2018–2019

research report

Ohio Supercomputer Center

An OH·TECH Consortium Member

Leadership

David Hudak, Ph.D.
Executive Director
(614) 247-8670 | dhudak@osc.edu

Doug Johnson
Chief Systems Architect and
Manager, HPC Systems
(614) 292-6286 | djohnson@osc.edu

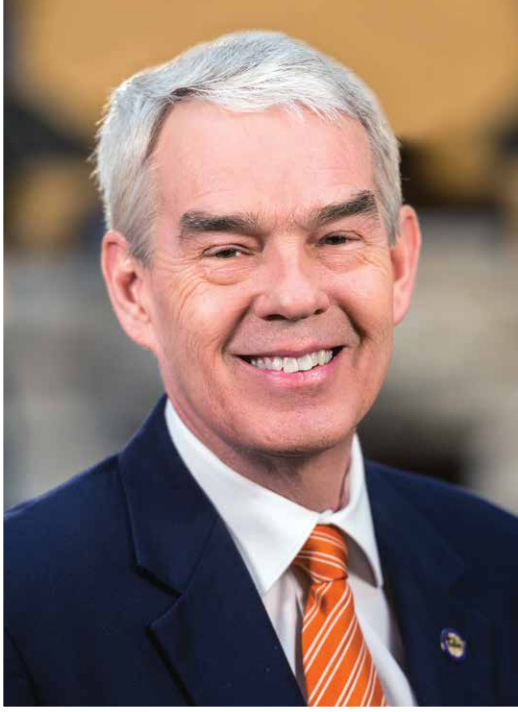
Karen Tomko, Ph.D.
Director, Research Software Applications
(614) 292-1091 | ktomko@osc.edu

Alan Chalker, Ph.D.
Director, Strategic Programs
(614) 247-8672 | alanc@osc.edu

Brian Guilfoos
Manager, HPC Client Services
(614) 292-2846 | guilfoos@osc.edu

Ohio Supercomputer Center (OSC) addresses the rising computational demands of academic and industrial research communities by providing a robust shared infrastructure and proven expertise in advanced modeling, simulation and analysis. OSC empowers scientists with the services essential to making extraordinary discoveries and innovations, partners with businesses and industry to leverage computational science as a competitive force in the global knowledge economy and leads efforts to equip the workforce with the key technology skills required for 21st century jobs.

Governed by the Chancellor of the Ohio Department of Higher Education (ODHE), the **Ohio Technology Consortium** (OH-TECH) serves as the technology and information division of ODHE. The consortium comprises a suite of widely respected member organizations collectively unsurpassed in any other state: OSC, OARnet and OhioLINK. The consortium drives efficiencies through common services provided to member organizations through the Shared Infrastructure and Consortia Services divisions.



> The services of the Ohio Supercomputer Center empower innovation and discovery across the state, and provide an essential educational resource for more than 3,000 students at over 25 Ohio universities in the vital science, technology, engineering, and mathematics (STEM) disciplines.
>
> — *Randy Gardner, Chancellor, Ohio Department of Higher Education*

osc.edu

PHOTO // Chancellor Randy Gardner directs ODHE and oversees the strategic initiatives of the OH-TECH and its member organizations in support of the state's technology infrastructure needs.

TABLE OF

contents

Tags:

● AREA OF SCIENCE
- Biohealth
- Data Analytics
- Environment
- Manufacturing

■ SERVICE/RESOURCE
- Artificial Intelligence/Machine Learning (AI/ML)
- Condo/Lease
- Education
- GPU Computing (GPU)
- High Performance Computing (HPC)
- High Throughput Computing (HTC)
- Huge Memory Node
- Project Storage
- Software Development

▲ INNOVATEOHIO SECTORS
- Aerospace
- Agribusiness
- Automotive
- Aviation
- Energy/Chemical
- Invention/Discovery*
- Manufacturing

InnovateOhio—the DeWine Administration's commitment to leading an aggressive, innovative path towards a better and stronger Ohio—has outlined these traditional areas of innovation strength throughout the state's history.

*(*Invention/Discovery describes InnovateOhio's Edison category.)*

Director's Letter

2018–19 Highlights

Systems Deployment

In 2018, we deployed the Center's most efficient supercomputer ever—the 10,560-core Dell/Intel Pitzer Cluster. This year, we also more than doubled our research data storage space.

Sponsored Activities

Our staff was actively involved with the National Science Foundation's Extreme Science and Engineering Discovery Environment (XSEDE), SetSM high-resolution digital elevation modeling, MVAPICH high-performance parallel programming libraries and Open OnDemand projects.

National Engagement

OSC staff participated in the Practice & Experience in Advanced Research Computing (PEARC) Conference, the Coalition for Academic Scientific Computation (CASC) and the MVAPICH Users Group meeting.

The Ohio Supercomputer Center makes a positive difference in the lives and careers of many Ohioans. That fundamental thought is behind much of what we work toward here, whether it's to support higher education, meet the demands of academic and commercial research or simply to improve access and productivity.

OSC serves an important role in the education of thousands of Ohio students. Throughout Ohio's colleges and universities, students—primarily graduate students—leverage the Center's systems to conduct computationally based research that serves as the core of their degree programs. Undergraduates are increasingly accessing OSC as part of their classroom instruction. In addition, middle and high school students participate in hands-on activities at the Center every summer as part of Summer Institute and Young Women's Summer Institute.

We recently surveyed students, faculty, scientists, engineers and clinicians who utilize our services, and they clearly stated that they need more computational and storage capacity. Last year, we deployed the Center's most efficient supercomputer ever—the 10,560-core Dell/Intel "Pitzer" Cluster—raising our Center-wide performance total to over 3 petaflops. This year, we more than doubled our existing research data storage and now offer our clients 14 petabytes of secure, reliable space.

With the classroom and research lab arrangements in place and the computational and storage systems deployed, it is imperative that we make all this powerful technology more accessible and productive. We are currently upgrading MyOSC, the Center's administrative portal, which gives clients the ability to adjust passwords and contact information, submit allocation proposals, manage project access, report funding and publications and run custom usage reports.

We also are developing the second generation of OnDemand, based on our NSF-funded, open-source software platform supporting web-based access to high performance computing (HPC) services. Open OnDemand is now being used at scores of facilities around the world. To help ensure that clients use these new technologies to the fullest, we have expanded our training efforts to feature even more workshops and one-on-one training opportunities.

If you are one of our new or established clients, I thank you for joining us at OSC in our endeavor to make an impact upon the lives of thousands of Ohioans. If you aren't yet a client, but think high performance computing and storage could benefit your academic or commercial work, please contact us and we'll get your project up and going!

David Hudak, Ph.D.
Executive Director

Client Profiles

The Ohio Supercomputer Center (OSC) provides statewide resources to empower Ohio students, faculty members, scientists, engineers and clinicians on their way to educational achievements, brilliant innovations and fascinating discoveries. These accomplishments lead to new products, services and breakthroughs in a vast array of scientific disciplines. Through partnerships with Ohio industries, supercomputing provides businesses with a competitive advantage. By educating Ohio's workforce in the key skills required for future jobs, OSC also helps move the needle ahead in economic development.

Academics

OSC supports educational activities at various types of institutions—two-year and four-year, public and private, large and small—across a broad range of disciplines and degree programs. Academic clients include students, faculty, staff and special interest groups, such as an introductory bioinformatics lab at The Ohio State University where future biologists and physicians access OSC systems to learn next-generation sequencing techniques and data-analysis methods from Yue Zhao, Ph.D. (see story, pg. 26).

Across Ohio, 27 colleges and universities leveraged OSC services for instructional purposes, either in the classroom or the laboratory. The largest areas of academic study among them were the natural sciences, engineering and technology.

A thorough, yet streamlined, peer review process is used to ensure quick turnaround of allocations requests, as well as to prioritize high-quality science and promote efficient utilization of resources.

Commercial

The AweSim program is OSC's industrial outreach initiative, designed to provide OSC's high performance computing (HPC) resources to businesses throughout Ohio. Company owners and management, engineers, industrial designers and IT professionals from a variety of businesses are learning more about the "HPC in the cloud" advantages available to them through AweSim.

AweSim staff members work with a team of engineering service providers who are experts in a variety of computer-aided engineering disciplines. With assistance from OSC, businesses of all sizes and across a variety of industries can use HPC for virtual product design, testing simulations and data-intensive projects.

This allows small- to mid-sized manufacturers to push the boundaries of innovation and be more competitive economically. AweSim provides affordable, accessible and scalable modeling and simulation on high performance computers via:

- online modeling and simulation applications
- educational resources
- industry-specific expertise and consultants.

Beyond providing critical shared statewide resources, OSC works to create a user-focused, user-friendly environment for our clients. In 2018–19, 4,419 distinct students, faculty and staff members across 27 Ohio universities and 47 companies accessed OSC's resources. As part of our expanding training programs, we had 462 people attend 30 individual training opportunities throughout the year.

client data

68 academic institutions

47 companies

4,419 clients

311 new projects created

30 training opportunities

462 trainees

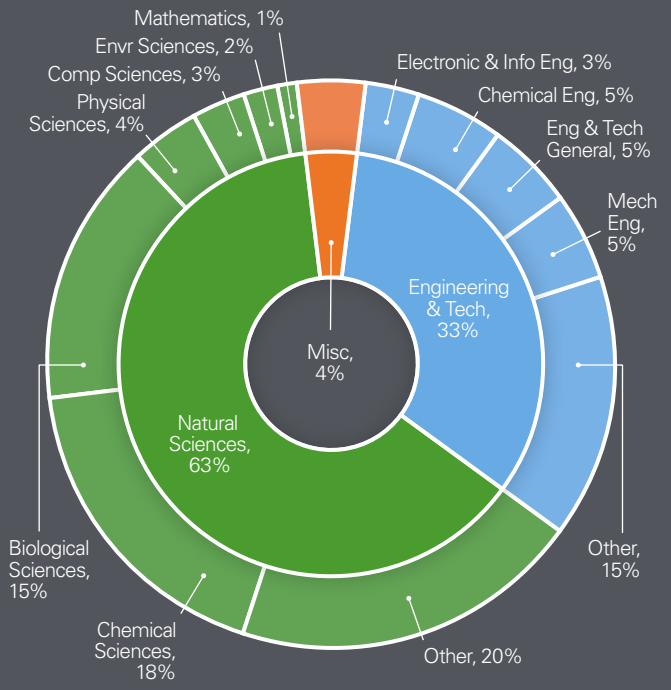
701 projects served

64 courses used OSC

all data: July 2018–June 2019

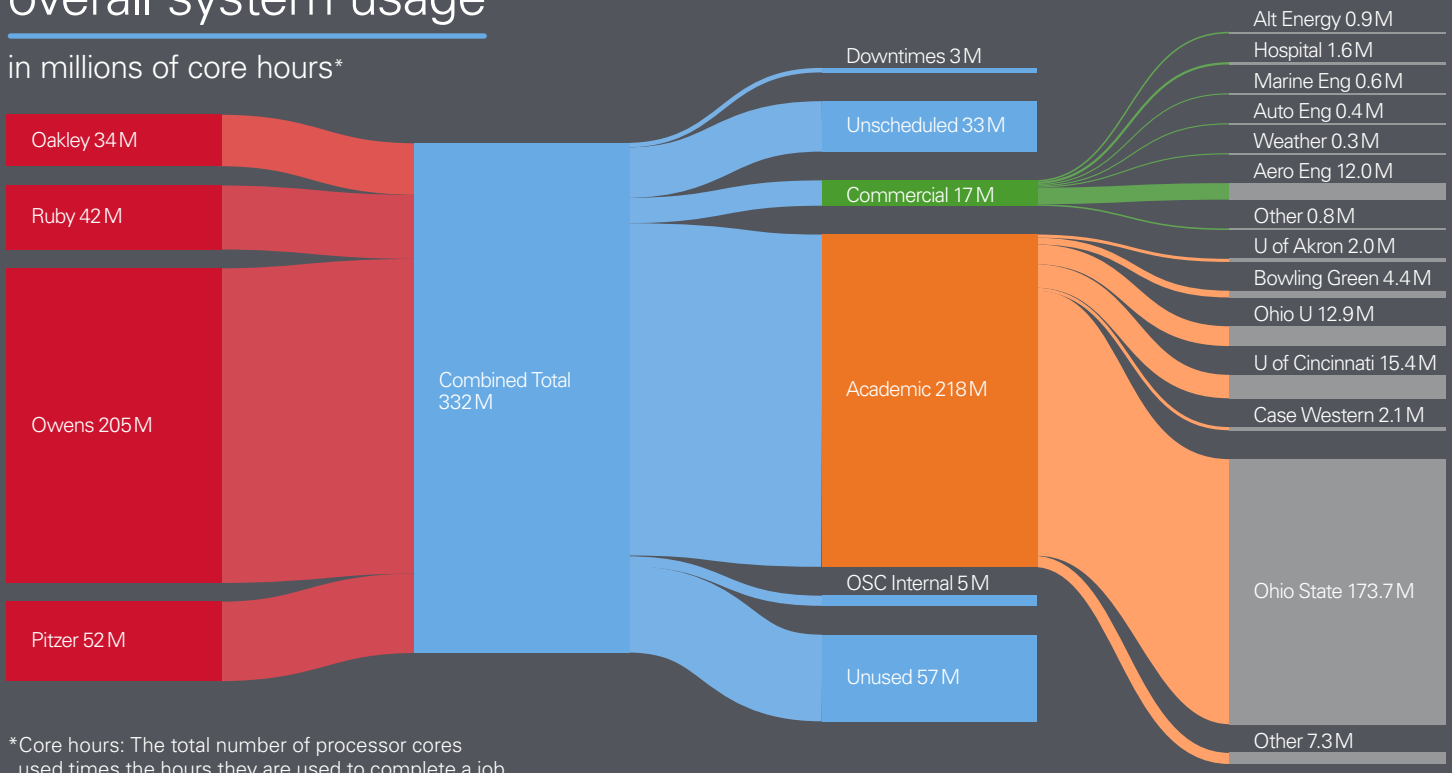
649 Ohio academic projects

usage by field of science



overall system usage

in millions of core hours*



*Core hours: The total number of processor cores used times the hours they are used to complete a job.

all data: July 2018–June 2019

case studies

Fungi Genomics

Slot group investigates coffee microbiome relationships

Jason Slot, associate professor of fungal evolutionary genomics at The Ohio State University, is performing research to ensure the longevity of one of the world's favorite crops: coffee. Specifically, Slot's group studies the genomics of fungi that live in coffee plants to understand their function and relationship within the plant and to better understand the plant's microbiome in general.

