitude 8.0 earthquake in Peru, fortunately, struck in a sparsely populated region.) These earthquakes occur under conditions where the standard physical model for earthquakes occurring in the crust breaks down, and because they do not occur on known faults, they can potentially be harder to forecast. Furthermore, data from these earthquakes show different frequency characteristics than shallow earthquakes, so it is important to understand how intermediate-depth earthquakes nucleate and how they vary from region to region, not only for scientific curiosity but also for earthquake engineering purposes. This work was done by applying a novel approach—simulating the physics of these earthquakes by utilizing the hypothesized mechanisms—and comparing the simulation data to real data, to better understand intermediate-depth earthquakes and to formulate a better fundamental model for them.

RESEARCH CHALLENGE

Intermediate-depth, intraslab earthquakes, which account for less than 1% of the world’s seismic activity, occur above the temperatures and pressures where the brittle failure found in shallow earthquakes is thought to be possible. Seismologists have generally agreed that the mechanism of initiation for intermediate-depth earthquakes is, thus, distinct from that of shallow earthquakes, and that it is highly likely some sort of dynamic weakening mechanism is involved. Data from intermediate-depth earthquakes also show characteristic differences from shallow earthquakes, such as high stress changes and low radiated efficiency. However, in the absence of good fundamental physical source models, most studies of intermediate-depth earthquakes utilize models extrapolated from the physics of shallow earthquakes. The aim of this project is to refine the source model for intermediate-depth events to better test the hypothesis of a dynamic weakening mechanism.

METHODS & CODES

In this project, the PI looked at waveform data from intermediate-depth earthquakes in Kyushu, Japan, which are available from the Hi-Net and Kik-Net seismometer and strong motion catalogs from the National Research Institute for Earth Science and Disaster Research. Previous work examined the scaling of seismic stress drop with the earthquake moment using the generalized circular crack model. However, that work determined that the results were harder to extrapolate to larger earthquakes, which appeared less cracklike.

The PI used the Support Operator Rupture Dynamics (SORD) code to synthesize earthquake waveforms. An entire fault surface was modeled by coupled nodes, where rupture is subject to local stress conditions and a friction law. Dynamic weakening mechanisms, which the researcher hypothesizes to be more important in intermediate-depth earthquakes, can be approximated by spatial variation of the simpler slip-weakening law. She modeled large earthquakes for which kinematic finite-source models exist and determined the physical feasibility of these model slip maps. The PI also tried to incorporate dynamic weakening mechanisms into the earthquake source to observe the impacts on waveforms and spectra.

RESULTS & IMPACT

The major result of this work thus far is the finding that slip map results from kinematic inversions (unconstrained by governing physics) can be physically impossible under conditions for dynamic rupture. Currently, physical models for intermediate-depth earthquakes are nonexistent; however, incorporating effects of dynamic weakening in these simulations is beginning to create better explanations for the atypical data retrieved from these earthquakes. Although analysis is ongoing, the technique of applying a physics-based simulation, enabled by high-performance computing, to study intermediate-depth earthquakes will help to bridge the gap between hypothesized mechanisms and the data. In addition, although rare, large intermediate-depth earthquakes can be very damaging, as the 2017 Chiapas event illustrates. Hence, having an accurate source model of these earthquakes for better ground motion prediction is also important to earthquake engineers.

WHY BLUE WATERS

Access to the Blue Waters system, where the SORD code is already optimized, has been pivotal to running the forward dynamic rupture simulations. The Blue Waters project staff were extremely helpful and prompt in their responses, which has greatly accelerated the progress of this research.

A fifth-year Ph.D. candidate in geophysics, Shanna Chu expects to graduate in March 2020 and is working under the direction of Gregory Beroza at Stanford University.



# ESCAPING FROM AN ULTRACOLD INFERNO: THE ULTRACOLD KRB DIMER REACTION

Micheline Soley, Harvard University  
2019–2020 Graduate Fellow

## EXECUTIVE SUMMARY

Ultracold chemistry, the study of reactions at temperatures below one millikelvin, offers an unprecedented opportunity for control over the outcomes of chemical reactions [1], and the ultracold molecules formed in these reactions have applications ranging from quantum computers [2] to investigations of fundamental constants of nature [3]. In this project, computational simulations of product formation in the ultracold potassium–rubidium (KRb) dimer reaction ( $[K_2Rb_2]^* \rightarrow K_2 + Rb_2$ ) were performed. Although ultracold reactions would typically be studied with quantum mechanics, computational simulation of time-dependent quantum mechanics is currently beyond reach for ultracold systems. Therefore, a combination of semiclassical and quantum mechanics have been employed on Blue Waters to bring simulation of ultracold chemical reactions within reach.

## RESEARCH CHALLENGE

Several properties of the ultracold KRb dimer reaction make it computationally intractable with state-of-the-art quantum techniques. Cold systems entail simulation of long lengthscales, which require larger memory allocations and more computational operations. Furthermore, although the incoming and outgoing molecules in the ultracold KRb dimer reaction are cold, the reaction passes through a hot intermediate phase (Fig. 1) that entails simulation of short lengthscales. Simulation of both long and short length scales demands multiscale and multiphysics techniques. These issues are compounded by the chaotic behavior of the molecules shown in Fig. 2, which is expensive to simulate. These concerns must be addressed in order to gain an understanding of ultracold reactions, which is vital to theorists and experimentalists in ultracold chemistry as well as physicists and chemists developing ultracold technologies.

## METHODS & CODES

To bring simulation of the ultracold KRb dimer reaction within reach, we make use of the fact that the reaction behaves like an ultracold inferno [4]. Although quantum mechanics is required to study the cold products, semiclassical mechanics can be used to study the extremely hot intermediate complex. Therefore, we combined quantum and semiclassical techniques to make investigation of the reaction possible. Blue Waters is then used to accelerate this research. To study the hot intermediate complex, Monte Carlo integration was performed with up to quintillions (1,000,000,000,000,000s) of sampling points on Blue Waters.

Parallelized random matrix and R-matrix theory code was then used on Blue Waters to predict the final distribution and rate of formation of the cold products.

## RESULTS & IMPACT

Preliminary results help confirm that parallelized semiclassical mechanics can be used to calculate the number of configurations efficiently in place of costly quantum mechanics. Analysis of the results of the ultracold KRb dimer reaction will directly inform ongoing experiments on the reaction, which will provide a deeper understanding of how chemical reactions occur at ultracold temperatures. This study is at the frontier of the field as it seeks to make simulation of reactions possible that are computationally intractable with existing time-dependent quantum

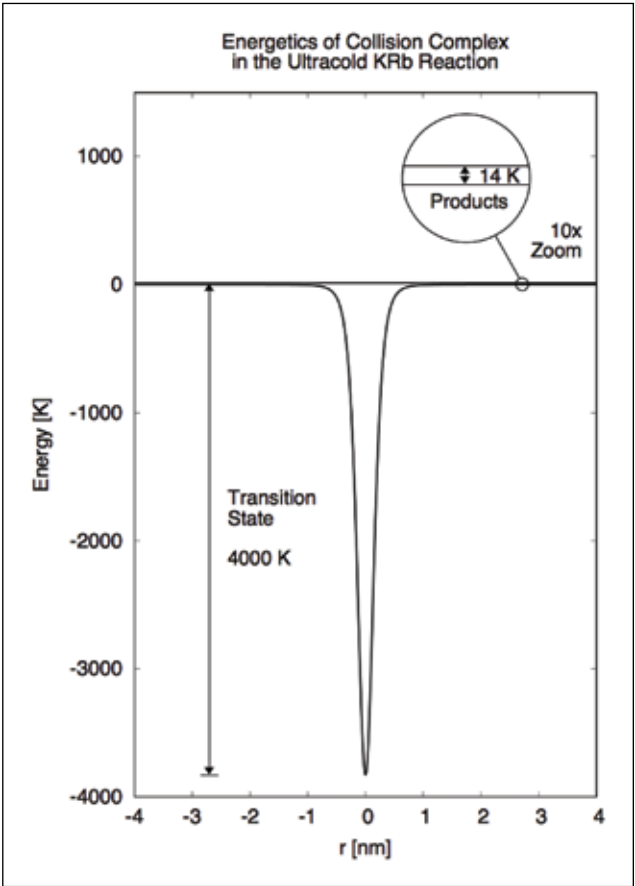


Figure 1: The ultracold KRb dimer reaction can be thought of as an ultracold inferno. In the reaction, a hot four-atom complex (4,000°K) breaks apart into two cold molecules (14°K). Magnification (10x) is shown to scale.

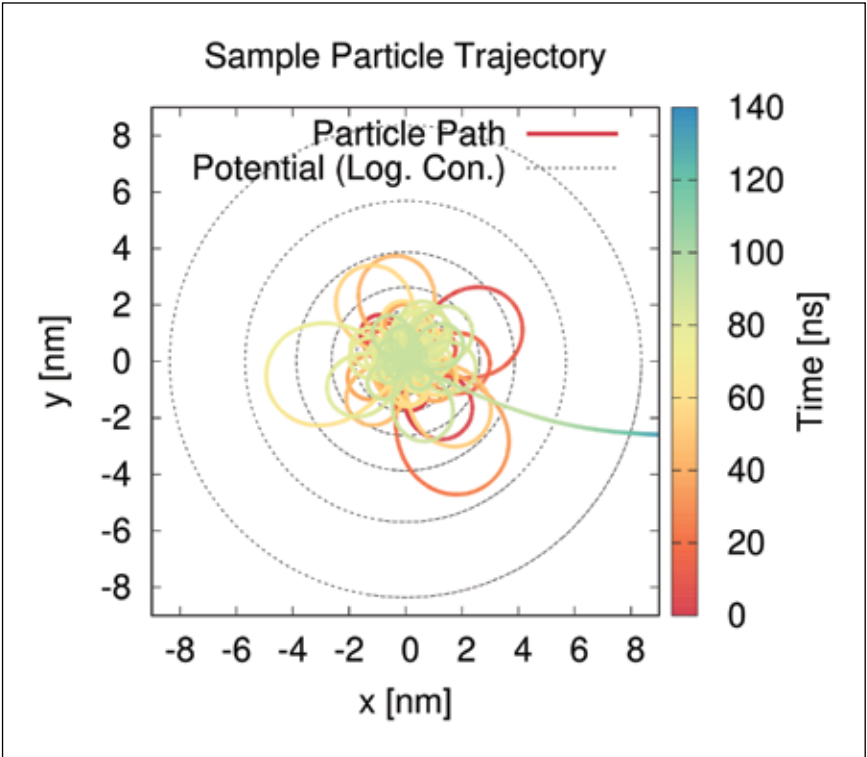


Figure 2: Computational simulation of the chaotic path of the molecules (multicolor line).

mechanical techniques. These techniques can be applied more generally to other chemical systems to help push the limits of reactions that can be studied in the field of computational chemistry. In collaboration with the results of the experimental Kang–Kuen Ni group, these computational results will help expedite development of key technologies such as quantum computers.

## WHY BLUE WATERS

The Monte Carlo integration and statistical R-matrix theory codes are computationally intensive. The Blue Waters super-

computer can significantly reduce the amount of time it takes to run these calculations and provide an opportunity to perform many tests simultaneously. This speed and flexibility make calculation of statistical R-matrix theory codes practical. In addition, the Blue Waters support staff facilitate acceleration of the program speed and the program development. The hope is that Blue Waters resources will help show that it is possible to simulate ultracold chemical reactions efficiently with a combination of semiclassical and quantum mechanics.

Micheline Soley, a fifth-year doctoral candidate in chemical physics at Harvard University, works under the direction of Eric J. Heller and expects to graduate in May 2020.



# THE IMPACTS OF HYDROMETEOR CENTRIFUGING ON TORNADO DYNAMICS: IMPROVING THE REALISM OF TORNADO SIMULATIONS

Ronald Stenz, University of North Dakota  
2016–2017 Graduate Fellow

## EXECUTIVE SUMMARY

Continued population growth in regions prone to tornadoes makes enhancing the understanding of these violent weather phenomena increasingly important. This research project attempts to improve the understanding of tornadoes by making simulations used to study these destructive and dangerous weather events more physically realistic. For the first time, the impacts that centrifuging of precipitation has on the vorticity budgets of these numerically simulated tornadoes will be quantified. Preliminary findings so far have been consistent with radar observations of tornadoes by removing an unrealistic buildup of precipitation in the vortex center (widely seen in current tornado simulations) of simulated vortices and tornadoes. Ongoing work uses numerous tornado simulations to evaluate the significance of the inclusion of precipitation centrifuging in tornado dynamics, as well as more generally studying how a tornado acquires its vorticity, or spin, in different environmental conditions.

## RESEARCH CHALLENGE

The primary research challenge being addressed is the lack of precipitation centrifuging in numerical simulations of tornadoes. In current simulations, precipitation follows the air flow, creating an unrealistic buildup of precipitation in the vortex center, which in turn creates a source of negative buoyancy that potentially limits the stretching of vorticity. In nature, as precipitation moves around a circulation such as a tornado, there is no force strong enough to keep the precipitation from moving outward, or being centrifuged, away from the circulation center. Observed tornadoes have a minimum of precipitation in the vortex center, while simulated tornadoes often have a relative maximum of precipitation in the vortex center. Addressing this challenge in improving the model’s realism requires the efficient calculation of millions of trajectories during a simulation and a solution that is numerically stable with other components of the model physics. Creating a centrifuging code that can work consistently with many different microphysical parameterizations is one of the challenging goals of this project.

With millions and sometimes billions of dollars of damage caused by tornadoes every year, along with the risk of fatalities or serious injuries from each tornado, a better understanding of these destructive weather events is needed to improve forecasting, preparedness, and mitigation of their impacts. By including the centrifuging of precipitation into the model used to learn about tornadoes, simulations become more consistent with what

is observed in nature. Research findings have shaped and will continue to shape forecasting methods and plans for preparedness and damage mitigation. Therefore, continued improvement in the understanding of tornadoes will provide results that can be used in operational settings, ultimately aiding those living in regions prone to tornadoes.

## METHODS & CODES

The widely used Cloud Model 1, which was designed for studying small-scale atmospheric phenomena such as thunderstorms [1] and has also been designed to run efficiently on supercomputers such as Blue Waters, was used for this research. To quantify the impacts of centrifuging on tornado dynamics, simulations were first performed without centrifuging. Just prior to the formation of a tornado, a checkpoint was employed, allowing the model to be run both with and without centrifuging from that point to determine the impact of the centrifuging of precipitation on tornado dynamics. To determine the magnitude of the centrifuging occurring, a centrifuging algorithm based on [2] used trajectories released within the simulation to calculate the curvature of the flow and ultimately how quickly precipitation will be centrifuged outward from the tornadic circulation. To quantify these impacts over a large sample size, atmospheric profiles of temperature, moisture, and wind from atmospheric soundings that were in close proximity to observed supercells [3] were used as the environmental conditions for our simulations of storms and their resulting tornadoes. A subset of these environments known to produce simulated tornadoes in previous research has been used for this study.

## RESULTS & IMPACT

Idealized simulations and full-scale storm simulations (with a resulting tornado) have been completed with and without centrifuging. In simulations without centrifuging, an unrealistic maximum of precipitation developed within the vortex core. However, after adding centrifuging, the precipitation in the vortex center was removed and a physically realistic precipitation minimum formed in the vortex center for both the idealized and full-scale tornado simulations. Similar to radar observations of tornadoes, the removal of precipitation from the vortex center was completed within several minutes in both types of simulations.

Optimization of this centrifuging algorithm is in progress, with the goal of sharing the findings and eventually the centrifuging code to allow future research to benefit from the improved real-

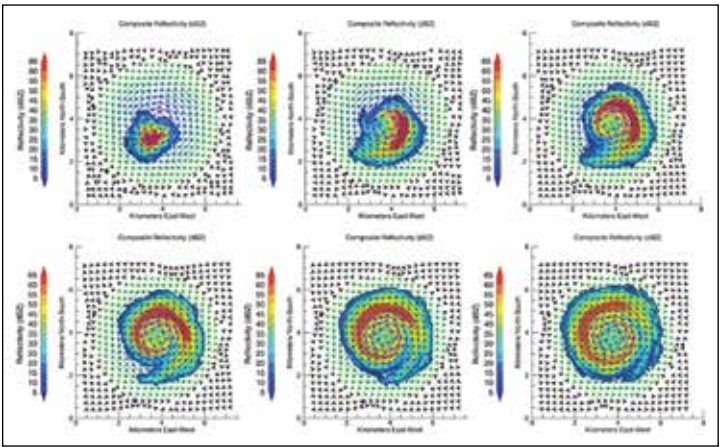


Figure 1: A time series of simulated reflectivity showing the centrifuging of precipitation away from an idealized vortex.

ism of tornado simulations. The work on the algorithm includes producing a centrifuging code that will work consistently with all microphysical parameterizations. Potential findings from this study on both the importance of centrifuging and also more general findings about how tornadoes work have the potential to improve future forecasting of tornadoes and also facilitate further research into understanding these deadly and destructive storms.

## WHY BLUE WATERS

Blue Waters was critical to this project because tornado simulations require thousands of computing cores and produce large amounts of data that must be stored and analyzed. Data generated in a typical simulation are in the order of tens and sometimes even hundreds of gigabytes per node. The computing power of Blue Waters, along with the available storage for the data, was a perfect match for this project. In addition, the technical and visualization support available greatly facilitated the accomplishment of our research goals.

Ronald Stenz, a sixth-year doctoral candidate in atmospheric sciences at the University of North Dakota, works under the direction of Matthew Gilmore. He expects to receive his degree in 2020.



# UNRAVELING FUNCTIONAL HOLE HOPPING PATHWAYS IN THE [Fe<sub>4</sub>S<sub>4</sub>]-CONTAINING DNA PRIMASE

Darius Teo, Duke University  
2018–2019 Graduate Fellow

## EXECUTIVE SUMMARY

This work has resulted in a Python module (EHPath.py) for characterizing charge hopping pathways in proteins and nucleic acids, which is the first computational tool that maps and ranks hopping pathways according to their mean residence time. The functionality of the module has been evaluated in several proteins including the oxygen-utilizing model enzyme cytochrome p450. Force field parameters that describe the high-potential iron–sulfur cluster have also been developed for molecular dynamics (MD) simulations. These two advances will enable the investigation of the role of an amino acid mutation found in gastric tumors in attenuating primase–DNA charge transfer and, in turn, primer handoff to polymerase  $\alpha$  in DNA replication.

## RESEARCH CHALLENGE

Recent work [1] suggests that primer handoff from the human DNA primase to polymerase  $\alpha$  (Pol $\alpha$ ), as part of the lagging strand synthesis in DNA replication, is driven by charge (electron or hole) transfer and the modulation of the redox states of the high-potential iron–sulfur (*i.e.*, [Fe<sub>4</sub>S<sub>4</sub>]<sup>2+/3+</sup>) clusters housed in the p58c and p180c domains of primase and Pol $\alpha$ , respectively. Experiments have investigated the efficiency of charge transfer between [Fe<sub>4</sub>S<sub>4</sub>]<sup>2+/3+</sup> in wild-type/mutant p58c and the protein-bound RNA/DNA duplex. Both Y345C and Y345F mutations in primase were found to reduce [Fe<sub>4</sub>S<sub>4</sub>]<sup>2+/3+</sup>-RNA/DNA charge transfer by approximately 60 to 95% [1]. The Y345C somatic mutation is of particular interest owing to its presence in gastric tumors [2].

In order to investigate the impact of mutation on charge transfer computationally, two challenges have to be met. First, a computational tool that can map and rank charge hopping pathways in proteins/nucleic acids has to be developed; this module can then be utilized to examine the hopping pathways between the RNA/DNA duplex and [Fe<sub>4</sub>S<sub>4</sub>]<sup>2+/3+</sup> in both the wild-type and mutant Y345C primase. Second, as there are no current force field parameters that can treat the high-potential [Fe<sub>4</sub>S<sub>4</sub>]<sup>2+/3+</sup> cluster, new parameters have to be developed. Once these two challenges are overcome, molecular dynamics (MD) simulations of the mutant and wild-type proteins can be performed; the trajectories obtained from these simulations would then help evaluate the possible attenuation of mutant primase binding to nucleic acid through Generalized Born/Poisson–Boltzmann methods, as well as the modulation of charge hopping pathways between the RNA/DNA duplex and [Fe<sub>4</sub>S<sub>4</sub>]<sup>2+/3+</sup>.

## METHODS & CODES

This work has resulted in a Python module (EHPath.py) that can map and rank charge hopping pathways in proteins/nucleic acids according to the mean residence time detailed in [3]. This code has also been used to evaluate hopping pathways in proteins of interest (manuscript submitted). The Python module is available at <https://github.com/etransfer/EHPath>. With regard to developing force field parameters to treat [Fe<sub>4</sub>S<sub>4</sub>]<sup>2+/3+</sup> for MD simulations, this work utilized broken-symmetry [4] density functional theory (BS–DFT) to optimize the geometries of the cluster in the two redox states (with relevant redox layer spin assignments), followed by force constant calculations using Seminario’s method [5]. Existing Lennard–Jones 6–12 parameters were used for the MD simulations. Partial atomic charges for the cluster were also derived. The force field parameters were derived and tested for robustness in MD simulations.

## RESULTS & IMPACT

High-potential Fe<sub>4</sub>S<sub>4</sub> clusters (in the 2+/3+ oxidation states) are important because they are commonly found in enzymes related to DNA replication and repair, including DNA primase and Pol $\alpha$ . Investigating the impact of the Y345C mutation in primase will advance the understanding of the primer handoff process driven by charge transfer as well as other cellular redox processes that are paramount in the regulation of major metabolic pathways. In addition, this will help inform the design of inhibitors that target such mutations. Furthermore, the newly developed force field parameters can be widely used by the research community for MD simulations of high-potential iron–sulfur cluster-containing proteins. Similarly, the Python module is available to the community for evaluating charge hopping pathways in proteins/nucleic acids that are relevant to other research areas.

## WHY BLUE WATERS

The Blue Waters fellowship has greatly advanced the progress of this research—both in terms of time and computation. Access to the top-notch computational power of the Blue Waters supercomputer has been very useful for running and completing BS–DFT and Hessian DFT calculations efficiently. The Blue Waters Point of Contact has also provided advice and engaged in helpful discussions through email exchanges and in-person meetings.

## PUBLICATIONS & DATA SETS

R. D. Teo, E. R. Smithwick, A. Migliore, and D. N. Beratan, “A single AT–GC exchange can modulate charge transfer-induced p53–DNA dissociation,” *Chem. Commun.*, vol. 55, no. 2, pp. 206–209, Nov. 2018, doi: 10.1039/C8CC09048C.

R. D. Teo, R. Wang, E. R. Smithwick, A. Migliore, M. Therien, and D. N. Beratan, “Mapping hole hopping escape routes in proteins,” *Proceedings of the National Academy of Sciences*, vol. 116, no. 32, pp. 15811–15816, Aug. 2019, doi: 10.1073/pnas.1906394116.

R. D. Teo, E. R. Smithwick, and A. Migliore, “2'-Deoxy-2'-fluoro-arabinonucleic acid: a valid alternative to DNA for biotechnological applications using charge transport,” *Phys. Chem. Chem. Phys.*, vol. 21, no. 41, pp. 22869–22878, Nov. 2019, doi: 10.1039/C9CP04805G.

Darius Teo is in the third year of a Ph.D. program in chemistry, working under the direction of David N. Beratan at Duke University. He expects to graduate in May 2020.



# THE TRANSPORT AND DYNAMICS OF WAVE-DRIVEN REEF JETS UNDER THE INFLUENCE OF ROTATION AND BOTTOM FRICTION

Walter Torres, Duke University  
2018–2019 Graduate Fellow

## EXECUTIVE SUMMARY

Predicting the fate of pollutants, heat, nutrients, carbon, and larvae in the coastal ocean is of acute ecological, commercial, and social importance—especially so on coral reef islands and atolls. On many reefs, jets arising from the interaction of reef topography and waves are responsible for exchanging water between the nearshore and open ocean, and so their dynamics are of particular interest.

This project involves a computational fluid dynamics study of an idealized coral reef island and demonstrates how the interaction among small-scale physical forcing (friction due to bottom roughness) and large-scale processes (e.g., the Coriolis force) modulates the behavior of wave-driven reef jets. Preliminary results show that lower bottom frictional regimes that are associated with degraded reef conditions increase the offshore export of the jets, simultaneously attenuating the relative importance of the Coriolis force that facilitates alongshore transport.

## RESEARCH CHALLENGE

Coral reefs are hotspots for marine biodiversity. Reefs provide habitat for a panoply of taxa while also providing vital ecosystem services such as food security, economic well-being, coastline protection, and they are also culturally significant heritage sites [1,2]. Unfortunately, coral reefs face global-scale threats such as

ocean warming and acidification; reefs worldwide have already experienced significant degradation, so it is paramount to understand how environmental processes affect coral health in order to inform ecological management efforts [3].

The resilience of a coral reef ecosystem to stressors is tightly entwined with the circulation field. Waves and currents replenish nutrients, transport coral and fish larvae between populations, moderate temperatures, and modify the coastal geomorphology [4]. Computational fluid dynamics modeling provides a way to investigate fundamental circulation processes on reefs that are otherwise analytically intractable, allowing us a deeper understanding of the physics underlying this complex multiscale system.

This study focuses specifically on the dynamics of wave-driven reef jets, which are common hydrodynamic features on reefs that arise owing to the interaction of reef topography and wave transformation in shallow water [5]. As surface gravity waves shoal and break, there is a vigorous shoreward input of energy, momentum, and mass; this is balanced by the presence of strong oceanward jets that form in the crenellations of the reef topography. These features can remain coherent over several kilometers yet are driven by wave-shoaling processes that happen over short spatial scales (10–100 m) in extremely shallow water (0.1–10 m). And so, this problem is inherently multiscale: a modeling challenge that demands a short timestep and fine spatial resolu-

tion in certain parts of the domain. Furthermore, reef jets often occur in proximity to one another, providing the opportunity for recirculation of scalars and particles under the right conditions, making it of interest to study an entire island system comprising an array of reef jets.

## METHODS & CODES

We modeled circulation on an idealized grid representing a coral reef island with a reef crest, inner lagoon, and a series of reef passes and reef flats. This annular domain was constructed in polar coordinates using variable grid spacing to conserve computation time. A uniform shoreward wave forcing was applied symmetrically to the domain on the outer boundary, with a closed inner boundary and periodic lateral boundary conditions. Results shown here are from a series of pilot numerical experiments that were carried out under permutations of bottom roughness and Coriolis force conditions (healthy rough reef vs. degraded smooth reef, 0° and 30°S latitude). Simulations used the Coupled–Ocean–Atmosphere–Wave–Sediment–Transport (COAWST) modeling system [6]. COAWST produces circulation and wave fields by coupling the ocean (Regional Ocean Modeling System) and wave (Simulating Waves in the Nearshore) models, which numerically solve the 3D primitive equations and 2D wave action equation, respectively. The wave-circulation coupling provided in COAWST was critical for simulating wave-driven reef jets.

## RESULTS & IMPACT

The results indicate that degraded reefs may be less retentive and experience shorter residence times owing to the decrease in bottom friction associated with the lower structural complexity of unhealthy coral. Preliminary model runs demonstrated that (unsurprisingly) the Coriolis force deflects the trajectory of the reef jets at some distance offshore, while jet centerline velocities’ magnitudes are weaker for healthy rough reefs owing to the large bottom friction. Stronger bottom friction also increased the relative importance of the Coriolis force in modifying the structure of the jet, as well as the size, speed, and coherence of the eddies shed from it. Surface waves also may influence the advection of these eddies, confining them nearer to shore via the Stokes drift mechanism. It is highly interesting that small-scale frictional processes on the very shallow back reef and reef crest have ramifications for the structure and evolution of kilometer-scale features such as jets and eddies. Future runs carried out over longer integration times along with detailed particle tracking studies will more clearly identify jet–jet interactions and recirculation patterns.

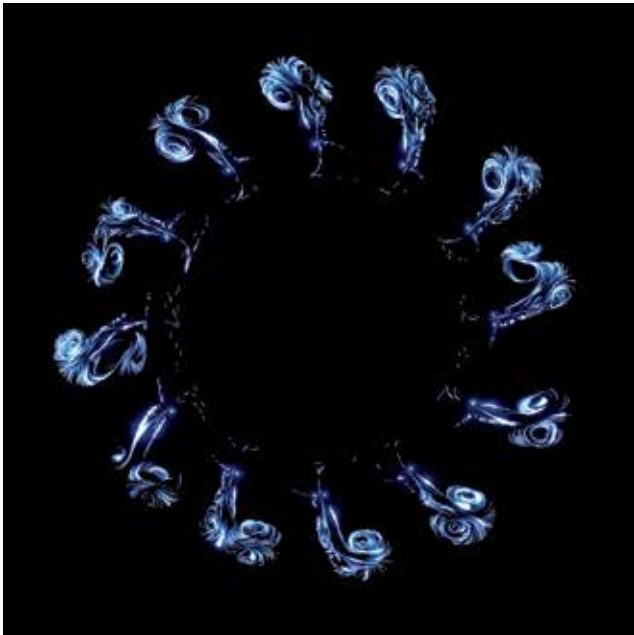


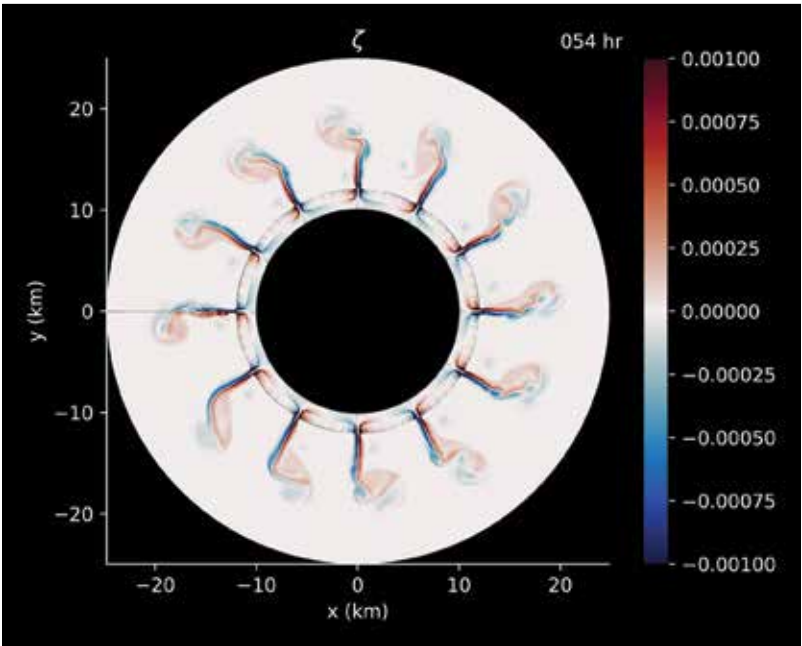
Figure 2: Particle trajectories illustrate the deflection of wave-driven reef jets in the idealized annulus domain owing to the earth’s rotation. This simulation was carried out at 45°S, and the scale is identical to Fig. 1.

