PURDUE UNIVERSITY | Rosen Center for Advanced Computing

High performance computing at the highest proven value

Annual Report 2021-2022

PURDUE UNIVERSITY | Rosen Center for Advanced Computing

Director's Welcome

Preston Smith, Executive Director, Rosen Center for Advanced Computing

Every day, our Rosen Center team works closely with top faculty from 60 departments, every academic college, and all 3 Purdue campuses, who turn to us to provide facilities and expertise to solve computational problems in support of their research and teaching. 2022 marked the 60th year of HPC at Purdue, and we're proud to carry on the legacy of our predecessors in the Computer Sciences Center founded alongside the nation's first CS department and the Purdue University Computing Center. Supporting the application of supercomputers since the CDC 6500 to Anvil today has changed greatly – it's no longer a world where a select few researchers utilize high-performance computing. Instead, applications ranging from statistical analysis, engineering simulations, climate and weather models, bioinformatics, and computer vision all require high-performance computing.

Beyond physics-based simulation, AI has now grown to be a top driver of computing needs, with applications in fields as diverse as communication, urban planning, agronomy, and more.

It is clear, then, that high-performance computing facilities are a core competency of an institution like Purdue, and without them, the ability of Purdue faculty to move the state of knowledge forward and impact society would be significantly limited. HPC facilities are needed to compete for new research, to conduct the science, and the training of students; and various measures of the impact of RCAC on Purdue's outputs are highlighted throughout this report. For example, 25% of AY 21-22's doctoral recipients were users of RCAC resources, demonstrating that HPC facilities are as crucial of a tool to educate future scientists as mass spectrometers, pipettes, and classrooms!

In February of 2022, the NSF-funded Anvil system entered production, beginning 5 years of operations where we will support researchers from across the nation via the NSF's ACCESS program. In just 8 months, Anvil has already supported 2,774 total users from 248 different institutions among 42 different fields of science. Our goal is for Anvil to serve as a complement to our community clusters, allowing Purdue PIs to run at greater scales than on campus, and for our research partners to leverage Anvil as part of their proposals' cyberinfrastructure to make us more competitive for future research. We also plan to use Anvil to impact the state and Purdue's industry partners, and provide it as another reason that industry partners choose to work with Purdue. In addition to working with our faculty and access to students, Anvil can be a competitive advantage for their modeling, simulation, and big data needs. More information about Anvil and its impact can be found on pages 4-6.

In 2022, HPC platforms continue to evolve, as do our service offerings. Architectures are increasingly heterogenous, and access modes are evolving. Community clusters, as we did on Anvil, have GPU subsystems, large memory nodes, composable offerings, and specialized filesystems. An NSF CC* award funded the "Geddes" composable platform, which provides a Kubernetes resource to run cloud-like workflows that previously could not fit on batch HPC. Geddes will be supplemented by a newly awarded grant for software defined storage in 2023. Interactive computing, using tools like Thinlinc Remote Desktops and Open OnDemand allow for easy access to desktops on the community clusters, and access to web-based tools in the browser.

AI continues to be an area of growth. Our Gilbreth Community Cluster saw incremental growth in 2021-22, and through late 2022 and 2023 we will see a new, more compelling allocation model and pricing structure and significant investment in capacity. Regulated research is critical for Purdue strategic areas in national defense and health, and to support these areas, our "Weber" community cluster supports export controlled research, and the 2022 community cluster "Negishi" will support data subject to NIH DBGaP controls.

Much of RCAC's staff and expertise, however, do not tie directly to high-performance computing — Research Software Engineering (RSE) is a growing line of business where software engineers trained in the research process support the development of scientific applications, science gateways, AR/VR solutions, and more. Led by Chief Scientist Carol Song, these RSEs are becoming crucial providers to apply cyberinfrastructure within major research projects. Some of the RSE initiatives and research projects that RCAC staff have partnered on are described in greater detail on pages 8,9, 14, 20 and 21.

Finally, we have been fortunate to grow our team in 2022, with 10 new hires completed so far, including a new Director of Scientific Applications, (Dr. Arman Pazouki), and new roles to support outreach, user training and education, and industry business development. We continue to recruit HPC engineers, research software engineers, and computational scientists in many specialties, including quantum and aerospace.

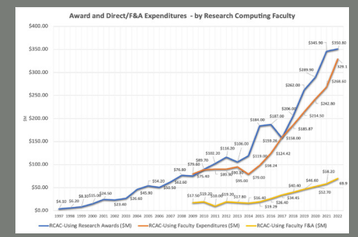
As always, thank you for your continued support and partnership with our center; we are honored to have the opportunity to be your cyberinfrastructure provider of choice.

Preston Smith
Executive Director,
Rosen Center for Advanced Computing

> **"It is clear, then, that high-performance computing facilities are a core competency of an institution like Purdue**

High Performance Computing is required by Purdue's most productive researchers

RCAC - faculty partners accounted for 61% of new grant awards in FY22, for a total of $350.8M



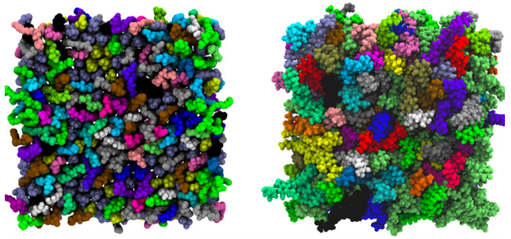
+25%

of AY 21-22's doctoral recipients were users of RCAC resources.

2022 marked the 60th year of HPC at Purdue

RESEARCH HIGHLIGHTS

Anvil accelerates sustainability and photosynthesis research



Top-view images of warm and cold adapted plant plasma membrane species. The membrane was prepared using CHARMM-GUI and the image is prepared using molecular visualization software VMD. Image courtesy of Daipayan Sarkar.