Coffee plants are plagued by several diseases, including coffee rust, caused by a pathogenic fungus. Interestingly enough, the fungi that live inside coffee plants have a symbiotic—mutually beneficial—relationship that could help fend off diseases, keeping both the coffee plants and the fungi alive. Slot's group is working to understand how this happens at a genomic level, as well as how the fungi stays alive inside the plant.

"What we propose to do is to identify chunks of the genome we call gene clusters that enable different species of fungi to live inside of the coffee trees, the coffee leaves actually," Slot said. "We also look for clusters of genes that enable the fungi to produce chemical defenses to defend the coffee plants against their disease organisms."

Slot's group sequences the fungal genome and uses the Ohio Supercomputer Center's (OSC) Owens Cluster to process the genetic data, looking for particular sets of genes in a process called annotation. Using a custom-developed software and programs, they can then compare genomes of different fungal species to identify different functions of the plants and build hypotheses for how the fungus performs its particular ecology.

"We're currently funded to ask a very basic question…and that's how plants can use their own genes to control what their microbiome looks like," Slot said.

While Slot may describe the research as basic, it is fundamental to many fields of biology, most of which have organisms and microorganisms interacting. Because of this, knowing how microbiomes work and how they can be altered to combat disease and stress, for example, is extremely important. As they develop this research, Slot's group is also building a library of potential fungi that could be used as biological control agents and uses high performance storage at OSC as well.

"Those that we find that are particularly associated with the coffee in its healthiest state could be potentially used by people in the coffee industry to be a natural pesticide—a living organism that's a pesticide," Slot said.

Slot's work is done in collaboration with Ana Alonso at the University of North Texas, and Priscila Chaverri at the University of Costa Rica/ University of Maryland. •

PROJECT LEAD // Jason Slot, Ph.D., The Ohio State University
RESEARCH TITLE // Evolutionary and ecological genomics of endophytic fungi of the coffee family
FUNDING SOURCE // National Science Foundation
WEBSITE // u.osu.edu/slot.1/research

PHOTO // Fernanda Rezende de Castro Moretti of Ana Alonso's lab collects metabolomic data (courtesy Ephrain Escudero). INSET // A fungus, *Mycena citricolor,* (courtesy Ephrain Escudero) causes disease on coffee plants.

Sentence Pair Modeling

Xu creates algorithms to boost understanding, generation

With the increased information-age use of voice interfaces and chatbots, Natural Language Processing (NLP) is becoming one of the world's most significant technologies, drawing from aspects of computer science, artificial intelligence and linguistics. The goal for NLP researchers, such as Wei Xu, an assistant professor of computer science and engineering at The Ohio State University, is to develop deep-learning techniques to enable computers to process or "understand" natural language in order to perform tasks like language translation and question answering.

For a computer, fully understanding and representing the meaning of spoken or digital language is extremely difficult. Human language is very complex, often carries encoded messaging and varies based on locale, culture and generation.

"We focus on designing algorithms for learning semantics from large datasets for natural language understanding and generation, in particular with stylistic variations," Xu said. "My research lies at the intersections of machine learning, natural language processing and social media."

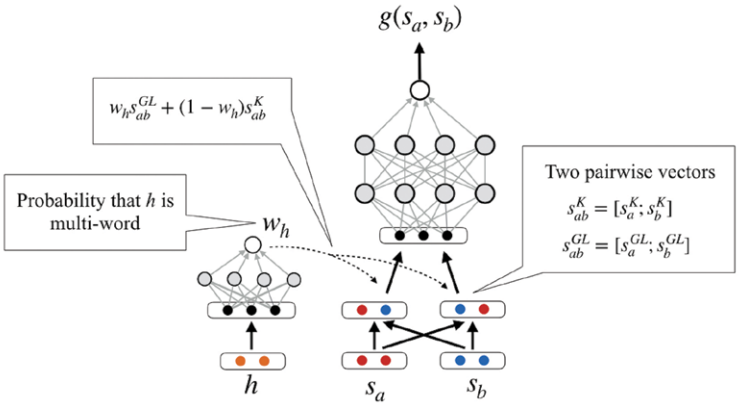
One of Xu's most recent research project, leveraging Ohio Supercomputer Center (OSC) services, involves developing a complex algorithm that evaluates sentence pair modeling, which is a fundamental technique underlying many NLP tasks. This measures the degree of equivalence in the underlying semantics of paired snippets of text; identifying whether two sentences express the same meaning; recognizing whether a hypothesis can be inferred from a premise; determining how well candidate sentences answer questions; and matching passages with an appropriate question and signaling the text region that contains the answer.

"To answer these questions and better understand different network designs, we systematically implemented five representative neural models and their variations on the popular PyTorch deep learning platform to analyze for sentence pair modeling," Xu explained.

In addition to using a small in-house GPU server to prototype their algorithms, her research team leveraged CPU and GPU nodes on OSC's Owens Cluster to process much larger modeling runs.

"We focused on identifying important network designs and presented a series of findings with quantitative measurements and in-depth analyses, releasing our implementations as an open-source toolkit to the research community."

Other research in Xu's lab involves NLP algorithms that match a complex sentence with a simpler version with nearly identical meaning, and, most recently, one that automatically divines the most likely meaning of a social-media hashtag; for example, quickly interpreting the hashtag #wingsofstrength as "Wings of Strength." •



PROJECT LEAD // Wei Xu, Ph.D., The Ohio State University
RESEARCH TITLE // Neural network models for paraphrase identification, semantic textual similarity, natural language inference, and question answering
FUNDING SOURCES // National Science Foundation, Defense Advanced Research Projects Agency
WEBSITE // cocoxu.github.io

FIGURE // Wei Xu and her colleagues are working to help computers better "understand" natural language through the development of algorithms for learning semantics. This chart outlines how multi-task learning is accomplished.

TAGS: ● DATA ANALYTICS ■ EDUCATION // HTC // HUGE MEMORY NODE ▲ INVENTION/DISCOVERY

Ordinal Outcomes

Archer group develops methods for predicting diagnoses

Science can be expressed in a myriad of different ways, but in the end, it all comes down to data, and methods are needed to appropriately interpret this data. Kellie Archer, Ph.D., professor and chair of the biostatistics division in the College of Public Health at The Ohio State University, leads a research group that develops statistical methods and computational algorithms for analyzing genomic data in oncology where oftentimes categorizing stages of the disease is relevant to tumor characteristics.

"In pathology, most of the diagnoses are recorded on an ordinal scale, and there's lots of different ordinal outcomes that people are familiar with, like stage of cancer," Archer said. "There's not a numerical relationship between pathological stages, but there's an order to them."

While many methods have been developed for continuous outcomes, such as quantitative traits like systolic blood pressure or level of prostate specific antigen (PSA), and for two-class outcomes (diseased or not diseased), there aren't as many for ordinal outcomes. Archer is trying to take some of the guesswork out of predicting these types of ordinal data where samples are hard to come by and there are a high number of variables involved.

"A lot of methods have been developed for continuous outcomes… but there's not as many for ordinal outcomes," Archer said. "One of the problems with a lot of statistical methods is that they were developed in situations where you have a large sample size relative to the number of variables that you're using to predict the outcome. Whereas in high-dimensional gene expression or other genomic data sets, you have the reverse problem."

Archer's group began with frequentist-based approaches to their statistical methods, which are useful in that they can identify variables important for modeling the ordinal outcome, but they do not provide p-values, standard error estimates or confidence intervals. Archer is now working on a Bayesian approach which incorporates probability distributions for the model parameters. This will provide more useful information for each predictor variable so that researchers will have better information about how important a variable is in the model.

The researchers are running their own software on the Owens Cluster at the Ohio Supercomputer Center to perform their statistical methods and process huge data sets. They are also using simulation studies to evaluate and continuously improve their methods.

"We design a lot of simulation studies where we can vary different scenarios: we vary the number of samples, we vary the number of variables, we vary the strength of the relationship between the variables and the outcome," Archer said. "Then we run our methods using a huge number of simulation studies to try to evaluate their performance under a wide variety of settings. And it would take forever if we just used our office computer to do that." •

PROJECT LEAD // Kellie Archer, Ph.D., The Ohio State University
RESEARCH TITLE // Informatic methods for predicting an ordinal response
FUNDING SOURCE // National Institutes of Health
WEBSITE // cph.osu.edu/people/karcher

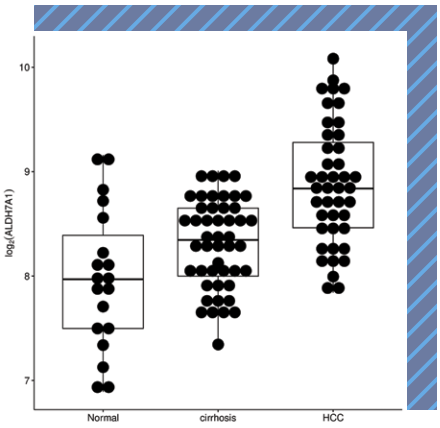


FIGURE // In examining genes involved in the progression from normal liver to cirrhosis to hepatocellular carcinoma (HCC), Kellie Archer's group identified several whose expression is either increasingly or decreasingly dysregulated. This figure represents one such gene, ALDH7A1.

Global Analytics

Spire enhances weather forecasting, transportation safety

Spire Global is one of the world's largest space-to-cloud analytics companies, providing global weather-forecasting services for the maritime and aviation industries as their satellites travel around Earth. Spire leverages the power of data analytics through high performance computing centers, such as the Ohio Supercomputer Center (OSC), and the world's largest fleet of multipurpose satellites to address customer needs, protect lives and save hundreds of millions of dollars.

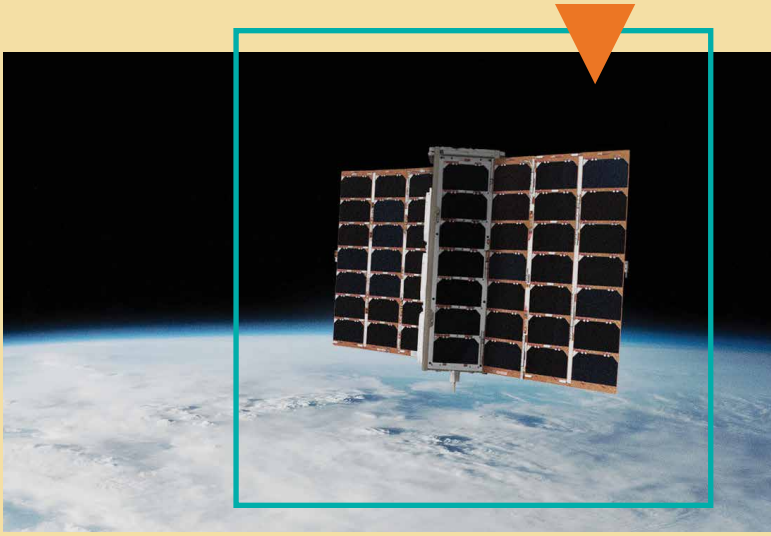
"There is not a single business or person on the planet which isn't affected by weather each day," Spire reminds people visiting their website.

Spire's constellation of nearly 100 football-sized satellites circle the globe, gathering large amounts of atmospheric, aviation and maritime data. Fusing this data with additional satellite and terrestrial data, it then procures millions of compute cycles and space on OSC's Owens Cluster and high performance storage systems. This feeds its proprietary system of weather models to create valuable weather predictions tailored to the operational needs of its customers.

"Because the data's always changing and the weather is always changing, the models always need tweaking or to be improved," said Timothy Brown, a software engineer for Spire. "We gather thousands of observations per day of atmospheric conditions—pressure, temperature and humidity—to just nudge the model a little bit. By having the actual observation, the forecast becomes even more valuable and accurate."

Spire also tracks location information of ships and planes, promoting fuel conservation, route planning efficiency and near real-time responses to global events like the recent incidents around the tanker Front Altair or Malaysia Airlines Flight 370 in 2014.

"We observe ships because of those disasters," said Brown. "The maritime industry later mandated installation of position broadcasters on ships after the Exxon Valdez incident. Aircraft have these tracking devices as well, and satellites can track both aircraft and ships in real-time."

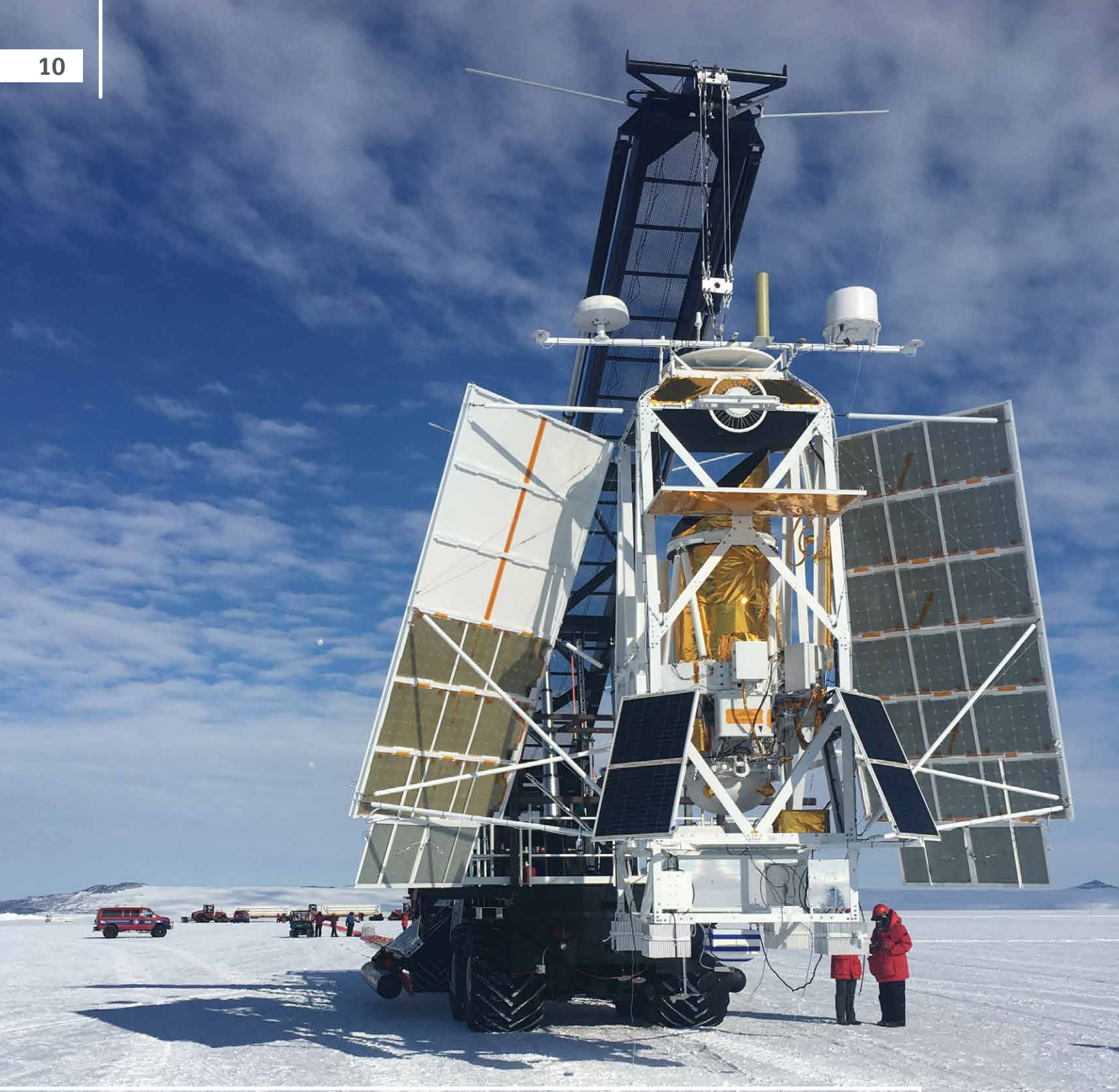


With exceptional compute power provided by centers like OSC, further applications around weather are possible, especially benefitting some of the more underserved communities globally.