This research has made progress in understanding fundamental exchange processes between the open ocean and nearshore reef environment, leveraging the high-resolution model simulations made possible by Blue Waters. The work has generated new hypotheses and predictions that will be evaluated *in situ* on Mo’orea, a coral reef island in French Polynesia, and will also aid the interpretation of ecological data being collected through the National Science Foundation’s Long Term Ecological Research (LTER) initiative, especially on factors affecting coral resilience such as larval recruitment, nutrient loading, and organismal behavior

## WHY BLUE WATERS

The Blue Waters supercomputing resource was essential in producing physically representative results; because we were able to achieve high spatial and temporal resolution for a coupled model over long integration times, the model captures the salient physics and time-evolution of barotropic reef jets. In addition, the Blue Waters support team provided outstanding and expedient technical support with software installation and module use.

Figure 1: A snapshot of relative vorticity ( $\zeta$ ) for the idealized annulus domain showing eddies shed off the wave-driven reef jets that begin to approach nearshore owing to Stokes drift via waves and/or entrainment. This simulation was carried out at 30°S with bottom roughness ( $z_0 = 10$  cm).



Walter Torres is a third-year Ph.D. candidate in marine science and conservation at Duke University, working under the direction of Jim Hench. He expects to graduate in the third quarter of 2021.



# SIMULATION OF BLUFF BODY STABILIZED FLAMES WITH PELEC: ADAPTIVELY RESOLVING TURBULENCE-COMBUSTION INTERACTIONS IN REAL-WORLD ENGINEERING PROBLEMS

Samuel Whitman, University of Colorado, Boulder  
2018-2019 Graduate Fellow

## EXECUTIVE SUMMARY

Gas turbine engines are widely used for propulsion and load-leveling applications, the latter being critical for incorporating intermittent renewable energy sources such as wind on the electrical grid. While efficiency, emissions, and flame stability are central to advances in turbine design, the extreme level of complexity continues to limit the accuracy of high-fidelity models and the control of gas turbine systems. This project uses the state-of-the-art high-performance computing (HPC) resources provided by Blue Waters to perform high-fidelity direct numerical simulations of turbulent premixed flames stabilized on bluff bodies. (By definition, bluff bodies are those that, because of their shape, have separated flow over a substantial part of their surface.) By studying the dynamics of stabilized flames, the researcher analyzed how high-fidelity models should be adapted to strongly turbulent compressible flows with complex chemistry in the presence of strong shear layers and examine mechanisms for enhancing flame stabilization while maintaining high efficiency and low emissions.

## RESEARCH CHALLENGE

As researchers and engineers work to improve the understanding and control of gas-engine systems, the complexity and extreme conditions found in these systems pose a significant challenge for computational approaches. To understand the challenges that engineers face and to better understand the interactions among turbulence, combustion, and recirculation-zone dynamics that impact stability, the project poses two important questions:

- How should large-eddy simulation (LES) modeling be adapted to highly turbulent compressible flows with complex chemis-

try in the presence of strong shear layers, where key physics can be left unresolved by the grid?

- What new mechanisms can be applied to better stabilize flames and prevent blowout while maintaining high efficiency and low emissions?

The first question is motivated by the current inability to model gas turbines in a physically accurate yet computationally efficient manner, while the second is central to turbine design and operation. In this project, these questions are addressed by performing high-fidelity direct numerical simulations (DNS) of turbulent premixed flames stabilized on bluff bodies. In particular, data produced by the high-resolution DNS are being used to analyze key dynamics associated with turbulence-flame interactions in bluff body configurations that are not well captured in high-fidelity LES, that also must be included in future low-fidelity models for practical turbine design, and that lead to instability and flame blowout.

## METHODS AND CODES

The researcher has adapted the next-generation compressible reacting flow solver PeleC to the Blue Waters HPC architecture. PeleC, under development at Lawrence Berkeley National Laboratory and the National Renewable Energy Laboratory, scales well, is fully parallelizable, and can be run using both the Message Passing Interface and OpenMP paradigms. The PeleC code uses built-in embedded boundary capabilities for structural modeling, which enables accurate implementation of different bluff body configurations for this work. Furthermore, PeleC incorporates Adaptive Mesh Refinement (AMR) via the AMReX suite, al-

lowing different regions of the simulation domain to be resolved at different levels of fidelity. This permits the addition of local refinement on areas of physical interest, fully resolving flame dynamics and turbulence-combustion interactions. Fig. 1 shows an example of AMReX and AMR at work, with refinement on vorticity. This dynamic refinement results in high-resolution simulations running at reduced computational cost when compared with traditional static meshes. This is particularly true for highly dynamical systems such as engines and turbines where the location of flames and vortex structures is often unpredictable and intermittent in time. The research also incorporates multi-step chemistry through Chemkin-type inputs in order to model chemical kinetic effects without loss of fidelity.

## RESULTS & IMPACT

This work used the PeleC exascale combustion code to simulate bluff body flow and successfully produced high-resolution simulations of nonreacting bluff body flow using AMR to provide localized resolution in the near-wake region of the bluff body. Localized refinement allows the simulations to physically resolve the

recirculation zone dynamics of the target experimental case at a level not previously examined. These simulations help to determine necessary parameters for computational modeling of this bluff body configuration, including simulation domain, boundary conditions, and resolution criteria necessary for accurate and efficient performance. The lessons learned have been used to design and instantiate reacting-flow simulations and, with these, to explore chemistry model performance and stability within the PeleC simulations running on Blue Waters.

## WHY BLUE WATERS

Blue Waters has been both formative and essential for this research. The allocation has allowed the researcher to begin using the highly scalable, adaptive PeleC code to tackle engineering challenges that would not otherwise be able to be considered: questions that are directly applicable to power generation, aviation, and the broader community. The staff have been both attentive and responsive, and ultimately were invaluable in getting the PeleC code up and running, helping with system-specific questions and advice for avoiding potential roadblocks along the way.

Samuel Whitman is a third-year Ph.D. candidate in mechanical engineering at the University of Colorado, Boulder. He expects to receive his degree in 2020 and has been working under the direction of Peter Hamlington and James G. Brasseur.

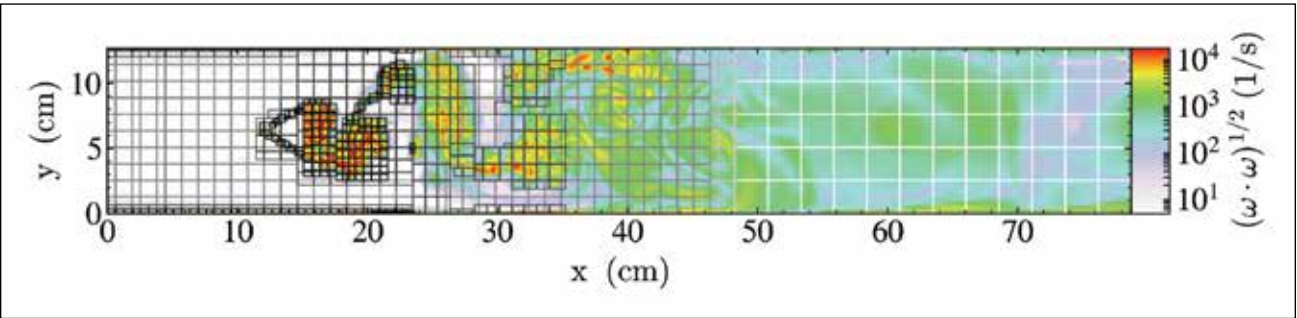


Figure 1: Flow around a triangular prism bluff body, with high-vorticity regions in red. Adaptive meshing provides localized refinement based on the vorticity. Here, each grayscale box shows three-dimensional grids that are between eight and 32 cells on each side. This localized refinement results in higher-resolution simulations at reduced computational cost.



# SCIENCE AND ENGINEERING TEAM ADVISORY COMMITTEE

The Science and Engineering Team Advisory Committee (SETAC) brings together a diverse group of scientists and engineers who represent the various science and engineering research teams using Blue Waters and the breadth of research powered by the Blue Waters system. The committee provides guidance and assessment to help the Blue Waters project deliver the best possible performance and services that will in turn assist research teams in achieving break-through results.

The SETAC makes recommendations on technical directions, strategies, and management while identifying potential challenges for petascale applications. As users themselves, the SETAC members also provide advice for solving common issues that arise from moving applications to Blue Waters and from system software at scale.

The SETAC members are nominated and the committee convenes three to four times per year and is available to the Blue Waters Project to provide guidance and advice as needed throughout the year.

## SETAC Members

### PETASCALE COMPUTING RESOURCE TEAMS

- **Paul Woodward**, Physics and Astrophysics, University of Minnesota (Chair)
- **Tom Cheatham**, Chemistry, University of Utah
- **David Ceperley**, Physics and Material Science, University of Illinois Urbana–Champaign
- **Tiziana Di Matteo**, Physics and Cosmology, Carnegie Mellon University
- **Paul Morin**, University of Minnesota
- **Susan Bates**, NCAR
- **Brian O'Shea**, Michigan State University
- **Manuela Campanelli**, Rochester Institute of Technology
- **Tom Jordan / Philip J. Maechling**, University of Southern California
- **Said Elghobashi**, University of CA Irvine
- **Nikolai Pogorelov**, University of Alabama, Huntsville

### GREAT LAKES CONSORTIUM FOR PETASCALE COMPUTATION TEAMS

- **H. Birali Runesha**, University of Chicago

### INDUSTRY TEAMS

- **Rick Arthur**, General Electric Global Research, Advanced Computing

### UNIVERSITY OF ILLINOIS AT URBANA–CHAMPAIGN TEAMS

- **Athol Kembal**, Astronomy, University of Illinois at Urbana–Champaign

# OTHER BLUE WATERS PROJECTS

The projects listed here had a Blue Waters allocation during this reporting period but did not submit a report for the project.

### Graphical Representation of Objects

John Cavazos

### Massive Galaxies and their Black Holes

Claude-Andre Faucher-Giguere

### CFD Software Development

Paul Fischer

### Parallel MLFMA

Levent Gurel

### Peptide Mutations Influenza Fusion

Peter Kasson

### Core-collapse Supernovae and their Ejecta

Eric Lentz

### 4-D Earth Simulation

Lijun Liu

### Studying Subduction Dynamics

Lijun Liu

### Cell Simulations of a Minimal Cell

Zaida Luthey-Schulten

### Lipids Control membrane Transporter Dynamics

Mahinthichaichan

### Gas & Galaxies in Enzo

Molly Peebles

### Neuroimaging with MR Elastography

Hillary Schwarb

### Modeling the Earth's Deep Interior

Xiaodong Song

### Modeling the Microbiome in IBD Patients

Rebecca Stumpf

### Magnetic Reconnection in Laser-Driven Plasmas: From Astrophysics to the Laboratory In Silico

Samuel Totorica

### Surface Defects on Hydrophobicity

Yang Zhang

### Protein Stability under Confinement

Yang Zhang



# REFERENCES

## SPACE SCIENCE

### Balsara, Dinshaw S.

[1] D. S. Balsara, “Wave propagation in molecular clouds,” *Astrophys. J.*, vol. 465, pp. 775–794, 1996.

[2] H.–B. Li and M. Houde, “Probing the turbulence dissipation range and magnetic field strengths in molecular clouds,” *Astrophys. J.*, vol. 677, p. 1151, 2008.

[3] D. A. Tilley and D. S. Balsara, “A two-fluid method for ambipolar diffusion,” *Mon. Notices Royal Astron. Soc.*, vol. 389, p. 1058, 2008.

[4] D.A. Tilley and D. S. Balsara, “Direct evidence for two-fluid effects in molecular clouds,” *Mon. Notices Royal Astron. Soc.*, vol. 406, pp. 1201–1207, 2010.

[5] D. A. Tilley and D. S. Balsara, “Two-fluid ambipolar diffusion for molecular clouds with realistic heating and cooling,” *Mon. Notices Royal Astron. Soc.*, vol. 415, pp. 3681–3692, 2011.

[6] D. A. Tilley, D. S. Balsara, and C. Meyer, “A numerical scheme and benchmark tests for two-fluid ambipolar diffusion,” *New Astron.*, vol. 17, pp. 368–376, 2012.

[7] S. Xu and A. Lazarian, “Damping of magnetohydrodynamic turbulence in partially ionized plasma: Implications for cosmic ray propagation,” *Astrophys. J.*, vol. 833, p. 215, 2016.

[8] C. Meyer, D. S. Balsara, B. Burkhardt, and A. Lazarian, “Observational diagnostics for two-fluid turbulence in molecular clouds as suggested by simulations,” *Mon. Notices Royal Astron. Soc.*, vol. 439, pp. 2197–2210, 2014.

[9] B. Burkhardt, A. Lazarian, D. S. Balsara, C. Meyer, and J. Cho, “Alfvenic turbulence beyond the ambipolar diffusion scale,” *Astrophys. J.*, vol. 805, pp. 118–126, 2015.

[10] S. Xu, S. Garain, D. S. Balsara, and A. Lazarian, “Turbulent dynamo in a weakly ionized medium,” *Astrophys. J.*, vol. 872, no. 62, pp. 1–12, 2019.

[11] D. S. Balsara, V. Florinski, S. Garain, S. Subramanyan, and K. F. Gurski, “Efficient, divergence-free high order MHD on 3D spherical meshes with optimal geodesic mapping,” *Mon. Notices Royal Astron. Soc.*, vol. 487, p. 1283, 2019.

### Campanelli, Manuela (1)

[1] D. B. Bowen, V. Mewes, S. C. Noble, M. J. Avara, M. Campanelli, and J. H. Krolik, “Quasi-periodicity of supermassive binary black hole accretion approaching merger,” *Astrophys. J.*, vol. 879, no. 2, p. 76, Jul. 2019.

[2] S. d’Ascoli, S. C. Noble, D. B. Bowen, M. Campanelli, J. H. Krolik, and V. Mewes, “Electromagnetic emission from supermassive binary black holes approaching merger,” *Astrophys. J.* vol. 865, no. 2, p. 140, Oct. 2018.

[3] D. B. Bowen, V. Mewes, M. Campanelli, S. C. Noble, J. H. Krolik, and M. Zilhão, “Quasi-periodic behavior of mini-disks in binary black holes approaching merger,” *Astrophys. J. Lett.*, vol. 853, no. 1, p. L17, Jan. 2018.

[4] S. Gawlowicz, “Model follows light to supermassive black holes,” Rochester Institute of Technology news. [Online]. Available: <https://www.rit.edu/news/model-follows-light-supermassive-black-holes?id=67992>, Accessed on: Aug. 7, 2019.

[5] J. Kazmierczak, “New simulation sheds light on spiraling supermassive black holes,” NASA news. [Online]. Available: <https://www.nasa.gov/feature/goddard/2018/new-simulation-sheds-light-on-spiraling-supermassive-black-holes>, Accessed on: Aug. 7, 2019.

[6] S. Kohler, “Oscillations from merging giants,” *Nova*. [Online]. Available: <https://aasnova.org/2019/07/29/featured-image-oscillations-from-merging-giants>, Accessed on: Aug. 7, 2019.

### Campanelli, Manuela (2)

[1] D. B. Bowen *et al.*, “Quasi-periodicity of supermassive binary black hole accretion approaching merger,” *Astrophys. J.*, vol. 879, no. 2, May 2, 2019, doi: 10.3847/1538-4357/ab2453.

[2] S. d’Ascoli, S. C. Noble, D. B. Bowen, M. Campanelli, and J. H. Krolik, “Electromagnetic emission from supermassive binary black holes approaching merger,” *Astrophys. J.*, vol. 865, p. 140, 2018, doi: 10.3847/1538-4357/aad8b4.

[3] D. B. Bowen, V. Mewes, M. Campanelli, S. C. Noble, and J. H. Krolik, “Quasi-periodic behavior of mini-disks in binary black holes approaching merger,” *Astrophys. J. Lett.*, vol. 853, p. L17, 2018, doi: 10.3847/2041-8213/aaa756.

### Draayer, Jerry

[1] T. Dytrych *et al.*, “Collective modes in light nuclei from first principles,” *Phys. Rev. Lett.*, vol. 111, pp. 252501–252501-5, 2013, doi: 10.1103/PhysRevLett.111.252501.

[2] K. D. Launey, T. Dytrych, and J. P. Draayer, “Symmetry-guided large-scale shell-model theory,” *Prog. Part. Nucl. Phys.*, vol. 89, pp. 101–106, 2016, doi: 10.1016/j.pnpnp.2016.02.001.

[3] D. Langr, T. Dytrych, K. D. Launey, and J. P. Draayer, “Accelerating many-nucleon basis generation for high performance computing enabled *ab initio* nuclear structure studies,” *Int. J. High Perform. Comp. Appl.*, vol. 33, no. 3, pp. 522–533, 2019, doi: 10.1177/1094342019838314.

[4] D. Langr, T. Dytrych, J. P. Draayer, K. D. Launey, and P. Tvrdík, “Efficient algorithm for representations of U(3) in U(N),” *Comput. Phys. Commun.*, 2019, doi: 10.1016/j.cpc.2019.05.018.

[5] T. Dytrych *et al.*, “Efficacy of the SU(3) scheme for *ab initio* large-scale calculations beyond the lightest nuclei,” *Comput. Phys. Commun.*, vol. 207, pp. 202–210, 2016, doi: 10.1016/j.cpc.2016.06.006; LSU3shell code, Louisiana State University, available under the GNU General Public License at the *git* repository: <https://sourceforge.net/projects/lsu3shell>.

[6] A. C. Dreyfuss, K. D. Launey, T. Dytrych, J. P. Draayer, and C. Bahri, “Hoyle state and rotational features in Carbon-12 within a no-core shell-model framework,” *Phys. Lett. B*, vol. 727, p. 511, 2013

[7] J. P. Draayer, K. D. Launey, and T. Dytrych, “Symmetry-adapted no-core shell model—the *ab initio* picture of nuclear collectivity,” in *Emergent Phenomena in Atomic Nuclei from Large-scale Modeling: A Symmetry-guided Perspective*, ed. K.D. Launey. World Scientific Publishing Co., 2017, pp. 229–263, doi: 10.1142/9789813146051\_0008.

### Gammie, Charles F.

[1] W. Benz, W. L. Slattery, and A. G. W. Cameron, “The origin of the moon and the single-impact hypothesis I,” *Icarus*, vol. 66, pp. 515–535, Jun. 1986.

[2] R. M. Canup, A. C. Barr, and D. A. Crawford, “Lunar-forming impacts: High-resolution SPH and AMR–CTH simulations,” *Icarus*, vol. 222, no. 1, pp. 200–219, Jan. 2013.

[3] H. Childs *et al.*, “VisIt: An end-user tool for visualizing and analyzing very large data,” in *High Performance Visualization—Enabling Extreme-Scale Scientific Insight*, E. W. Bethel, H. Childs, and C. Hansen, Eds. New York, NY: CRC Press, 2012, pp. 357–372.

[4] C. F. Gammie, W.–T. Liao, and P. M. Ricker, “A hot big bang theory: Magnetic fields and the early evolution of the protolunar disk,” vol. 828, no. 1, p. 58, Sept. 2016.

[5] P. D. Mullen and C. F. Gammie, “Magnetized models of giant impacts,” in preparation, 2019.

[6] J. M. Stone, K. Tomida, C. J. White, and K. G. Felker, “The Athena++ adaptive mesh refinement framework: Magnetohydrodynamic solvers,” in preparation, 2019.

### Hawley, John F.

[1] S. A. Balbus and J. F. Hawley, “A powerful local shear instability in weakly magnetized disks: I. Linear analysis,” *Astrophys. J.*, vol. 376, pp. 214–222, 1991.

[2] S. A. Balbus and J. F. Hawley, “Instability, turbulence, and enhanced transport in accretion disks,” *Rev. Mod. Phys.*, vol. 70, pp. 1–53, 1998.

[3] J. H. Krolik and J. F. Hawley, “A steady-state alignment front in an accretion disk subject to a Lense–Thirring torque,” *Astrophys. J.*, vol. 806, 2015, doi: 10.1088/0004-637X/806/1/141.

[4] J. F. Hawley and J. M. Stone, “MOCCT: A numerical technique for astrophysical MHD,” *Comp. Phys. Comm.*, vol. 89, pp. 127–148, 1995.

[5] C. R. Evans and J. F. Hawley, “Simulation of general relativistic magnetohydrodynamic flows: A constrained transport method,” *Astrophys. J.*, vol. 332, pp. 659–677, 1998.

[6] K. Sorathia, J. H. Krolik, and J. F. Hawley, “MHD simulation of a disk subjected to Lense–Thirring precession,” *Astrophys. J.*, vol. 777, p. 21, 2013.

[7] J. C. B. Papaloizou and C. Terquem, “On the dynamics of tilted discs around young stars,” *Mon. Not. Roy. Astr. Soc.*, vol. 274, pp. 987–1001, 1995.

[8] R. Nealon, D. J. Price, and C. J. Nixon, “On the Bardeen–Petterson effect in black hole accretion discs,” *Mon. Not. Roy. Astr. Soc.*, vol. 448, p. 1526, 2015.

### Huerta, Eliu (1)

[1] M. Doi, M. Fukugita, and S. Okamura, “Morphological classification of galaxies using simple photometric parameters,” *Mon. Notices Royal Astron. Soc.*, vol. 264, p. 832, 1993.

[2] D. B. Wijesinghe, A. M. Hopkins, B. C. Kelly, N. Welikala, and A. J. Connolly, “Morphological classification of galaxies and its relation to physical properties,” *Mon. Notices Royal Astron. Soc.*, vol. 404, p. 2077, 2010, arXiv:1001.5322.

[3] D. J. Eisenstein *et al.*, “SDSS-III: Massive spectroscopic surveys of the distant universe, the Milky Way, and extra-solar planetary systems,” *Astron. J.*, vol. 142, no. 3, p. 72, 2011.

[4] C. J. Conselice, “The relationship between stellar light distributions of galaxies and their formation histories,” *ApJS*, vol. 147, no. 1, pp. 1–28, 2003.

[5] O. Lahav *et al.*, “Galaxies, human eyes and artificial neural networks,” *Science*, vol. 267, p. 859, 1995.

[6] Dark Energy Survey Collaboration *et al.*, “The Dark Energy Survey: More than dark energy—an overview,” *Mon. Notices Royal Astron. Soc.*, vol. 460, p. 1270, 2016.

[7] LSST Dark Energy Science Collaboration, “Large Synoptic Survey Telescope: Dark Energy Science Collaboration,” *ArXiv e-prints*, 2012. [Online]. Available: arXiv:1211.0310

[8] Keras, “Keras: The Python Deep Learning library,” 2018. [Online]. Available: <https://keras.io/>

[9] M. Abadi *et al.*, “TensorFlow: Large-scale machine learning on heterogeneous distributed systems,” *ArXiv preprint*, 2016. [Online]. Available: arXiv:1603.04467

[10] F. Chollet, “Xception: Deep learning with depth-wise separable convolutions,” *ArXiv e-prints*, 2016. [Online]. Available: arXiv:1610.02357

[11] J. Deng, W. Dong, R. Socher, L.–J. Li, K. Li, and L. Fei–Fei, “ImageNet: A large-scale hierarchical image database,” in *Proc. CVPR*, 2009.

[12] S. Kornblith, J. Shlens, and Q. V. Le, “Do better ImageNet models transfer better?” 2018. [Online]. Available: arXiv:1805.08974



[13] NCSA Gravity Group, “Pattern learning recognition in convolutional neural networks,” Jun. 2019. [Online]. Available: <https://www.youtube.com/watch?v=1F3q7M8QjTQ>

Huerta, Eliu (2)

[1] D. Johnson, E. A. Huerta, and R. Haas, “Python Open source Waveform ExtractoR (POWER): An open source, Python package to monitor and post-process numerical relativity simulations,” *Classical Quantum Gravity*, vol. 35, p. 027002, Jan. 2018.

[2] E. A. Huerta *et al.*, “Eccentric, nonspinning, inspiral, Gaussian-process merger approximant for the detection and characterization of eccentric binary black hole mergers,” *Phys. Rev. D*, vol. 97, p. 024031, 2018.

[3] E. A. Huerta *et al.*, “Physics of eccentric binary black hole mergers: A numerical relativity perspective,” *Phys. Rev. D*, vol. 100, p. 064003, Sept. 2019.

Huerta, Eliu (3)

[1] J. Strader, L. Chomiuk, T. J. Maccarone, J. C. A. Miller–Jones, and A. C. Seth, “Two stellar-mass black holes in the globular cluster M22,” *Nature*, vol. 490, pp. 71–73, 2012.

[2] C. J. Hailey *et al.*, “A density cusp of quiescent X-ray binaries in the central parsec of the galaxy,” *Nature*, vol. 556, 2018, doi: 10.1038/nature25029.

Kemball, Athol J.

[1] LSST Science Collaborations, *LSST Science Book Version 2.0 | The Large Synoptic Survey Telescope*. [Online]. Available: <https://www.lsst.org/scientists/scibook>

[2] P. E. Dewdney, P. J. Hall, R. T. Schilizzi, and J. L. W. Lazio, “The Square Kilometre Array,” *Proc. IEEE*, vol. 97, no. 8, pp. 1482–1496, Aug. 2009.

[3] D.G. Koch *et al.*, “Kepler mission design, realized photometric performance, and early science,” *Astron. Astrophys.*, vol. 713, no. 2, pp. L79–86, Apr. 2010.

[4] J.–M. Alimi *et al.*, “First-ever full observable universe simulation,” in *2012 International Conference for High Performance Computing, Networking, Storage and Analysis (SC 2012)*. IEEE Conf., Salt Lake City, UT, USA: IEEE, doi: 10.1109/SC.2012.58.

[5] A. J. Kemball, R. M. Crutcher, R. Hasan, “A component-based framework for radio-astronomical imaging software systems,” *Software Pract. Exper.*, vol. 38, no. 5, pp. 493–507, Apr. 2008.

Levin, Deborah A. (1)

[1] S. Sawant, O. Tumuklu, R. Jambunathan, and D. A. Levin, “Application of adaptively refined unstructured grids in DSMC to shock wave simulations,” *Comput. Fluids*, vol. 170, pp. 197–212, 2018.

[2] S. Sawant, O. Tumuklu, V. Theofilis, and D. A. Levin, “Linear instability of shock-dominated laminar hypersonic separated flows,” presented at IUTAM Transition 2019, Imperial College, London, UK [Online]. Available: <https://youtu.be/cdQJFuk-ShvE>

[3] S. Sawant, O. Tumuklu, V. Theofilis, and D. A. Levin, “Study of spanwise perturbations in hypersonic shock-wave/boundary layer interactions on a double wedge,” AIAA Aviation 2019 Forum, Dallas, Tex., USA, Jun. 17–21, Paper 2012–3442, doi: 10.2514/6.2019-3442.

Levin, Deborah A. (2)

[1] R. Jambunathan and D. A. Levin, “CHAOS: An octree-based PIC–DSMC code for modeling of electron kinetic properties in a plasma plume using MPI–CUDA parallelization,” *J. Comput. Phys.*, vol. 373, pp. 571–604, 2018.

[2] R. Jambunathan and D. A. Levin, *IEEE Transactions*, in review, 2019.

[3] N. Nuwal and D. A. Levin, “Kinetic modeling of parasitic currents on spacecraft surfaces due to ambient space plasmas,” presented at the AIAA Propulsion and Energy 2019 Forum, AIAA Paper No. 2019–4147.

Menanteau, Felipe

[1] C. Chang *et al.*, “Dark Energy Survey Year 1 results: Curved-sky weak lensing mass map,” *Mon. Notices Royal Astron. Soc.*, vol. 475, no. 3, pp. 3165–3190, 2018.

Mösta, Philipp

[1] B. P. Abbott *et al.*, “Multi-messenger observations of a binary neutron star merger,” *Astrophys. J. Lett.*, vol. 848, no. 2, p. L12, 2017, doi: 10.3847/2041-8213/AA91C9.

[2] D. Radice, L. Rezzolla, and F. Galeazzi, “High-order fully general-relativistic hydrodynamics: new approaches and tests,” *Class. Quant. Grav.*, vol. 31, p. 075012, 2014, doi: 10.1088/0264-9381/31/7/075012.

[3] P. Mösta *et al.*, “GRHydro: A new open source general-relativistic magnetohydrodynamics code for the Einstein Toolkit,” *Class. Quant. Grav.*, vol. 31, p. 015005, 2014, doi: 10.1088/0264-9381/31/1/015005.

[4] F. Löffler *et al.*, “The Einstein Toolkit: A community computational infrastructure for relativistic astrophysics,” *Class. Quant. Grav.*, vol. 29, p. 115001, 2012, doi: 10.1088/0264-9381/29/11/115001.

Norman, Michael L.

[1] D. Tytler *et al.*, “The effect of large-scale power on simulated spectra of the Ly $\alpha$  forest,” *Mon. Notices Royal Astron. Soc.*, vol. 393, p. 723, 2009, doi: 10.1111/j.1365-2966.2008.14196.x.