"By using our satellites to read signals from GPS satellites—signals that are not passing through the atmosphere but reflecting off the earth—we can deduce measurements like soil moisture at a specific location," said Brown. "And, therefore, we can feed that into climatological models to determine if this growing season is going to be good or bad. So, for example, you could help predict a famine in Africa." •

PROJECT LEAD // Timothy Brown, Spire Global Inc.
RESEARCH TITLE // Weather forecasting and radio occulation data
WEBSITE // weather.spire.com

ILLUSTRATION // Spire Global maintains a constellation of nearly 100 small satellites circling the globe to gather large amounts of atmospheric, aviation and maritime data. This data is fused with additional satellite and terrestrial data and processed on OSC systems.



Neutrino Interactions

Connolly team seeks evidence in large data sets

The field of ultra-high energy (UHE) cosmic neutrino experiments has entered an exciting phase, according to Amy Connolly, Ph.D., a physics professor at The Ohio State University. Scientists there have accumulated enough data to reach a mature understanding of detector designs in the radio environment in Antarctica. They are well positioned for a potential discovery and poised to determine the future of the field for decades to come, according to Connolly.

"We search for the highest energy neutrinos in the universe as the only probes of cosmic distances at their energy regime, with experiments designed to detect them through a radio signature," said Connolly, also a researcher at the Center for Cosmology and AstroParticle Physics (CCAPP) at Ohio State. "Ultra-high energy (UHE) neutrino interactions probe much higher center-of-mass energies than in collisions produced at the Large Hadron Collider and thus could point to new physics."

Leveraging storage systems and each of the high performance computing clusters at the Ohio Supercomputer Center (OSC), her group leads simulations for two neutrino detection projects: the Antarctic Impulsive Transient Array (ANITA) balloon project and the Askaryan Radio Array (ARA) project.

The ANITA project is an international, NASA-funded stratospheric balloon payload detector program that over four flights has scanned the Antarctic ice for UHE neutrinos from its near-40-kilometer altitude. The ARA project is an array of five stations of antennas deep underground that measures the enhanced radio-frequency radiation emitted during the interaction of neutrinos with the Antarctic ice. Two next-generation experiments of each type have been proposed, named Payload for Ultrahigh Energy Observations (PUEO—not coincidentally the name of a subspecies of the Hawaiian short-eared owl) and Radio Neutrino Observatory (RNO), respectively.

"We utilize a supercomputer to analyze large data sets to perform leading searches for UHE neutrinos, run sophisticated simulations to design future experiments with optimal neutrino sensitivity and perform theoretical calculations to interpret the implications of our results," Connolly said. "That's why it's important that we can use OSC, because we can run lots of jobs in parallel and be able to examine whole sets of data in a reasonable amount of time."

During the bulk of the ARA analysis period, a small subset (a "burn sample" of approximately 10 percent) of the data is analyzed on the Owens and Pitzer Clusters in order to estimate backgrounds and find an optimal set of cuts. Once cuts are finalized, then the entire data set is run through to find the outcome of the search. For ANITA, Connolly's group uses the Owens Cluster to analyze the full dataset.

"We are also taking the lead on exploring new ideas that could potentially have a large payoff for the field, namely using genetic algorithms to improve neutrino detectors and a potentially game-changing technique to use radar to detect neutrino-induced particle cascades."

Connolly initiated the genetic-algorithms research to improve existing UHE neutrino experiments by leveraging this form of machine learning. The project, known as GENETIS (Genetically Evolving NEuTrIno TeleScopes), is led by undergraduate students and involves linking genetic algorithms together with state-of-the-art time domain modeling and the simulations of experiments in one complete loop.

The novel radar technique came to Ohio State with the recent arrival of CCAPP Postdoctoral Fellow, and now Ohio State Presidential Postdoctoral Scholar, Steven Prohira, Ph.D. Prohira's technique involves transmitting a continuous wave signal from within the ice in order to look for reflections from neutrino-induced particle cascades. This will require running simulations of the particle showers that neutrinos induce, the characteristics of the received signal after reflection from the cascade and its dependence on a variety of characteristics. •

PROJECT LEAD // Amy L. Connolly, Ph.D., The Ohio State University
RESEARCH TITLE // Searching for ultra-high energy cosmic neutrinos and designing next-generation detectors at the OSC
FUNDING SOURCES // National Aeronautics and Space Administration, National Science Foundation
WEBSITE // u.osu.edu/connolly

PHOTO // Amy Connolly is among a group of scientists searching for the existence of highest energy neutrinos, gathering data from such detectors as the international, NASA-funded stratospheric balloon payload detector known as the ANITA project and flown high over Antarctica.

TAGS: ● BIOHEALTH ■ HTC // HUGE MEMORY NODE // PROJECT STORAGE ▲ INVENTION/DISCOVERY

Risk Assessment

Goel group develops protocols to reduce ACL injuries

Sports enthusiasts know when a player is sidelined holding their knee, that "ACL" are the season-ending letters no one wants to hear. According to the American Academy of Orthopedic Surgeons, there are over 127,000 anterior cruciate ligament tears in the United States each year. Despite preventative measures and identifying ways to reduce the risk of an anterior cruciate ligament (ACL) tear or rupture, the rate of ACL reconstruction cases has increased over the last 30 years. The University of Toledo's (UT) Distinguished University Professor Vijay Goel has set out to reduce the rates of ACL tears with a novel ACL injury risk assessment protocol.

Goel and his team developed their protocol through in vivo, in vitro and in silico testing, using both cadaver and simulations to test several types of strains on the ligament. This project was pursued in collaboration with Dr. Tim Hewett, Ph.D., and his team from the Mayo Clinic.

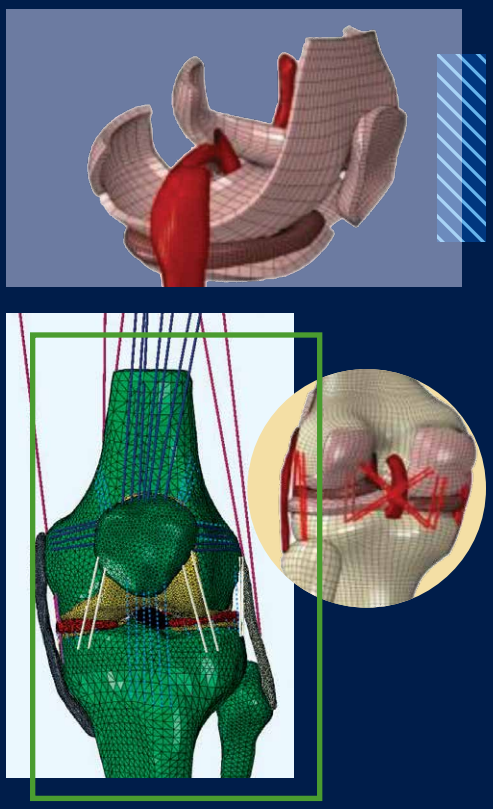
"It is generally agreed that risk factors for and casual factors of ACL injury can be anatomic, neuromuscular or biomechanical in nature," Goel said. "Despite multiple ACL injury studies that have successfully identified these risk factors, only moderate progress has been made toward clinically applicable integration of multiple risk factors into risk assessment protocols."

Goel's group is bridging this gap with the simulation approach and a finite element analysis that accounts for violent movements in all three planes of motion on the knee. They hope to develop a risk continuum that accounts for knee abduction movement, the knee moving away from the body's midline; anterior tibial shear force, pushing against the front of the lower leg; and internal tibial rotation movement, the lower leg twisting toward the body midline.

The team uses ABAQUS Dynamic to run large amounts of data for each model on the Owens Cluster at the Ohio Supercomputer Center (OSC).

"All our jobs must be run in parallel and with multiple nodes to speed up the process and get the desired results," Goel said. "Without the power of OSC, we would not be able to develop this protocol in the time we want to."

Goel is also the co-director of UT's Engineering Center for Orthopedic Research Excellence (E-CORE), where students are using high-tech approaches, including accessing high performance computing resources, to improve patient's quality of life. Goel received the Ohio Faculty Council Technology Commercialization Award from the Ohio Department of Higher Education last year for his technology commercialization efforts. •



PROJECT LEAD // Vijay Goel, Ph.D., University of Toledo
RESEARCH TITLE // Finite element models of the knee joint to investigate ACL injury mechanisms
FUNDING SOURCES // National Institutes of Health, Mayo Clinic
WEBSITE // utoledo.edu/engineering/ecore

ILLUSTRATIONS // Multiple studies have successfully identified risk factors for ACL injuries, but according to Vijay Goel, very few risk factors are integrated into risk assessment protocols. Therefore, he and his team are bridging this gap with a simulation approach and finite element analysis.

Toxin Diversity

Daly, partners study genetic variation of anemone venom

The Ohio State University's Marymegan Daly and her research partners probed the depths of Monterey Bay to collect samples of the tube-dwelling sea anemone, samples that are allowing the scientists to generate and analyze the transcriptomes of these ancient animals and reveal the diversity of toxins within their venom.

"Venom mediates interactions between sea anemones and the rest of their communities: it helps them to defend against predators, acquire prey and support symbiotic partners," said Daly, professor of evolution ecology and organismal biology. "Very little is known about the diversity of genes responsible for venom in sea anemones. Understanding the evolution of venom will impact our ability to predict and interpret the function of sea anemone venom in a variety of ecological and evolutionary contexts."

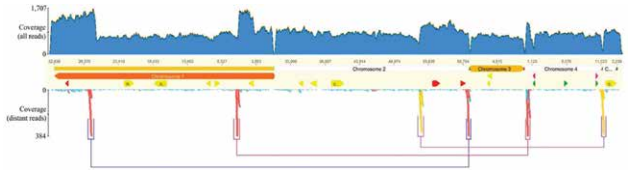
Daly and her colleagues attempted to use standard gene-sequencing software to generate the genomic information for the toxin study. However, many of them could not distinguish between sequencing errors and unique genetic variations. They finally found that the Geneious bioinformatics software platform could accurately make most of those differentiations. With the software in place, they knew that desktop computers and even the small local cluster were inadequate for the degree of computation that would be required. They turned their attention to the Ohio Supercomputer Center (OSC).

"We use our local cluster as a kind of intermediate sandbox," Daly said. "And then once we're ready to scale up and do comprehensive analysis then we use OSC's supercomputers. We feel like we've done our due diligence that way."

Early on, Daly's group recognized that in the process of seeking the venom RNA code, they had at the same time generated an entire RNS sequence dataset. "This was like filtering all of the water in a swimming pool for just one leaf," Daly explained. "What do you do with the rest of the water?"

They determined that the RNA of the tube anemone does what few other species' mitochondrial genomes have been shown to do. Defying the classic doughnut shape it "should" be in, the anemone RNA instead is arranged in several fragmented pieces. The biology of this simple relative of coral and the jellyfish turned out to be more complex than expected, possessing the largest mitochondrial genome on record—81,000 base pairs, or pieces of genetic information, compared to 17,000 base pairs found in human mitochondrial DNA. •

PROJECT LEAD // Marymegan Daly, Ph.D., The Ohio State University
RESEARCH TITLE // Using geneious to reveal venom gene copies hidden in sea anemone de novo transcriptomes
FUNDING SOURCE // National Science Foundation
WEBSITE // eeob.osu.edu/people/daly.66



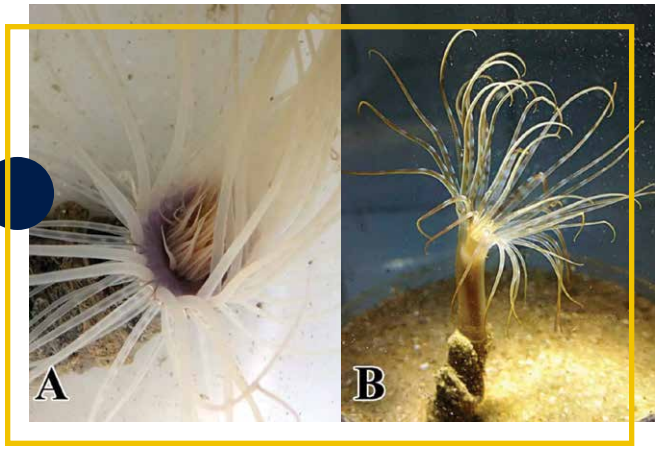
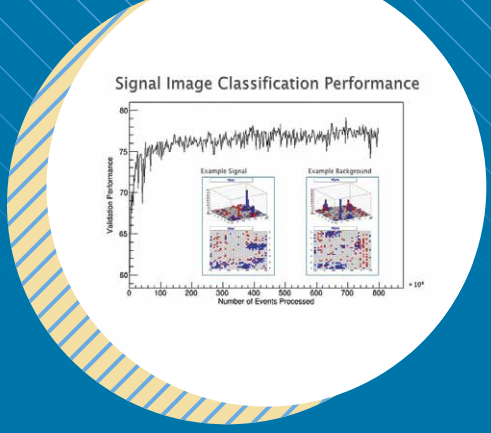


FIGURE // The mapped read of mitochondrial genome (*Pachycerianthus magnus*) generated by Marymegan Daly and her partners. PHOTOS // The species of Ceriantharia included in Daly's study: (A) *Pachycerianthus magnus* (Nakamoto, 1919); (B) *Isarachnanthus nocturnus* (den Hartog, 1977).



Processing Power

Hughes leverages HPC in research, classroom, business

Richard Hughes, Ph.D., has an impressive resume in the world of physics: He was a leading member of the team that discovered top quark, contributed to the search for dark matter in the Milky Way halo, was an integral part of the discovery and study of the Higgs Boson at the Large Hadron Collider—oh, and has founded his own company and is also a professor of physics at The Ohio State University. Typically, Ohio Supercomputer Center (OSC) users fall into one of three categories: academic research, classroom use or industrial. It should come as no surprise that Hughes has leveraged OSC for all three purposes.

Hughes first started using OSC a decade ago when his high-energy physics research required more resources than the local department resources could provide. Several years later, Hughes took an interest in machine learning and, again, needed more resources than he had locally.

"A student and I started to look at various tools that were available at OSC and whether they would work for us, and it seemed like they would," Hughes said. "That was incredibly successful. That's where I really learned how to attack machine learning problems at a really large scale. I got a lot of experience with using GPUs which I had never used before."

OSC also provided a license for the MATLAB software that Hughes could use, as well as Python. As he gained more experience with machine learning through his own research at OSC, Hughes was tapped for an outside consulting project by his sister Stephanie Hughes, an expert in strategy and competitive intelligence and professor and chair of the management department within the Haile US Bank College of Business at Northern Kentucky University.

Together, the two founded Chapel Hill North (CHN) Analytics, a company that does text-based analytics. Hughes' machine-learning expertise translated well, from looking at millions of records containing numbers to millions of documents with text. Again, Hughes began running low on processing power on his local computing system and started wondering what he could do to manage the data—now from an industry standpoint.

"I knew that I could do a lot of the things I wanted to do at OSC, but I couldn't do them because it wasn't research work," Hughes said. "I just started searching the (OSC) website wondering if they had some sort of industry program—and they did!"

Hughes worked closely with OSC's Chase Eyster, business development manager, and Alan Chalker, Ph.D., director of strategic programs, to ensure he was using the right resources for his company so he could focus on the business rather than the computational side.

FIGURE // Richard Hughes and a student leveraged OSC resources to do image classification.

"It seemed too good to be true. I found out that not only is it something they encourage, they're incredibly helpful," Hughes said. "It really is an advantage for Ohio companies to have access to this kind of computing power."

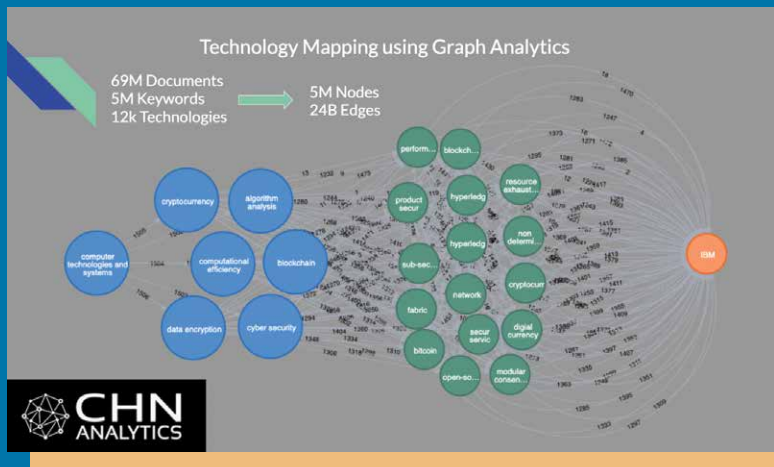
Hughes is also introducing the future of Ohio's companies to this computing power in his classroom. The head of the physics department at Ohio State approached Hughes about creating an introduction to machine learning course with practical applications in mind. His original plan was to have students install a Python environment on their laptops, but he realized that could create a headache with all the different platforms students use.

"As I got closer to the startup of the class, I began to panic," Hughes said. "It's usually a complete disaster trying to install software on all these different platforms. I thought, you know, maybe I could do it at OSC, but then the problem is I have to teach students how to operate within the OSC environment."

Hughes started the class using a Google service, which was not ideal. Then a colleague told him about OSC's OnDemand platform. Hughes immediately switched the class over, with help from OSC's education and training support team. Hughes could focus on the material and not have to worry about installing programs or the environment.

"I have to admit I was a little worried because you know I'm thinking this could be a complete train wreck, but once again I had so much support from OSC," Hughes said. "I'm going to teach this class again in the fall, and I'm going to start out from day one with the OnDemand setup. I'm really looking forward to it, because it made my life as an instructor incredibly easy." •

> It really is an advantage for Ohio companies to have access to this kind of computing power. It seemed too good to be true. I found out that not only is it something they encourage, they're incredibly helpful.
>
> — *Richard Hughes, Ph.D.*



PROJECT LEAD // Richard Hughes, Ph.D., The Ohio State University
RESEARCH TITLES // Big Data analytics in physics; The search for new Higgs channels at the LHC using the CMS detector
FUNDING SOURCES // The Ohio State University, Department of Energy, National Science Foundation
WEBSITE // asc.ohio-state.edu/hughes.319/hughes.html

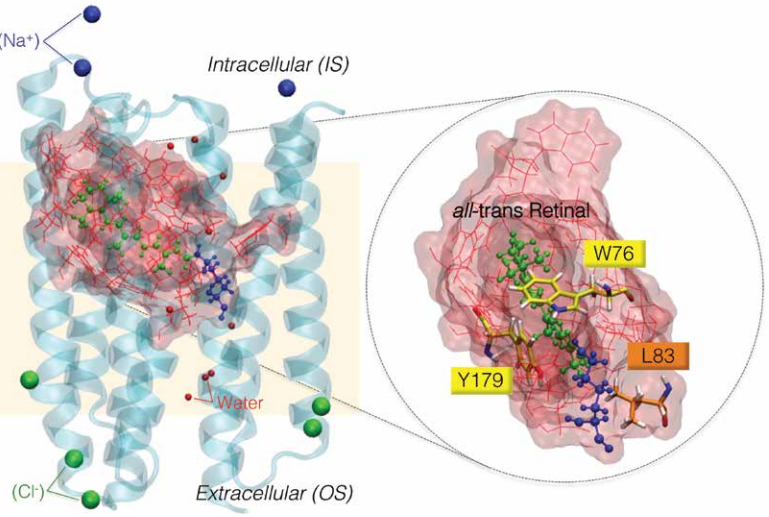
FIGURE // Richard Hughes and sister Stephanie Hughes, co-founders of Chapel Hill North Analytics, have a team there conducting text classification work exemplified by the illustration above, Technology Mapping using Graph Analytics.

Biological Photoreceptors

Olivucci lab models fluorescence of rhodopsin mutants

In the emerging field of optogenetics, scientists are working to develop light-responsive proteins (photoreceptors) that will allow them to observe the nerve impulses in the brain or to control specific cellular features, such as metabolic pathways, gene expression and ion channels.

Massimo Olivucci, Ph.D., research professor and director of the Laboratory for Computational Photochemistry and Photobiology at Bowling Green State University is leveraging Ohio Supercomputer Center (OSC) services to study the optogenetic potential of *Anabaena* sensory rhodopsin, a light-sensing protein found in a cyanobacterium, or blue-green algae. Rhodopsin proteins are present in almost every organism featuring "vision" in a basic or complex form.



"We are trying to understand the details of molecular mechanisms allowing biological organisms to capture and use light energy and how this knowledge could then be exploited to design new materials," Olivucci said. "For this reason, we use computers to model the light-triggered dynamics of biological photoreceptors at atomic-level resolution."

Olivucci's group is also modeling the dynamics of rhodopsin mutants—in principle, thousands of mutants—in order to find molecular modifications that will produce high levels of fluorescence. More specifically, the lab is screening mutant models to isolate a few promising variants that biologists then attempt to produce in the lab. In a sense, the group is attempting to simulate biological evolution under a selective pressure for maximizing fluorescence. To facilitate this computationally intensive process, Olivucci is leasing a small group of very powerful nodes on OSC's Owens Cluster—what is known in the supercomputing industry as a "condo."

"Given the large number of calculations we need to run, we need extremely fast processors," Olivucci said. "Even more importantly, these are very long calculations, and the condo arrangement with OSC offers us the opportunity to run jobs for an entire week."

Olivucci also points out that the number of variants they are screening is far beyond that which an individual is able to process manually in a reasonable timeframe. Therefore, his group is continuously working to enhance their Automatic Rhodopsin Modeling (ARM) software to build and evaluate thousands of computer models of variants.

Additionally, this research is providing students on his team with extremely valuable experience. Pharmaceutical companies today, for instance, design new drugs by simulating drug interactions within a computer model before investing enormous amounts of time and money testing them in the laboratory or on live subjects. •

PROJECT LEAD // Massimo Olivucci, Ph.D., Bowling Green State University
RESEARCH TITLE // Investigating *Anabaena* sensory rhodopsin mutants for the design of novel optogenetic tools
FUNDING SOURCE // National Science Foundation
WEBSITE // ccmaol1.chim.unisi.it

ILLUSTRATION // Massimo Olivucci conducts research in the emerging field of optogenetics. These figures show a computational quantum-mechanical/molecular mechanical-model generated by his Automatic Rhodopsin Modeling software for *Anabaena* sensory rhodopsin wild type.

Big Data Challenges

Production portal allows researchers to focus on science

Researchers and clinicians at The Ohio State University's Wexner Medical Center rely upon the staff members of the Biomedical Informatics Shared Resource (BISR) to address their high-throughput, high-dimensional biological data-analysis needs, such as for next-generation sequencing research. The gigabytes of data generated by these technologies present enormous challenges in storage, computational resources, analysis and interpretation. Moreover, many of the latest technology advances designed to solve big data challenges have not yet been widely adopted by the research and clinical communities.

In collaboration with specialists at the Ohio Supercomputer Center (OSC), BISR staff members have developed an interactive analysis reporting system delivered through OSC's existing OnDemand platform, a secure, scalable, open-source software environment that gives clients web-based access to high performance computing services.

"The cost of supporting this service using OnDemand is lower than building or managing a stand-alone solution," said Eric Franz, lead engineer for OSC's Gateways group. "The existing OnDemand installation already has built-in solutions for authentication, authorization, accounting and monitoring that can be leveraged by any OnDemand app."

"Most importantly, this approach succeeds in shifting the responsibility for ensuring the production availability of the service from BISR to OSC, which enables us to focus more on our domain of work," said Venkat S. Gadepalli, Ph.D., a research scientist with BISR.

App developers chose to employ the R programming language, considered an ideal tool for data wrangling, analysis and visualization, and Shiny, an R-based library that makes it easy for researchers without the expertise of a software developer to build interactive web apps and reports.

BISR first considered self-hosting Shiny but was concerned about the cost of setting up authentication. The staff also found that authorizing dataset access is a problem that Shiny Server currently does not address. And, because BISR's results data was already stored on OSC systems, it just made more sense for OSC to host BISR's Shiny visualization service.

"BISR is actively evaluating and iteratively refining the features of the Shiny report app, which is the main focus for the analyst's and researcher's user experience," explained Maciej Pietrzak, Ph.D., technical director of BISR and a research assistant professor in the biomedical informatics department at Ohio State. "In the next 18 months, we expect to expand the number of users by actively engaging the RNAseq Shiny report app with relevant projects and developing other apps." ●

PROJECT LEAD // Maciej Pietrzak, Ph.D., The Ohio State University
RESEARCH TITLE // Scaling R Shiny Apps to Multiple Concurrent Users in a Secured HPC Environment Using Open OnDemand
FUNDING SOURCE // National Science Foundation
WEBSITE // medicine.osu.edu/bmi/people/research/maciej_pietrzak/Pages/index.aspx

FIGURE // In collaboration with OSC specialists, Maciej Pietrzak and his colleagues have developed an interactive analysis reporting system delivered through OSC's existing OnDemand platform, allowing researchers to focus more on their clinical domain of work, such as prostate cancer.

Process Automation

TotalSim simulation data management app speeds results

In the auto racing world, if you blink, you fall behind.

With decades spent in racecar engineering and design, nobody knows that better than the TotalSim US team, which relies on the high performance computing (HPC) resources at the Ohio Supercomputer Center (OSC) to produce invaluable computational fluid dynamics (CFD) data for the auto racing teams and other industrial companies they partner with.

The resulting problem has been the tidal wave of data and how to present the information, which has been becoming an unwieldy and time-consuming headache. So, TotalSim innovated.

Technical Director Naethan Eagles and President Ray Leto dreamed up TS Results—a simulation data management app that efficiently processes and collates CFD simulation data and allows TotalSim to share the results quickly in a standardized format.

"We're using about 2,000 cores a day at OSC, 300 days a year," Leto said. "That's a ton of data. You can get buried in it! While you're trying to come up with an answer, you may miss your deadline or an opportunity to do more work."

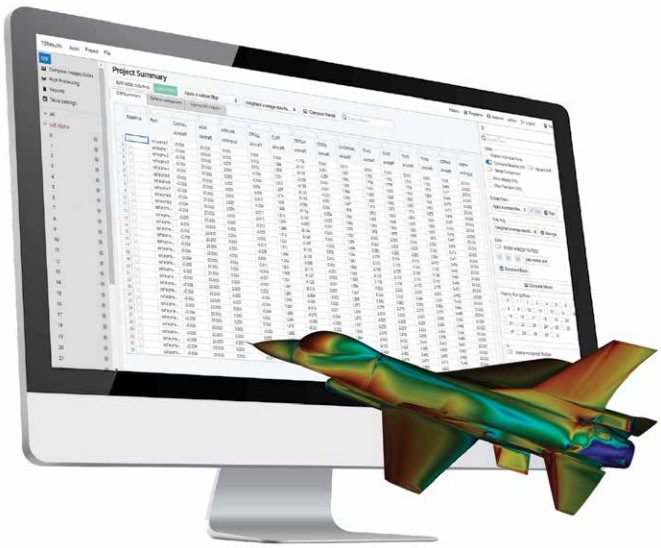
In the auto racing world, where winning and losing literally hinges on hundredths of a second, there's no room for missed deadlines or hesitation. The same mindset can be applied across various industries where rapid design cycles are becoming the norm.