[2] G. Bryan *et al.*, “Enzo: An adaptive mesh refinement code for astrophysics,” *Astrophys. J. Suppl.*, vol. 211, p. 19, 2014, doi: 10.1088/0067-0049/211/2/19.

[3] J. O. Bordner and M. L. Norman, “Computational cosmology and astrophysics on adaptive meshes using Charm++,” presented at the SC’18 PAW–ATM Workshop. [Online]. Available: <https://arxiv.org/abs/1810.01319>

[4] C. Burstedde, L. Wilcox, and O. Ghattas, “p4est: Scalable algorithms for parallel adaptive mesh refinement on forests of octrees,” *SIAM J. Sci. Comput.*, vol. 33, no. 3, pp. 1103–1133, 2011, doi: 10.1137/100791634.

O’Shea, Brian W.

[1] Bryan, G. *et al.*, “Enzo: An adaptive mesh refinement code for astrophysics,” *Astrophys. J.*, vol. 211, p. 19, 2014.

[2] C. Brummel-Smith *et al.*, “Enzo: An adaptive mesh refinement code for astrophysics (Version 2.6),” *J. Open Source Softw.*, doi: 10.21105/joss.01636.

[3] M. Turk *et al.*, “yt: A multi-code analysis toolkit for astrophysical simulation data,” *Astrophys. J. Suppl.*, vol. 192, p. 9, 2011.

Petravick, Donald

[1] B. Flaugher *et al.*, “The Dark Energy Camera,” *Astron. J.*, vol. 150, no. 5, p. 150, Nov. 2015.

[2] A. Dey *et al.*, “Overview of the DESI legacy surveys,” *Astrophys. J.*, vol. 157, no. 5, p. 168, May 2018.

[3] The Dark Energy Survey Collaboration, “The Dark Energy Survey,” Oct. 2005. [Online]. Available: <https://arxiv.org/abs/astro-ph/0510346>

[4] E. Schlafly *et al.*, “The DECam plane survey: Optical photometry of two billion objects in the southern galactic plane,” *Astrophys. J. Suppl.*, vol. 234, no. 2, p. 39, Feb. 2018.

[5] E. Morganson *et al.*, “The Dark Energy Survey image processing pipeline,” *Pub. Astr. Soc. Pac.*, vol. 130, no. 989, p. 074501, Jul. 2018.

[6] G. B. Bernstein *et al.*, “Instrumental response model and detrending for the Dark Energy Camera,” *Pub. Astr. Soc. Pac.*, vol. 129, no. 981, p. 114502, Nov. 2017.

Pogorelov, Nikolai

[1] D. J. McComas *et al.*, “Seven years of imaging the global heliosphere with IBEX,” *Astrophys. J. Suppl.*, vol. 229, no. 2, p. 41, 2017.

[2] D. J. McComas *et al.*, “Interstellar pickup ion observations to 38 AU,” *Astrophys. J. Suppl.*, vol. 233, p. 8, 2017.

[3] J. C. Kasper *et al.*, “Solar Wind Electrons Alphas and Protons (SWEAP) investigation: Design of the solar wind and coronal plasma instrument suite for solar probe plus,” *Space Sci. Rev.*, vol. 204, pp. 131–186, 2016.

[4] M. Zhang *et al.*, “Determination of plasma, pickup ion, and suprathermal particle spectrum in the solar wind frame of reference,” *Astrophys. J.*, vol. 871, p. 60, 2019.

[5] F. Fraternale, N. V. Pogorelov, J. D. Richardson, and D. Tordella, “Magnetic turbulence spectra and intermittency in the heliosheath and in the local interstellar medium,” *Astrophys. J.*, vol. 874, p. 40, 2019.

Quinn, Thomas

[1] L. V. Kale and S. Krishnan, “Charm++: Parallel programming with message-driven objects,” in *Parallel Programming using C++*, 1996, pp. 175–213.

[2] H. Menon *et al.*, “Adaptive techniques for clustered N-body cosmological simulations,” *Comput. Astrophys. Cosmol.*, vol. 2, 2015.

[3] M. Tremmel, F. Governato, M. Volonteri, and T. Quinn, “Off the beaten path: a new approach to realistically model the orbital decay of supermassive black holes in galaxy formation simulations,” *Mon. Notices Royal Astron. Soc.*, vol. 451, pp. 1868–1874, 2015.

[4] Tremmel, M. *et al.*, “The Romulus cosmological simulations: a physical approach to the formation, dynamics and accretion models of SMBHs,” *Mon. Notices Royal Astron. Soc.*, vol. 470, pp. 1121–1139, 2017.

[5] Anderson, L. *et al.*, “The little galaxies that could (reionize the universe): predicting faint end slopes and escape fractions at  $z > 4$ ,” *Mon. Notices Royal Astron. Soc.*, vol. 468, pp. 4077–4092, 2017.

Shapiro, Stuart L.

[1] B. Paczynski, “Gamma-ray bursters at cosmological distances,” *Astrophys. J.*, vol. 308, pp. L43, Sept. 1986.

[2] D. Eichler, M. Livio, T. Piran, and D. N. Schramm, “Nucleosynthesis, neutrino bursts, and gamma-rays from coalescing neutron stars,” *Nature*, vol. 340, pp. 126, Jul. 1989.

[3] R. Narayan, B. Paczynski, and T. Piran, “Gamma-ray bursts as the death throes of massive binary stars,” *Astrophys. J. Lett.*, vol. 395, pp. L83, Aug. 1992.

[4] V. Paschalidis, M. Ruiz, and S. L. Shapiro, “Relativistic simulations of black hole–neutron star coalescence: The jet emerges,” *Astrophys. J. Lett.*, vol. 806, pp. L14, Feb. 2015.

[5] B. D. Lackey and L. Wade, “Reconstructing the neutron-star equation of state with gravitational-wave detectors from a realistic population of inspiralling binary neutron stars,” *Phys. Rev.*, vol. D91, p. 043002, Feb. 2015.

[6] *GCN Circular*, May 5, 2019. [Online]. Available: <https://gc.gsfc.nasa.gov/gcn3/24411.gcn3>

[7] M. Ruiz, S. L. Shapiro, and A. Tsokaros, “Jet launching from binary black hole–neutron star mergers: Dependence on black hole spin, binary mass ratio, and magnetic field orientation,” *Phys. Rev.*, vol. D98, p. 123017, Dec. 2018.

[8] M. Ruiz, A. Tsokaros, V. Paschalidis, and S. L. Shapiro, “Effects of spin on magnetized binary neutron star mergers and jet launching,” *Phys. Rev.*, vol. D99, p. 084032, Mar. 2019.



[9] Z. Etienne, Y. Liu, and S. Shapiro, “Relativistic magnetohydrodynamics in dynamical spacetimes: A new adaptive mesh refinement implementation,” *Phys. Rev.*, vol. D82, p. 084031, Oct. 2010.

[10] M. Ruiz, R. Lang, V. Paschalidis, and S. L. Shapiro, “Binary neutron star mergers: A jet engine for short gamma-ray bursts,” *Astrophys. J. Lett.*, vol. 824, p. L6, Jun. 2016.

[11] P. Narayana *et al.*, “The 3rd Fermi GBM gamma-ray burst catalog: The first six years,” *Astrophys. J. Suppl.*, vol. 223, p. 28, Mar. 2016.

[12] R. Blandford and R. Znajek, “Electromagnetic extraction of energy from Kerr black holes,” *Mon. Not. Roy. Astron. Soc.*, vol. 179, p. 442, May 1977.

Tchekhovskoy, Alexander

[1] M. Liska *et al.*, Disc tearing and Bardeen-Petterson alignment in GRMHD simulations of highly tilted thing accretion discs, submitted to *Mon. Notices Royal Astron. Soc.*, 2019, arXiv:1904.08428.

[2] M. Liska, C. Hesp, A. Tchekhovskoy, A. Ingram, M. van der Klis, and S. Markoff, “Formation of precessing jets by tilted black hole discs in 3D general relativistic MHD simulations,” *Mon. Notices Royal Astron. Soc.*, vol. 474, p. L81, 2018.

Toth, Gabor

[1] G. Toth *et al.*, “Adaptive numerical algorithms in space weather modeling,” *J. Comp. Phys.*, vol. 231, pp. 870, 2012, doi: 10.1016/j.jcp.2011.02.006.

Turk, Matthew J.

[1] H.-Y. Schive, J. A. ZuHone, M. J. Goldbaum, M. J. Turk, M. Gaspari, C.-Y. Cheng, “GAMER-2: a GPU-accelerated adaptive mesh refinement code—accuracy, performance, and scalability,” *Mon. Notices Royal Astron. Soc.*, vol. 481, no. 4, pp. 4815–4840, Dec. 2018.

[2] B. Smith *et al.*, “GRACKLE: a chemistry and cooling library for astrophysics,” *Mon. Notices Royal Astron. Soc.*, vol. 466, pp. 2217–2234, Apr. 2017.

GEOSCIENCE

Corcoran, Jennifer

[1] M.-J. Noh and I. M. Howat, “Automated stereo-photogrammetric DEM generation at high latitudes: Surface Extraction with TIN-based Search-space Minimization (SETSM) validation and demonstration over glaciated regions,” *Glsci. Remote Sens.*, vol. 52, no. 2, pp. 198–217, 2015, doi: 10.1080/15481603.2015.1008621.

[2] B. Huberty, *Vegetation Surface Canopy Digital Elevation Models*, DOI Remote Sensing Activities 2017, U.S. Department of Interior. [Online]. Available: <https://eros.usgs.gov/doi-remote-sensing-activities/2017/fws/vegetation-surface-canopy-digital-elevation-models>

[3] B. Huberty *et al.*, “Wetland identification and change detection using a multi-sensor, multi-frequency remote sensing approach,” presented at Earth Observation Summit 2017, Montreal, Canada.

Diao, Chunyuan

[1] J. T. Morissette *et al.*, “A tamarisk habitat suitability map for the continental United States,” *Front. Ecol. Environ.*, vol. 4, no. 1, pp. 11–17, 2006.

[2] E. Zavaleta, “Valuing ecosystem services lost to Tamarix invasion in the United States,” in *Invasive Species in a Changing World*, H. A. Mooney and R. J. Hobbs, Eds. Washington, D.C., USA: Island Press, 2000, pp. 261–300.

[3] C. Diao and L. Wang, “Incorporating plant phenological trajectory in exotic saltcedar detection with monthly time series of Landsat imagery,” *Remote Sensing Environ.*, vol. 182, pp. 60–71, 2016.

[4] C. Diao and L. Wang, “Temporal partial unmixing of exotic salt cedar using Landsat time series,” *Remote Sensing Lett.*, vol. 7, no. 5, pp. 466–475, 2016.

[5] J. Silván-Cárdenas and L. Wang, “Retrieval of subpixel Tamarix canopy cover from Landsat data along the Forgotten River using linear and non-linear spectral mixture models,” *Remote Sensing Environ.*, vol. 114, no. 8, pp. 1777–1790, 2010.

[6] J. M. Friedman, J. E. Roelle, and B. S. Cade, “Genetic and environmental influences on leaf phenology and cold hardiness of native and introduced riparian trees,” *Int. J. Biometeorol.*, vol. 55, no. 6, pp. 775–787, 2011.

[7] J. M. Di Tomaso, “Impact, biology, and ecology of saltcedar (*Tamarix spp.*) in the Southwestern United States,” *Weed Technol.*, vol. 12, no. 2, pp. 326–336, 1998.

Di Girolamo, Larry

[1] National Research Council, *Earth Science and Applications from Space: National Imperatives for the Next Decade and Beyond*. Washington, D.C., USA: Nat. Acad. Press, 2007.

[2] R. Wolfe and H. Ramapriyan, “Scaling the pipe: NASA EOS Terra data systems at 10,” in *2010 IEEE Int. Geoscience and Remote Sens. Symp.*, pp. 1300–1303.

[3] G. Zhao, L. Di Girolamo, D. J. Diner, C. J. Bruegge, K. J. Mueller, and D. L. Wu, “Regional changes in Earth’s color and texture as observed from space over a 15-year period,” *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 7, pp. 4240–4249, Jul. 2016.

Dominguez, Francina

[1] D. P. Dee *et al.*, “The ERA-interim reanalysis: Configuration and performance of the data assimilation system,” *Q. J. Royal Meteorol. Soc.*, vol. 137, pp. 553–597, 2011, doi: 10.1002/qj.828.

[2] INPE—Instituto Nacional de Pesquisas Espaciais Projeto Prodes, Projeto Prodes: Monitoramento da Floresta Amazônica brasileira por satélite. [Online]. Available: <http://www.obt.inpe.br/prodes/index.php>, Accessed on: Jun. 24, 2019.

[3] D. Insua-Costa, and G. Miguez-Macho, “A new moisture tagging capability in the Weather Research and Forecasting model: Formulation, validation and application to the 2014 Great Lake-effect snowstorm,” *Earth Syst. Dyn.*, vol. 9, pp. 167–185, 2018, doi: 10.5194/esd-9-167-2018.

[4] J. A. Martinez, F. Dominguez, and G. Miguez-Macho, “Impacts of a groundwater scheme on hydroclimatological conditions over southern South America,” *J. Hydrometeorol.*, vol. 17, pp. 2959–2978, 2016, doi: 10.1175/JHM-D-16-0052.1.

[5] G. A. Miguez-Macho, A. Rios-Entenza, and F. Dominguez, “The Impact of soil moisture and evapotranspiration fluxes on the spring water cycle in the Iberian Peninsula: A study with moisture tracers in WRF,” AGU Fall Meeting Abstracts, 2013.

[6] B. Soares *et al.*, “Modelling conservation in the Amazon basin,” *Nature*, vol. 440, pp. 520–523, 2006, doi: 10.1038/nature04389.

[7] R. J. van der Ent, H. H. G. Savenije, B. Schaeffli, and S. C. Steele-Dunne, “Origin and fate of atmospheric moisture over continents,” *Water Resour. Res.*, vol. 46, p. W09525, 2010, doi: 10.1029/2010WR009127.

[8] W. Weng, M. K. Lüdeke, D. C. Zemp, T. Lakes, and J. P. Kropp, “Aerial and surface rivers: Downwind impacts on water availability from land use changes in Amazoniam,” *Hydrol. Earth Syst. Sci.*, vol. 22, no. 1, pp. 911–927, 2018, doi: 10.5194/hess-22-911-2018.

[9] D. C. Zemp *et al.*, “On the importance of cascading moisture recycling in South America,” *Atmos. Chem. Phys.*, vol. 14, pp. 13337–13359, 2014, doi: 10.5194/acp-14-13337-2014.

[10] H. R. da Rocha *et al.*, “Patterns of water and heat flux across a biome gradient from tropical forest to savanna in Brazil,” *J. Geophys. Res.*, vol. 114, p. G00B12, 2009, doi: 10.1029/2007JG000640.

Guan, Kaiyu (1)

[1] Y. Ryu *et al.*, “Integration of MODIS land and atmosphere products with a coupled-process model to estimate gross primary productivity and evapotranspiration from 1 km to global scales,” *Global Biogeochem. Cycles*, vol. 25, 2011.

[2] C. Jiang and Y. Ryu, “Multi-scale evaluation of global gross primary productivity and evapotranspiration products derived from Breathing Earth System Simulator (BESS),” *Rem. Sensing Environ.*, vol. 186, pp. 528–547, 2016.

[3] Y. Luo, K. Guan, and J. Peng, “STAIR: A generic and fully-automated method to fuse multiple sources of optical satellite data to generate a high-resolution, daily and cloud-/gap-free surface reflectance product,” *Rem. Sensing Environ.*, vol. 214, pp. 87–99, 2018.

Guan, Kaiyu (2)

[1] C. Lesk, P. Rowhani, and N. Ramankutty, “Influence of extreme weather disasters on global crop production,” *Nature*, vol. 529, pp. 84–87, 2016, doi: 10.1038/nature16467.

[2] D. B. Lobell *et al.*, “The critical role of extreme heat for maize production in the United States,” *Nat. Clim. Change*, vol. 3, no. 5, pp. 497–501, 2013, doi: 10.1038/nclimate1832.

[3] C. Rosenzweig *et al.*, “Assessing agricultural risks of climate change in the 21st century in a global gridded crop model intercomparison,” *Proc. Natl. Acad. Sci. USA*, vol. 111, no. 9, pp. 3268–3273, 2014, doi: 10.1073/pnas.1222463110.

[4] W. Schlenker and M. J. Roberts, “Nonlinear temperature effects indicate severe damages to U.S. crop yields under climate change,” *Proc. Natl. Acad. Sci. USA*, vol. 106, no. 37, pp. 15,594–15,598, 2009, doi: 10.1073/pnas.0906865106.

[5] Y. L. Everingham *et al.*, “Enhanced risk management and decision-making capability across the sugarcane industry value chain based on seasonal climate forecasts,” *Agric. Syst.*, vol. 74, no. 3, pp. 459–477, 2002, doi: 10.1016/S0308-521X(02)00050-1.

[6] J. W. Hansen and M. Indeje, “Linking dynamic seasonal climate forecasts with crop simulation for maize yield prediction in semi-arid Kenya,” *Agric. Forest Meteorol.*, vol. 125, nos. 1–2, pp. 143–157, 2004, doi: 10.1016/j.agrformet.2004.02.006.

[7] O. Isengildina-Massa, S. H. Irwin, D. L. Good, and J. K. Gomez, “The impact of situation and outlook information in corn and soybean futures markets: Evidence from WASDE reports,” *J. Agric. Applied Econ.*, vol. 40, no. 01, pp. 89–103, 2008, doi: 10.1017/S1074070800027991.

[8] B. P. Kirtman *et al.*, “The North American multi-model ensemble: phase-1 seasonal-to-interannual prediction; phase-2 toward developing intraseasonal prediction,” *Bull. Am. Meteorol. Soc.*, vol. 95, pp. 585–601, 2014.

Lasher-Trapp, Sonia

[1] G. H. Bryan and H. Morrison, “Sensitivity of a simulated squall line to horizontal resolution and parameterization of microphysics,” *Mon. Wea. Rev.*, vol. 140, pp. 202–225, 2012.

[2] E. R. Mansell, C. L. Ziegler, and E. C. Bruning, “Simulated electrification of a small thunderstorm with two-moment bulk microphysics,” *J. Atmos. Sci.*, vol. 67, pp. 171–194, 2010.

[3] D. H. Moser and S. Lasher-Trapp, “The influence of successive thermals on entrainment and dilution in a simulated cumulus congestus,” *J. Atmos. Sci.*, vol. 74, pp. 375–392, 2017.



Morin, Paul

[1] M.–J. Noh and I. M. Howat, “Automated stereo-photogrammetric DEM generation at high latitudes: Surface Extraction with TIN-based Search-space Minimization (SETSM) validation and demonstration over glaciated regions,” *GISci. Rem. Sensing*, vol. 52, no. 2, pp. 198–217, 2015.

[2] M.–J. Noh and I. M. Howat, “The Surface Extraction from TIN based Search-space Minimization (SETSM) algorithm,” *ISPRS J. Photogramm. Remote Sens.*, vol. 129, pp. 55–76, 2017.

Nesbitt, Stephen W.

[1] G. H. Bryan and J. M. Fritsch, “A benchmark simulation for moist nonhydrostatic numerical models,” *Mon. Weather Rev.*, vol. 130, no. 1992, pp. 2917–2928, 2002.

Pavlis, Nikolaos

[1] A. Jackson, A. R. T. Jonkers, and M. R. Walker, “Four centuries of geomagnetic secular variation from historical records,” *Philos. Trans. Royal Soc. A*, vol. 358, no. 1768, 2000, doi: 10.1098/rsta.2000.0569.

[2] W. Jiang and W. Kuang, “An MPI based MoSST core dynamics model,” *Phys. Earth Planet. Inter.*, vol. 170, nos. 1–2, 2008, doi: 10.1016/j.pepi.2008.07.020.

[3] M. Korte, C. Constable, F. Donadini, and R. Holme, “Reconstructing the Holocene geomagnetic field,” *Earth Planet. Sci. Lett.*, vol. 312, pp. 497–505, 2011.

[4] W. Kuang *et al.*, “MoSST\_DAS: The first generation geomagnetic data assimilation framework,” *Commun. Comp. Phys.*, vol. 3, pp. 85–108, 2008.

[5] W. Kuang and B. F. Chao, “Geodynamo modeling and core–mantle interactions,” in *Earth’s Core: Dynamics, Structure, Rotation*, V. Dehant, K. C. Creager, S.–I. Ichiro, K. S. Zatman, Eds. Geodynamics Series, vol. 31, AGU Monograph, Washington, D.C., USA, 2003, pp. 193–212.

[6] W. Kuang and J. Bloxham, “Numerical modeling of magnetohydrodynamic convection in a rapidly rotating spherical shell: weak and strong field dynamo action,” *J. Comp. Phys.*, vol. 153, pp. 51–81, 1999.

[7] N. Olsen, M. Manda, T. J. Sabaka, and L. Toffner–Clausen, “The CHAOS-3 geomagnetic field model and candidates for the 11th generation IGRF,” *Earth Planets Space*, vol. 62, no. 10, 2010.

[8] T. J. Sabaka, N. Olsen, and M. E. Purucker, “Extending comprehensive models of the Earth’s magnetic field with Orsted and CHAMP data,” *Geophys. J. Int.*, vol. 159, pp. 521–547, 2004.

[9] Z. Sun and W. Kuang, “An ensemble algorithm based component for geomagnetic data assimilation,” *Terr. Atmos. Ocean Sci.*, 2015, doi: 10.3319/TAO.2014.08.19.05.

[10] Z. Sun, A. Tangborn, and W. Kuang, “Data assimilation in a sparsely observed one-dimensional modeled MHD system,” *Nonlin. Processes Geophys.*, vol. 14, pp. 181–192, 2007.

[11] A. Tangborn and W. Kuang, “Impact of archeomagnetic field model data on modern era geomagnetic forecasts,” *Phys. Earth Planet. Inter.*, 2017, doi: 10.1016/j.pepi.2017.11.002.

[12] A. Tangborn and W. Kuang, “Geodynamo model and error parameter estimation using geomagnetic data assimilation,” *Geophys. J. Int.*, 2015, doi: 10.1093/gji/ggu409.

Riemer, Nicole

[1] T. F. Stocker *et al.*, “Technical summary—Climate Change 2013: The Physical Science Basis” in *Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge, UK: Cambridge Univ. Press, 2013, pp. 33–115, 2013.

[2] N. Riemer and M. West, “Quantifying aerosol mixing state with entropy and diversity measures,” *Atmos. Chem. Phys.*, vol. 13, pp. 11423–11439, 2013, doi: 10.5194/acp-13-11423-2013.

Tabor, Clay

[1] R. Seager *et al.*, “Model projections of an imminent transition to a more arid climate in southwestern North America,” *Science*, vol. 316, no. 5828, pp. 1181–1184, 2007.

[2] N. S. Diffenbaugh, D. L. Swain, and D. Touma, “Anthropogenic warming has increased drought risk in California,” *Proc. Nat. Acad. Sci. USA*, vol. 112, no. 13, pp. 3931–3936, 2015.

[3] J. S. Munroe and B. J. Laabs, “Temporal correspondence between pluvial lake highstands in the Southwestern US and Heinrich Event 1,” *J. Quat. Sci.*, vol. 28, no. 1, pp. 49–58, 2013.

[4] J. L. Oster, D. E. Ibarra, M. J. Winnick, and K. Maher, “Steering of westerly storms over western North America at the last glacial maximum,” *Nat. Geosci.*, vol. 8, no. 3, p. 201, 2015.

[5] J. W. Hurrell *et al.*, “The Community Earth System Model: A framework for collaborative research,” *Bull. Am. Meteorol. Soc.*, vol. 94, no. 9, pp. 1339–1360, 2013.

[6] E. Brady *et al.*, “The connected isotopic water cycle in the Community Earth System Model version 1,” *J. Adv. Modeling Earth Systems*, 2019, doi: 10.1029/2019MS001663.

[7] C. R. Tabor *et al.*, “Interpreting precession-driven  $\delta^{18}\text{O}$  variability in the South Asian monsoon region,” *J. Geophys. Res.: Atmos.*, vol. 123, no. 11, pp. 5927–5946, 2018.

[8] C. Members, “Climatic changes of the last 18,000 years: Observations and model simulations,” *Science*, vol. 241, no. 4869, pp. 1043–1052, 1988.

[9] M. Löfverström and J. Liakka, “On the limited ice intrusion in Alaska at the LGM,” *Geophys. Res. Lett.*, vol. 43, no. 20, pp. 11–030–11–038, 2016.

[10] M. Lyle *et al.*, “Out of the tropics: the Pacific, Great Basin Lakes, and Late Pleistocene water cycle in the western United States,” *Science*, vol. 337, no. 6102, pp. 1629–1633, 2012.

[11] J. M. Lora, J. L. Mitchell, C. Risi, and A. E. Tripati, “North Pacific atmospheric rivers and their influence on western North America at the Last Glacial Maximum,” *Geophys. Res. Lett.*, vol. 44, no. 2, pp. 1051–1059, 2017.

[12] M. Kageyama *et al.*, “The PMIP4 contribution to CMIP6—Part 4: Scientific objectives and experimental design of the PMIP4-CMIP6 Last Glacial Maximum experiments and PMIP4 sensitivity experiments,” *Geosci. Model Dev.*, vol. 10, no. 11, pp. 4035–4055, 2017.

Vidale, John

[1] R. Graves *et al.*, “CyberShake: A physics-based seismic hazard model for Southern California,” *Pure Appl. Geophys.*, vol. 168, no. 3, pp. 367–381, 2011.

[2] M. P. Moschetti *et al.*, “Integrate urban-scale seismic hazard analyses with the U.S. National Seismic Hazard Model,” *Seismol. Res. Lett.*, vol. 89, no. 3, pp. 967–970, 2018, doi: 10.1785/0220170261.

[3] C. Crouse, T. H. Jordan, K. R. Milner, C. A. Goulet, S. Callaghan, and R. W. Graves, “Site-specific MCER response spectra for Los Angeles Region based on 3D numerical simulations and the NGA West2 equations,” presented at the 11th Nat. Conf. Earthquake Eng., Los Angeles, CA, USA, Jun. 25–29, 2018, Paper 518.

[4] C. A. Cornell, “Engineering seismic risk analysis,” *Bull. Seismol. Soc. Am.*, vol. 58, pp. 1583–1606, 1968.

[5] M. D. Petersen *et al.*, “Documentation for the 2014 update of the United States national seismic hazard maps,” U.S. Geological Survey Open-File Rep. 2014–1091, 2014, doi: 10.3133/ofr20141091.

[6] E. H. Field, T. H. Jordan, and C. A. Cornell, “OpenSHA: A developing community-modeling environment for seismic hazard analysis,” *Seismol. Res. Lett.*, vol. 74, no. 4, pp. 406–419, 2003.

[7] P. Small *et al.*, “The SCEC Unified Community Velocity Model software framework,” *Seismolog. Res. Lett.*, vol. 88, no. 5, 2017, doi: 10.1785/0220170082.

[8] T. Tu, H. Yu, L. Ramírez–Guzmán, J. Bielak, O. Ghattas, K.–L. Ma, and D. R. O’Hallaron, “From mesh generation to scientific visualization: an end-to-end approach to parallel supercomputing,” in *Proc. ACM/IEEE Int. Conf. for High Perf. Comput., Networking, Storage and Analysis*, IEEE Computer Society, 2006.

[9] Y. Cui *et al.*, “Scalable earthquake simulation on petascale supercomputers,” in *Proc. ACM/IEEE Int. Conf. for High Perf. Comput., Networking, Storage and Analysis*, New Orleans, LA, USA, Nov. 13–19, 2010, doi: 10.1109/SC.2010.45.

[10] D. Thain, T. Tannenbaum, and M. Livny, “Distributed computing in practice: The Condor experience,” *Concur. Comput. Pract. Exp.*, vol. 17, no. 2–4, pages 323–356, Feb.–Apr., 2005.

[11] E. Deelman *et al.*, “Pegasus: A workflow management system for science automation,” *Future Gen. Comput. Syst.*, vol. 46, p. 17–35, 2015.

[12] D. Roten, K. B. Olsen, S. M. Day, and Y. Cui, (2017), “Quantification of fault-zone plasticity effects with spontaneous rupture simulations,” *Pure Appl. Geophys.*, pp. 1–23, 2017, doi: 10.1007/s00024-017-1466-5.

[13] Y. Cui, and SCEC Collaboration, “Regional scale earthquake simulations on OLCF Titan and NCSA Blue Waters,” presented at Int. Workshop Persp. GPU Comp. Sci., Sapienz Università di Roma, Sept. 26–28, 2016.

West, Matthew

[1] D. Dockery and A. Pope, “Epidemiology of acute health effects: summary of time-series studies,” in *Particles in Our Air: Concentrations and Health Effects*, Cambridge, MA, USA: Harvard Univ. Physics Dept., 1996, pp. 123–147.

[2] T. F. Stocker *et al.*, “Technical summary—Climate Change 2013: The Physical Science Basis” in *Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge, UK: Cambridge Univ. Press, 2013, pp. 33–115, 2013.

[3] N. Riemer, M. West, R. A. Zaveri, and R. C. Easter, “Simulating the evolution of soot mixing state with a particle-resolved aerosol model,” *J. Geophys. Res.: Atmos.*, vol. 114, no. D9, 2009, doi: 10.1029/2008JD011073.

[4] W. Skamarock *et al.*, “A description of the advanced research WRF version 3,” *NCAR Tech. Note*, NCAR/TN-475+STR, 2008.

[5] EPA, “Preparation of Emissions Inventories for the Version 6.3, 2011 Emissions Modeling Platform Technical Support Document,” 2016.

[6] N. Riemer and M. West, “Quantifying aerosol mixing state with entropy and diversity measures,” *Atmos. Chem. Phys.*, vol. 13, pp. 11423–11439, 2013, doi: 10.5194/acp-13-11423-2013.

Wuebbles, Donald J.