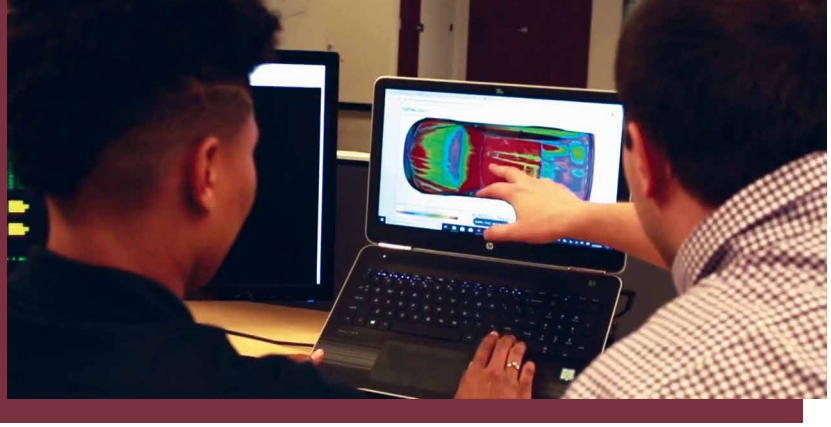
While the data TotalSim was producing was incredibly insightful, the problem came in post-processing. Although TotalSim had pioneered an automated CFD process including creation of images and movies, the team was still manually gathering results, creating charts and graphs, and writing reports and PowerPoint presentations, before sending it off to a customer.

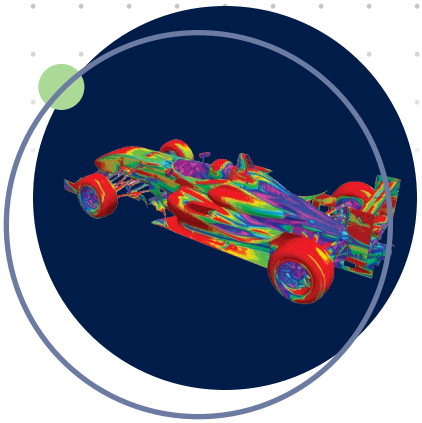
As a result, the TotalSim team began looking for a way to more fully automate the process. They found a solution in OSC's AweSim program. AweSim's app eco-structure was designed to give industrial partners an end-to-end workflow that included post processing of HPC results. TotalSim, a founding partner of AweSim, extended their contribution to the post processing and collaboration piece to create the TS Results web app.

"We just made it a standalone app," Leto said. "Out of the AweSim program, TS Results really started to take off. Now it's moved into something we use internally on a daily basis, and we keep adding outside customers."

Images, movies and data; it's all collected and published into the TS Results web app that is part of OSC's OnDemand platform—an open-source web portal that provides advanced web and graphical interfaces for HPC centers. TotalSim's clients can see results as soon as a simulation is done. It allows for more collaboration and ultimately, better decision-making.







While the bulk of TotalSim's HPC usage is in the world of auto racing and automotive design using TotalSim's automated OpenFOAM CFD process, the team also uses TS Results for clients who leverage other CFD software (like STAR CCM+ or FUN3D). It is part of TotalSim's many CFD Web Apps: TS Auto for automotive and motorsports; TS Aero, for aviation simulations; TS Truck, for semi-truck aerodynamic simulations; and other customized apps in various industry verticals.

Leto said TotalSim also has developed TS Results into a commercially viable product for engineers either doing CFD simulations on their own internal clusters or at any HPC center. TotalSim recently received a Phase I, federal Small Business Innovation Research (SBIR) program award to demonstrate how TS Results could be adapted for use by Air Force researchers to increase their productivity.

The real payoff is time. Leto said his team spends up to two-thirds less time on post-processing because of the app and that one of his main clients said it has doubled their efficiency by allowing them more time to do analysis.

The app is continuing to evolve. TotalSim added math functions and more plotting and reporting routines to the 2.0 version, and the group has its sights set on doing much more.

"Now that we've set up these platforms, the next layer on top is more analytics, maybe applying optimization and machine learning to the design processes to get better insight by using different analysis techniques while getting even more efficient," Leto said. •

> Out of the AweSim program, TS Results really started to take off. Now it's moved into something we use internally on a daily basis, and we keep adding outside customers.
>
> — *Ray Leto, President of Totalsim*

PROJECT LEAD // Ray Leto, President, TotalSim
RESEARCH TITLE // TotalSim's TS Results delivers
FUNDING SOURCE // TotalSim, Department of Defense
WEBSITE // totalsim.us/tsresults

PHOTOS // TotalSim US produces computational fluid dynamics data for auto racing teams and other companies. Though Ray Leto's firm had automated the process, it still included some manual tasks until they leveraged OSC's AweSim app eco-structure to create the TS Results web app.

TAGS: ● BIOHEALTH ■ HTC ▲ INVENTION/DISCOVERY

Antibiotic Resistance

Van den Akker's research seeks treatment alternatives

Antibiotic resistance causes 23,000 deaths and two million infections each year in the United States, according to the Centers for Disease Control and Prevention. Across the globe, antibiotic resistance is growing, and it is a threat to everything from food production to healthcare to saved lives.

"Bacterial infections are arguably some of the most serious threats to humankind to date," shared Focco van den Akker, Ph.D., an associate professor of biochemistry at Case Western Reserve University. "The options for treating infections have dwindled substantially, and the reports of antibiotic resistant pathogens are increasing at an alarming rate."

With the help of the Ohio Supercomputer Center (OSC), van den Akker is trying to unravel the mysteries of certain bacterial enzymes that block medications from working. His discoveries could lead to new ways to treat bacterial infections.

One of his recent studies focuses on the lytic translgycosylase enzyme from a major foodborne pathogen called *Campylobacter jejuni*. This pathogen is the second highest cause of hospitalization by bacterial infection amongst adults over 65 years old. Using powerful OSC computational and storage resources, insights from van den Akker's research could lead to new treatments for this pathogen.

"Our computational and experimental results reveal how this important bacterial enzyme works at a molecular level to modify its bacterial cell wall," van den Akker said. "The bacterial cell wall is a known and current drug target for many antibiotics such as penicillin. Gaining molecular insights into these cell-wall processes could lead to the development of new antibiotics, including new classes of antibiotics, to combat the current antibiotic resistance problem."

Van den Akker said that gaining insights into how the lytic transglycosylase enzyme functions would lead to a broader understanding of similar enzymes from other bacterial pathogens that stop medications from working.

"Such compounds could lead to new therapeutic opportunities to treat bacterial infections," he said.

Running Nanoscale Molecular Dynamics (NAMD) software on OSC's Owens Cluster, van den Akker performed long, intense simulations to perform calculations.

"The length of our simulations was up to one millionth of a second, which for an all-atom simulation is a very long simulation that took one month of calculations using 28 processors at a time," van den Akker said. "The software and web-based user interface of OSC are about as good as it gets, I was very impressed." •

PROJECT LEAD // Focco van den Akker, Ph.D., Case Western Reserve University
RESEARCH TITLE // Molecular dynamics simulations and inhibitor screening of the lytic transglycosylase protein
FUNDING SOURCE // Case Western Reserve University
WEBSITE // case.edu/med/biochemistry/vandenAkker_lab

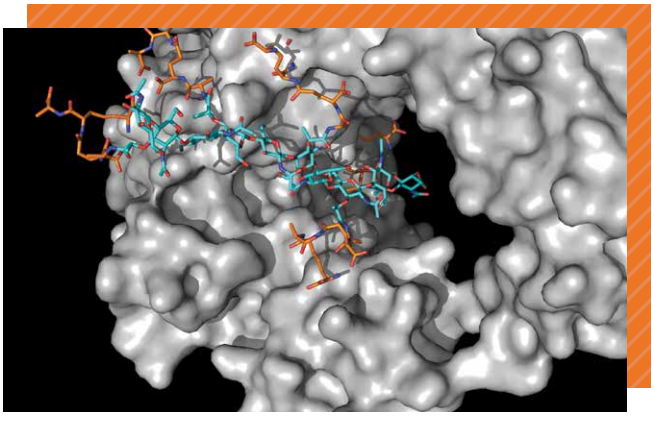


ILLUSTRATION // Sntibiotic resistance is growing worldwide. Focco van den Akker is trying to unravel the mysteries of certain bacterial enzymes that block medications. In this illustration, a peptidoglycan strand enters the hole of the *Campylobacteri jejuni* lytic transglycosylase to access its active site.

Maritime Viruses

Sullivan lab studies samples from Arctic soils, global seas

Matthew Sullivan, Ph.D., and the Ohio Supercomputer Center (OSC) have teamed up to give scientists insight into how to better study viruses found in a variety of communities. This information could prove invaluable to understanding everything from what's going on inside our bodies to how we might combat climate change.

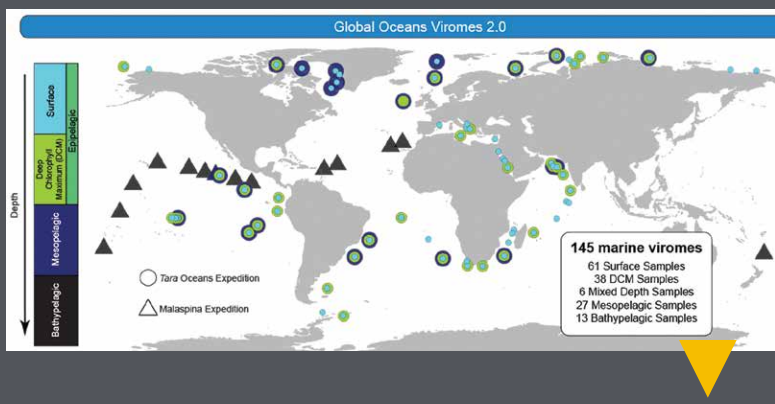
Sullivan's lab develops datasets and informational tools that help scientists peer into viruses found in soils, oceans, humans and numerous other environments. This year, a computational research scientist in Sullivan's lab, Benjamin Bolduc, Ph.D., used each of OSC's three clusters—Pitzer, Owens and Ruby—to create a large-scale analysis of viruses found in thawing Arctic soils and the global oceans. The research began when Sullivan was part of the three-year Tara Oceans Expedition, in which more than 200 experts ventured into the world's oceans, including tropical, temperate, Antarctic and Arctic waters, to collect samples of the many ocean inhabitants, including viruses and bacteria.

He and Bolduc then harnessed the high performance computing power at OSC to conduct two studies.

"Both studies provide critical baselines for scientists to try to evaluate the role of soils and the oceans in climate change as well as a roadmap for how to study viruses in complex communities," said Sullivan, a professor of microbiology and civil, environmental and geodetic engineering at The Ohio State University.

The first study examined thawing soils from northern latitudes. While the larger group from the expedition is working to discover how microbes are responding to this thaw, Sullivan assessed how viruses in these soils might be involved.

"We found that there were thousands of viruses," he said, "that they infected many key carbon cycling microbes, and that they contained genes that were critical to carbon cycling metabolisms."



The second study focused on viral communities in the global oceans, where the more than 200,000 new viral species were discovered and mapped at latitude and depth.

"This highlighted the Arctic as a biodiversity hotspot and provided the first look at diversity within the species as opposed to just studying the number of species," Sullivan said. "This foundational look at viral ecology in the oceans will provide baselines for countless future studies."

Sullivan's lab used microbiome apps that his team either developed or brought to OSC. These apps take terabytes of raw sequence data and help Sullivan develop new methods to simplify the data. The result was what he calls an ecogenomic toolkit—for each virus ("iVirus") and microbe ("iMicrobe")—and both toolkits are available at OSC.

"It's a best-in-the-world type toolkit, and we could not even put these datasets together without the power of OSC," Sullivan said. •

PROJECT LEAD // Matthew Sullivan, Ph.D., The Ohio State University
RESEARCH TITLE // Ecological impacts and drivers of viral communities in the global oceans
FUNDING SOURCES // The Ohio State University, National Science Foundation
WEBSITE // u.osu.edu/viruslab

FIGURE // Matthew Sullivan and Benjamin Bolduc studied samples from the world's oceans to provide critical baselines for scientists evaluating the climate-change role of soils and oceans, as well as a roadmap for studying viruses in complex communities.

TAGS: ● BIOHEALTH ■ HTC ▲ INVENTION/DISCOVERY

Biomolecular Machines

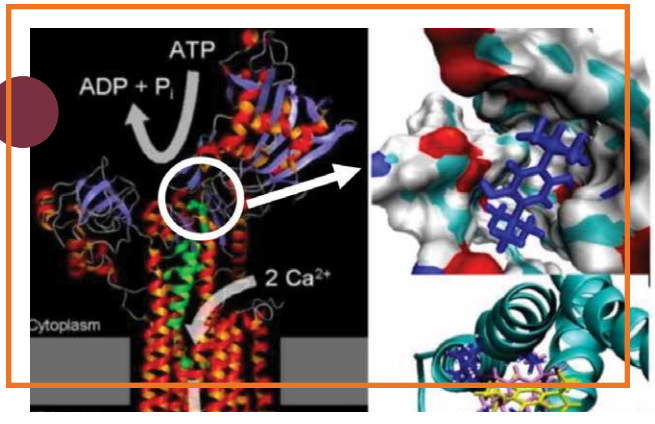
Jayasinghe group models binding to create better drugs

The cornerstone of an effective therapeutic drug development program is a rock-solid computational protocol that accurately and efficiently illustrates how molecules interact within the medicine and inside the human body. That information can be used to help fight and cure disease.

But it first requires someone to get deep into the weeds to study the physics of how drug-like molecules bind together with molecular machines such as proteins. Manori Jayasinghe, Ph.D., an associate professor of physics at the University of Cincinnati Blue Ash College, is such a person.

Jayasinghe researches computational molecular biophysics and, as one might expect, that work is computationally intense. That is why Jayasinghe uses the world-class storage and high performance computing systems of the Ohio Supercomputer Center (OSC).

Jayasinghe's current research involves modeling "the structure, dynamics and processes of biomolecular machines, like proteins or DNA," as she works toward deep insight into how the subunits of "large biomolecular systems cooperate to work like a real machine to perform their intended action in the living cell."



In order to develop an understanding of the structure, dynamics and processes of proteins and DNA, Jayasinghe performs numerous molecular dynamics simulations on the Oakley and Owens clusters at OSC. To accomplish that work, she relies on a variety of software packages, such as CHARMM, Python, Gromacs, Gaussian and MatLab.

"Parallel processing, such as the CHARMM program we use for our simulations, efficiently use high-bandwidth, low-latency, communication between cores to distribute computational tasks to multiple processors," Jayasinghe said. "We can get results five, six times faster that way than with single processor computations."

As those results get back to Jayasinghe, she can identify high-potency, drug-like compounds that bind with targeted proteins and help drug developers in their synthesis process. One recent project on which she teamed with OSC involved unraveling the reactions within proteins that are the target for anti-cancer medicines in prostate cancer therapy.

That information will help Jayasinghe develop computational methods to estimate the effectiveness of drug potency on specific proteins.

"That protocol will be used to study other inhibitor/enzyme systems in the future," Jayasinghe said. "This knowledge at the molecular level and the inhibition mechanisms (which slow or prevent chemical reactions) will serve as a great resource of information to characterize mechanisms of enzyme inhibition in general." •

PROJECT LEAD // Manori Jayasinghe, Ph.D., University of Cincinnati Blue Ash College
RESEARCH TITLE // Evaluation of absolute binding free energies of BHQ (2,5 di-tert-butylhydroquinone) analogues binding of human Ca2+ ion transporter (SERCA)
FUNDING SOURCE // University of Cincinnati
WEBSITE // linkedin.com/in/manori-jayasinghe-945b9a32

ILLUSTRATIONS // In these illustrations, Manori Jayasinghe identified high potency, drug-like compounds that bind with targeted proteins and help drug developers in the synthesis process.

Funding Allocations

Kwon builds economic model of family decision making

Ask any parent worldwide what the most important thing in life is, and the answer will likely be his or her children. Allocations of public funds for families with children vary wildly from country to country, and besides affecting a family's bottom line these programs could affect a child's development as well. Sungjoon Kwon, an advisee under David Blau in the economics department at The Ohio State University, is studying the effect of child care subsidies and income transfers on child cognitive and non-cognitive skill development, as well as implications for how public funds are allocated.

Kwon built an economic model of family decisions regarding employment, child care and the allocation of resources to children. Using high-speed data processing through the Ohio Supercomputer Center (OSC), he simulated the effects of child care subsidies versus the effects of income transfers. His model employed data gathered from the Longitudinal Study of Australian Children.

"This task requires strong computing power because the model solution is obtained after evaluating millions of data points for each set of candidate model parameters, and the model estimation requires obtaining solutions for hundreds of sets of candidate model parameters," Kwon said. "The use of OSC helped to save me a lot of time because multiple cores can be used to solve the model at multiple sets of model parameters at the same time."

Kwon's simulation results suggested the optimal public policy on aiding families with children allocates about seven percent of government spending to child care subsidies with the rest going to income transfers. The simulation showed that subsidies have a positive effect on cognitive skill development but a negative effect on non-cognitive skill development. Income transfers showed positive effects on both types of skill development.

"This is quite surprising because child care subsidies usually promote use of formal child care," Kwon said. "However, my results show that formal child care is less productive in producing non-cognitive skill compared to maternal child care at preschool ages, so substituting formal child care for maternal child care harms child non-cognitive skill development."

Kwon conceded that since the project used data from only Australia, the results are likely to be different for data from different countries. For example, he said, it is known that European countries have high-quality formal child care, so simulation results could show that it is optimal to allocate more to child care subsidies than income transfers. •

PROJECT LEAD // David Blau, Ph.D., The Ohio State University
RESEARCH TITLE // Parental choice of child care, child development and the effect of income transfers and child care subsidies
FUNDING SOURCE // The Ohio State University
WEBSITE // economics.osu.edu/people/blau.12

		Difference/ Percentage Change from Benchmark	
	Benchmark	Child Care Subsidy	Income Transfer
	(1)	(2)	(3)
Government Spending per Week	-	$ 514,006	$ 514,006
Child Care Subsidy Rate or Income Transfer Rate	-	100%	$ 82
Cognitive Skill at Ages 10-11	0.0000	0.0111	0.0032
Non-Cognitive Skill at Ages 10-11	0.0000	-0.0113	0.0049

Note: Benchmark is a scenario without child care subsidy and income transfer. Child Care Subsidy and Income Transfer are scenarios with no maternal work requirement and income test. The subsidy rate is 100% in Child Care Subsidy scenario. The values in Column 1 are averages in Benchmark. Cognitive and Non-Cognitive skills are restandardized, using Benchmark's means and standard deviations. The symbol '$' indicates the Australian dollar.

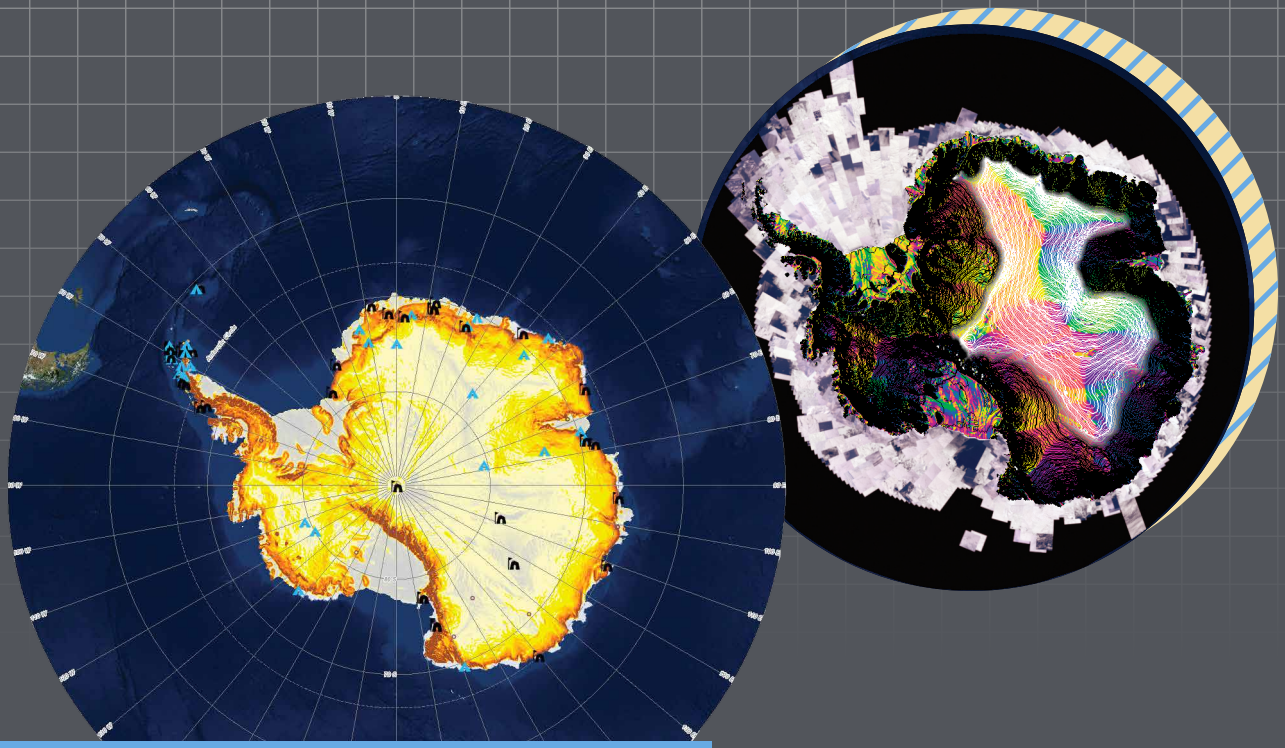
Table 4: Comparison between Child Care Subsidy and Income Transfer

	Child Care Subsidy	Income Transfer
Government Spending per Week	$ 51,739	$ 694,515
	6.9%	93.1%
Rate	61%	$ 258
Maternal Work Requirement	No	Yes (more than 0 hour)
Income Cutoff	$ 1,515	$ 2,590

Note: Total government spending per week is 746,254 AUD. The symbol '$' indicates the Australian dollar. The weight on cognitive skill (ω) is 0.57.

Table 5: Optimal Mix of Child Care Subsidy and Income Transfer

TABLES // Sungjoon Kwon used many tables, two of which are shown here, to show different considerations in trying to reach the optimal public policy on aiding families with children.



Digital Elevation

Howat, project team construct hi-res maps of Antarctica

Until 2018, scientists could find better terrain maps of Mars than they had of Antarctica.

Now, with software engineering, code optimization and parallel software development assistance from the Ohio Supercomputer Center (OSC), they have "the highest-resolution terrain map by far of any continent," according to Ian Howat, Ph.D.

Howat is a professor of earth sciences and director of the Byrd Polar and Climate Research Center at The Ohio State University, as well as the leader of a mapping project called The Reference Elevation Model of Antarctica (REMA).

"At this resolution, you can see almost everything. We can actually see variations in the snow in some places. We will be able to measure changes in the surface of the continent over time," said Howat. "We will see changes in snow cover, changes in the motion of ice, be able to monitor river discharge, flooding and volcanoes and to see the thinning of glaciers."

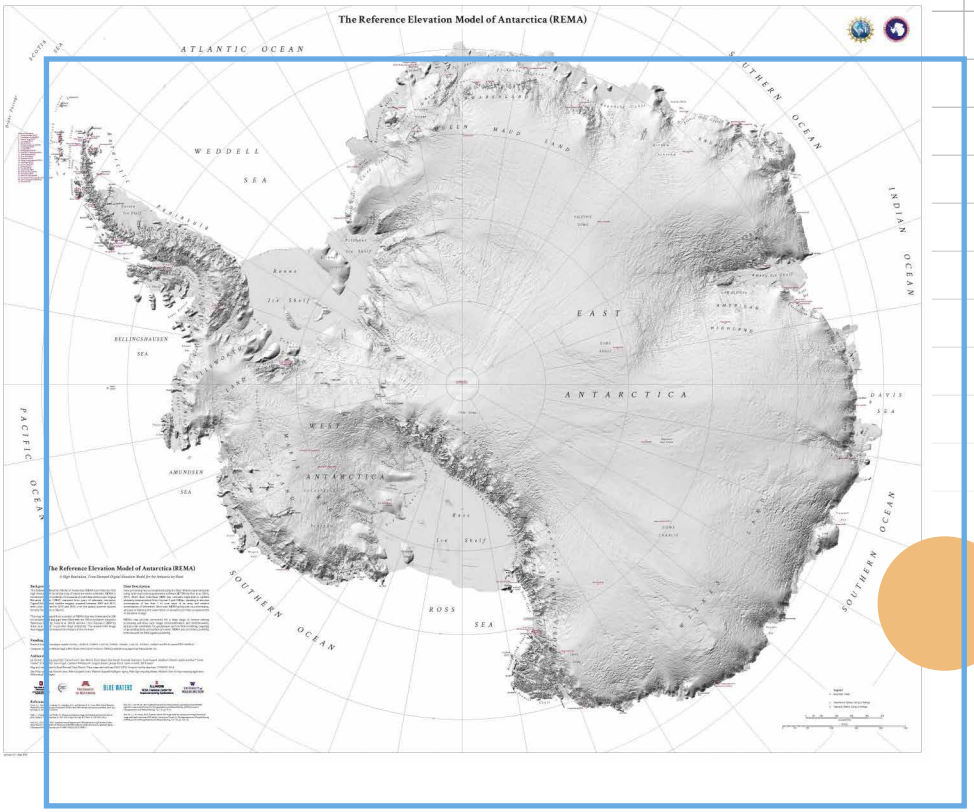
The model's total file size is more than 150 terabytes, or 150,000 gigabytes, which is precise enough to allow scientific teams to plan some trips over the continent's treacherous terrain.

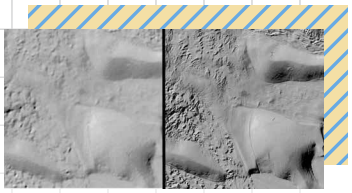
"We had to start from scratch to build this. The software had to filter the data, process it and turn it into a refined product for the scientific and broader community to use," Howat said.

The project began with images taken from a constellation of polar-orbiting satellites. Using the open-source Surface Extraction from TIN-based Searchspace Minimization (SETSM) software, developed at Ohio State by Howat and Myoung-Jong Noh, Ph.D., the images were processed into Digital Elevation Models. For this data-intensive process, Howat and his colleagues used the petascale Blue Waters supercomputer located at the National Center for Supercomputing Applications in Illinois.

"Dr. Howat has used OSC systems for earlier terrain

PHOTOS // Two views from the Reference Elevation Model of Antarctica viewer: a slope map (left) and an aspect map (right).





mapping projects, including one that examined the melting Greenland ice sheet in 2014 and another that aided disaster relief assistance following Nepal's 2015 earthquake," said Karen Tomko, Ph.D., OSC's director of research software applications and manager of the scientific applications group.

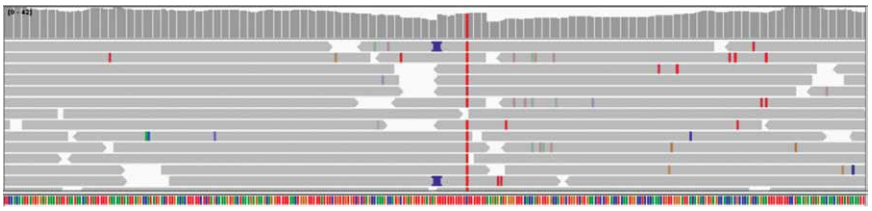
Members of Tomko's group (Judith Gardiner, Ph.D., and Samuel Khuvis, Ph.D.) have been collaborating on software development with Howat and Noh, co-principal investigator of the grant. "The REMA project was better suited for a petascale system, and we were pleased to help them scale the code accordingly," Tomko said.

Other collaborators on the project included the Polar Geospatial Center at the University of Minnesota. •

> At this resolution, you can see almost everything. We can actually see variations in the snow in some places.
>
> — *Ian Howat, Ph.D., professor of earth sciences and director of Byrd Polar and Climate Research Center at The Ohio State University*

PROJECT LEAD // Ian Howat, Ph.D., The Ohio State University
RESEARCH TITLE // Automated, high resolution terrain generation for XSEDE
FUNDING SOURCES // National Geospatial-Intelligence Agency, National Science Foundation
WEBSITE // byrd.osu.edu/people/howat.4

ILLUSTRATION // The Reference Elevation Model of Antarctica created by Ian Howat and his research partners now represents the highest-resolution terrain map of any continent on Earth. IMAGES // Comparing older surface imaging and the new map shows dramatically improved detail.