[1] J. F. Lamarque *et al.*, “CAM-chem: Description and evaluation of interactive atmospheric chemistry in the Community Earth System Model,” *Geosci. Model Dev.*, vol. 5, no. 2, p.369, 2012.

[2] S. Tilmes *et al.*, “Description and evaluation of tropospheric chemistry and aerosols in the Community Earth System Model (CESM1. 2),” *Geosci. Model Dev.*, vol. 8, pp.1395–1426, 2015.

Zhang, Xiangdong

[1] D. P. Dee *et al.*, “The ERA-interim reanalysis: Configuration and performance of the data assimilation system,” *Quart. J. Roy. Meteor. Soc.*, vol. 137, pp. 553–597, 2011, doi: 10.1002/qj.828.

[2] M. Steele, R. Morley, and W. Ermold, “PHC: A global ocean hydrography with a high quality Arctic Ocean,” *J. Clim.*, vol. 14, pp. 2079–2087, 2001.



[3] X. Zhang, J. E. Walsh, J. Zhang, U. S. Bhatt, and M. Ikeda, "Climatology and interannual variability of Arctic cyclone activity: 1948–2002," *J. Clim.*, vol. 17, pp. 2300–2317, 2004.

PHYSICS & ENGINEERING

Aluru, Narayana R. (1)

[1] K. D. Kreuer, A. Fuchs, and J. Maier, "H/D isotope effect of proton conductivity and proton conduction mechanism in oxides," *Solid State Ion.*, vol. 77, pp. 157–162, 1995.

[2] W. Münch, K. D. Kreuer, G. Seifertli, and J. Majer, "A quantum molecular dynamics study of proton diffusion in SrTiO<sub>3</sub> and CaTiO<sub>3</sub>," *Solid State Ion.*, vol. 125, pp. 38–45, 1999.

[3] Y. Yamazaki, F. Blanc, Y. Okuyama, L. Buannic, J. C. Lucio–Vega, C. P. Grey, and S. M. Haile, "Proton trapping in yttrium-doped barium zirconate," *Nat. Mater.*, vol. 12, pp. 647–651, 2013.

[4] B. Merinov and W. Goddard III, "Proton diffusion pathways and rates in Y-doped BaZrO<sub>3</sub> solid oxide electrolyte from quantum mechanics," *J. Chem. Phys.*, vol. 130, p. 194707, 2009.

[5] A. Braun and Q. Chen, "Experimental neutron scattering evidence for proton polaron in hydrated metal oxide proton conductors," *Nat. Commun.*, vol. 8, p. 15830, 2017.

[6] Y. Jing, H. Matsumoto, and N. R. Aluru, "Mechanistic insights into hydration of solid oxides," *Chem. Mater.*, vol. 30, pp. 138–144, 2018.

[7] G. Kresse and J. Furthmüller, "Efficiency of *ab-initio* total energy calculations for metals and semiconductors using a plane-wave basis set," *Comput. Mater. Sci.*, vol. 6, pp. 15–50, 1996.

[8] G. Kresse and D. Joubert, "From ultrasoft pseudopotentials to the projector augmented-wave method," *Phys. Rev. B: Condens. Matter Mater. Phys.*, vol. 59, p. 1758, 1999.

[9] G. Kresse and J. Furthmüller, "Efficient iterative schemes for *ab initio* total-energy calculations using a plane-wave basis set," *Phys. Rev. B: Condens. Matter Mater. Phys.*, vol. 54, p. 11169, 1996.

[10] J. P. Perdew, K. Burke, and M. Ernzerhof, "Generalized gradient approximation made simple," *Phys. Rev. Lett.*, vol. 77, p. 3865, 1996.

[11] G. Henkelman, B. P. Uberuaga, and H. Jónsson, "A climbing image nudged elastic band method for finding saddle points and minimum energy paths," *J. Chem. Phys.*, vol. 113, pp. 9901–9904, 2000.

Aluru, Narayana R. (2)

[1] D. B. Wells *et al.*, "Assessing graphene nanopores for sequencing DNA," *Nano Lett.*, vol.12, no. 8, pp. 4117–4123, Jul. 2012.

[2] D. Branton *et al.*, "The potential and challenges of nanopore sequencing," *Nat. Biotech.*, vol. 26, pp. 1146–1153, Oct. 2008.

Aluru, Narayana R. (3)

[1] A. Hospital, J. R. Goñi, M. Orozco, and J. L. Gelpi, "Molecular dynamics simulations: Advances and applications," *Adv. Appl. Bioinform. Chem.*, vol. 8, Nov. 2015.

[2] C. G. Mayne, J. Saam, K. Schulten, E. Tajkhorshid, and J. C. Gumbart, "Rapid parameterization of small molecules using the Force Field Toolkit," *J. Comput. Chem.*, vol. 34, no. 34, Nov. 2013.

[3] R. L. Henderson, "A uniqueness theorem for fluid pair correlation functions," *Phys. Lett. A*, vol. 49, no. 34, Sept. 1974.

[4] J. Byers, "The physics of data," *Nat. Phys.*, vol. 3, Jul. 2017.

Araya, Guillermo

[1] G. Araya, L. Castillo, C. Meneveau, and K. Jansen, "A dynamic multi-scale approach for turbulent inflow boundary conditions in spatially evolving flows," *J. Fluid Mech.*, vol. 670, pp. 518–605, 2011.

[2] T. Lund, X. Wu, and K. Squires, "Generation of turbulent inflow data for spatially-developing boundary layer simulations," *J. Comput. Phys.*, vol. 140, no. 2, pp. 233–258, 1998.

[3] C. H. Whiting, K. E. Jansen, and S. Dey, "Hierarchical basis in stabilized finite element methods for compressible flows," *Comp. Meth. Appl. Mech. Engng.*, vol. 192, no. 47–48, pp. 5167–5185, 2003.

[4] K. E. Jansen, "A stabilized finite element method for computing turbulence," *Comp. Meth. Appl. Mech. Engng.*, vol. 174, pp. 299–317, 1999.

[5] J. C. R. Hunt, A. A. Wray, and P. Moin, "Eddies, streams, and convergence zones in turbulent flows," in *Studying Turbulence Using Numerical Simulation Databases, 2: Proceedings of the 1988 Summer Program*, pp. 193–208, 1988.

Bodony, Daniel J.

[1] M. Diener, S. White, L. V. Kale, M. Campbell, D. J. Bodony, and J. B. Freund, "Improving the memory access locality of hybrid MPI applications," presented at the EuroMPI '17 Proc. 24th European MPI Users' Group Meeting.

[2] M. Diener, D. J. Bodony, and L. Kale, "Accelerating scientific applications on heterogeneous systems with HybridOMP," *Lecture Notes in Computer Science: High Performance Computing for Computational Science—VECPAR 2018*, vol. 11333, 2019.

[3] M. Diener, D. J. Bodony, and L. Kale, "Heterogeneous Computing with OpenMP and Hydra," accepted for publication in *Concurrency Comput.: Practice Exp.*, 2019.

Brehm, Christoph

[1] M. Murakami and K. Kikuyama, "Turbulent flow in axially rotating pipes," *J. of Fluids Eng.*, vol. 102, no. 1, pp. 97–103, 1980.

[2] K. Kikuyama, M. Murakami, K. Nishibori, and K. Maeda, "Flow in an axially rotating pipe: A calculation of flow in the saturated region," *Bull. JSME*, vol. 26, no. 214, pp. 506–513, 1983.

[3] K.Nishibori, K. Kikuyama, and M. Murakami, "Laminarization of turbulent flow in the inlet region of an axially rotating pipe," *JSME Inter. J.*, vol. 30, no. 260, pp. 255–262, 1987.

[4] G. Reich and H. Beer, "Fluid flow and heat transfer in an axially rotating pipe I. Effect of rotation on turbulent pipe flow," *Inter. J. Heat Mass Trans.*, vol. 32, no. 3, pp. 551–562, 1989.

[5] A. Feiz, M. Ould-Rouis, and G. Lauriat, "Turbulence statistics in a fully developed rotating pipe flow," *J. Enhanced Heat Trans.*, vol. 12, no. 3, 2005.

[6] P. F. Fischer, "An overlapping Schwarz Method for spectral element solution of the incompressible Navier–Stokes equations," *J. Comput. Physics*, vol. 133, pp. 84–101, 1997.

[7] G. K. El Khoury, P. Schlatter, A. Noorani, P. F. Fischer, G. Brethouwer, and A. V. Johansson, "Direct numerical simulation of turbulent pipe flow at moderately high Reynolds numbers," *Flow, Turbul. Combust.*, vol. 91, no. 3, pp. 475–495, 2013.

[8] N. Ashton, J. Davis, and C. Brehm, "Assessment of the elliptic blending Reynolds stress model for a rotating turbulent pipe flow using new DNS data," AIAA Aviat. Aeronaut. Forum and Expo., Dallas, TX, USA, Jun. 17–21, 2019, AIAA Paper 2019.

[9] P. Orlandi and M. Fatica, "Direct simulations of turbulent flow in a pipe rotating about its axis," *J. Fluid Mech.*, vol. 343, pp. 43–72, 1997.

[10] P. Orlandi and D. Ebstein, "Turbulent budgets in rotating pipes by DNS," *Inter. J. Heat Fluid Flow*, vol. 21, no. 5, pp. 499–505, 2000.

[11] P. Orlandi and M. Fatica, "Direct simulations of turbulent flow in a pipe rotating about its axis," *Oceanographic Lit. Rev.*, vol. 7, no. 45, p. 1257, 1998.

[12] C. Brehm, J. Davis, S. Ganju, and S. Bailey, "A numerical investigation of the effects of rotation on turbulent pipe flows," 11th Inter. Symp. Turbulence and Shear Flow Phenomena (TSFP11), Southampton, UK, Jul. 30–Aug. 2, 2019.

[13] J. Davis, S. Ganju, N. Ashton, S. Bailey, and C. Brehm, "A DNS study to investigate turbulence suppression in rotating pipe flows," AIAA Aviat. Aeronaut. Forum and Expo. Dallas, TX, USA, Jun. 17–21, 2019, AIAA Paper 2019.

Browne, Oliver M. F.

[1] O. M. F. Browne, A. P. Haas, H. F. Fasel, and C. Brehm, "An efficient linear wavepacket tracking method for hypersonic boundary-layer stability prediction," *J. Comput. Phys.*, vol. 380, pp. 243–268, 2019.

[2] O. M. F. Browne, A. P. Haas, H. F. Fasel, and C. Brehm, "An efficient strategy for computing wave-packets in high-speed boundary layers," presented at the AIAA Conf., Denver, CO, USA, 2017

[3] O. M. F. Browne, S. M. A. Al Hasnine, and C. Brehm, "Towards modeling and simulation of particulate interactions with high-speed transitional boundary-layer flows," presented at the AIAA Conf., Dallas, TX, USA, 2019.

[4] O. M. F. Browne, A. P. Haas, H. F. Fasel, and C. Brehm, "An efficient nonlinear wavepacket tracking method for hypersonic boundary-layer flows," under consideration, 2019.

[5] P. MacNeice, K. M. Olson, C. Mobarry, R. de Fainchtein, and C. Packer, "PARAMESH: A parallel adaptive mesh refinement community toolkit," *Comput. Phys. Commun.*, vol. 126, no. 3, pp. 330–354, 2000.

Ceperley, David M.

[1] P. M. Celliers *et al.*, *Science*, vol. 361, p. 677, 2018.

[2] K. Chang, "Settling arguments about hydrogen with 168 giant lasers," The New York Times, Aug. 16, 2018. [Online]. Available: <https://www.nytimes.com/2018/08/16/science/metallic-hydrogen-lasers.html>, Accessed on: Jun. 20, 2019.

Chew, Huck Beng

[1] S. Bagchi, C. Ke, and H. B. Chew, "Oxidation effect on the shear strength of graphene on aluminum and titanium surfaces," *Phys. Rev. B*, vol. 98, Art. no. 174106, 2018.

[2] W. Qu, S. Bagchi, X. Chen, H. B. Chew, and C. Ke, "Bending and interlayer shear moduli of ultrathin boron nitride nanosheet," *J. Phys. D: Appl. Phys.*, in press, 2019.

Clark, Bryan

[1] D. Luo and B. K. Clark, "Backflow transformations via neural networks for quantum many-body wave functions," *Phys. Rev. Lett.*, vol. 122, no. 22, p. 226401, 2019.

[2] R. P. Feynman and M. Cohen, "Energy spectrum of the excitations in liquid helium," *Phys. Rev.*, vol. 102, p. 1189, 1956.

[3] L. Savary and L. Balents, "Quantum spin liquids: a review," *Rep. Prog. Phys.*, vol. 80, no. 1, p. 016502, 2016.

[4] D. A. Huse, R. Nandkishore, and V. Oganesyan, "Phenomenology of fully many-body-localized systems," *Phys. Rev. B*, vol. 90, no. 17, p. 174202, 2014.

[5] D. Reig-i-Plessis *et al.*, "Deviation from the dipole-ice model in the spinel spin-ice candidate MgEr<sub>2</sub>Se<sub>6</sub>," *Phys. Rev. B*, vol. 99, p. 134438, 2019.

[6] D. Luo and B. K. Clark, "Backflow transformations via neural networks for quantum many-body wave functions," *Phys. Rev. Lett.*, vol. 122, no. 22, p. 226401, 2019.

[7] G. Carleo and M. Troyer, "Solving the quantum many-body problem with artificial neural networks," *Science*, vol. 355, no. 6325, p. 602–606, 2017.



[8] E. Chertkov and B. K. Clark, “Computational inverse method for constructing spaces of quantum models from wave functions,” *Phys. Rev. X*, vol. 8, no. 3, p. 031029, 2018.

[9] Z. Zhu and S. R. White, “Spin liquid phase of the  $S=1/2$  Heisenberg model on the triangular lattice,” *Phys. Rev. B*, vol. 92, no. 4, p. 041105, 2015.

[10] X. Yu, D. Luo, and B.K. Clark. “Beyond many-body localized states in a spin-disordered Hubbard model with pseudo-spin symmetry,” *Phys. Rev. B*, vol. 98, no. 11, p. 115106, 2018.

Duan, Lian

[1] G. S. Jiang and C. W. Shu, “Efficient Implementation of Weighted ENO schemes,” *J. Comput. Phys.*, vol. 126, no. 1, pp. 202–228, 1996.

[2] E. M. Taylor, M. Wu, and M. P. Martín, “Optimization of nonlinear error sources for weighted non-oscillatory methods in direct numerical simulations of compressible turbulence,” *J. Comput. Phys.*, vol. 223, no. 1, pp. 384–397, 2006.

[3] J. Williamson, “Low-storage Runge–Kutta schemes,” *J. Comput. Phys.*, vol. 35, no. 1, pp. 48–56, 1980.

[4] S. Xu and M. P. Martín, “Assessment of inflow boundary conditions for compressible turbulent boundary layers,” *Phys. Fluids*, vol. 16, no. 7, pp. 2623–2639, 2004.

[5] E. Toubert and N. D. Sandham, “Oblique shock impinging on a turbulent boundary layer: Low-frequency mechanisms,” presented at 38th Fluid Dynamics Conf. and Exhibit, Seattle, WA, Jun. 23–26, 2008, AIAA Paper 2008–4170.

El-Khadra, Aida X.

[1] G. W. Bennett *et al.*, “Final Report of the muon E821 anomalous magnetic moment measurement at BNL,” *Phys. Rev. D*, vol. 73, p. 072003, 2006.

[2] J. Grange *et al.*, “Muon (g-2) Technical Design Report,” FERMILAB-FN-0992-E, FERMILAB-DE-SIGN-2014-02, p. 666, arXiv:1501.06858.

[3] T. Blum, *et al.*, “The muon (g-2) theory value: Present and future,” arXiv:1311.2198.

[4] C. T. H. Davies *et al.*, “Hadronic-vacuum-polarization contribution to the muon’s anomalous magnetic moment from four-flavor lattice QCD,” arXiv:1902.04223.

[5] C. DeTar. *The MILC code*. [Online]. Available: [http://www.physics.utah.edu/~detar/milc/milc\\_qcd.html](http://www.physics.utah.edu/~detar/milc/milc_qcd.html)

[6] *USQCD Collaboration software*. [Online]. Available: <http://usqcd-software.github.io/>

[7] M. A. Clark, R. Babich, K. Barros, R. C. Brower, and C. Rebbi, “Solving lattice QCD systems of equations using mixed precision solvers on GPUs,” *Comput. Phys. Commun.*, vol. 181, p. 1517, 2010.

[8] *QUDA software library*. [Online]. Available: <https://github.com/lattice/quda>

Ertekin, Elif

[1] U.S. Energy Information Agency, *Annual Energy Outlook 2018*.

[2] J. Yan *et al.*, “Materials descriptors for predicting thermoelectric performance,” *Energy Environ. Sci.*, vol. 8, pp. 983–994, 2015.

[3] TEDesignLab. [Online]. Available: <http://www.tedesignlab.org>

[4] B. R. Ortiz *et al.*, “Carrier density control in  $\text{Cu}_x\text{HgGeTe}_4$  and discovery of  $\text{Hg}_x\text{GeTe}_4$  via phase boundary mapping,” *J. Mater. Chem. A*, vol. 7, pp. 621–631, 2019, doi: 10.1039/C8TA10332A.

[5] B. R. Ortiz *et al.*, “Ultralow thermal conductivity in diamond-like semiconductors: Selective scattering of phonons from antisite defects,” *Chem. Mater.*, vol. 30, no. 10, pp. 3395–3409, May 2018.

[6] T. Zhu and E. Ertekin, “Mixed phononic and non-phononic transport in hybrid lead halide perovskites: Glass-crystal duality, dynamical disorder, and anharmonicity,” *Energy Environ. Sci.*, vol. 12, pp. 216–229, Jan. 2019, doi: 10.1039/C8EE02820F.

Freund, J.

[1] H. Kolla, E. R. Hawkes, A. R. Kerstein, N. Swaminathan, and J. H. Chen, “On velocity and reactive scalar spectra in turbulent premixed flames,” *J. Fluid Mech.*, vol. 754, pp. 456–487, 2014.

[2] P. E. Hamlington, A. Y. Poludnenko, and E. S. Oran, “Interactions between turbulence and flames in premixed reacting flows,” *Phys. Fluids*, vol. 32, p. 125111, 2011.

[3] J. F. MacArt, T. Greda, and M. E. Mueller, “Effects of combustion heat release on velocity and scalar statistics in turbulent premixed jet flames at low and high Karlovitz numbers,” *Combust. Flame*, vol. 191, pp. 468–485, 2018.

[4] D. Veynante, A. Trounev, K. N. C. Bray, and T. Mantel, “Gradient and counter-gradient scalar transport in turbulent premixed flames,” *J. Fluid Mech.*, vol. 332, pp. 263–293, 1997.

[5] O. Desjardins, G. Blanquart, G. Balarac, and H. Pitsch, “High order conservative finite difference scheme for variable density low Mach number turbulent flows,” *J. Comput. Phys.*, vol. 227, no. 15, pp. 7125–7159, 2008.

[6] J. F. MacArt and M. E. Mueller, “Semi-implicit iterative methods for low Mach number turbulent reacting flows: Operator splitting versus approximate factorization,” *J. Comput. Phys.*, vol. 326, pp. 569–595, 2016.

García, Marcelo H.

[1] J. M. Mier, “Experimental analysis of flow and turbulence characteristics in oscillatory boundary layers from LDV measurements,” Ph.D. dissertation, Civil and Env. Eng., Univ. of Illinois at Urbana–Champaign, Urbana, IL, 2014.

[2] P.F. Fischer, J.W. Lottes, and S.G. Kerkemeier. *NEK5000*. [Online]. Available: <https://nek5000.mcs.anl.gov/>, Accessed on: Jun. 13, 2019.

[3] M. I. Cantero, S. Balachandar, and M. H. García, “An Eulerian–Eulerian model for gravity currents driven by inertial particles,” *Intl. J. Multiph. Flow*, Vol. 34, pp. 484–501, 2008.

[4] M.H. García, “Chapter 2: Sediment transport and morphodynamics,” in *Sedimentation Engineering*. Reston, VA, USA: ASCE Manuals and Reports on Engineering Practice, No. 110, ASCE/EWRI, 2008.

[5] F. Pedocchi, M. I. Cantero, and M. H. García, “Turbulent kinetic energy balance of an oscillatory boundary layer in the transition to the fully turbulent regime,” *J. Turb.*, vol. 12, no. 32, pp. 1–27, 2011.

[6] S. Carstensen, B. M. Sumer, and J. Fredsøe, “Coherent structures in wave boundary layers. Part 1. Oscillatory motion,” *J. Fluid. Mech.*, Vol. 646, pp. 169–206, 2010.

[7] S. Carstensen, B. M. Sumer, and J. Fredsøe, “A note on turbulent spots over a rough bed in wave boundary layer,” *Phys. of Fluids*, vol. 4, p. 115104, 2010.

[8] F. Pedocchi and M. H. García, “Friction coefficient for oscillatory flow: The rough-smooth turbulence transition,” *J. of Hydraul. Res.*, vol. 47, no. 4, pp. 438–444, 2009.

Huff, Kathryn

[1] A. Lindsay, G. Ridley, A. Rykhlevskii, and K. Huff, “Introduction to Moltres: An application for simulation of molten salt reactors,” *Ann. Nucl. Energy*, vol. 114, pp. 530–540, Apr. 2018, doi: 10.1016/j.anucene.2017.12.025.

[2] A. Lindsay and K. Huff, “Moltres: Finite element based simulation of molten salt reactors,” *J. Open Source Softw.*, vol. 3, no. 21, pp. 1–2, Jan. 2018, doi: 10.21105/joss.00298.

[3] G. Ridley, A. Lindsay, and K. Huff, “An Introduction to Moltres, an MSR multiphysics code,” in *Trans. Amer. Nucl. Soc.*, Washington, DC, USA, 2017.

[4] G. Ridley, A. Lindsay, M. Turk, and K. Huff, “Multiphysics analysis of molten salt reactor transients,” University of Illinois at Urbana–Champaign, Urbana, IL, Undergraduate Report UTUC–ARFC-2017-01, Aug. 2017.

[5] A. Rykhlevskii, A. Lindsay, and K. D. Huff, “Full-core analysis of thorium-fueled molten salt breeder reactor using the SERPENT 2 Monte Carlo code,” in *Trans. of the Amer. Nucl. Soc.*, Washington, DC, USA, 2017.

[6] A. Rykhlevskii, J. W. Bae, and K. Huff, “arfc/salt-proc: Code for online reprocessing simulation of molten salt reactor with external depletion solver SERPENT,” *Zenodo*, Mar. 2018, doi: 10.5281/zenodo.1196455.

[7] A. Rykhlevskii, “Advanced online fuel reprocessing simulation for thorium-fueled molten salt breeder reactor,” M.S. thesis, Univ. of Illinois at Urbana–Champaign, 2018.

[8] D. Gaston, C. Newman, G. Hansen, and D. Lebrun-Grandie, “MOOSE: A parallel computational framework for coupled systems of nonlinear equations,” *Nucl. Eng. Des.*, vol. 239, no. 10, pp. 1768–1778, Oct. 2009, doi: 10.1016/j.nuceng-des.2009.05.021.

[9] B. S. Kirk, J. W. Peterson, R. H. Stogner, and G. F. Carey, “libMesh: a C++ library for parallel adaptive mesh refinement/coarsening simulations,” *Eng. Comp.*, vol. 22, no. 3–4, pp. 237–254, Dec. 2006, doi: 10.1007/s00366-006-0049-3.

[10] S. Balay *et al.*, *PETSc Users Manual*, 2015. [Online]. Available: <http://www.mcs.anl.gov/petsc>

Jain, Prashant K.

[1] P. Giannozzi *et al.*, “QUANTUM ESPRESSO: A modular and open-source software project for quantum simulations of materials,” *J. Phys. Condens. Matter*, vol. 21, no. 39, p. 395502, 2009.

[2] T. Hesjedal and Y. Chen, “Engineered heterostructures,” *Nat. Mater.*, vol.16 no. 1, pp. 3–4, 2017.

Johnsen, Eric

[1] C. E. Brennen, *Cavitation and bubble dynamics*. Cambridge, UK: Cambridge Univ. Press, 2013.

[2] W. Lauterborn and T. Kurz, “Physics of bubble oscillations,” *Rep. Prog. Phys.*, vol. 73, p. 106501, 2010.

[3] R. E. Arndt, “Cavitation in fluid machinery and hydraulic structures,” *Annu. Rev. Fluid Mech.*, vol. 34, pp. 273–328, 1991.

[4] A. J. Coleman *et al.*, “Acoustic cavitation generated by an extracorporeal shockwave lithotripter,” *Ultrasound Med. Bio.*, vol. 13, pp. 69–76, 1987.

[5] Y. Tomita and A. Shima, “Mechanisms of impulsive pressure generation and damage pit formation by bubble collapse,” *J. Fluid Mech.*, vol. 169, pp. 535–564, 1986.

[6] S. A. Beig and E. Johnsen, “Maintaining interface equilibrium conditions in compressible multiphase flows using interface capturing,” *J. Comput. Phys.*, vol. 302, pp. 548–566, 2015.

[7] S. A. Beig, B. Aboulhasanzadeh, and E. Johnsen, “Temperatures produced by inertially collapsing bubbles near rigid surfaces,” *J. Fluid Mech.*, vol. 852, pp. 105–125, 2018.

[8] D. Fuster and T. Colonius, “Modeling bubble clusters in compressible liquids,” *J. Fluid Mech.*, vol. 688, pp. 352–389, 2011.

Johnson, Harley T.

[1] P. Ghale and H. T. Johnson, “A sparse matrix–vector multiplication based algorithm for accurate density matrix computations on systems of millions of atoms,” *Comput. Phys. Comm.*, vol. 227, pp. 17–26, 2018.

[2] S. Balay *et al.*, *PETSc Users Manual*, 2015. [Online]. Available: <http://www.mcs.anl.gov/petsc>



[3] L. Jay, H. Kim, Y. Saad, and J. Chelikowsky, “Elec-  
tronic structure calculations for plane-wave codes  
without diagonalization,” *Comput. Phys. Comm.*,  
vol. 118, no. 1, pp. 21–30, 1999.

[4] A. Niklasson, C. Tymczak, and M. Challacombe,  
“Trace resetting density matrix purification in O(N)  
self-consistent-field theory,” *J. Chem. Phys.*, vol. 118,  
no. 19, pp. 8611–8620, 2003.

[5] H. H. Rosenbrock, “Some general implicit pro-  
cesses for the numerical solution of differential  
equations,” *Comput. J.*, vol. 5, no. 4, pp. 329–330,  
1963.

[6] C. Moler and C. Van Loan, “Nineteen dubious ways  
to compute the exponential of a matrix,” *SIAM Rev.*,  
vol. 20, no. 4, pp. 801–836, 1978.

[7] C. Moler and C. Van Loan, “Nineteen dubious ways  
to compute the exponential of a matrix, twenty-five  
years later,” *SIAM Rev.*, vol. 45, no. 2, pp. 3–49,  
2003.

[8] V. V. Dobrovitski and H. A. De Raedt, “Efficient  
scheme for numerical simulations of the spin-bath  
decoherence,” *Phys. Rev. E*, vol. 67, no. 5, p. 056702,  
2003.

[9] B. A. Shadwick and W. F. Buell, “Unitary inte-  
gration: A numerical technique preserving the  
structure of the quantum Liouville equation,” *Phys.*  
*Rev. Lett.*, vol. 79, no. 26, p. 5189, 1997.

Koric, Seid

[1] S. Koric and A. Gupta, “Sparse matrix factorization  
in the implicit finite element method on petascale  
architecture,” *Comput. Methods in Appl. Mech. Eng.*,  
vol. 302, p. 281, Apr. 2016.

[2] C. Ashcraft *et al.*, “Increasing the scale of LS–  
DYNA implicit analysis,” presented at 15th Inter.  
LS–DYNA Conf., Detroit, MI, USA, Jun. 2018.

[3] C. Ashcraft *et al.*, “Running jet engine models on  
thousands of processors with LS–DYNA implicit,”  
presented at the 12th European LS–DYNA Conf.,  
Koblenz, Germany, May 2019.

[4] T. A. Simons *et al.*, “Reducing the time-to-solution  
for finite element analysis of gas turbine engines,”  
presented at the AIAA Propulsion and Energy  
Forum and Exposition, Indianapolis, IN, USA, Aug.  
2019.

Leburton, Jean–Pierre

[1] A. Tubbs and A. Nussenzweig, “Endogenous DNA  
damage as a source of genomic instability in cancer,”  
*Cell*, vol. 168, no. 4, pp. 644–656, 2017.

[2] W. Humphrey, A. Dalke, and K. Schulten, “VMD:  
Visual molecular dynamics,” *J. Mol. Graphics*, vol.  
14, no. 1, pp. 33–38, 1996.

[3] N. Foloppe and J. A. D. Mackerell, “All-atom  
empirical force field for nucleic acids: I. Param-  
eter optimization based on small molecule and  
condensed phase macromolecular target data,” *J.*  
*Comput. Chem.*, vol. 21, no. 2, pp. 86–104, 2000.

[4] A. Aksimentiev, J. B. Heng, G. Timp, and K. Schul-  
ten, “Microscopic kinetics of DNA translocation  
through synthetic nanopores,” *Biophys. J.*, vol. 87,  
no. 3, pp. 2086–2097, 2004.

[5] A. Girdhar, C. Sathe, K. Schulten, and J.–P. Lebur-  
ton, “Graphene quantum point contact transistor  
for DNA sensing,” *Proc. Nat. Acad. Sci. USA*, vol.  
110, no. 42, pp. 16748–16753, 2013.

[6] A. Sarathy and J.–P. Leburton, “Electronic conduc-  
tance model in constricted MoS<sub>2</sub> with nanopores,”  
*Appl. Phys. Lett.*, vol. 108, no. 5, p. 053701, 2016.

[7] N. B. M. Athreya, A. Sarathy, and J.–P. Leburton,  
“Large scale parallel DNA detection by two-dimen-  
sional solid-state multipore systems,” *ACS Sensors*,  
vol. 3, no. 5, pp. 1032–1039, 2018.

Mashayek, Farzad

[1] D. A. Kopriva and J. H. Kolias, “A conservative  
staggered-grid Chebyshev multidomain method for  
compressible flows,” *J. Comput. Phys.*, vol. 125, no.  
1, pp. 244–261, Apr. 1996.

[2] D. A. Kopriva, “A staggered-grid multidomain  
spectral method for the compressible Navier–  
Stokes equations,” *J. Comput. Phys.*, vol. 143, no. 1,  
pp. 125–158, Jun. 1998.

[3] G. B. Jacobs, D. A. Kopriva, and F. Mashayek,  
“Validation study of a multidomain spectral code for  
simulation of turbulent flows,” *AIAA J.*, vol. 43, no.  
6, pp. 1256–1264, May 2005.

Matalon, Moshe

[1] M. Matalon and B. J. Matkowsky, “Flames as gasdy-  
namic discontinuities,” *J. Fluid Mech.*, vol. 124, pp.  
239–259, 1982.

[2] M. Matalon, C. Cui, and J. K. Bechtold, “Hydro-  
dynamic theory of premixed flames: Effects of  
stoichiometry, variable transport coefficients and  
arbitrary reaction orders,” *J. Fluid Mech.*, vol. 487,  
pp.179–210, 2003.

[3] A. Patyal and M. Matalon, “Nonlinear development  
of hydrodynamically-unstable flames in three-di-  
mensional laminar flows,” *Combust. Flame*, vol. 195,  
pp. 128–139, 2018.

Raman, Venkat

[1] K. Y. Cho, J. R. Codoni, B. A. Rankin, J. Hoke, and F.  
Schauer, “High-repetition-rate chemiluminescence  
imaging of a rotating detonation engine,” presented  
at the 54th AIAA Aerospace Sciences Meeting,  
San Diego, CA, USA, Jan. 4–8, 2016, AIAA Paper  
2016–1648.

[2] P. A. Cocks, A. T. Holley, and B. A. Rankin, “High  
fidelity simulations of a non-premixed rotating  
detonation engine,” presented at the 54th AIAA  
Aerospace Sciences Meeting, San Diego, CA, USA,  
Jan. 4–8, 2016, AIAA Paper 2016–0125.

[3] B. A. Rankin *et al.*, “Chemiluminescence imaging  
of an optically accessible non-premixed rotating  
detonation engine,” *Combust. Flame*, vol. 176, pp.  
12–22, 2017.

Riedl, Caroline

[1] The DOE/NSF Nuclear Science Advisory Com-  
mittee Working Group, *The Frontiers of Nuclear  
Science: A Long-Range Plan*, Dec. 2007, arX-  
iv:0809.3137.

[2] COMPASS Collaboration, “First measurement of  
transverse-spin-dependent azimuthal asymmetries  
in the Drell–Yan process,” *Phys. Rev. Lett.*, vol. 119,  
p. 112002, 2017.

[3] H. Riahi *et al.*, “FTS3: Quantitative monitoring,” *J.*  
*Phys.*, Conference Series 664, p. 062051, 2015.

[4] J. Allison *et al.*, “Facilities and methods: Geant4—A  
simulation toolkit,” *Nucl. Instrum. Methods Phys.*  
*Res. A*, vol. 506, pp. 250–303, 2003.

[5] The XXVII International Workshop on Deep  
Inelastic Scattering and Related Topics (DIS 2019),  
Apr. 8–12, 2019, Torino, Italy.

Schleife, André (1)

[1] M. Xu, T. Liang, M. Shi, and H. Chen, “Graphene-  
like two-dimensional materials,” *Chem. Rev.*, vol.  
113, no. 5, pp. 3766–3798, May 2013, doi: 10.1021/  
cr300263a.

[2] U. Bangert *et al.*, “Ion implantation of graphene—  
Toward IC compatible technologies,” *Nano Lett.*,  
vol. 13, no. 10, Oct. 2013, doi: 10.1021/nl402812y.

[3] L. Vicarelli, S. Heerema, C. Dekker, and H.  
Zandbergen, “Controlling defects in graphene for  
optimizing the electrical properties of graphene  
nanodevices,” *ACS Nano*, vol. 9, no. 4, Apr. 2015,  
doi: 10.1021/acsnano.5b01762.

[4] A. Schleife, E. Draeger, Y. Kanai, and A. Correa,  
“Plane-wave pseudopotential implementation of ex-  
plicit integrators for time-dependent Kohn–Sham  
equations in large-scale simulations,” *J. Chem. Phys.*,  
vol. 137, no. 22, Oct. 2012, doi: 10.1063/1.4758792.

[5] C. Ullrich, *Time-Dependent Density-Functional  
Theory: Concepts and Applications*. Oxford, UK,  
Oxford Univ. Press, 2013. [Online]. Available: doi:  
10.1093/acprof:oso/9780199563029.001.0001

[6] S. Abhyankar, J. Brown, E. Constantinescu, D.  
Ghosh, B. Smith, and H. Zhang, “PETSc/TS: A  
modern scalable ODE/DAE solver library,” Jun.  
2018, arXiv:1806.01437 [math.NA].

[7] S. Balay, W. Gropp, L. McInnes, and B. Smith,  
“PETSc Users Manual,” Tech. Rep. ANL–95/11—Re-  
vision 3.10, Argonne National Laboratory, 2018.

Schleife, André (2)

[1] S. A. Maier *et al.*, “Local detection of electromag-  
netic energy transport below the diffraction limit in  
metal nanoparticle plasmon waveguides,” *Nature*,  
vol. 2, no. 4, p. 220, 2003.

[2] D. P. Fromm, A. Sundaramurthy, P. J. Schuck, G.  
Kino, and W. E. Moerner, “Gap-dependent optical  
coupling of single ‘bowtie’ nanoantennas resonant  
in the visible,” *Nano Lett.*, vol. 4, no. 5, pp. 957–961,  
2004.

[3] S. Prayakarao *et al.*, “Tunable VO<sub>2</sub>/Au hyperbolic  
metamaterial,” *App. Phys. Lett.*, vol. 109, no. 6, p.  
061105, 2016.

[4] S. Kawata, A. Ono, and P. Verma, “Subwavelength  
colour imaging with a metallic nanolens,” *Nat.*  
*Photonics*, vol. 2, no. 7, p. 438, 2008.

[5] N. Engheta, A. Salandrino, and A. Alu, “Circuit  
elements at optical frequencies: Nanoinductors,  
nanocapacitors, and nanoresistors,” *Phys. Rev. Lett.*,  
vol. 95, no. 9, p. 095504, 2005.

[6] T. Stakenborg and L. Lagae, “Gold nanoring as a  
sensitive plasmonic biosensor for on-chip DNA de-  
tection,” *App. Phys. Lett.*, vol. 100, no. 17, p. 173114,  
2012.

[7] J. B. Khurgin and A. Boltasseva, “Reflecting upon  
the losses in plasmonics and metamaterials,” *MRS  
Bull.*, vol. 37, no. 8, p. 768779, 2012.

[8] G. Kresse and J. Furthmüller, “Efficient iterative  
schemes for *ab initio* total-energy calculations  
using a plane-wave basis set,” *Phys. Rev. B*, vol. 54,  
no. 16, pp. 11169–11186, 1996.

[9] A. Jain *et al.*, “The Materials Project: A materials  
genome approach to accelerating materials innova-  
tion,” *APL Mat.*, vol. 1, no. 1, p. 011002, 2013.

[10] S. P. Ong *et al.*, “Python Materials Genomics  
(pymatgen): A robust, open-source python library  
for materials analysis,” *Comp. Mat. Sci.*, vol. 68, pp.  
314–319, Feb. 2013.

[11] F. Pedregosa *et al.*, “Scikit-learn: Machine learn-  
ing in Python,” *J. Mach. Learn. Res.*, vol. 12, pp.  
2825–2830, 2011.

Thomas, Brian G.

[1] World Steel Association, “Table 4: Production of  
continuously cast steel,” in *Steel Statistical Year-  
book 2018*. Brussels, Belgium: World Steel Assoc.,  
2018, pp. 9–12.

[2] S.–M. Cho and B. G. Thomas, “Electromagnetic  
forces in continuous casting of steel slabs,” *Metals  
(Spec. Issue: Cont. Casting)*, vol. 9, no. 4, pp. 1–38,  
2019, doi: 10.3390/met9040471.

[3] S.–M. Cho, B. G. Thomas, and S.–H. Kim, “Effect  
of nozzle port angle on transient flow and surface  
slag behavior during continuous steel-slab casting,”  
*Metall. Mater. Trans. B*, vol. 50B, no. 1, pp. 52–76,  
2019, doi: 10.1007/s11663-018-1439-9.

[4] H. Yang, S. P. Vanka, and B. G. Thomas, “A hybrid  
Eulerian–Eulerian discrete-phase model of turbu-  
lent bubbly flow,” *J. Fluids Eng.*, vol. 140, no. 10, p.  
101202, 2018, doi: 10.1115/1.4039793.

[5] H. Yang, S. P. Vanka, and B. G. Thomas, “Modeling  
of argon gas behavior in continuous casting of steel,”  
in *CFD Modeling and Simulation in Materials Pro-  
cessing 2018*, Warrendale, PA, USA: Springer, 2018,  
pp. 119–131.

[6] H. Yang, S. P. Vanka, and B. G. Thomas, “Modeling  
of argon gas behavior in continuous casting of steel,”  
*JOM*, vol. 70, no. 10, pp. 2148–2156, 2018, doi:  
10.1007/s11837-018-2997-7.



[7] K. Jin, S. P. Vanka, and B. G. Thomas, “Large eddy simulations of electromagnetic braking effects on argon bubble transport and capture in a steel continuous casting mold,” *Metall. Mater. Trans. B*, vol. 49B, no. 3, pp. 1360–1377, 2018, doi: 10.1007/s11663-018-1191-1.

[8] S. P. Vanka, A. F. Shinn, and K. C. Sahu, “Computational fluid dynamics using graphics processing units: Challenges and opportunities,” in *Proc. ASME 2011 IMECE Conf.*, pp. 429–437.

[9] J. Sengupta, B. G. Thomas, H. Shin, G. Lee, and S. Kim, “Mechanism of hook formation during continuous casting of ultra-low carbon steel slabs,” *Metall. Mater. Trans. A*, vol. 37A, pp. 1597–1611, 2006.

[10] H. Shin, S. Kim, B. G. Thomas, G. Lee, J. Park, and J. Sengupta, “Measurement and prediction of lubrication, powder consumption, and oscillation mark profiles in ultra-low carbon steel slabs,” *ISIJ Int.*, vol. 46, pp. 1635–1644, 2006.

Tinoco Lopez, Rafael O.

[1] E. W. Koch, “Beyond light: Physical, geological, and geochemical parameters as possible submersed aquatic vegetation habitat requirements,” *Estuaries*, vol. 24, no. 1, pp. 1–17, 2001.

[2] H. M. Nepf, “Hydrodynamics of vegetated channels,” *J. Hydraul. Res.*, vol. 50, no. 3, pp. 262–79, 2012.

[3] H. M. Nepf, “Flow and transport in regions with aquatic vegetation,” *Ann. Rev. Fluid Mech.*, vol. 44, pp. 123–142, 2012.

[4] R. Tinoco and G. Cocco, “A laboratory study on sediment resuspension within arrays of rigid cylinders,” *Adv. Water Res.*, vol. 92, pp. 1–9, 2016.

[5] P. F. Fischer, J. W. Lottes, and S. G. Kerkemeier, *NEK5000 webpage*. (2008). [Online]. Available: <https://nek5000.mcs.anl.gov>

[6] M. Deville, P. F. Fischer, and E. Mund, *High-Order Methods for Incompressible Fluid Flow*. Cambridge, England: Cambridge Univ. Press, 2002.

[7] S. Dutta, M. I. Cantero, and M. H. Garcia, “Effect of self-stratification on sediment diffusivity in channel flows and boundary layers: A study using direct numerical simulations,” *Earth Surf. Dyn.*, vol. 2, no. 2, p. 419, 2014.

Toussaint, Kimani

[1] Y. Hernandez *et al.*, “High-yield production of graphene by liquid-phase exfoliation of graphite,” *Nat. Nanotechnol.*, vol. 3, no. 9, p. 563, 2008.

[2] M. Lotya *et al.*, “Liquid phase production of graphene by exfoliation of graphite in surfactant/water solutions,” *J. Am. Chem. Soc.*, vol. 131, no. 10, pp. 3611–3620, 2009.

[3] V. Sresht, A. A. H. Pádua, and D. Blankschtein, “Liquid-phase exfoliation of phosphorene: Design rules from molecular dynamics simulations,” *ACS Nano.*, vol. 9, no. 8, pp. 8255–8268, 2015.

Trinkle, Dallas R.

[1] S. L. Frederiksen, K. W. Jacobsen, K. S. Brown, and J. P. Sethna, “Bayesian ensemble approach to error estimation of interatomic potentials,” *Phys. Rev. Lett.*, vol. 93, p. 165501, 2004.

[2] P. Zhang and D. R. Trinkle, “Database optimization for empirical interatomic potential models,” *Model. Simul. Mater. Sci. Eng.*, vol. 23, p. 065011, 2015, doi: 10.1088/0965-0393/23/6/065011.

[3] P. Zhang and D. R. Trinkle, “A modified embedded atom method potential for interstitial oxygen in titanium,” *Comp. Mater. Sci.*, vol. 124, pp. 204–210, 2016, doi: 10.1016/j.commatsci.2016.07.039.

Wagner, Lucas K.

[1] L. K. Wagner, M. Bajdich, and L. Mitas, “QWalk: A quantum Monte Carlo program for electronic structure,” *J. Comput. Phys.*, vol. 228, pp. 3390–3404, 2009.

Wang, Zhi Jian

[1] S. Y. Han, R. Taghavi, and S. Farokhi, “Passive control of supersonic rectangular jets through boundary layer swirl,” *Int. J. Turbo Jet-Eng.*, vol. 30, no. 2, pp. 199–216, Jun. 2013.

[2] Z. J. Wang *et al.*, “Towards industrial large eddy simulation using the FR/CPR method,” *Comput. Fluids*, vol. 156, pp. 579–589, Oct. 2017.

[3] M. J. Lighthill, “On sound generated aerodynamically. I. General theory,” *Proc. R. Soc. Lond. A*, vol. 222, pp. 564–587, 1952.

[4] H. T. Huynh, “A flux reconstruction approach to high-order schemes including discontinuous Galerkin methods,” presented at the 18th AIAA Comput. Fluid Dyn. Conf., Miami, Florida, USA, Jun. 25–28, 2007.

[5] Z. J. Wang and H. Gao, “A unifying lifting collocation penalty formulation including the discontinuous Galerkin, spectral volume/difference methods for conservation laws on mixed grids,” *J. Comput. Phys.*, vol. 228, pp. 8161–8186, 2009.

[6] C. K. W. Tam, “Supersonic jet noise,” *Ann. Rev. Fluid Mech.*, vol. 27, pp. 17–43, 1995.

Wentzcovitch, Renata M.

[1] D. B. Zhang, T. Sun, and R. M. Wentzcovitch, “Phonon quasiparticles and anharmonic free energy in complex systems,” *Phys. Rev. Lett.*, vol. 112, 2014, doi: 10.1103/PhysRevLett.112.058501.

[2] K. Sarkar, M. Topsakal, N. A. W. Holzwarth, and R. M. Wentzcovitch, “Evolutionary optimization of PAW data-sets for accurate high pressure simulations,” *J. Computat. Phys.*, vol. 347, pp. 39–55, 2017, doi:10.1016/j.jcp.2017.06.032.

Xu, Bin

[1] H. F. Jansen and A. J. Freeman, “Total-energy full-potential linearized augmented-plane-wave method for bulk solids: Electronic and structural properties of tungsten,” *Phys. Rev. B*, vol. 30, p. 561, 1984.

[2] M. Weinert, E. Wimmer, and A. J. Freeman, “Total-energy all-electron density functional method for bulk solids and surfaces,” *Phys. Rev. B*, vol. 26, pp. 4571, 1982.

Xu, Zhen

[1] C. E. Brennen, *Cavitation and Bubble Dynamics*, Cambridge Univ. Press, 2013.

[2] T. Deplancke, O. Lame, J. Y. Cavaille, M. Fivel, M. Riondet, and J. P. Franc, “Outstanding cavitation erosion resistance of Ultra High Molecular Weight Polyethylene (UHMWPE) coatings,” *Wear*, vol. 328–329, pp. 301–308, 2015.

[3] J. P. Franc, M. Riondet, A. Karimi, and G. L. Chahine, “Material and velocity effects on cavitation erosion pitting,” *Wear*, vol. 274–275, pp. 248–259, 2012.

[4] *Advanced Experimental and Numerical Techniques for Cavitation Erosion Prediction*, K.–H. Kim, G. Chahine, J.–P. Franc, and A. Karimi, Eds., Springer, 2014.

[5] S. A. Beig and E. Johnsen, “Maintaining interface equilibrium conditions in compressible multiphase flows using interface capturing,” *J. Comput. Phys.*, vol. 302, pp. 548–566, 2015.

[6] S. A. Beig, B. Aboulhasanzadeh, and E. Johnsen, “Temperatures produced by inertially collapsing bubbles near rigid surfaces,” *J. Fluid Mech.*, vol. 852, pp. 105–125, 2018.

[7] M. Rodriguez and E. Johnsen, “A high-order, finite-difference approach for numerical simulations of shocks interacting with interfaces separating different linear viscoelastic materials,” *J. Comput. Phys.*, vol. 379, pp. 70–90, 2019.

[8] M. Rodriguez, K. G. Powell, and E. Johnsen, “A high-order accurate AUSM+-up approach for simulations of compressible multiphase flows with linear viscoelasticity,” *Shock Waves*, vol. 29, no. 5, pp. 717–734, 2019.

Yan, Yonghua

[1] C. Lee and R. Li, “Dominant structure for turbulent production in a transitional boundary layer,” *J. Turbul.*, vol. 8, no. 55, p. N55, Jan. 2007.

[2] F. F. Grinstein, L. G. Margolin, and W. J. Rider, *Implicit Large Eddy Simulation*. Cambridge, UK: Cambridge Univ. Press, 2011.

[3] S. P. Spekreijse, “Elliptic grid generation based on Laplace equations and algebraic transformations,” *J. Comput. Phys.*, vol. 118, no. 1, pp. 38–61, 1995.

[4] Y. Yan, L. Chen, Q. Li, and C. Liu, “Numerical study on microramp vortex generation for supersonic ramp flow control at Mach 2.5,” *Shock Waves*, vol. 27, no. 1, pp 79–96, 2016.

[5] Y. Yan, C. Chen, P. Lu, and C. Liu, “Study on shock wave–vortex ring interaction by the micro vortex generator controlled ramp flow with turbulent inflow,” *Aerosp. Sci. Technol.*, vol. 30, no. 1, pp. 226–231, 2013.

[6] Y. Yan, C. Chen, H. Fu, and C. Liu, “DNS study on  $\Lambda$ -vortex and vortex ring formation in flow transition at Mach number 0.5,” *J. Turbul.*, vol. 15, no. 1, pp. 1–21, Jan. 2014.

[7] L. Jiang, H. Shan, C. Liu, and M. R. Visbal, “Non-reflecting boundary conditions for DNS in curvilinear coordinates,” in *Recent Advances in DNS and LES: Proc. of the 2nd AFOSR Conf.*, New Brunswick, NJ, USA, Jun. 7–9, 1999, pp. 219–233.

Yeung, Pui–Kuen

[1] U. Frisch, *Turbulence. The Legacy of A.N. Kolmogorov*. Cambridge, UK: Cambridge Univ. Press, 1995.

[2] P. K. Yeung, K. R. Sreenivasan, and S. B. Pope, “Effects of finite spatial and temporal resolution on extreme events in direct numerical simulations of incompressible isotropic turbulence,” *Phys. Rev. Fluids*, vol. 3, p. 064603, 2018.

[3] V. Yakhot and K. R. Sreenivasan, “Anomalous scaling of structure functions and dynamic constraints on turbulence simulations,” *J. Stat. Phys.*, vol. 121, pp. 823–841, 2005.

[4] A. A. Migdal, “Loop equation and area law in turbulence,” *Int. J. Mod. Phys. A*, vol. 9, pp. 1197–1238, 1994.

[5] K. R. Sreenivasan, A. Juneja, and A. K. Suri, “Scaling properties of circulation in moderate-Reynolds-number turbulent wakes,” *Phys. Rev. Lett.*, vol. 75, pp. 433–436, 1995.

[6] N. Cao, S. Chen, and K. R. Sreenivasan, “Properties of velocity circulation in three-dimensional turbulence,” *Phys. Rev. Lett.*, vol. 76, pp. 616–619, 1996.

[7] P. K. Yeung, X. M. Zhai, and K. R. Sreenivasan, “Extreme events in computational turbulence,” *Proc. Nat. Acad. Sci. USA*, vol. 112, pp. 12633–12638, 2015.

[8] K. P. Iyer, K. R. Sreenivasan, and P. K. Yeung, “Circulation in high Reynolds number isotropic turbulence is a bifractal,” in revision, *Phys. Rev. X*, 2019.

[9] A. N. Kolmogorov, “The local structure of turbulence in an incompressible fluid with very large Reynolds numbers,” *Dokl. Akad. Nauk SSSR*, vol. 30, pp. 301–305, 1941.

[10] J. C. R. Hunt, O. M. Phillips, and D. Williams, Eds., *Turbulence and Stochastic Processes: Kolmogorov’s Ideas 50 Years On*. Cambridge, UK: Cambridge Univ. Press, 1991.

[11] K. R. Sreenivasan, “Fractals and multifractals in fluid turbulence,” *Ann. Rev. Fluid Mech.*, vol. 23, pp. 539–600, 1991.

Zhang, Yang

[1] G. Azimi, R. Dhiman, H.–M. Kwon, A. T. Paxson, and K. K. Varanasi, “Hydrophobicity of rare-earth oxide ceramics,” *Nat. Mater.*, vol. 12, 2013.



[2] G. Kresse and J. Furthmüller, “Efficient iterative schemes for *ab initio* total-energy calculations using a plane-wave basis set,” *Phys. Rev. B—Condens. Matter Mater. Phys.*, vol. 54, p. 11169, 1996.

COMPUTER SCIENCE & ENGINEERING

Cox, Donna

[1] A. Hitchcock, C. N. Hunter, and M. Sener, “Determination of cell doubling times from the return-on-investment time of photosynthetic vesicles based on atomic detail structural models,” *J. Phys. Chem. B*, vol. 121, no. 15, 2017, doi: 10.1021/acs.jpcb.6b12335.

[2] J. Stone *et al.*, “Atomic detail visualization of photosynthetic membranes with GPU-accelerated ray tracing,” *Parallel Comput.*, vol. 55, 2016, doi: 10.1016/j.parco.2015.10.015.

[3] M. Cartron *et al.*, “Integration of energy and electron transfer processes in the photosynthetic membrane of *Rhodobacter sphaeroides*,” *Biochim. Biophys. Acta—Bioenerg.*, vol. 1837, no. 10, 2014, doi: 10.1016/j.bbabo.2014.02.003.

[4] S. J. Lock *et al.*, “The Origin of the Moon Within a Terrestrial Synestia,” *J. Geophys. Res.: Planets*, 2018, doi: 10.1002/2017JE005333.

[5] M. Norman *et al.*, “Cosmic Bubble Bath: First Galaxies Reionize the Universe.”

[6] GEOS-5 modeled cloud simulation, NASA/GSFC—Global Modeling and Assimilation Office (GMAO).

[7] Blue Marble Next Generation Earth Imagery, Reto Stöckli, NASA Earth Observatory (NASA Goddard Space Flight Center).

[8] Population Density, NASA Earth Observatory, Socioeconomic Data and Applications Center (SEDAC) at Columbia University.

[9] United States Geological Survey, Natural Resources Canada, Mexico National Institute of Statistics and Geography, *North American Atlas—Roads*. [Online]. Available: [https://nationalmap.gov/small\\_scale/mld/road00l.html](https://nationalmap.gov/small_scale/mld/road00l.html).

Duursma, Iwan M.

[1] M. Sugino, Y. Ienaga, N. Tokura, and T. Kasami, “Weight distribution of (128, 64) Reed–Muller code (Corresp.),” *IEEE Trans. Inf. Theory*, vol. 17, no. 5, pp. 627–628, Sept. 1971.

[2] O. Milenkovic, S. T. Coffey, and K. J. Compton, “The third support weight enumerators of the doubly-even, self-dual [32,16,8] codes,” *IEEE Trans. Inf. Theory*, vol. 49, no. 3, pp. 740–746, 2003.

[3] I. M. Duursma, “Weight distributions of geometric Goppa codes,” *Trans. Amer. Math. Soc.*, vol. 351, no. 9, pp. 3609–3639, 1999.

[4] Y. Desaki, T. Fujiwara, and T. Kasami, “The weight distributions of extended binary primitive BCH codes of length 128,” *IEEE Trans. Inf. Theory*, vol. 43, no. 4, pp. 1364–1371, 1997.

[5] D. Britz, T. Britz, K. Shiromoto, and H. K. Sorensen, “The higher weight enumerators of the doubly-even, self-dual [48,24,12] code,” *IEEE Trans. Inf. Theory*, vol. 53, no. 7, pp. 2567–2571, 2007.

Gropp, William

[1] C. Zhang, M. Rabiee, E. Sayyari, and S. Mirarab, “ASTRAL-III: polynomial time species tree reconstruction from partially resolved gene trees,” *BMC Bioinformatics*, vol. 19 (suppl. 6), article 153, May 2018, doi: 10.1186/s12859-018-2129-y.

[2] A. Stamatakis, “RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies,” *Bioinformatics*, vol. 30, no. 9, pp. 1312–1313, May 2014, doi: 10.1093/bioinformatics/btu033.

Jha, Shantenu

[1] V. Balasubramanian *et al.*, “ExTASY: Scalable and flexible coupling of MD simulations and advanced sampling techniques,” *Proc. 2016 IEEE 12th Int. Conf. e-Science*, 2017.

[2] E. Hruska *et al.*, “Quantitative comparison of adaptive sampling methods for protein dynamics,” *J. Chem. Phys.*, vol. 149, 2018, doi: 10.1063/1.5053582.

[3] E. Hruska *et al.*, “Extensible and scalable adaptive sampling on supercomputers,” 2019, ArXiv: 1907.06954.

Olson, Luke N. (1)

[1] A. Bienz and L. N. Olson. (2019). *RAPtor: parallel algebraic multigrid*, v1.0. [Online]. Available: <https://github.com/raptor-library/raptor>, Accessed on: Jul. 31, 2019.

[2] A. Bienz and L. N. Olson. (2019). *Node-Aware MPI*. [Online]. Available: [https://github.com/bienz2/Node\\_Aware\\_MPI](https://github.com/bienz2/Node_Aware_MPI), Accessed on: Jul. 31 2019.

[3] A. Bienz, W. D. Gropp, L. N. Olson, “Node aware sparse matrix–vector multiplication,” *J. Parallel Distrib. Comput.*, vol. 130, pp. 166–178, 2019, doi: 10.1016/j.jpdc.2019.03.016.

Olson, Luke (2)

[1] L. Spies *et al.*, *Tausch*, 2019. [Online]. Available: <https://bitbucket.org/luspi/tausch>

[2] D. Moulton, L. Olson, and A. Reisner, *Cedar*, 2019. [Online]. Available: <https://github.com/cedar-framework/cedar>

[3] A. Bienz and L. Olson, *RAPtor: parallel algebraic multigrid*, version 1.0, 2019. [Online]. Available: <https://github.com/raptor-library/raptor>

[4] A. Reisner, L. Olson, and D. Moulton, “Scaling structured multigrid to 500K+ cores through coarse-grid redistribution,” *SIAM J. Sci. Comput.*, 2018, doi: 10.1137/17M1146440.

Rahman, Rajib

[1] M. Veldhorst *et al.*, “A two-qubit logic gate in silicon,” *Nature*, vol. 526, pp. 410–414, Oct. 2015.

Snir, Marc

[1] C. Wang, N. Dryden, F. Cappello, and M. Snir, “Neural network based silent error detector,” in *IEEE International Conf. Cluster Computing (Cluster 2018)*, Belfast, UK, Sept. 2018.

Warnow, Tandy

[1] N. Wickett *et al.*, “Phylotranscriptomic analysis of the origin and diversification of land plants,” *Proc. Nat. Acad. Sci. USA*, vol. 111, no. 45, pp. E4859–E4868, Nov. 11, 2014, doi: 10.1073/pnas.1323926111.

[2] E. Jarvis *et al.*, “Whole-genome analyses resolve early branches in the tree of life of modern birds,” *Science*, pp. 1320–1333, Dec. 12, 2014, doi: 10.1126/science.1253451.

[3] W. P. Maddison, “Gene trees in species trees,” *Syst. Biol.*, vol. 46, pp. 523–536, 1997, doi: 10.1093/sysbio/46.3.523.

[4] E. K. Molloy and T. Warnow, “NJMerge: A Generic Technique for Scaling Phylogeny Estimation Methods and Its Application to Species Trees,” in *Comparative Genomics: RECOMB-CG 2018, Lecture Notes in Computer Science*, M. Blanchette and A. Ouangraoua, Eds. New York, NY, USA: Springer, vol. 11183, 2018.

[5] E. K. Molloy and T. Warnow, “TreeMerge: A new method for improving the scalability of species tree estimation methods,” *Bioinformatics*, special issue for Intelligent Systems for Molecular Biology 2019, vol. 35, no. 14, pp. i417–i426, Jul. 2019.

[6] T. Le, A. Sy, E. K. Molloy, Q. Zhang, S. Rao, and T. Warnow, “Using INC within divide-and-conquer phylogeny estimation,” in *Proc. Int. Conf. Algorithms for Comput. Biol.*, 2019, pp. 167–178, doi: 10.1007/978-3-030-18174-1\_12.