Next-gen Sequencing

Zhao leverages HPC in classroom, lab

With assistance from the Ohio Supercomputer Center (OSC), Research Scientist Yue Zhao, Ph.D., leads several classroom and lab sessions for an introductory bioinformatics course taught at The Ohio State University by Lijun Cheng, Ph.D., assistant Professor of biomedical informatics. The class introduces students to basic topics of bioinformatics, including sequence analyses, proteomics, microarrays and associated genomics and genetics databases. In her sessions, Zhao introduces future biologists and physicians to next-generation sequencing techniques and data-analysis methods, such as RNA-seq analysis, differential gene expression analysis and mutation calling.

Zhao's class sessions require up to 10 software packages, so she reached out for some assistance from OSC's physical Help Desk location on campus, which Zhao subsequently visited numerous times leading up to her class sessions. The experts there were able to help her get all her software installed and operational in short order.

"The staff working at OSC helped me set up a class project and assigned student pass codes," said Zhao. "I am delighted the way the students in the classroom can get approved almost immediately and are allowed to manage their own accounts."

Zhao also leverages the Center's Owens Cluster to help conduct her own sequencing research into therapies to combat prostate cancer, a project led by Xiaoqi Liu, Ph.D., professor and chair of toxicology and cancer biology at the University of Kentucky, and Lang Li, Ph.D., professor and chair of biomedical informatics at Ohio State. Initial therapies that many doctors employ against prostate cancer attempt to lower levels of 'male hormones'—mostly testosterone and androstenedione—called androgens. For patients whose cancer keeps growing despite therapies, doctors employ a hormone therapy that targets a protein—an androgen receptor—that regulates the transfer of testosterone and related hormones into the cell's nucleus.

A second-generation androgen receptor drug, enzalutamide, has shown encouraging clinical effectiveness in patients whose cancer doesn't respond to the initial androgen-lowering treatments. Yet, despite early success, a majority eventually develop a resistance. Zhao is utilizing a computational tool called MAGeCK to identify important genes from the CRISPR-Cas9 knockout screens and to then examine the molecular events behind the resistance development.

"Our study demonstrates a new research strategy to investigate the mechanism underlying the drug resistance at whole-genome level," said Zhao. "Genes and signaling networks identified through CRISPR-Cas9 screening further expand our understanding of the androgen signaling network and provide valuable insight on new therapy development." •

PROJECT LEAD // Yue Zhao, Ph.D., The Ohio State University
RESEARCH TITLE // Genome CRISPR-Cas9 knockout screens reveal new mechanism of enzalutamide resistance in prostate cancer
FUNDING SOURCE // The Ohio State University
WEBSITE // medicine.osu.edu/bmi/people/research/yue_zhao/Pages/index.aspx

FIGURE // Yue Zhao helps teach an introductory bioinformatics course, where she instructs students on next-generation sequencing techniques. This image is used to show students how to examine a sequence using HaplotypeCaller.

Material Properties

Ghazisaeidi investigating behavior of high entropy alloys

By combining state-of-the-art computational methods with theoretical considerations, a research group led by Maryam Ghazisaeidi at The Ohio State University is studying the connection between microscopic physical phenomena and macroscopic mechanical behavior of engineering materials.

"Our research is at the intersection of materials science, physics and mechanics," said Ghazisaeidi, Ph.D., an associate professor in the materials science and engineering department. "We use atomic-scale computations—electronic structures and classical potentials—coupled to larger length-scale continuum and statistical mechanics to improve and predict the properties of existing and new materials."

Recently, Ghazisaeidi's group undertook a project to better understand the remarkable mechanical properties of high entropy alloys (HEAs) by investigating their fundamental deformation mechanisms—micro-scale processes, such as dislocation and stacking fault structures and energies, that allow it to change its internal structure, shape or volume. HEAs are formed at the atomic level of equal or near-equal concentrations of multiple metallic elements, crystalize as solid solutions and don't take a secondary form, or phase, such as a gas or liquid.

The Ghazisaeidi group is leveraging Ohio Supercomputer Center (OSC) systems to model these attributes in concentrated multi-element alloys (CMAs), of which the single-phase HEAs are an interesting subset. CMAs are fundamentally different from those in dilute alloys—those with a much more varied concentration of constituent elements. The group also looks to develop new computational tools that enable dislocation modeling in CMAs with the quantum mechanical modelling method known as density functional theory (DFT). The group employs the DFT code VASP—currently the most widely used plane-wave, first-principles code, to perform the required calculations.

"Electronic structure calculations of dislocations typically require larger simulation sizes and accurate treatments of boundary conditions," Ghazisaeidi noted. "In this project, we take a first-principles computational approach using recently-developed capabilities to address two questions concerning these multicomponent alloys: What is the correct way of thinking about structural defects? What is the solid-solution strengthening mechanism?"

To answer these questions, Ghazisaeidi and her team are logging substantial computation time on OSC's Ruby and Owens Clusters, as well as leveraging the Center's high performance storage resources.

"Existing approaches to modeling material behavior generally rely on phenomenological constitutive relations that must be calibrated experimentally, and therefore lack predictive capability," Ghazisaeidi said. "Connecting the accurate atomic-scale study of deformation to the overall mechanical properties will potentially transform alloy design by replacing trial-and-error approaches with quantitative predictive models." ●

PROJECT LEAD // Maryam Ghazisaeidi, Ph.D., The Ohio State University
RESEARCH TITLE // Understanding novel characteristics of defects in concentrated solid solutions from first principles calculations
FUNDING SOURCE // National Science Foundation
WEBSITE // ghazisaeidigroup.engineering.osu.edu

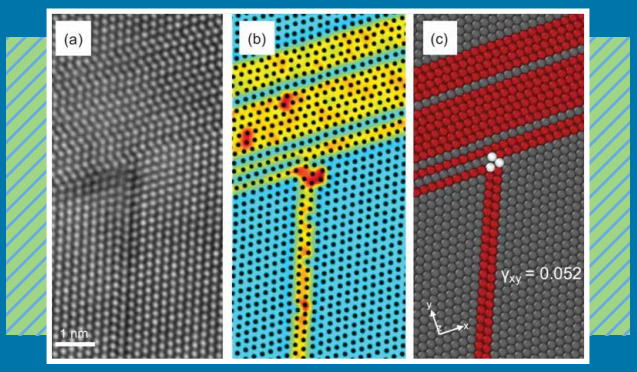
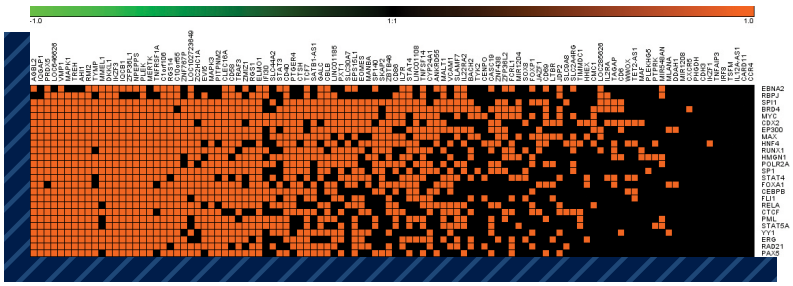


FIGURE // Maryam Ghazisaeidi's group studies the mechanical properties of high entropy alloys. This figure compares experimental results of deformation mechanisms in a Chromium-Cobalt-Nickel sample (left, center) with an atomistic simulation in a model system (right).

Gene Regulation

Weirauch research team investigates disease risk variants

"The origins or causes of diseases determine how they manifest illness and what we can do about them," explained Matthew T. Weirauch, Ph.D., an associate professor in the Center for Autoimmune Genomics and Etiology at Cincinnati Children's Hospital and the pediatrics department at the University of Cincinnati College of Medicine. "Of the known genetic risk variants for diseases, about 90 percent likely alter gene regulation, but almost nothing is known about how they do this."



Weirauch's research team at Cincinnati Children's is studying "how they do this" by analyzing how these genes alter disease risk and interact with the environment. With that knowledge, he believes, scientists might be able to leverage these mechanisms to develop specific therapies and preventive measures for a variety of conditions, such as multiple sclerosis, rheumatoid arthritis, type 1 diabetes, inflammatory bowel disease and lupus.

Using high performance computing and storage systems at the Ohio Supercomputer Center, the team created the Regulatory Element Locus Intersector (RELI) algorithm for estimating the significance of the overlap between disease-associated variants and large collections of transcription factor (TF) genome-wide binding datasets. TFs are proteins that control when and where in the body genes get "turned on" and "turned off." In this regard, they are key players in translating the DNA-encoded genome into RNA, which then creates the proteins that make cells (and life) work.

The team has discovered 2,264 associations of TF binding sites with multiple disease-associated genome regions related to 93 human disease or physical conditions. Its findings potentially identify molecular participants in the regulatory mechanisms that initiate disease, their contributing genetic mutations and the cell types involved in the developmental pathways of disease.

Prior to the advent of high-throughput genotyping and sequencing, it was widely assumed that an understanding of the cause of a disease could be achieved by studying one genetic variant at a time. It is now appreciated that most diseases are generally a manifestation of the additive action of many variants tipping the scales towards increased disease risk.

"We have created an innovative computational method that leverages the wide availability of data describing TF binding interactions with the genome to deliver a solution analyzing multiple genetic locations simultaneously," said Weirauch. "We expect that this project represents an important initial step on a long journey exploring the association between the binding of TFs and the onset of many human diseases." •

PROJECT LEAD // Matthew Weirauch, Ph.D., Cincinnati Children's Hospital Medical Center
RESEARCH TITLE // Gene regulation in disease genetics
FUNDING SOURCES // National Institutes of Health, Alliance for Lupus Research, Cincinnati Children's Hospital Research Fund
WEBSITE // cincinnatichildrens.org/bio/w/matthew-weirauch

FIGURE // This table depicts the overlap between multiple sclerosis-associated variants at each locus and the 25 most significant proteins. Strikingly, Matthew Weirauch's RELI algorithm found that 83 datasets strongly overlap MS variants.

services

Statewide Users Group

The demand for high performance computing in Ohio is relentless, and it does not discriminate by field. At the Ohio Supercomputer Center's (OSC) Statewide Users Group (SUG) conferences, clients in fields spanning everything from astrophysics to linguistics gathered to share research highlights and hear updates about the Center's direction and role in supporting science across Ohio.

SUG is a volunteer group composed of the scientists and engineers who provide OSC leadership with program and policy advice and direction to ensure a productive environment for research.

Statewide Users Group conferences in September 2018 and April 2019 placed OSC clients face-to-face with OSC representatives and experts and brought the research being done on their supercomputer clusters to the forefront of both groups.

The September conference featured a keynote address from Russ Fromkin, Americas HPC and HPDA sales director for Intel Corporation.

Fromkin and his team are responsible for helping technical computing and data analytics customers implement cutting-edge technologies developed by Intel. Fromkin joined Intel Corporation in 1999 as part of an acquisition of Dialogic Corporation.

Attendees at the April conference heard a keynote address from Santosh Rao, senior technical director at NetApp, a hybrid cloud data services and data management company. Rao is responsible for NetApp's Artificial Intelligence (AI) and data engineering products and solutions business. He works closely with the AI solution ecosystem across GPU, computer, software and channel partners, as well as customers worldwide.

Both conferences also featured a variety of topics and poster and flash-talk competitions. Nearly 40 participants competed in each conference, with winners receiving 5,000 resource units of time on OSC systems and runners-up receiving 2,500 resource units. All competitors received 1,000 resource units.

PHOTO // Intel's Russ Fromkin delivered the keynote address for the Fall 2018 SUG meeting. INSET // The Spring 2019 SUG meeting featured a keynote address presented by NetApp's Santosh Rao.

SUG conference winners

Fall 2018 Flash Talks

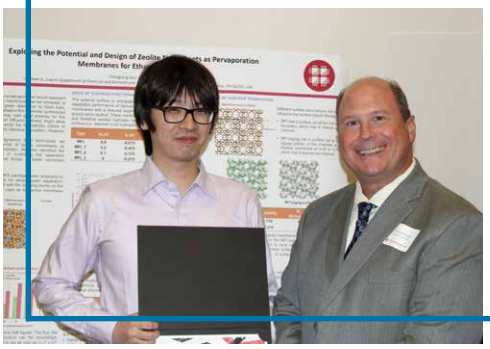
The Ohio State University's Dylan Wood won the flash-talk competition for his presentation titled "Modeling of Failure Mechanisms for Flood Control Systems in a Hydrodynamic Storm Surge Model."

Taking the runner-up position, also from Ohio State, was Megan Smith with "Disentangling the Process of Speciation Using Machine Learning."

Fall 2018 Posters

The poster competition displayed a wide range of disciplines, including first-place winner Changlong Zou, from Ohio State, for his poster titled "Computational Studies of Zeolite Nanosheets as Pervaporation Membranes for Ethanol Extraction."

Second place went to Ohio State's Heather Hansen for her poster titled "Using the Ohio Supercomputer Cluster to Measure Developmental Changes in Connectivity between the Amygdala Subnuclei and Occipit."



Spring 2019 Flash Talks

Ohio State's Arif Hossain won first place in the flash-talk competition for his presentation, titled "Sweeping Jet Film Cooling and Impingement Cooling for Gas Turbine Heat Transfer Application."

Taking the runner-up position, also from Ohio State, was Alexandria Volkening with "Forecasting Elections with Mathematical Models of Contagion Spread."

Spring 2019 Posters

Zhiping Zhong from Ohio State won first place with his poster, titled "Viruses Potentially Enhance Hosts Cold- and Salt-Tolerance."

Second place went to Ohio State's Harper McMinn-Sauder for a poster titled "Measuring Honey Bee Utilization of Conservation Reserve Program (CRP) Pollinator Plantings Using DNA Metabarcoding."

PHOTOS // Dylan Wood won the flash talk competition at the Fall 2018 SUG Conference, and Changlong Zou placed first in the poster competition. At the Spring 2019 meeting, Arif Hossain took top honors in the flash talks, while Zhiping Zhong won among the posters.

Education Services

Graduate Education

Computationally based research programs at several of the state's universities—many of which leverage the Ohio Supercomputer Center's (OSC) computational and storage services—play an integral role in the education of many of Ohio's graduate students. Also, poster and flash-talk competitions held at the Center twice each year (see pg. 31) allow college students and faculty members the opportunity to publicly present their work—practice they need to help earn the science communications skills and highly sought-after degrees they seek.

Summer K–12 Programs

Since 1989, Ohio high school students and middle school girls have received extraordinary STEM education through first-hand application of high performance computing (HPC) and computational modeling during OSC's annual Summer Institute (SI) and the Young Women's Summer Institute (YWSI).