[7] A. Stamatakis, “RA x ML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies,” *Bioinformatics*, vol. 30, no. 9, pp. 1312–1313, 2014.

[8] S. Mirarab, R. Reaz, M. S. Bayzid, T. Zimmermann, M. S. Swenson, and T. Warnow, “ASTRAL: genome-scale coalescent-based species tree estimation,” *Bioinformatics*, vol. 30, no. 17, pp. i541–i548, 2014.

[9] S. Mirarab and T. Warnow, “ASTRAL-II: coalescent-based species tree estimation with many hundreds of taxa and thousands of genes,” *Bioinformatics*, vol. 31, no. 12, pp. i44–i52, 2015.

BIOLOGY, CHEMISTRY & HEALTH

Aksimentiev, Aleksei (1)

[1] J.K. Johnson *et al.*, “Capsid conformational sampling in HK97 maturation visualized by X-ray crystallography and cryo-EM,” *Structure*, vol. 14, no. 11, pp. 1655–1665, 2006.

[2] R. L. Duda *et al.*, “Structure and energetics of encapsidated DNA in bacteriophage HK97 studied by scanning calorimetry and cryo-electron microscopy,” *J. Mol. Bio.*, vol. 391, no. 2, pp. 471–483, 2009.

[3] J. R. Perilla and K. Schulten, “Physical properties of the HIV-1 capsid from all-atom molecular dynamics simulations,” *Nature Commun.*, vol. 8, p. 15959, 2017.

[4] D.S. Larsson, L. Liljas, D. van der Spoel, “Virus capsid dissolution studied by microsecond molecular dynamics simulations,” *PLOS Comput. Biol.*, vol. 8, no. 5, p. e1002502, 2012.

[5] D. Wendell *et al.*, “Translocation of double-stranded DNA through membrane-adapted phi29 motor protein nanopores,” *Nat. Nanotechnol.*, vol. 4, no. 11, p. 765, 2009.

[6] J. Yoo, and A. Aksimentiev, “The structure and intermolecular forces of DNA condensates,” *Nucleic Acids Res.*, vol. 44, no. 5, pp. 2036–2046, 2016.

[7] L. Kalé *et al.*, “NAMD2: greater scalability for parallel molecular dynamics,” *J. Comput. Phys.*, vol. 151, no. 1, pp. 283–312, 1999.

[8] A. Evilevitch *et al.*, “Osmotic pressure inhibition of DNA ejection from phage,” *Proc. Natl. Acad. Sci. USA*, vol. 100, no. 16, pp. 9292–9295, 2003.

Aksimentiev, Aleksei (2)

[1] L. Restrepo–Pérez, C. Joo, and C. Dekker, “Paving the way to single-molecule protein sequencing,” *Nat. Nanotechnol.*, 2018, doi: 10.1038/s41565-018-0236-6.

[2] P. Edman, “A method for the determination of amino acid sequence in peptides,” *Arch. Biochem.*, 1949.

[3] H. Steen and M. Mann, “The ABC’s (and XYZ’s) of peptide sequencing,” *Nat. Rev. Mol. Cell Biol.*, 2004, doi: 10.1038/nrm1468.

[4] J. Comer *et al.*, “The adaptive biasing force method: Everything you always wanted to know but were afraid to ask,” *J. Phys. Chem. B*, 2015, doi: 10.1021/jp506633n.

[5] E. Darve and A. Pohorille, “Calculating free energies using average force,” *J. Chem. Phys.*, 2001, doi: 10.1063/1.1410978.

[6] J. C. Phillips *et al.*, “Scalable molecular dynamics with NAMD,” *J. Comput. Chem.*, 2005, doi: 10.1002/jcc.20289.

[7] J. Wilson, K. Sarthak,W. Si, L. Gao, and A. Aksimentiev, “Rapid and accurate determination of nanopore ionic current using a steric exclusion model,” *ACS Sensors*, vol. 4, pp. 634–644, 2019.



Aksimentiev, Aleksei (3)

[1] M. Langecker *et al.*, “Synthetic lipid membrane channels formed by designed DNA nanostructures,” *Science*, vol. 338, no. 6109, pp. 932–936, 2012.

[2] J. R. Burns, E. Stulz, and S. Howorka, “Self-assembled DNA nanopores that span lipid bilayers,” *Nano Lett.*, vol. 13, no. 6, pp. 2351–2356, 2013.

[3] J. R. Burns, A. Seifert, N. Fertig, and S. Howorka, “A biomimetic DNA-based channel for the ligand-controlled transport of charged molecular cargo across a biological membrane,” *Nat. Nanotechnol.*, vol. 11, no. 2, p. 152, 2016.

[4] L. Lei, and H.–C. Wu, “DNA-based nanopore sensing,” *Angew. Chem. Inter. Ed.*, vol. 55, no. 49, pp. 15216–15222, 2016.

[5] A. Ohmann *et al.*, “A synthetic enzyme built from DNA flips 10<sup>7</sup> lipids per second in biological membranes,” *Nat. Commun.*, vol. 9, no. 1, p. 2426, 2018.

[6] J. C. Phillips *et al.*, “Scalable molecular dynamics with NAMD,” *J. Comput. Chem.*, vol. 26, no. 16, pp. 1781–1802, 2005.

[7] K. Hart, N. Fologpe, C. M. Baker, E. J. Denning, L. Nilsson, and A. D. MacKerell, “Optimization of the Charmm additive force field for DNA: Improved treatment of the Bi/Bii conformational equilibrium,” *J. Chem. Theory Comput.*, vol. 8, pp. 348–362, 2012.

[8] J. Yoo and A. Aksimentiev, “New tricks for old dogs: Improving the accuracy of biomolecular force fields by pair-specific corrections to non-bonded interactions,” *Phys. Chem. Chem. Phys.*, vol. 20, pp. 8432–8449, 2018.

[9] W. Humphrey, A. Dalke, and K. Schulten, “VMD: Visual molecular dynamics,” *J. Mol. Graph.*, vol. 14, no. 1, pp. 33–38, 1996.

[10] D. R. Roe and T. E. Cheatham III, “PTRAJ and CPPTRAJ: Software for processing and analysis of molecular dynamics trajectory data,” *J. Chem. Theory Comput.*, vol. 9, no. 7, pp. 3084–3095, 2013.

Amaro, Rommie

[1] C.–Y. Wu *et al.*, “Influenza A surface glycosylation and vaccine design,” *Proc. Natl. Acad. Sci. USA*, vol. 114, no. 2, pp. 280–285, Jan. 2017.

[2] M. O. Altman *et al.*, “Human influenza A virus hemagglutinin glycan evolution follows a temporal pattern to a glycan limit,” *MBio*, vol. 10, no. 2, pp. 1–15, Apr. 2019.

[3] J. D. Durrant, R. M. Bush, and R. E. Amaro, “Microsecond molecular dynamics simulations of influenza neuraminidase suggest a mechanism for the increased virulence of stalk-deletion mutants,” *J. Phys. Chem. B*, vol. 120, no. 33, pp. 8590–8599, Aug. 2016.

[4] R. E. Amaro *et al.*, “A computational assay that explores the hemagglutinin/neuraminidase functional balance reveals the neuraminidase secondary site as a novel anti-influenza target,” *ACS Cent. Sci.*, vol. 4, no. 11, pp. 1570–1577, Nov. 2018.

[5] A. Harris *et al.*, “Influenza virus pleiomorphy characterized by cryoelectron tomography,” *Proc. Natl. Acad. Sci. USA*, vol. 103, no. 50, pp. 19123–19127, Dec. 2006.

[6] R. Danne *et al.*, “doGlycans—Tools for preparing carbohydrate structures for atomistic simulations of glycoproteins, glycolipids, and carbohydrate polymers for GROMACS,” *J. Chem. Inf. Model.*, vol. 57, no. 10, pp. 2401–2406, Oct. 2017.

[7] J. C. Phillips *et al.*, “Scalable molecular dynamics with NAMD,” *J. Comput. Chem.*, vol. 26, no. 16, pp. 1781–1802, Dec. 2005.

[8] J. Huang and A. D. MacKerell, “CHARMM36 all-atom additive protein force field: Validation based on comparison to NMR data,” *J. Comput. Chem.*, vol. 34, no. 25, pp. 2135–2145, Sept. 2013.

Bernardi, Rafael C.

[1] L. F. Milles, K. Schulten, H. E. Gaub, and R. C. Bernardi, “Molecular mechanism of extreme mechanostability in a pathogen adhesion,” *Science*, vol. 359, no. 6383, pp. 1527–1533, Mar. 2018.

Dickerson, Julie

[1] A. Dobin *et al.*, “STAR: Ultrafast universal RNA-seq aligner,” *Bioinformatics*, vol. 29, pp. 15–21, 2013.

[2] M. Pertea, D. Kim, G. M. Pertea, J. T. Leek, and S. L. Salzberg, “Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown,” *Nat. Protoc.*, vol. 11, pp. 1650–1667, 2016.

Dill, Ken A.

[1] A. Perez, J. L. MacCallum, and K. A. Dill, “Accelerating molecular simulations of proteins using Bayesian inference on weak information,” *Proc. Natl. Acad. Sci. USA*, vol. 112, no. 38, pp. 11846–11851, 2015.

[2] J. L. MacCallum, A. Perez, and K. A. Dill, “Determining protein structures by combining semi-reliable data with atomistic physical models by Bayesian inference,” *Proc. Natl. Acad. Sci. USA*, vol. 112, no. 22, pp. 6985–6990, 2015.

[3] A. Perez, J. A. Morrone, E. Brini, J. L. MacCallum, and K. A. Dill, “Blind protein structure prediction using accelerated free-energy simulations,” *Sci. Adv.*, vol. 2, no.11, p. e1601274, 2016.

[4] A. Perez, F. Sittel, G. Stock, and K. Dill, “Meld-path efficiently computes conformational transitions, including multiple and diverse paths,” *J. Chem. Theory Comput.*, vol.14, no. 4, pp. 2109–2116, 2018.

[5] P. Eastman *et al.*, “OpenMM 7: Rapid development of high performance algorithms for molecular dynamics,” *PLoS Comput. Biol.*, vol. 13, no. 7, p. e1005659, 2017.

[6] J. C. Robertson *et al.*, “NMR-assisted protein structure prediction with MELD x MD,” *Proteins*, in revision, 2019.

[7] J. A. Maier, C. Martinez, K. Kasavajhala, L. Wickstrom, K. E. Hauser, and C. Simmerling, “ff14sb: Improving the accuracy of protein side chain and backbone parameters from ff99sb,” *J. Chem. Theory Comput.*, vol. 11, no. 8, pp. 3696–3713, 2015.

[8] H. Nguyen, D. R. Roe, and C. Simmerling, “Improved generalized born solvent model parameters for protein simulations,” *J. Chem. Theory Comput.*, vol. 9, no. 4, pp. 2020–2034, 2013.

[9] J. C. Robertson, A. Perez, and K. A. Dill, “MELD x MD folds nonthreadables, giving native structures and populations,” *J. Chem. Theory Comput.*, vol. 14, no.12, pp. 6734–6740, 2018.

[10] J. A. Morrone, A. Perez, J. L. MacCallum, and K. A. Dill, “Computed binding of peptides to proteins with MELD-accelerated molecular dynamics,” *J. Chem. Theory Comput.*, vol. 13, no. 2, pp. 870–876, 2017.

[11] J. A. Morrone *et al.*, “Molecular simulations identify binding poses and approximate affinities of stapled  $\alpha$ -helical peptides to MDM2 and MDMX,” *J. Chem. Theory Comput.*, vol. 13, no. 2, pp. 863–869, 2017.

[12] E. Brini, D. Kozakov, and K. A. Dill, “Predicting protein dimer structures using MELD x MD,” *J. Chem. Theory Comput.*, 2019.

Caetano–Anollés, Gustavo

[1] J. Espadaler, E. Querol, F. X. Aviles, and B. Oliva, “Identification of function-associated loop motifs and application to protein function prediction,” *Bioinformatics*, vol. 22, no. 18, pp. 2237–43, Jul. 2006.

[2] J. A. Marsh and S. A. Teichmann, “Protein flexibility facilitates quaternary structure assembly and evolution,” *PLoS Biol.*, vol. 12, no. 5, p. e1001870, May 2014.

[3] M. F. Aziz, K. Caetano-Anollés, and G. Caetano-Anollés, “The early history and emergence of molecular functions and modular scale-free network behavior,” *Sci. Rep.*, p. 25058, Apr. 2016.

[4] A. D. Goldman, J. A. Baross, and R. Samudrala, “The enzymatic and metabolic capabilities of early life,” *PLoS One*, vol. 7, no. 9, p. e39912, Jan. 2012.

[5] L. Skjærven, X.-Q. Yao, G. Scarabelli, and B. J. Grant, “Integrating protein structural dynamics and evolutionary analysis with Bio3D,” *BMC Bioinformatics*, vol. 15, no. 1, p. 399, Jan. 2014.

[6] G. Csárdi, “The igraph software package for complex network research,” *InterJournal*, Complex Systems, p. 1695, 2006.

[7] M. L. Kendall, M. Boyd, and C. Colijn, “PhyloTop: Calculating topological properties of phylogenies,” 2018. [Online]. Available: <https://rdrr.io/cran/phyloTop>.

[8] K. M. Kim *et al.*, “Protein domain structure uncovers the origin of aerobic metabolism and the rise of planetary oxygen,” *Structure*, vol. 20, no. 1, pp. 67–76, Jan. 2012.

[9] N. A. O’Leary *et al.*, “Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation,” *Nucleic Acids Res.*, vol. 44, no. D1, pp. D733–D745, Jan. 2016.

[10] Y. Bao *et al.*, “National Center for Biotechnology Information viral genomes project,” *J. Virol.*, vol. 78, no. 14, pp. 7291–7298, Jul. 2004.

[11] J. Gough and C. Chothia, “SUPERFAMILY: HMMs representing all proteins of known structure. SCOP sequence searches, alignments and genome assignments,” *Nucleic Acids Res.*, vol. 30, no. 1, pp. 268–272, Jan. 2002.

[12] J. Mittenenthal, D. Caetano-Anollés, and G. Caetano-Anollés, “Biphasic patterns of diversification and the emergence of modules,” *Front. Genet.*, vol. 3, p. 147, Aug. 2012.

[13] C. Colijn and J. Gardy, “Phylogenetic tree shapes resolve disease transmission patterns,” *Evol. Med. Public Health*, vol. 2014, no. 1, pp. 96–108, Jun. 2014.

[14] G. Plazzotta, C. Kwan, M. Boyd, and C. Colijn, “Effects of memory on the shapes of simple outbreak trees,” *Sci. Rep.*, vol. 6, no. 1, p. 21159, Aug. 2016.

Clancy, Colleen E.

[1] J. I. Vandenberg, M. D. Perry, M. J. Perrin, S. A. Mann, Y. Ke, and A. P. Hill, “hERG K<sup>+</sup> channels: structure, function, and clinical significance,” *Physiol. Rev.*, vol. 92, no. 3, pp. 1393–478, Jul 2012.

[2] E. R. Behr and D. Roden, “Drug-induced arrhythmia: pharmacogenomic prescribing?,” *Eur. Heart J.*, vol. 34, no. 2, pp. 89–95, Jan. 2013.

[3] W. Jiang *et al.*, “Generalized scalable multiple copy algorithms for molecular dynamics simulations in NAMD,” *Comput. Phys. Comm.*, vol. 185, no. 3, pp. 908–916, 2014.

[4] W. Wang and R. MacKinnon, “Cryo-EM structure of the open human ether-á-go-go-related K<sup>+</sup> channel hERG,” *Cell*, vol. 169, no. 3, pp. 422–430, e10, Apr. 20, 2017.

[5] D. A. Kopfer *et al.*, “A molecular switch driving inactivation in the cardiac K<sup>+</sup> channel HERG,” *PLoS One*, vol. 7, no. 7, p. e41023, 2012.

[6] V. Pau, Y. Zhou, Y. Ramu, Y. Xu, and Z. Lu, “Crystal structure of an inactivated mutant mammalian voltage-gated K(+) channel,” *Nat. Struct. Mol. Biol.*, vol. 24, no. 10, pp. 857–865, Oct. 2017.

[7] M. J. Perrin, P. W. Kuchel, T. J. Campbell, and J. I. Vandenberg, “Drug binding to the inactivated state is necessary but not sufficient for high-affinity binding to human ether- á-go-go-related gene channels,” *Mol. Pharmacol.*, vol. 74, no. 5, pp. 1443–1452, Nov. 2008.

[8] V. Vijayvergiya, S. Acharya, J. Poulos, and J. Schmidt, “Single channel and ensemble hERG conductance measured in droplet bilayers,” *Biomed. Microdevices*, vol. 17, no. 1, p. 12, Feb. 2015.

[9] M. Weerapura, T. E. Hebert, and S. Nattel, “Dofetilide block involves interactions with open and inactivated states of HERG channels,” *Pflügers Arch.*, vol. 443, no. 4, pp. 520–531, Feb. 2002.



[10] P.–C. Yang *et al.*, "A computational pipeline to predict cardiotoxicity: From the atom to the rhythm," *bioRxiv*, p. 635433, 2019.

El–Kebir, Mohammed

[1] J. N. Weinstein *et al.*, "The cancer genome atlas pan-cancer analysis project," *Nat. Genet.*, vol. 45, no. 10, pp. 1113–1120, 2013, doi: 10.1038/ng.2764.

Gazzola, Mattia

[1] M. Gazzola, P. Chatelain, W. van Rees, and P. Koumoutsakos, "Simulations of single and multiple swimmers with non-divergence free deforming geometries," *J. Computat. Phys.*, vol. 230, pp. 7093–7114, 2011.

[2] M. Gazzola, B. Hejazialhosseini, and P. Koumoutsakos, "Reinforcement learning and wavelet adapted vortex methods for simulations of self-propelled swimmers," *SIAM J. Sci. Comput.*, vol. 36, pp. B622–B639, 2014.

[3] M. Gazzola, L. Dudte, A. McCormick, and L. Mahadevan, "Forward and inverse problems in the mechanics of soft filaments," *Royal Soc. Open Sci.*, vol. 5, no. 6, 2018, doi: 10.1098/rsos.171628.

[4] Parthasarathy, T., Chan, F., & Gazzola, M. (2019). Streaming-enhanced flow-mediated transport. *Journal of Fluid Mechanics*, 878, 647–662. doi:10.1017/jfm.2019.643

[5] Zhang, X., Chan, F.K., Parthasarathy, T. et al. Modeling and simulation of complex dynamic musculoskeletal architectures. *Nat Commun* 10, 4825 (2019). <https://doi.org/10.1038/s41467-019-12759-5>

[6] G. J. Pagan–Diaz *et al.*, "Simulation and fabrication of stronger, larger and faster walking biohybrid machines," *Adv. Funct. Mat.*, vol. 28, no. 23, 2018, doi: 10.1002/adfm.201801145.

[7] Zhang, Xiaotian & Nuethong, Sittinon & Pagan, Gelson & Gazzola, Mattia. (2019). Neuromuscular actuation of biohybrid motile bots. *Proceedings of the National Academy of Sciences*. 116. 201907051. 10.1073/pnas.1907051116.

Hadden, Jodi A. (1)

[1] J. A. Hadden and J. R. Perilla, "All-atom simulations of viruses," *Curr. Opin. Virol.*, vol. 31, no. 8, pp. 82–91, Aug. 2018, doi: 10.1016/j.coviro.2018.08.007.

[2] J. R. Perilla, J. A. Hadden, B. C. Goh, C. G. Mayne, and K. Schulten, "All-atom molecular dynamics of virus capsids as drug targets," *J. Phys. Chem. Lett.*, vol. 7, no. 10, pp. 1836–1844, Apr. 2016, doi: 10.1021/acs.jpclett.6b00517.

[3] J. C. Phillips *et al.*, "Scalable molecular dynamics with NAMD," *J. Comput. Chem.*, vol. 26, no. 16, pp. 1781–1802, Dec. 2005, doi: 10.1002/jcc.20289.

[4] J. A. Hadden, J. R. Perilla, C. J. Schlicksup, B. Venkatakrishnan, A. Zlotnick, and K. Schulten, "All-atom molecular dynamics of the HBV capsid reveals insights into biological function and cryo-EM resolution limits," *eLife*, no. 7, 2018, doi: 10.7554/eLife.32478.

[5] L. Ruan, J. A. Hadden, and A. Zlotnick, "Assembly properties of hepatitis B virus core protein mutants correlate with their resistance to assembly-directed antivirals," *J. Virol.*, vol. 92, no. 20, Sept. 2018, doi: 10.1128/JVI.01082-18.

Hadden, Jodi A. (2)

[1] J. A. Hadden and J. R. Perilla, "All-atom simulations of viruses," *Curr. Opin. Virol.*, vol. 31, no. 8, pp. 82–91, Aug. 2018, doi: 10.1016/j.coviro.2018.08.007.

[2] J. R. Perilla, J. A. Hadden, B. C. Goh, C. G. Mayne, and K. Schulten, "All-atom molecular dynamics of virus capsids as drug targets," *J. Phys. Chem. Lett.*, vol. 7, no. 10, pp. 1836–1844, Apr. 2016, doi: 10.1021/acs.jpclett.6b00517.

[3] J. C. Phillips *et al.*, "Scalable molecular dynamics with NAMD," *J. Comput. Chem.*, vol. 26, no. 16, pp. 1781–1802, Dec. 2005, doi: 10.1002/jcc.20289.

[4] J. A. Hadden, J. R. Perilla, C. J. Schlicksup, B. Venkatakrishnan, A. Zlotnick, and K. Schulten, "All-atom molecular dynamics of the HBV capsid reveals insights into biological function and cryo-EM resolution limits," *eLife*, no. 7, 2018, doi: 10.7554/eLife.32478.

[5] L. Ruan, J. A. Hadden, and A. Zlotnick, "Assembly properties of hepatitis B virus core protein mutants correlate with their resistance to assembly-directed antivirals," *J. Virol.*, vol. 92, no. 20, Sept. 2018, doi: 10.1128/JVI.01082-18.

Hirata, So

[1] S. Hirata, A. E. Doran, P. J. Knowles, and J. V. Ortiz, "One-particle many-body Green's function theory: Algebraic recursive definitions, linked-diagram theorem, irreducible-diagram theorem, and general-order algorithms," *J. Chem. Phys.*, vol. 147, p. 044108, 2018.

[2] S. Y. Willow, K. S. Kim, and S. Hirata, "Stochastic evaluation of second-order Dyson self-energies," *J. Chem. Phys.*, vol. 138, p. 164111, 2013.

[3] A. E. Doran and S. Hirata, "Monte Carlo second- and third-order many-body Green's function methods with frequency-dependent, non-diagonal self-energy," *Journal of Chemical Theory and Computation*. PMID 31580066 DOI: 10.1021/acs.jctc.9b00693 .

[4] C. M. Johnson, A. E. Doran, S. L. Ten-no, and S. Hirata, "Monte Carlo explicitly correlated second-order many-body Green's function theory," *J. Chem. Phys.*, vol. 149, p. 174112, 2018.

[5] S. Y. Willow, M. R. Hermes, K. S. Kim, and S. Hirata, "Convergence acceleration of parallel Monte Carlo second-order many-body perturbation calculations using redundant walkers," *J. Chem. Theory Comput.*, vol. 9, pp. 4396–4402, 2013.

[6] A. E. Doran and S. Hirata, "Monte Carlo MP2 on many graphical processing units," *J. Chem. Theory Comput.*, vol. 12, pp. 4821–4832, 2016.

Jiang, Tao

[1] C. Hartzell, I. Putzier, and J. Arreola, "Calcium-activated chloride channels," *Annu. Rev. Physiol.*, vol. 67, no. 1, pp. 719–758, Mar. 2005.

[2] N. Pedemonte and L. J. Galletta, "Structure and function of TMEM16 proteins (anoctamins)," *Physiol. Rev.*, vol. 94, no. 2, pp. 419–459, Apr. 2014.

[3] C. Duran and H. C. Hartzell, "Physiological roles and diseases of TMEM16/anoctamin proteins: are they all chloride channels?" *Acta. Pharmacol. Sin.*, vol. 32, no. 6, pp. 685–692, Jun. 2011.

[4] B. C. Suh and B. Hille, "PIP<sub>2</sub> is a necessary cofactor for ion channel function: how and why?" *Annu. Rev. Physiol.*, vol. 37, no. 1, pp. 175–195, Jun. 2008.

[5] S. B. Hansen, "Lipid agonism: the PIP<sub>2</sub> paradigm of ligand-gated ion channels," *Biochim. Biophys. Acta.*, vol. 1851, no. 5, pp. 620–628, May 2015.

[6] D. E. Logothetis *et al.*, "Phosphoinositide control of membrane protein function: a frontier led by studies on ion channels," *Annu. Rev. Physiol.*, vol. 77, no. 1, pp. 81–104, Feb. 2015.

[7] B. Hille, E. J. Dickson, M. Kruse, O. Vivas, and B. C. Suh, "Phosphoinositides regulate ion channels," *Biochim. Biophys. Acta.*, vol. 1851, no. 6, pp. 844–856, Jun. 2015.

[8] J. J. De Jesus–Perez *et al.*, "Phosphatidylinositol 4,5-bisphosphate, cholesterol, and fatty acids modulate the calcium-activated chloride channel TMEM16A (ANO1)," *Biochim. Biophys. Acta. Mol. Cell. Biol. Lipids*, vol. 1863, no. 3, pp. 299–312, Mar. 2018.

[9] C. M. Ta, K. E. Acheson, N. J. G. Rorsman, R. C. Jongkind, and P. Tammaro, "Contrasting effects of phosphatidylinositol 4,5-bisphosphate on cloned TMEM16A and TMEM16B channels," *Br. J. Pharmacol.*, vol. 174, no. 18, pp. 2984–2999, Sept. 2017.

[10] H. A. Pritchard, N. Leblanc, A. P. Albert, and I. A. Greenwood, "Inhibitory role of phosphatidylinositol 4,5-bisphosphate on TMEM16A-encoded calcium-activated chloride channels in rat pulmonary artery," *Br. J. Pharmacol.*, vol. 171, no. 18, pp. 4311–4321, Sept. 2014.

[11] C. Paulino, V. Kalienkova, A. K. M. Lam, Y. Neldner, and R. Dutzler, "Activation mechanism of the calcium-activated chloride channel TMEM16A revealed by cryo-EM," *Nature*, vol. 552, no. 7685, pp. 421–425, Dec. 2017.

[12] J. V. Vermaas, T. V. Pogorelov, and E. Tajkhorshid, "Extension of the Highly Mobile Membrane Mimetic to transmembrane systems through customized in silico solvents," *J. Phys. Chem. B*, vol. 121, no. 15, pp. 3764–3776, Apr. 2017.

[13] J. C. Phillips *et al.*, "Scalable molecular dynamics with NAMD," *J. Comput. Chem.*, vol. 26, no. 16, pp. 1781–1802, Dec. 2005.

Khalili-Araghi, Fatemeh

[1] M. Furuse, H. Sasaki, and S. Tsukita, "Manner of interaction of heterogeneous claudin species within and between tight junction strands," *J. Cell Biol.*, vol. 147, no. 4, pp. 891–903, Nov. 1999, doi: 10.1083/jcb.147.4.891.

[2] C. R. Weber *et al.*, "Claudin-2Dependent paracellular channels are dynamically gated," *eLife*, vol. 4, no. e09906, Nov. 2015, doi: 10.7554/eLife.09906.001.

[3] H. Suzuki *et al.*, "Crystal structure of a claudin provides insight into the architecture of tight junctions," *Science*, vol. 344, no. 6181, pp. 304–307, Apr. 2014, doi: 10.1126/science.1248571.

[4] H. Suzuki, K. Tani, A. Tamura, S. Tsukita, and Y. Fujiyoshi, "Model for the architecture of claudin-based paracellular ion channels through tight junctions," *J. Mol. Biol.*, vol. 427, no. 2, pp. 291–297, Jan. 2015, doi: 10.1016/j.jmb.2014.10.020.

[5] P. Samanta *et al.*, "Molecular determination of claudin-15 organization and channel selectivity," *J. Gen. Physiol.*, vol. 150, no. 7, pp. 949–968, Jun. 2018, doi: 10.1085/jgp.201711868.

Makri, Nancy

[1] R. Lambert and N. Makri, "Quantum-classical path integral. I. Classical memory and weak quantum nonlocality," *J. Chem. Phys.*, vol. 137, Art. no. 22A552, 2012.

[2] R. Lambert and N. Makri, "Quantum-classical path integral. II. Numerical methodology," *J. Chem. Phys.*, vol. 137, Art. no. 22A553, 2012.

[3] N. Makri, "Quantum-classical path integral: A rigorous approach to condensed-phase dynamics," *Int. J. Quantum Chem.* (invited perspective), vol. 115, pp. 1209–1214, 2015.

[4] P. L. Walters and N. Makri, "Iterative quantum-classical path integral with dynamically consistent state hopping," *J. Chem. Phys.*, vol. 144, Art. no. 044108, 2016.

[5] N. Makri, "Blip-summed quantum-classical path integral with cumulative memory," *Faraday Discuss.*, vol. 195, pp. 81–92, 2016.

[6] P. L. Walters and N. Makri, "Quantum-classical path integral simulation of ferrocene-ferrocenium charge transfer in liquid hexane," *J. Phys. Chem. Lett.*, vol. 6, pp. 4959–4965, 2015.

[7] F. Wang and N. Makri, "Quantum-classical path integral with a harmonic treatment of the back-reaction," *J. Chem. Phys.*, vol. 150, Art. no. 184102, 2019.

[8] N. Makri, "Communication: Modular path integral: Quantum dynamics via sequential necklace linking," *J. Chem. Phys.*, vol. 148, Art. no.101101, 2018.

[9] N. Makri, "Modular path integral methodology for real-time quantum dynamics," *J. Chem. Phys.*, vol. 149, Art. no. 204108, 2018.

[10] S. Kundu and N. Makri, "Modular path integral for discrete systems with non-diagonal couplings," *J. Chem. Phys.*, vol. 151, Art. no. 074110, 2019.



Masud, Arif

[1] M. Anand, J. Kwack, and A. Masud, “A new generalized Oldroyd-B model for blood flow in complex geometries,” *Int. J. Eng. Sci.*, vol. 72, pp. 78–88, 2013.

[2] P. Chen, T. J. Truster, and A. Masud, “Interfacial stabilization at finite strains for weak and strong discontinuities in multi-constituent materials,” *Comput. Methods Appl. Mech. Eng.*, vol. 328, pp. 717–751, 2018.

[3] J. Kwack and A. Masud, “A three-field formulation for incompressible viscoelastic fluids,” *Int. J. Eng. Sci.*, vol. 48, pp. 1413–1432, 2010.

[4] J. Kwack, A. Masud, and K. R. Rajagopal, “Stabilized mixed three-field formulation for a generalized incompressible Oldroyd-B model,” *Int. J. Numer. Methods Fluids*, vol. 83, no. 9, pp. 704–734, 2017.

[5] T. J. Truster and A. Masud, “Primal interface formulation for coupling multiple PDEs: A consistent derivation via the Variational Multiscale method,” *Comput. Methods Appl. Mech. Eng.*, vol. 268, pp. 194–224, 2014.

[6] T. J. Truster and A. Masud, “A unified mixture formulation for density and volumetric growth of multi-constituent solids in tissue engineering,” *Comput. Methods Appl. Mech. Eng.*, vol. 314, pp. 222–268, 2017.

[7] T. J. Truster and A. Masud, “A discontinuous/continuous Galerkin method for modeling of interphase damage in fibrous composite systems,” *Comput. Mech.*, vol. 52, no. 3, pp. 499–514, 2013.

Moore, Jeffery S.

[1] G. G. Glenner and C. W. Wong, “Alzheimer’s disease: Initial report of the purification and characterization of a novel cerebrovascular amyloid protein,” *Biochem. Biophys. Res. Commun.*, vol. 120, pp. 885–890, 1984.

[2] Y. Song, E. G. Moore, Y. Guo, and J. S. Moore, “Polymer–peptide conjugates disassemble amyloid fibrils in a molecular-weight dependent manner,” *J. Am. Chem. Soc.*, vol. 139, pp. 4298–4301, 2017.

[3] Y. Song, P.–N. Cheng, L. Zhu, E. G. Moore, and J. S. Moore, “Multivalent macromolecules redirect nucleation-dependent fibrillar assembly into discrete nanostructures,” *J. Am. Chem. Soc.*, vol. 136, pp. 5233–5236, 2014.

[4] J. C. Phillips *et al.*, “Scalable molecular dynamics with NAMD,” *J. Comput. Chem.*, vol. 26, no. 16, pp. 1781–1802, 2005.

[5] W. Humphrey, A. Dalke, and K. Schulten, “VMD: Visual molecular dynamics,” *J. Mol. Graph.*, vol. 14, no. 1, pp. 33–38, 1996.

Moradi, Mahmoud (baqz)

[1] I. T. Paulsen and R. A. Skurray, “The POT family of transport proteins,” *Trends Biochem. Sci.*, vol. 19, no. 10, p. 404, 1994.

[2] K. Ito *et al.*, “Analysing the substrate multispecificity of a proton-coupled oligopeptide transporter using a dipeptide library,” *Nat. Commun.*, vol. 4, pp. 2502, 2013.

[3] S. Newstead, “Molecular insights into proton coupled peptide transport in the PTR family of oligopeptide transporters,” *Biochim. Biophys. Acta, Gen. Subj.*, vol. 1850, pp. 488–499, 2015.

[4] D. E. Smith, B. Cleméncon, and M. A. Hediger, “Proton-coupled oligopeptide transporter family SLC15: Physiological, pharmacological and pathological implications,” *Mol. Aspects Med.*, vol. 34, pp. 323–336, 2013.

[5] J. Rautio *et al.*, “Prodrugs: design and clinical applications,” *Nat. Rev. Drug Discovery*, vol. 7, pp. 255–270, 2008.

[6] N. Solcan, *et al.* “Alternating access mechanism in the POT family of oligopeptide transporters,” *EMBO J.*, vol. 31, pp. 3411–3421, 2012.

[7] S. Newstead *et al.*, “Crystal structure of a prokaryotic homologue of the mammalian oligopeptide-proton symporters, PepT1 and PepT2,” *EMBO J.*, vol. 30, pp. 417–426, 2011.

[8] F. Guettou *et al.*, “Structural insights into substrate recognition in proton-dependent oligopeptide transporters,” *EMBO Rep.*, vol. 14, pp. 804–810, 2013.

[9] J. A. Lyons *et al.*, “Structural basis for polyspecificity in the POT family of proton-coupled oligopeptide transporters,” *EMBO Rep.*, vol. 15, 886–893, 2014.

[10] S. Doki *et al.*, “Structural basis for dynamic mechanism of proton-coupled symport by the peptide transporter POT,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 110, pp. 11343–11348, 2013.

[11] K. Immadisetty, J. Hettige, and M. Moradi, “What can and cannot be learned from molecular dynamics simulations of bacterial proton-coupled oligopeptide transporter GkPOT?” *J. Phys. Chem. B.*, vol. 121, no. 15, pp. 3644–3656, 2017, doi: 10.1021/acs.jpcc.6b09733.

[12] M. Moradi and E. Tajkhorshid, “Computational recipe for efficient description of large-scale conformational changes in biomolecular systems,” *J. Chem. Theory Comput.*, vol. 10, pp. 2866–2880, 2014.

[13] M. Moradi and E. Tajkhorshid, “Mechanistic picture for conformational transition of a membrane transporter at atomic resolution,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 110, pp. 18916–18921, 2013.

[14] M. Moradi, G. Enkavi, and E. Tajkhorshid, “Atomic-level characterization of transport cycle thermodynamics in the glycerol-3-phosphate:phosphate antiporter,” *Nat. Commun.*, vol. 6, pp. 8393, 2015.

[15] A. Fakharzadeh and M. Moradi, “Effective Riemannian diffusion model for conformational dynamics of biomolecular systems,” *J. Phys. Chem. Lett.*, vol. 7, pp. 4980–4987, 2016.

[16] D. Ogden, K. Immadisetty, and M. Moradi, “Conformational transition pathways in major facilitator superfamily transporters,” *bioRxiv* 708289, 2019, doi: 10.1101/708289

Moradi, Mahmoud

[1] B. Martinac, M. Buechner, A. H. Delcour, J. Adler, and C. Kung, “Pressure-sensitive ion channel in *Escherichia coli*,” *Proc. Nat. Acad. Sci. USA*, vol. 84, pp. 2297–2301, 1987.

[2] E. Haswell, R. Phillips, and D. Rees, “Mechanosensitive channels: What can they do and how do they do it?” *Structure*, vol. 19, pp. 1356–1369, 2011.

[3] I. Iscla and P. Blount, “Sensing and responding to membrane tension: The bacterial MscL channel as a model system,” *Biophys. J.*, vol. 103, pp. 169–174, 2012.

[4] I. Iscla, R. Wray, S. Wei, B. Posner, and P. Blount, “Streptomycin potency is dependent on MscL channel expression,” *Nat. Commun.*, vol. 5, p. 4891, 2014.

[5] A. Kocer, M. Walko, W. Mieijberg, and B. L. Feringa, “A light-actuated nanovalve derived from a channel protein,” *Science*, vol. 309, pp. 755–758, 2005.

[6] A. Kocer, M. Walko, and B. L. Feringa, “Synthesis and utilization of reversible and irreversible light-activated nanovalves derived from the channel protein MscL,” *Nat. Prot.*, vol. 2, pp. 1426–1437, 2007.

[7] J. Pacheco–Torres *et al.*, “Image guided drug release from pH-sensitive ion channel-functionalized stealth liposomes into an *in vivo* glioblastoma model,” *Nanomedicine: NBM*, vol. 11, pp. 1345–1354, 2015.

[8] L.–M. Yang *et al.*, “Engineering a pH-sensitive liposomal MRI agent by modification of a bacterial channel,” *Small*, vol. 14, Art. no. 1704256, 2017.

[9] S. Steinbacher, R. B. Bass, P. Strop, and D. C. Rees, “Structures of the prokaryotic mechanosensitive channels MscL and MscS,” *Current Topics in Membranes, Part A*, vol. 58, pp. 1–24, 2007.

[10] E. Perozo, D. M. Cortes, P. Somporpisut, A. Kloda, and B. Martinac, “Open channel structure of MscL and the gating mechanism of mechanosensitive channels,” *Nature*, vol. 418, pp. 942–948, 2002.

[11] M. Louhivuori, H. J. Risselada, E. van der Giessen, and S. J. Marrink, “Release of content through mechanosensitive gates in pressurized liposomes,” *Proc. Nat. Acad. Sci. USA*, vol. 107, pp. 19856–19860, 2010.

[12] M. Moradi and E. Tajkhorshid, “Computational recipe for efficient description of large-scale conformational changes in biomolecular systems,” *J. Chem. Theory Comput.*, vol. 10, pp. 2866–2880, 2014.

[13] M. Moradi and E. Tajkhorshid, “Mechanistic picture for conformational transition of a membrane transporter at atomic resolution,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 110, pp. 18916–18921, 2013.

[14] M. Moradi, G. Enkavi, and E. Tajkhorshid, “Atomic-level characterization of transport cycle thermodynamics in the glycerol-3-phosphate:phosphate antiporter,” *Nat. Commun.*, vol. 6, p. 8393, 2015.

[15] A. Fakharzadeh and M. Moradi, “Effective Riemannian diffusion model for conformational dynamics of biomolecular systems,” *J. Phys. Chem. Lett.*, vol. 7, pp. 4980–4987, 2016.

[16] K. Immadisetty, A. Polasa, R. Shelton, and M. Moradi, “Elucidating the molecular basis of pH activation of an engineered mechanosensitive channel,” *bioRxiv* 707794, 2019, doi: 10.1101/707794.

Peterson, Joseph R.

[1] S. C. Partridge *et al.*, “MRI measurements of breast tumor volume predict response to neoadjuvant chemotherapy and recurrence-free survival,” *Am. J. Roentgenol.*, vol. 184, no. 6, pp. 1774–81, Jun. 2005.

[2] N. M. Hylton *et al.*, “Neoadjuvant chemotherapy for breast cancer: Functional tumor volume by MR imaging predicts recurrence-free survival-results from the ACRIN 6657/CALGB 150007 I-SPY 1 trial,” *Radiology*, vol. 279, no. 1, pp. 44–55, Dec. 2016.

[3] X. Li *et al.*, “Multiparametric magnetic resonance imaging for predicting pathological response after the first cycle of neoadjuvant chemotherapy in breast cancer,” *Investig. Radiol.*, vol. 50, no. 4, pp. 195–204, Apr. 2015.

Pogorelov, Taras V.

[1] R. D. Cannon *et al.*, “*Candida albicans* drug resistance another way to cope with stress,” *Microbiol.*, vol. 153, no. 10, pp. 3211–3217, Oct. 2007.

[2] J. Mora–Duarte *et al.*, “Comparison of caspofungin and amphotericin B for invasive candidiasis,” *N. Engl. J. Med.*, vol. 347, no. 25, pp. 2020–2029, Dec. 2002.

[3] J. C. Phillips *et al.*, “Scalable molecular dynamics with NAMD,” *J. Comput. Chem.*, vol. 26, no. 16, pp. 1781–1802, Dec. 2005.

[4] C. L. Mendes *et al.*, “Deploying a large petascale system: The Blue Waters experience,” *Procedia Comput. Sci.*, vol. 29, pp. 198–209, Jun. 2014.

[5] G. H. Bauer, T. Hoefler, W. T. Kramer, and R. A. Fiedler, “Analyses and modeling of applications used to demonstrate sustained petascale performance on Blue Waters,” presented at the CUG 2012, Stuttgart, Germany.

[6] M. J. Arcario, Y. Z. Ohkubo, and E. Tajkhorshid, “Capturing spontaneous partitioning of peripheral proteins using a biphasic membrane-mimetic model,” *J. Phys. Chem. B*, vol. 115, no. 21, pp. 7029–7037, Jun. 2011.

[7] J. V. Vermaas, T. V. Pogorelov, and E. Tajkhorshid, “Extension of the highly mobile membrane mimetic to transmembrane systems through customized *in silico* solvents,” *J. Phys. Chem. B*, vol. 121, no. 15, pp. 3764–3776, Feb. 2017.

[8] Y. Z. Ohkubo, T. V. Pogorelov, M. J. Arcario, G. A. Christensen, and E. Tajkhorshid, “Accelerating membrane insertion of peripheral proteins with a novel membrane mimetic model,” *Biophys. J.*, vol. 102, no. 9, pp. 2130–2139, May 2012.

[9] T. V. Pogorelov, J. V. Vermaas, M. J. Arcario, and E. Tajkhorshid, “Partitioning of amino acids into a model membrane: capturing the interface,” *J. Phys. Chem. B*, vol. 118, pp. 1481–1492, 2014.



[10] Y. Qi *et al.*, “CHARMM–GUI HMMM builder for membrane simulations with the highly mobile membrane-mimetic model,” *Biophys. J.*, vol. 109, no. 10, pp. 2012–2022, Nov. 2015.

[11] T.M. Anderson *et al.*, “Amphotericin forms an extramembranous and fungicidal sterol sponge,” *Nat. Chem. Biol.*, vol. 10, no. 5, pp. 400–406, May 2014.

Prather, Kimberly

[1] T. H. Bertram, R. E. Cochran, V. H. Grassian, and E. A. Stone, “Sea spray aerosol chemical composition: elemental and molecular mimics for laboratory studies of heterogeneous and multiphase reactions,” *Chem. Soc. Rev.*, vol. 47, no. 7, pp. 2374–2400, Apr. 2018.

[2] L. Martínez, R. Andrade, E. G. Birgin, and J. M. Martínez, “PACKMOL: A package for building initial configurations for molecular dynamics simulations,” *J. Comput. Chem.*, vol. 30, no. 13, pp. 2157–2164, Oct. 2009.

[3] J. Huang and A. D. Mackerell, “CHARMM36 all-atom additive protein force field: Validation based on comparison to NMR data,” *J. Comput. Chem.*, vol. 34, no. 25, pp. 2135–2145, 2013.

[4] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. L. Klein, “Comparison of simple potential functions for simulating liquid water,” *J. Chem. Phys.*, vol. 79, no. 2, pp. 926–935, Jul. 1983.

[5] J. C. Phillips *et al.*, “Scalable molecular dynamics with NAMD,” *J. Comput. Chem.*, vol. 26, no. 16, pp. 1781–1802, Dec. 2005.

Shukla, Diwakar (1)

[1] K. H. Liu, C. Y. Huang, and Y. F. Tsay, “CHL1 is a dual-affinity nitrate transporter of Arabidopsis involved in multiple phases of nitrate uptake,” *Plant Cell*, vol. 11, no. 5, pp. 865–874, May 1999.

[2] C.H. Ho, S. H. Lin, H.C. Hu, and Y. F. Tsay, “CHL1 functions as a nitrate sensor in plants,” *Cell*, vol. 138, no. 6, pp. 1184–1194, Sept. 2009.

[3] J. Sun *et al.*, “Crystal structure of the plant dual-affinity nitrate transporter NRT1.1,” *Nature*, vol. 507, no. 7490, pp. 73–77, Mar. 2014.

[4] J. L. Parker and S. Newstead, “Molecular basis of nitrate uptake by the plant nitrate transporter NRT1.1,” *Nature*, vol. 507, no. 7490, pp. 68–72, Mar. 2014.

[5] D. Case *et al.*, *AMBER14*, 2014.

[6] R. T. McGibbon *et al.*, “MDTraj: A Modern Open Library for the Analysis of Molecular Dynamics Trajectories,” *Biophys. J.*, vol. 109, no. 8, pp. 1528–1532, Oct. 2015.

[7] D. R. Roe and T. E. Cheatham III, “PTRAJ and CPPTRAJ: software for processing and analysis of molecular dynamics trajectory data,” *J. Chem. Theory Comput.*, vol. 9, no. 7, pp. 3084–3095, 2013.

[8] K. A. Beauchamp *et al.*, “MSMBuilder2: Modeling Conformational Dynamics at the Picosecond to Millisecond Scale,” *J. Chem. Theory Comput.*, vol. 7, no. 10, pp. 3412–3419, Oct. 2011.

Shukla, Diwakar (2)

[1] M. Berger, J. A. Gray, and B. L. Roth, “The expanded biology of serotonin,” *Annu. Rev. Med.*, vol. 60, pp. 355–366, 2009.

[2] J. A. Coleman *et al.*, “Serotonin transporter–ibogaine complexes illuminate mechanisms of inhibition and transport,” *Nature*, vol. 569, no. 7754, p.141, 2019.

[3] D. Case *et al.*, AMBER 2018, University of California, San Francisco.

Shukla, Diwakar (3)

[1] A. J. Adams, S. D. Banister, L. Irizarry, J. Trecki, M. Schwartz, and R. Gerona, “‘Zombie’ outbreak caused by the synthetic cannabinoid AMB–FUBINACA in New York,” *New Engl. J. Med.*, vol. 376, no. 3, pp. 235–242, 2017.

[2] S. W. Jung, A. E. Cho, and W. Yu, “Exploring the ligand efficacy of cannabinoid receptor 1 (CB1) using molecular dynamics simulations,” *Sci. Rep.*, vol. 8, no. 1, Art. no. 13787, 2018.

[3] L. Li *et al.*, “Opposite effects of cannabinoid CB1 and CB2 receptors on antipsychotic clozapine-induced cardiotoxicity,” *Br. J. Pharmacol.*, vol. 176, no. 7, pp. 890–905, 2019.

Soltesz, Ivan

[1] N. Carnevale and M. Hines, *The NEURON Book*. Cambridge, UK: Cambridge Univ. Press, 2006.

[2] M. Hines and N. Carnevale, “Computer modeling methods for neurons,” *The Handbook of Brain Theory and Neural Networks*, M. Arbib, Ed. Cambridge, MA, USA: MIT Press, 1995, p. 226–230.

[3] M. Hines and N. Carnevale, “Translating network models to parallel hardware in NEURON,” *J. Neurosci. Methods*, vol. 169, pp. 425–455, 2008.

[4] Y. Senzai and G. Buzsaki, “Physiological properties and behavioral correlates of hippocampal granule cells and mossy cells,” *Neuron*, vol. 93, no. 3, pp. 691–704.e5, 2017, doi: 10.1016/j.neuron.2016.12.011.

[5] A. D. Bui *et al.*, “Dentate gyrus mossy cells control spontaneous convulsive seizures and spatial memory,” *Science*, vol. 359, no. 6377, pp. 787–790, 2018, doi: 10.1126/science.aan4074

[6] M. Bezaire *et al.*, “Interneuronal mechanisms of hippocampal *theta* oscillations in a full-scale model of the rodent CA1 circuit,” *eLife*, vol. 5, p. e18566, 2016, doi: 10.7554/eLife.18566.

[7] L. de Almeida, M. Idiart, and J. E. Lisman, “The input-output transformation of the hippocampal granule cells: from grid cells to place fields,” *J. Neurosci.*, vol. 29, no. 23, pp. 7504–7512, 2009, doi: 10.1523/JNEUROSCI.6048-08.2009.

[8] J. E. Lisman, “Relating hippocampal circuitry to function: recall of memory sequences by reciprocal dentate–CA3 interactions,” *Neuron*, vol. 22, no. 2, pp. 233–242, 1999.

Sotomayor, Marcos

[1] O. J. Harrison *et al.*, “The extracellular architecture of adherens junctions revealed by crystal structures of type I cadherins,” *Structure*, vol. 19, pp. 244–256, 2011.

[2] C. Marcozzi, I. D. Burdett, R. S. Buxton, and A. I. Magee, “Coexpression of both types of desmosomal cadherin and plakoglobin confers strong intercellular adhesion,” *J. Cell Sci.*, 1998.

[3] S. Getsios *et al.*, “Coordinated expression of desmoglein 1 and desmocollin 1 regulates intercellular adhesion,” *Differentiation*, vol. 72, pp. 419–433, 2004.

[4] A. Al–Amoudi and A. S. Frangakis, “Structural studies on desmosomes,” *Biochem. Soc. Trans.*, 2008, doi: 10.1042/bst0360181.

[5] A. Al–Amoudi *et al.*, “The three-dimensional molecular structure of the desmosomal plaque,” *Proc. Natl. Acad. Sci. USA*, vol. 108, pp. 6480–6485, 2011.

[6] D. Garrod, “Desmosomes *in vivo*,” *Dermatol. Res. Pract.*, Art. no. 212439, 2010.

[7] O. J. Harrison *et al.*, “Structural basis of adhesive binding by desmocollins and desmogleins,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 113, pp. 7160–7165, 2016.

[8] W. Humphrey, A. Dalke, and K. Schulten, “VMD: Visual molecular dynamics,” *J. Mol. Graph.*, vol. 14, pp. 33–38, 1996.

[9] J. C. Phillips *et al.*, “Scalable molecular dynamics with NAMD,” *J. Comput. Chem.*, vol. 26, pp. 1781–1802, 2005.

Srinivasan, Ashok

[1] D. Helbing and P. Molnar, “Social force model for pedestrian dynamics,” *Phys. Rev. E*, vol. 51, no. 5, p. 4282, 1995.

[2] M. Nikolić, M. Bierlaire, B. Farooq, and M. de Lapparent, “Probabilistic speed–density relationship for pedestrian traffic,” *Transp. Res. Part B: Methodol.*, vol. 89, pp. 58–81, 2016.

[3] S. Namilae, A. Srinivasan, A. Mubayi, M. Scotch, and R. Pahle, “Self-propelled pedestrian dynamics model: Application to passenger movement and infection propagation in airplanes,” *Physica A*, vol. 465, pp. 248–260, 2017.

[4] S. Chunduri, M. Ghaffari, M. S. Lahijani, A. Srinivasan, and S. Namilae, “Parallel low discrepancy parameter sweep for public health policy,” presented at the IEEE/ACM Int. Symp. Cluster, Cloud, and Grid Comput., 2018

Sutton, Brad P.

[1] 2019 AIM Advocacy Forum Advocates Guide, Alzheimer’s Impact Movement, Mar. 31, 2019. [Online]. Available: <https://alzimpact.org/media/serve/id/5c9b7251489fe>

[2] J. L. Holtrop and B. P. Sutton, “High spatial resolution diffusion weighted imaging on clinical 3 T MRI scanners using multislab spiral acquisitions,” *J Med. Imaging*, vol. 3, no. 2, p. 023501, 2016, doi: 10.1117/1.JMI.3.2.023501.

[3] C. Johnson *et al.*, “Brain MR elastography with multiband excitation and nonlinear motion-induced phase error correction,” *Proc. 24th Ann. Meeting of the Int. Society Magnetic Resonance in Medicine*, Singapore, 2016, p. 1951.

[4] A. Cerjanic *et al.*, “PowerGrid: An open source library for accelerated iterative magnetic resonance image reconstruction,” *Proc. 24th Ann. Meeting of the Int. Society Magnetic Resonance in Medicine*, Singapore, 2016.

[5] J. A. Fessler and B. P. Sutton, “Nonuniform Fast Fourier Transforms using min-max interpolation,” *IEEE Trans. Signal Process.*, vol. 51, no. 2, pp. 560–74, 2003.

Tajkhorshid, Emad (1)

[1] S. M. Sedlak, L. C. Schendel, H. E. Gaub, and R. C. Bernardi, “Streptavidin/biotin: tethering geometry defines unbinding mechanics,” *Sci. Adv.*, <https://advances.sciencemag.org/content/6/13/eaay5999>, 2019.

[2] S. M. Sedlak *et al.*, “Direction matters: Monovalent streptavidin/biotin complex under load,” *Nano Lett.*, vol. 19, no. 6, pp. 3415–3421, Jun. 2019.

[3] C. Schoeler *et al.*, “Ultrastable cellulosome-adhesion complex tightens under load,” *Nat. Commun.*, vol. 5, no. 5635, Dec. 2014.

[4] R. C. Bernardi *et al.*, “Mechanisms of nanonewton mechanostability in a protein complex revealed by molecular dynamics simulations and single-molecule force spectroscopy,” *J. Am. Chem. Soc.*, vol. 141, no. 37, pp. 14752–14763, Sept. 2019.

[5] T. Verdorfer *et al.*, “Combining *in vitro* and *in silico* single-molecule force spectroscopy to characterize and tune cellulosomal scaffoldin mechanics,” *J. Am. Chem. Soc.*, vol. 139, no. 49, pp. 17841–17852, Dec. 2017.

Tajkhorshid, Emad (2)

[1] A. Patwardhan *et al.*, “Building bridges between cellular and molecular structural biology,” *eLife*, vol. 6, p. e25835, 2017.

[2] K. Kurihara *et al.*, “A recursive vesicle-based model protocell with a primitive model cell cycle,” *Nat. Comm.*, vol. 6, p. 8352, 2015.

[3] N. Ichihashi and T. Yomo, “Positive roles of compartmentalization in internal reactions,” *Curr. Op. Cell Bio.*, vol. 22, pp. 12–17, 2014.

[4] S. A. Shaikh *et al.*, “Visualizing functional motions of membrane transporters with molecular dynamics simulations,” *Biochemistry*, vol. 52, pp. 569–587, 2013.



[5] J. Li, P.-C. Wen, M. Moradi, and E. Tajkhorshid, “Computational characterization of structural dynamics underlying function in active membrane transporters,” *Curr. Op. in Struct. Bio.*, vol. 31, pp. 96–105, 2015.

[6] C. G. Mayne *et al.*, “The cellular membrane as a mediator for small molecule interaction with membrane proteins,” *Bio. Biophys. Acta—Biomemb.*, vol. 1058, pp. 2290–2304, 2016.

[7] J. L. Baylon *et al.*, “Atomic-level description of protein-lipid interactions using an accelerated membrane model,” *Bio. Biophys. Acta—Biomemb.*, vol. 1858, pp. 1573–1583, 2016.

[8] M. Moradi, G. Enkavi, and E. Tajkhorshid, “Atomic-level characterization of transport cycle thermodynamics in the glycerol-3-phosphate:phosphate antiporter,” *Nat. Comm.*, vol. 6, p. 8393, 2015.

[9] J. C. Phillips *et al.*, “Scalable molecular dynamics with NAMD,” *J. Comp. Chem.*, vol. 26, pp. 1781–1802, 2005.

[10] J. B. Klauda *et al.*, “Update of the CHARMM all-atom additive forcefield for lipids: Validation on six lipid types,” *J. Phys. Chem. B*, vol. 114, pp. 7830–7843, 2010.

[11] J. Huang *et al.*, “CHARMM36m: An improved forcefield for folded and intrinsically disordered proteins,” *Nat. Meth.*, vol. 14, pp. 71–73, 2017.

[12] T. Darden, D. York, and L. Pederson, “Particle mesh Ewald: An Nlog(N) method for Ewald sums in large systems,” *J. Chem. Phys.*, vol. 98, p. 10089, 1993.

[13] E. A. Rakhmanov, E. B. Saff, and Y. M. Zhou, “Minimal discrete energy on the sphere,” *Math. Res. Lett.*, vol. 1, pp. 647–662, 1994.

[14] T. Tarnai and Z. Gáspár, “Multi-symmetric close packings of equal spheres on the spherical surface,” *Acta Cryst. A*, vol. 43, pp. 612–616, 1987.

[15] E. B. Saff and A. B. J. Kuijlaars, “Distributing many points on a sphere,” *Math. Int.*, vol. 19, no. 1, pp. 5–11, 1997.

Tajkhorshid, Emad (3)

[1] D. Sirohi *et al.*, “The 3.8 Å resolution cryo-EM structure of Zika virus,” *Science*, vol. 352, pp. 467–470, 2016.

[2] M. Sevvana *et al.*, “Refinement and analysis of the mature Zika virus cryo-EM structure at 3.1 Å resolution,” *Structure*, vol. 26, pp. 1169–1177, 2018.

[3] J. C. Phillips *et al.*, “Scalable molecular dynamics with NAMD,” *J. Comput. Chem.*, vol. 26, pp. 1781–1802, 2005.

[4] J. B. Klauda *et al.*, “Update of the CHARMM all-atom additive forcefield for lipids: Validation on six lipid types,” *J. Phys. Chem. B*, vol. 114, pp. 7830–7843, 2010.

[5] J. Huang *et al.*, “CHARMM36m: An improved forcefield for folded and intrinsically disordered proteins,” *Nat. Methods*, vol. 14, pp. 71–73, 2017.

[6] T. Darden, D. York and L. Pederson, “Particle mesh Ewald: An Nlog (N) method for Ewald sums in large systems,” *J. Chem. Phys.*, vol. 98, p. 10089, 1993.

[7] L. L. Swift, “Assembly of very low density lipoproteins in rat liver: a study of nascent particles recovered from the rough endoplasmic reticulum,” *J. Lipid Res.*, vol. 36, pp. 395–406, 1995.

[8] M. A. Martin–Acebes *et al.*, “The composition of West Nile virus lipid envelope unveils a role of sphingolipid metabolism on flavivirus biogenesis,” *J. Virol.*, vol. 88, pp. 12041–12054, 2014.

Voth, Gregory A.

[1] *Global HIV & AIDS statistics—2018 fact sheet*. [Online]. Available: <https://www.unaids.org/en/resources/fact-sheet>, Accessed on: May 27, 2019.

[2] S. Izvekov and G. A. Voth, “A multiscale coarse-graining method for biomolecular systems,” *J. Phys. Chem. B*, vol. 109, no. 7, pp. 2469–2473, Feb. 2005.

[3] S. Izvekov and G. A. Voth, “Multiscale coarse graining of liquid-state systems,” *J. Chem. Phys.*, vol. 123, no. 13, p. 134105, Oct. 2005.