SI is OSC's unique and intense two-week residential summer program for high school students entering their sophomore, junior or senior year. YWSI, initiated in 1999, is a weeklong program for Ohio's middle school girls to gain valuable computer, math, science and engineering skills through hands-on, collaborative projects.

Classroom Accounts

OSC staff assist faculty and student researchers by making classroom accounts available for professors who want to incorporate HPC resources into courses. In FY2019, 19 departments at 17 universities taught 64 courses to more than 2,200 students by leveraging OSU services. Two researchers in this report—Richard Hughes (pg. 14) and Yue Zhao (pg. 26)—note the value that access to classroom accounts bring to their instructional approach.

Staff Expertise

OSC staff members often collaborate with faculty members or staff colleagues on research or education projects, sometimes travelling across Ohio or beyond.

For example, Katharine Cahill, Ph.D., an education and training specialist at OSC, will be mentoring a colleague from Wofford College in South Carolina as they work to boost a federal program aimed at expanding a capable and innovative advanced digital-resource workforce across the country. Under the same program, Cahill served as a mentor last year to a colleague at the University of Tulsa. Cahill is the education lead for the National Science Foundation's (NSF) Extreme Science and Engineering Discovery Environment (XSEDE) program at OSC, where she aids in computational science curriculum development.

Cahill also is coordinating the Computational and Data Science Curriculum Exchange in collaboration with colleagues across the country. The three-year, OSC-led NSF grant program, funded in Summer 2018, seeks to grow the country's research workforce with individuals trained in computational and data science, particularly through curriculum taught at institutions faced with resource limitations and focused on minority inclusion.

Training Services

Workshops

Staff members at the Ohio Supercomputer Center (OSC) regularly reach out across the state to clients and other potential users, offering workshops, one-on-one classes, web-based training, helpful materials, consulting services and updates. Many of the workshops hosted at OSC offices focus on in-house systems and services on topics such as Getting Started at OSC, Containers for Research Computing and Optimization & Performance Tuning.

Other training opportunities feature workshops delivered remotely to OSC by peer facilities or national organizations, such as a two-day workshop on Message Passing Interface (MPI) presented by the National Science Foundation's Extreme Science and Engineering Discovery Environment (XSEDE) program and the Blue Waters Petascale Institute's webcast from the National Center for Supercomputing Applications at the University of Illinois.

Office Hours

For answers to specific questions about research done on OSC systems or simply for advice on getting started, clients can talk directly to an OSC expert. See the story about how Yue Zhao took advantage of office hours (pg. 26).

The Center holds office hours via phone, Monday through Friday, 9 a.m.–5 p.m. For clients on or near The Ohio State University's campus, in-person office hours are currently held Wednesday and Friday afternoon, 3–5 p.m. at Pomerene Hall and every other Tuesday afternoon, 1–4 p.m. at the Research Commons. No appointment is needed.

Consulting

As professionals in high performance computing (HPC) and software engineering, OSC's staff has deep expertise in developing and deploying software that runs efficiently and correctly on large-scale cluster computing platforms. They are available to advise on advanced approaches to debug or optimize code.

Technical Support

OSC also provides clients with training support in the form of an online FAQ, a web-based knowledge base and a remote help desk. Staffed by members of OSC's HPC Client Services group, OSC Help features Level-1 basic support and Level-2 advanced support Monday through Friday, 9 a.m.–5 p.m. These services can help clients resolve issues, such as debugging, software installation and workflow improvement, provide onboarding for new users or deliver system status updates. OSC Help is available by email at oschelp@osc.edu or by phone at 800-686-6472 (toll-free) or 614-292-1800 (local).

PHOTO // Zhi-Qiang You, Ph.D., an OSC scientific applications engineer, leads a workshop on Containers for Research Computing.

Software Services

The Ohio Supercomputer Center (OSC) staff has experience with several computing languages, programming models, numerical libraries and development tools for parallel/threaded computing and data analysis. They are available for consulting or collaboration to help tackle client coding problems or reduce the runtime for analyses.

The Scientific Applications group at OSC helps to optimize and scale codes for academic research groups studying various themes, such as:

- Next-generation high performance computing communication mechanisms, led by Dhabaleswar Panda, Ph.D., professor and distinguished scholar of computer science at The Ohio State University
- Automated high-resolution terrain generation, led by Ian Howat, Ph.D., director of the Byrd Polar & Climate Research Center, both at Ohio State
- Strategic, high-fidelity computational aerodynamics capabilities for the United States Air Force Research Laboratory, led by Datta Gaitonde, Ph.D., Ohio Research Scholar and John Glenn Chair of the mechanical and aerospace department at Ohio State

To enable an ever-widening set of user communities to productively carry out its science, the Center maintains a broad assortment of applications—more than 170 different software packages with about 20 being licensed packages.

OSC staff members recently developed a continuous integration-based framework to automatically install and test this large collection of applications.

Top Applications

by CPU Hours
- LAMMPS
- VASP
- OpenFOAM
- WRF
- Gromacs
- Python/ Jupyter
- R/Rstudio
- Gaussian
- QChem
- StarCCM
- NAMD

by Number of Users
- Python/Jupyter
- Matlab
- R/Rstudio
- Gaussian
- LAMMPS
- VASP
- Bowtie

Top Application Areas

by CPU Hours
- **Molecular Dynamics** (LAMMPS, Gromacs, NAMD)
- **Quantum Chemistry** (VASP, Gaussian, QChem)
- **Computational Fluid Dynamics** (OpenFOAM, StarCCM, WRF, Converge, user codes)
- **Data Analytics** (Python/Jupyter, R/Rstudio)

by Number of Users
- **Data Analytics** (Python/Jupyter, R/Rstudio, Matlab)
- **Quantum Chemistry** (VASP, Gaussian)
- **Molecular Dynamics** (LAMMPS)
- **Bioinformatics** (Bowtie)

Gateways Development

The expert gateways development team at the Ohio Supercomputer Center (OSC) helps clients create custom web interfaces to simplify the use of powerful high performance computing (HPC) resources, allowing data to be the focus. Simplifying workflows allows researchers to dive into finding solutions and making breakthroughs.

OnDemand

OnDemand is an open-source, web-based portal that allows HPC centers to offer their users seamless, flexible access to HPC computing and storage services. Users can upload and download files, as well as create, edit, submit and monitor jobs. OSC runs a customized version of the software, known as OSC OnDemand.

"Up until this point, clients had to install software, use a cryptic file editor and learn batch system commands," said David Hudak, Ph.D., executive director of OSC and the primary investigator for the grant to build the portal. "Our objective with the OnDemand project has been to build a portal that will allow them to use HPC without a big learning curve. It's about lowering the barrier of entry to the world of supercomputing."

In 2015, OSC was awarded a National Science Foundation grant (NSF #1534949) to develop an open-source version of the application that could be shared with other HPC facilities. The 1.0 version of Open OnDemand was launched in 2017.

OPEN OnDemand

A 2019 follow-up project award, Open OnDemand 2.0 (NSF #1835725), will enhance resource utilization visibility and extend to more resource types and institutions. Other developments will include enhancing the web portal, integrating the XD Metrics OnDemand tool (XDMoD) developed at the University of Buffalo Center for Computational Research, extending the portal to provide methods of access for other science domains and improving the scaling of the system.

The Open OnDemand portal already has been successfully downloaded by more than 200 HPC facilities around the world, at least 135+ of them located in the United States.

MyOSC

The OSC client portal, MyOSC, was developed by in-house and contract engineers and deployed in Fall 2018 as a full replacement of OSC's former accounting and account management infrastructure (including app.osc.edu and an earlier version of MyOSC). It maintains the essential functions of the earlier system but offers OSC clients many more powerful features.

With MyOSC, clients can now self-sign up for accounts, manage user access to research projects, report funding and publications relevant to their use of OSC services, annotate jobs and run custom reports on usage.

"The project dashboard provides much more visibility for the clients, giving them access to reports that show a client what usage looks like across OSC's various clusters, the types of resources on those clusters and usage over time," explains Brian Guilfoos, OSC's HPC client services manager.

Several new features were added to MyOSC to assist OSC staff members. The new portal links to OSC's existing identity management infrastructure, provides flexible and robust user-defined custom reports, is extendable for future process changes and requirements, and provides industry standard data security.

Hardware Services

Cluster Computing

The Ohio Supercomputer Center (OSC) offers a fully scalable center with mid-range machines to match those found at National Science Foundation centers and other national labs. Collectively, OSC supercomputers provide a peak computing performance of more than three petaflops.

In November, OSC engineers and Dell EMC specialists deployed the Center's newest, most efficient supercomputer system, the liquid-cooled, 10,560-core Dell/Intel Xeon Pitzer Cluster. The system is named for Russell M. Pitzer, a co-founder of the center and emeritus professor of chemistry at The Ohio State University. The Pitzer Cluster powers a wide range of research, from understanding the human genome to mapping the global spread of viruses.

Preparations were being made over the past year to install a new system in 2020 to replace the Ruby Cluster, which is nearly five years old. Responding to the computational needs and wishes of clients, the new system is intended to increase overall computational capacity at the center and lower queue wait times.

Research Data Storage

OSC has partnered with IBM in 2019 to expand the center's high performance computing (HPC) storage capacity by 8.6 petabytes for a total file-storage capacity of about 14 petabytes over several file systems. Engineers are working to develop a remote data backup site for the home and project directories at a location geographically distant from the Center to ensure data security and availability.

OSC also offers HPC clients more than 10 petabytes of backup tape storage, plus the potential for another 10 petabytes through the purchase and installation of additional tapes.

OSC engineers also are working to further develop a protected data environment that currently meets International Traffic in Arms Regulations (ITAR) and Export Administration Regulations (EAR). The environment is expected to meet Health Insurance Portability and Accountability Act of 1996 (HIPAA) and other regulatory regimes in the near future.

production capacity

236 M+
core hours consumed

82.6%
average HPC system utilization

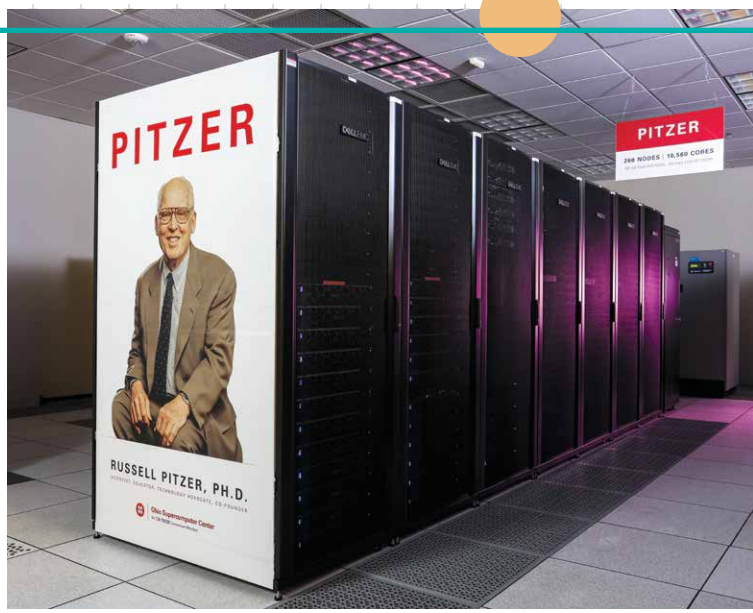
4.8 M+
computational jobs

99.1%
availability*

70.7%
average storage system utilization

3.7 PB
data stored

*includes planned outages



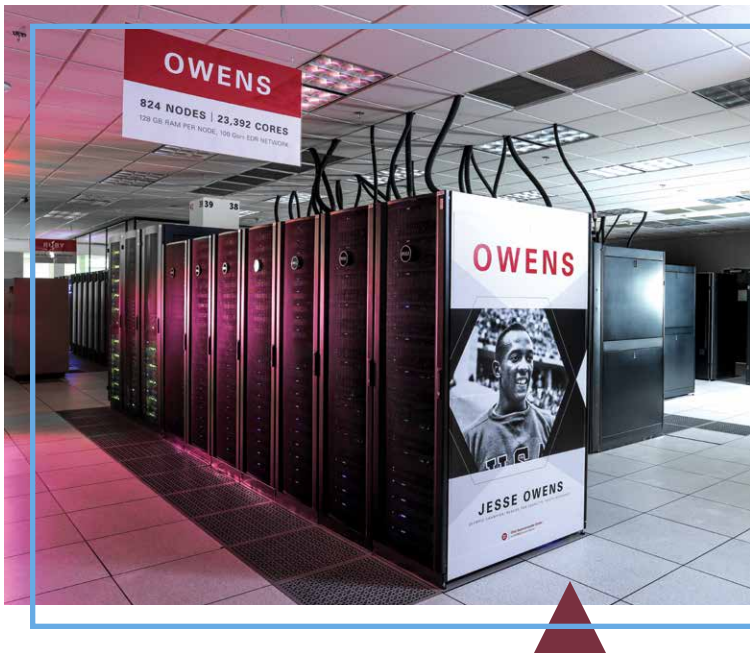
Supercomputers

Students, faculty members, scientists, engineers and clinicians depend upon on these key OSC systems:

Pitzer Cluster

A 10,560-core Dell/Intel Xeon machine (~1,300 TF):

- 40 cores and 192 GB of memory per node
- 4 nodes have 3.0 TB of memory and 80 cores
- 32 nodes have 2 NVIDIA Tesla V100 GPUs and 384 GB of memory



Owens Cluster

A 23,392-core Dell/Intel Xeon machine (~1,600 TF):

- 28 cores and 128 GB of memory per node
- 16 nodes have 1.5 TB of memory and 48 cores
- 160 nodes have NVIDIA Tesla P100 GPUs

Ruby Cluster

A 4,800-core HP/Intel Xeon machine (~144 TF):

- 20 cores and 64 GB of memory per node
- One node has 1 TB of memory and 32 cores
- 20 nodes have NVIDIA Tesla K40 GPUs

2.2 PB
data transferred

65%
jobs started within 30 minutes

PHOTOS // OSC system photos by Jodi Miller

twitter.com/osc

facebook.com/ohiosupercomputercenter

oh-tech.org/blog

linkedin.com/company/ohio-supercomputer-center

osc.edu

Ohio Supercomputer Center

An OH·TECH Consortium Member

OH-TECH is the technology and information division of the Ohio Department of Higher Education.

1224 Kinnear Road | Columbus, OH 43212 | ph: (614) 292-9248