[4] W. G. Noid *et al.*, “The multiscale coarse-graining method. I. A rigorous bridge between atomistic and coarse-grained models,” *J. Chem. Phys.*, vol. 128, no. 24, p. 244114, Jun. 2008.

[5] G. A. Voth, *Coarse-Graining of Condensed Phase and Biomolecular Systems*. CRC Press, 2008.

SOCIAL SCIENCE, ECONOMICS, & HUMANITIES

Cai, Yongyang

[1] Y. Cai, W. Brock, A. Xepapadeas, and K. L. Judd, “Climate policy under spatial heat transport: cooperative and noncooperative regional outcomes,” working paper, 2019.

[2] Y. Cai, K. L. Judd, and T.S. Lontzek, “The social cost of carbon with economic and climate risks,” *J. Political Econ.*, 2019, doi: 10.1086/701890.

[3] Y. Cai, K. L. Judd, G. Thain, and S. Wright, “Solving dynamic programming problems on a computational grid,” *Comput. Econ.*, vol. 45, no. 2, pp. 261–284, 2015.

[4] Y. Cai and K. L. Judd, “Climate policy with carbon capture and storage in the face of economic risks and climate target constraints,” working paper, 2019.

Downie, J. Stephen

[1] H.L. Goodall, Jr., *Writing Qualitative Inquiry: Self, Stories, and Academic Life*. New York, NY, USA: Routledge. 2008, p. 30.

[2] S. Bhattacharyya, P. Organisciak, and J. S. Downie, “A fragmentizing interface to a large corpus of digitized text,” *Interdisciplinary Science Rev.*, vol. 40, no. 1, pp. 61–77, Mar. 2015.

[3] G. Lukács, “Narrate or Describe?” *Writer and Critic and Other Essays by Georg Lukács*, A. D. Kahn, Ed. New York, NY, USA: Grosset and Dunlap, 1971, pp. 110–148.

[4] H. White, “The value of narrativity in the representation of reality,” *Crit. Inq.*, vol. 7, pp. 5–27, Autumn 1980.

[5] B. Brown, *A Sense of Things: The Object Matter of American Literature*. Chicago, IL, USA: Univ. Chicago Press, 2004.

[6] C. Schmitt, “Interpret or Describe?” *Representations*, vol. 135, no. 1, Summer 2016.

[7] Maojin & Jensen, Eric & Beitzel, Steve & Argamon, Shlomo. (2004). “Choosing the Right Bigrams for Information Retrieval.” *Proc. Meeting Inter. Fed. Classification Societies*, pp 531-540, Jul. 2004. 10.1007/978-3-642-17103-1\_50

[8] T. Underwood *et al.*, *Word frequencies in English-language literature*, 1700–1922 (0.2) [Data set]. Hathi Trust Research Center, doi: 10.13012/J8JW8BSJ.

[9] S. Wang and C. Manning, “Baselines and bigrams: Simple, good sentiment and text classification,” in *Proc. 50th Ann. Meeting Assoc. Comput. Linguistics*, ACL, 2012.

Ye, Mao

[1] C. Yao and M. Ye, “Why trading speed matters: A tale of queue rationing under price controls,” *Rev. Financial Stud.*, vol. 31, pp. 2157–2183, 2018.

[2] S. Li, X. Wang, and M. Ye, “Who provides liquidity, and when?” working paper, 2019.

[3] X. Li, M. Ye, and M. Zheng, “Market structure and corporate payout policy: Evidence from a controlled experiment,” working paper, 2019.

[4] M. Ye, M. Zheng, and W. Zhu, “Does the secondary financial market affect firm investment? Evidence from a controlled experiment, and more,” [https://papers.ssrn.com/sol3/Papers.cfm?abstract\\_id=3254585](https://papers.ssrn.com/sol3/Papers.cfm?abstract_id=3254585), 2019.

MATHEMATICAL & STATISTICAL SCIENCES

Cho, Wendy K. Tam

[1] B. Hutson, “NCSA scientist employs supercomputer simulations in Ohio gerrymandering case,” news release. [Online]. Available: <https://bit.ly/2WHmJmR>, Accessed on Sept. 16, 2019.

GRADUATE FELLOWS

Agee, Elizabeth

[1] D. Leitner *et al.*, “A dynamic root system growth model based on L-systems,” *Plant Soil*, vol. 332, nos.1–2, pp. 177–192, 2010.

[2] G. E. Hammond *et al.*, “Evaluating the performance of parallel subsurface simulators: An illustrative example with PFLOTRAN,” *Water Resour. Res.*, vol. 50, no. 1, pp. 208–228, 2014.

[3] V. Couvreur *et al.*, “A simple three-dimensional macroscopic root water uptake model based on the hydraulic architecture approach,” *Hydrol. Earth Syst. Sci.*, vol. 16, no. 8, pp. 2957–2971, 2012.

Applebaum, Elaad

[1] S. Benincasa, J. Wadsley, H. Couchman, and B. Keller, “The anatomy of a star-forming galaxy: pressure-driven regulation of star formation in simulated galaxies,” *Mon. Notices Royal Astron. Soc.*, vol. 462, no. 3, pp. 3053–3068, Nov. 2016, doi: 10.1093/mnras/stw1741.

[2] V. Semenov, A. Kravtsov, and N. Gnedin, “How galaxies form stars: the connection between local and global star formation in galaxy simulations,” *Astrophys. J.*, vol. 861, no. 1, p. 4, Jul. 2018, doi: 10.3847/1538-4357/aac6eb.

[3] H. Menon *et al.*, “Adaptive techniques for clustered N-body cosmological simulations,” *Comput. Astrophys. Cosm.*, vol. 2, no. 1, Mar. 2015, doi: 10.1186/s40668-015-0007-9.

[4] J. Wadsley, B. Keller, and T. Quinn, “Gasoline2: a modern smoothed particle hydrodynamics code,” *Mon. Notices Royal Astron. Soc.*, vol. 471, no. 2, pp. 2357–2369, Oct. 2017, doi: 10.1093/mnras/stx1643.

[5] F. Munshi *et al.*, “Dancing in the dark: uncertainty in ultrafaint dwarf galaxy predictions from cosmological simulations,” *Astrophys. J.*, vol. 874, no. 1, p. 40, Mar 2019, doi: 10.3847/1538-4357/ab0085.

Barber, Katelyn

[1] T. P. Lane, R. D. Sharman, S. B. Trier, R. G. Fovell, and J. K. Williams, “Recent advances in the understanding of near-cloud turbulence,” *Bull. Amer. Meteor. Soc.*, vol. 93, pp. 499–515, Apr. 2012.

[2] P. F. Lester, *Turbulence: A New Perspective for Pilots*. Englewood, CO, USA: Jeppesen Sanderson, 1994, pp. 212.

[3] Federal Aviation Administration, “Official guide to basic flight information and ATC procedures,” *Aeronautical Information Manual*, 2017, ch. 7, pp. 435–539.

[4] W. C. Skamarock and J. B. Klemp, “A time-split nonhydrostatic atmospheric model for weather and forecasting applications,” *J. Comp. Phys.*, vol. 227, pp. 3465–3485, Mar. 2008.

[5] K. A. Barber, W. Deierling, G. L. Mullendore, C. Kessinger, R. Sharman, and D. Munoz–Esparza, “Properties of convectively induced turbulence over developing oceanic convection,” *Mon. Wea. Rev.*, <https://journals.ametsoc.org/doi/pdf/10.1175/MWR-D-18-0409.1>, 2019.

[6] R. M. Endlich, “The mesoscale structure of some regions of clear-air turbulence,” *J. Appl. Meteor.*, vol. 3, pp. 261–276, Jun. 1964.



[7] N. N. Ahmad and F. H. Proctor, “Estimation of eddy dissipation rates from mesoscale model simulations,” NASA Langley, Hampton, VA, USA, NASA Tech. Rep. AIAA 0429, Apr. 2012.

[8] R. Frehlich and R. Sharman, “Estimates of turbulence from numerical weather prediction model output with applications to turbulence diagnosis and data assimilation,” *Mon. Wea. Rev.*, vol. 132, pp. 2308–2324, Oct. 2004.

[9] T. P. Lane, R. D. Sharman, T. L. Clark, and H. M Hsu, “An investigation of turbulence generation mechanisms above deep convection,” *J. Atmos. Sci.*, vol. 60, pp. 1297–1321, May 2003.

Brooks, Maureen T.

[1] D. Laffoley *et al.*, “The Protection and Management of the Sargasso Sea: The Golden Floating Rainforest of the Atlantic Ocean,” *Summary Science and Supporting Evidence Case*, Sargasso Sea Alliance, pp. 1–44, 2011.

[2] V. Smetacek and A. Zingone, “Green and golden seaweed tides on the rise,” *Nature*, vol. 504, pp. 84–88, 2013.

[3] E. Chassignet *et al.*, “US GODAE: Global Ocean Prediction with the HYbrid Coordinate Ocean Model (HYCOM),” *Oceanography*, vol. 22, no. 2, pp. 64–75, 2009.

[4] K. Fennel, R. Hetland, Y. Feng, and S. DiMarco, A coupled physical–biological model of the Northern Gulf of Mexico shelf: model description, validation and analysis of phytoplankton variability, *Biogeosciences*, vol. 8, no. 7, pp. 1881–1899, 2006.

[5] *Ocean Color*, National Aeronautics and Space Administration Goddard Space Flight Center, 2018. [Online]. Available: <http://oceancolor.gsfc.nasa.gov>

Butsky, Iryna

[1] M. Tremmel *et al.*, “Introducing ROMULUSC: a cosmological simulation of a galaxy cluster with unprecedented resolution,” *Mon. Notices Royal Astron. Soc.*, vol. 483, no. 3, pp. 3336–3362, 2019, doi: 10.1093/mnras/sty3336.

[2] L. V. Kale and S. Krishnan, “Charm++: Parallel programming with message-driven objects,” *Parallel Programming using C++*, 1996, pp. 175–213.

[3] H. Menon *et al.*, “Adaptive techniques for clustered N-body cosmological simulations,” *Comput. Astrophys. Cosmol.*, vol. 2, 2015.

[4] M. Turk *et al.*, “yt: A multi-code analysis toolkit for astrophysical simulation data,” *ApJS*, vol. 192, 2010.

[5] C. Hummels, B. Smith, and D. Silvia, “Trident: A universal tool for generating synthetic spectra from astrophysical simulations,” *ApJ*, vol. 847, no. 1, p. 59.

[6] J. Burchett *et al.*, “A deep search for faint galaxies associated with very low-redshift C IV absorbers. II. Program design, absorption-line measurements, and absorber statistics,” *ApJ*, vol. 815, p. 91.

Cieri, Robert

[1] C. G. Farmer, “The evolution of unidirectional pulmonary airflow,” *Physiology*, vol. 30, no. 4, pp. 260–272, 2015.

[2] R. L. Cieri and C. G. Farmer, “Computational fluid dynamics reveals a unique net-unidirectional pulmonary airflow pattern in the savannah monitor (*Varanus exanthematicus*),” in review, 2019.

[3] R. L. Cieri and C. G. Farmer, “Unidirectional pulmonary airflow in vertebrates: a review of structure, function, and evolution,” *J. Comp. Physiol. B*, vol. 186, no. 5, pp. 541–552, 2016.

[4] C. G. Farmer and K. Sanders, “Unidirectional airflow in the lungs of alligators,” *Science*, vol. 327, no. 5963, pp. 338–340, Jan. 2010.

[5] R. L. Cieri, B. A. Craven, E. R. Schachner, and C. G. Farmer, “New insight into the evolution of the vertebrate respiratory system and the discovery of unidirectional airflow in iguana lungs,” *Proc. Natl. Acad. Sci. USA*, vol. 111, no. 48, pp. 17218–17223, 2014.

[6] E. R. Schachner, R. L. Cieri, J. P. Butler, and C. G. Farmer, “Unidirectional pulmonary airflow patterns in the savannah monitor lizard,” *Nature*, vol. 506, pp. 367–370, Dec. 2013.

[7] C. G. Farmer, “Unidirectional flow in lizard lungs: a paradigm shift in our understanding of lung evolution in Diapsida,” *Zoology*, vol. 188, no. 5, pp. 299–301, 2015.

[8] C. J. Clemente, P. C. Withers, and G. G. Thompson, “Metabolic rate and endurance capacity in Australian varanid lizards (Squamata: Varanidae: Varanus),” *Biol. J. Linn. Soc.*, vol. 97, no. 3, pp. 664–676, Jun. 2009.

[9] C. G. Farmer, “Similarity of crocodilian and avian lungs indicates unidirectional flow is ancestral for archosaurs,” *Integr. Comp. Biol.*, pp. 1–10, 2015.

Clement, Matthew

[1] K. Haisch, E. Lada, and C. Lada, “Disk frequencies and lifetimes in young clusters,” *ApJ*, vol. 553, no. 2, 2001, doi: 10.1086/320685.

[2] R. Gomes *et al.*, “Origin of the cataclysmic Late Heavy Bombardment period of the terrestrial planets,” *Nature*, vol. 435, pp. 466–469, 2005, doi: 10.1038/nature03676.

[3] H. Levison *et al.*, “Origin of the structure of the Kuiper Belt during a dynamical instability in the orbits of Uranus and Neptune,” *Icarus*, vol. 196, no. 1, 2008, doi: 10.1016/j.icarus.2007.11.035.

[4] N. Kaib and J. Chambers, “The fragility of the terrestrial planets during a giant-planet instability,” *Mon. Notices Royal Astron. Soc.*, vol. 455, no. 4, 2016, doi: 10.1093/mnras/stv2554.

[5] J. Chambers, “Late-stage planetary accretion including hit-and run collisions and fragmentation,” *Icarus*, vol. 224, no. 1, 2013, doi: 10.1016/j.icarus.2013.02.015.

[6] S. Grimm and J. Stadel, “The GENGA code: Gravitational encounters in N-body simulations with GPU acceleration,” *ApJ*, vol. 796, no. 1, 2014, doi: 10.1088/0004-637X/796/1/23.

[7] J. Chambers, “Making more terrestrial planets,” *Icarus*, vol. 152, no. 1, 2001, doi: 10.1006/icar.2001.6639.

[8] S. Raymond *et al.*, “Building the terrestrial planets: Constrained accretion in the inner solar system,” *Icarus*, vol. 203, no. 2, 2009, doi: 10.1016/j.icarus.2009.05.016.

[9] N. Dauphas and A. Pourmand., “Hf-W-Th evidence for rapid growth of Mars and its status as a planetary embryo,” *Nature*, vol. 473, 2011, doi: 10.1038/nature10077.

[10] K. Walsh *et al.*, “A low mass for Mars from Jupiter’s early gas-driven migration,” *Nature*, vol. 475, 2011, doi: 10.1038/nature10201.

[11] J. Chambers, “A hybrid symplectic integrator that permits close encounters between massive bodies,” *Mon. Notices Royal Astron. Soc.*, vol. 304, no. 4, 1999, doi: 10.1046/j.1365-8711.1999.02379.x.

Cook, Salme

[1] United Nations Environment Program, *Marine and Coastal Ecosystems and Human Wellbeing: A Synthesis Report Based on the Findings of the Millennium Ecosystem Assessment*. UNEP: Nairobi, Kenya, 2006.

Glines, Forrest

[1] H. Li, G. Lapenta, J. M. Finn, S. Li, and S. A. Colgate, “Modeling the large-scale structures of astrophysical jets in the magnetically dominated limit,” *Astrophys. J.*, vol. 643, no. 1, p. 92, 2006.

[2] I. S. Butsky and T. R. Quinn, “The role of cosmic-ray transport in shaping the simulated circumgalactic medium,” *Astrophys. J.*, vol. 868, no. 2, p. 108, Nov. 2018.

[3] P. Grete, B. W. O’Shea, K. Beckwith, W. Schmidt, and A. Christlieb, “Energy transfer in compressible magnetohydrodynamic turbulence,” *Phys. of Plasmas*, vol. 24, no. 9, p. 092311, Sept. 2017.

[4] H. Carter Edwards, C. R. Trott, and D. Sunderland, “Kokkos: Enabling manycore performance portability through polymorphic memory access patterns,” *J. Parallel Distrib. Comput.*, vol. 74, no. 12, pp. 3202–3216, Dec. 2014.

[5] C. J. White, J. M. Stone, and C. F. Gammie, “An Extension of the Athena++ code framework for GRMHD based on advanced Riemann solvers and staggered-mesh constrained transport,” *Astrophys. J. Suppl. Series*, vol. 225, p. 22, Aug. 2016.

Lansford, Joshua

[1] G. Pilania, C. Wang, X. Jiang, S. Rajasekaran, and R. Ramprasad, “Accelerating materials property predictions using machine learning,” *Sci. Rep.*, vol. 3, p. 2810, 2013.

[2] A. Heyden, A. T. Bell, and F. J. Keil, “Efficient methods for finding transition states in chemical reactions: Comparison of improved dimer method and partitioned rational function optimization method,” *J. Chem. Phys.*, vol. 123, p. 224101, 2005.

[3] W. Kohn, “Nobel Lecture: Electronic structure of matter—wave functions and density functionals,” *Rev. Mod. Phys.*, vol. 71, pp. 1253–1266, 1999.

[4] M. Rupp, “Machine learning for quantum mechanics in a nutshell,” *Int. J. Quantum Chem.*, vol. 115, pp. 1058–1073, 2015.

[5] A. A. Peterson, “Acceleration of saddle-point searches with machine learning,” *J. Chem. Phys.*, vol. 145, p. 074106, 2016.

[6] Z. D. Pozun, K. Hansen, D. Sheppard, M. Rupp, K.–R. Müller, and G. Henkelman, “Optimizing transition states via kernel-based machine learning,” *J. Chem. Phys.*, vol. 136, p. 174101, 2012.

[7] K. Toyoura *et al.*, “Machine-learning-based selective sampling procedure for identifying the low-energy region in a potential energy surface: A case study on proton conduction in oxides,” *Phys. Rev. B*, vol. 93, p. 054112, 2016.

[8] G. Kresse and J. Hafner, “*Ab initio* molecular dynamics for liquid metals,” *Phys. Rev. B*, vol. 47, pp. 558–561, Jan. 1993.

[9] T. A. Manz and N. G. Limas, “Introducing DDEC6 atomic population analysis: Part 1. Charge partitioning theory and methodology,” *RSC Adv.*, vol. 6, pp. 47771–47801, 2016.

Marsac, Kara

[1] EPA, “Profile of the oil and gas extraction industry,” EPA, 310-R-99-006, 2000.

[2] D. L. Shaffer, L. H. Arias Chavez, M. Ben–Sasson, S. Romero–Vargas Castrillón, N. Y. Yip, and M. Elimelech, “Desalination and reuse of high-salinity shale gas produced water: drivers, technologies, and future directions,” *Env. Sci. Tech.*, vol. 47, no. 17, pp. 9569–9583, 2013.

[3] J. Veil, M. G. Puder, D. Elcock, and R. J. Redweik, Jr., “White Paper Describing Produced Water from Production of Crude Oil, Natural Gas, and Coal Bed Methane,” 2004, doi: 10.2172/821666.

[4] A. Fakhru’l–Razi, A. Pendashteh, L. C. Abdullah, D. R. A. Biak, S. S. Madaeni, and Z. Z. Abidin, "Review of technologies for oil and gas produced water treatment," *J. of Hzds. Mats.*, vol. 170, no. 2–3, pp. 530–551, 2009.

[5] P. Lichtner *et al.*, “PFLOTTRAN user manual,” 2013. [Online]. Available: <https://documentation.pfлотran.org>

[6] K. Marsac and A. K. Navarre–Sitchler, “Geochemical modeling investigations of brines as a fresh water alternative for hydraulic fracturing,” in review, 2019.



Soley, Micheline

[1] R. V. Krems, “Molecules near absolute zero and external field control of atomic and molecular dynamics,” *Int. Rev. Phys. Chem.*, vol. 24, pp. 99–118, 2005.

[2] D. DeMille, “Quantum computation with trapped polar molecules,” *Phys. Rev. Lett.*, vol. 88, p. 067901, Jan. 2002.

[3] J. J. Hudson, B. E. Sauer, M. R. Tarbutt, and E. A. Hinds, “Measurement of the electron electric dipole moment using YbF molecules,” *Phys. Rev. Lett.*, vol. 89, p. 023003, Jun. 2002.

[4] T. Zelevinsky, S. Kotochigova, and J. Ye, “Precision test of mass-ratio variations with lattice-confined ultracold molecules,” *Phys. Rev. Lett.*, vol. 100, p. 043201, 2008.

[5] M. B. Soley and E. J. Heller, “Classical approach to collision complexes in ultracold chemical reactions,” *Phys. Rev. A*, vol. 98, no. 5, pp. 052702, Nov. 2018.

Stenz, Ronald

[1] G. H. Bryan and J. M. Fritsch, “A benchmark simulation for moist nonhydrostatic numerical models,” *Mon. Wea. Rev.*, vol. 130, pp. 2917–2928, 2002.

[2] M. S. Gilmore, R. Davies–Jones, J. Straka, E. Rasmussen, and R. Wilhelmson, “Centrifugal precipitation transport in tornadic supercells: An algorithm consistent for use with bulk microphysics schemes,” *Proc. 14th Intl. Conf. Clouds and Precipitation*, vol. 82, nos. 1–2, 2004, pp. 1654–1656.

[3] R. L. Thompson, R. Edwards, J. A. Hart, K. L. Elmore, and P. M. Markowski, “Close proximity soundings within supercell environments obtained from the rapid update cycle,” *Wea. Forecasting*, vol. 18, pp. 1243–1261, 2003.

Teo, Darius

[1] E. O’Brien *et al.*, “The [4Fe4S] cluster of human DNA primase functions as a redox switch using DNA charge transport,” *Science*, vol. 355, no. 6327, p. eaag1789, Feb. 2017.

[2] S. A. Forbes *et al.*, “COSMIC: mining complete cancer genomes in the catalogue of somatic mutations in cancer,” *Nucleic Acids Res.*, vol. 39, no. suppl\_1, pp. D945–D950, Jan. 2011.

[3] R. D. Teo, B. J. G. Rousseau, E. R. Smithwick, R. Di Felice, D. N. Beratan, and A. Migliore, “Charge transfer between [4Fe4S] proteins and DNA is unidirectional. Implications for biomolecular signaling,” *Chem.*, vol. 5, no. 1, pp. 122–137, Jan. 2019.

[4] L. Noodleman, D. A. Case, and A. Aizman, “Broken symmetry analysis of spin coupling in iron–sulfur clusters,” *J. Am. Chem. Soc.*, vol. 110, no. 4, pp. 1001–1005, Feb. 1988.

[5] J. M. Seminario, “Calculation of intramolecular force fields from second-derivative tensors,” *Int. J. Quantum Chem.*, vol. 60, no. 7, pp. 1271–1277, Mar. 1996.

Torres, Walter

[1] C. R. Darwin, *The Structure and Distribution of Coral Reefs. Being the First Part of the Geology of the Voyage of the Beagle*. London: Smith Elder and Co., 1842.

[2] F. Moberg and C. Folke, “Ecological goods and services of coral reef ecosystems,” *Ecol. Econom.*, vol. 29, no. 2, pp. 215–233, 1999.

[3] T. P. Hughes *et al.*, “Climate change, human impacts, and the resilience of coral reefs,” *Science*, vol. 301, p. 929, 2009.

[4] S. G. Monismith, “Hydrodynamics of coral reefs,” *Annu. Rev. Fluid Mech.*, vol. 39, pp. 37–55, 2006.

[5] I. R. Young, K. P. Black, and M. L. Heron, “Circulation in the Ribbon Reef region of the Great Barrier Reef,” *Cont. Shelf Res.*, vol. 14, 1994.

[6] J. C. Warner, B. Armstrong, R. He, and J. B. Zambon, “Development of a Coupled Ocean–Atmosphere–Wave–Sediment Transport (COAWST) modeling system,” *Ocean Model.*, vol. 35, pp. 230–244, 2010.



# INDEX

## A

Agee, Elizabeth 332  
Aksimentiev, Aleksei 238, 240, 242  
Aluru, Narayana 120, 122, 124  
Amaro, Rommie 244  
Anisimov, Victor 314  
Applebaum, Elaad 334  
Araya, Guillermo 126

## B

Balsara, Dinshaw 14  
Barber, Katelyn 336  
Bernardi, Rafael 246  
Bernholc, Jerry 128  
Bodony, Daniel 130  
Brehm, Christoph 132  
Brooks, Maureen 338  
Browne, Oliver 134  
Burrows, Adam 16  
Butsky, Iryna 340

## C

Caetano-Anolles, Gustavo 248  
Cai, Yongyang 320  
Campanelli, Manuela 18, 20  
Carrasco Kind, Matias 22  
Ceperley, David 136  
Chew, Huck Beng 138  
Cho, Wendy 328  
Chu, Shanna 361  
Cieri, Robert 342  
Clancy, Colleen 250  
Clark, Bryan 140  
Clement, Matthew 344  
Cook, Salme 346  
Corcoran, Jennifer 78  
Cox, Donna 212

## D

Di Girolamo, Larry 80  
Di Matteo, Tiziana 24  
Diao, Chunyuan 82  
Dickerson, Julie 252  
Dill, Ken 254  
Dominguez, Fancina 84  
Downie, J. Stephen 322  
Draayer, Jerry 26

Duan, Lian 142  
Duursma, Iwan 214

## E

El-Kebir, Mohammed 315  
El-Khadra, Aida 144  
Emerick, Andrew 348  
Ertekin, Elif 146

## F

Ferguson, Andrew 256  
Freund, Jonathan 148

## G

Gammie, Charles 28  
Garcia, Marcelo 150  
Gazzola, Mattia 258  
Glines, Forrest 350  
Gnedin, Nickolay 30  
Gregg, Patricia 86  
Gropp, William 216, 218  
Guan, Kaiyu 88, 90  
Gurvich, Alexander 352

## H

Hadden–Perilla, Jodi 260, 262  
Hawley, John 32  
Hays, Jennifer 354  
Hirata, So 264  
Hooberman, Benjamin 152  
Hopkins, Philip 34  
Hudson, Matthew 266  
Huerta Escudero, Eliu 36, 38, 40  
Huff, Kathryn 154

## I

Iyer, Ravishankar 220

## J

Jain, Prashant 156  
Jha, Shantenu 222  
Jiang, Tao 268  
Johnsen, Eric 158  
Johnson, Harley 160

## K

Kemball, Athol 42  
Khalili-Araghi, Fatemeh 270  
Koric, Seid 162

## L

Lansford, Joshua 356  
Lasher-Trapp, Sonia 92  
LeBauer, David 272

Leburton, Jean-Pierre 164  
Levin, Deborah 44, 46  
Liu, Yi-Hsin 48

## M

Makri, Nancy 274  
Marsac, Kara 358  
Mashayek, Farzad 166  
Masud, Arif 276  
Matalon, Moshe 168  
Menanteau, Felipe 50  
Moesta, Philipp 52  
Moore, Jeffery 278  
Moradi, Mahmoud 280, 282  
Morin, Paul 94

## N

Neubauer, Mark 170  
Nesbitt, Stephen 96  
Norman, Michael 54

## O

O'Shea, Brian 56  
Olson, Luke 224, 226  
Orf, Leigh G. 98

## P

Pavlis, Nikolaos 100  
Perilla, Juan 284  
Peterson, Joseph 286  
Petravick, Donald 58  
Pogorelov, Nikolai 60  
Pogorelov, Taras 316  
Prather, Kimberly 288  
Pratt, Jane 62

## Q

Quinn, Thomas 64

## R

Rahman, Rajib 172  
Raman, Venkat 174  
Ricker, Paul 66  
Riedl, Caroline 176  
Riemer, Nicole 102  
Rosato, Nicole 360  
Roux, Benoit 290

## S

Schleife, André 178, 180  
Shanahan, Phiala 208  
Shapiro, Stuart 68  
Shukla, Diwakar 292, 294, 296

Sirignano, Justin 234  
Snir, Marc 228  
Soley, Micheline 362  
Solomonik, Edgar 230  
Soltesz, Ivan 298  
Sotomayor, Marcos 300  
Srinivasan, Ashok 302  
Stenz, Ronald 364  
Sutton, Brad 304

## T

Tabor, Clay 104  
Taflin, Dave 235  
Tajkhorshid, Emad 306, 308, 310  
Tchekhovskoy, Alexander 70  
Teo, Darius 366  
Teukolsky, Saul A. 75  
Thomas, Brian 182  
Tinoco Lopez, Rafael 184  
Torres, Walter 368  
Toth, Gabor 72  
Toussaint, Kimani 186  
Trapp, Robert 106  
Trinkle, Dallas 188  
Turk, Matthew 74

## V

Vidale, John 108  
Voth, Gregory 312  
Vuković, Lela 190

## W

Wagner, Lucas 192  
Wang, Zhi Jian 194  
Warnow, Tandy 232  
Wentzcovitch, Renata 110  
West, Matthew 112  
Whitman, Samuel 370  
Wuebbles, Donald J. 114

## X

Xu, Bin 196  
Xu, Zhen 198

## Y

Yan, Jinhui 202  
Yan, Yonghua 200  
Ye, Mao 324  
Yeung, Pui-Kuen 204

## Z

Zhang, Xiangdong 116  
Zhang, Yang 206



computational  
different  
protein  
future  
single  
provides  
several  
key  
including  
simulations  
provide  
significant  
understand  
requires  
cell  
molecular  
show  
one  
black  
To  
turbulence  
years  
climate  
local  
design  
materials  
scaling  
two  
current  
Science  
study  
problem  
critical  
node  
particles  
simulatedevolution  
perform  
us  
structures  
cores  
new  
required  
gas  
process  
software  
larger  
support  
performed  
on  
successful  
equations  
modeling  
coupled  
processing  
GPU  
performance  
algorithms  
experiments  
proteins  
transport  
field  
number  
great  
code  
first  
A  
best  
free  
WATERS  
turbulentfunction  
impact  
gain  
e  
ly  
ross  
models  
efficient  
lex  
ce  
ies





NCSA | National Center for  
Supercomputing Applications

*Blue Waters is supported by the  
National Science Foundation*