Purdue's Anvil supercomputer is helping a Michigan State research team study processes associated with photosynthesis and plant metabolism, and apply those insights to facilitate more efficient energy conversion to meet today's sustainability challenges.

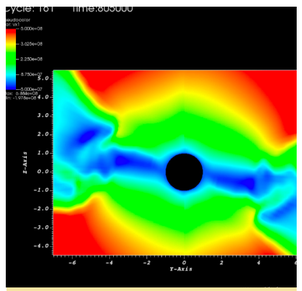
Josh Vermaas, an assistant professor at Michigan State University's DOE Plant Research Laboratory and his postdocs Daipayan Sarkar and Martin Kulke used Anvil to perform molecular dynamics simulations to study the movement of small molecules, such as how carbon dioxide crosses a plant plasma membrane during photosynthesis.

"Because of Anvil's state-of-the-art GPUs, we've gained four times better performance compared to our previous benchmarks," says Sarkar.

"Our new molecular simulation engines can make GPUs talk to each other and how fast a simulation can go is limited by how much they can communicate," adds Vermaas. "The key fact of Anvil for us is that it has four A100 GPUs. They're faster than the GPUs we have locally and because they're all linked together, we can use all of them at once."

Professor developing numerical tools for astrophysical modeling

with the help of Anvil supercomputer



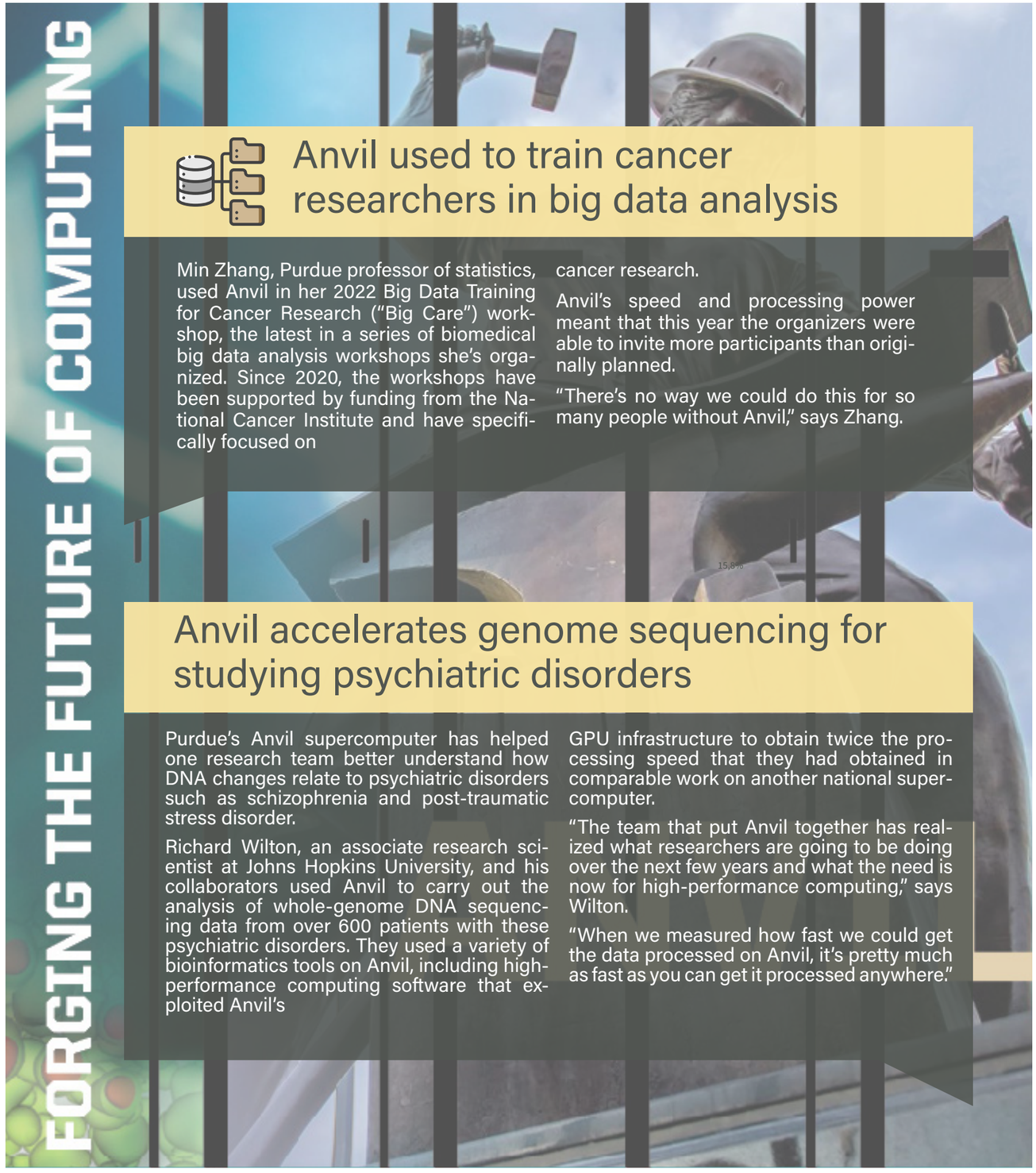
3D MHD simulation of a massive star wind. Image courtesy of Asif ud-Doula

Asif ud-Doula, an associate professor of physics at Penn State Scranton, develops numerical tools to model the stellar winds of massive stars. He uses three-dimensional magnetohydrodynamics (MHD), a very computationally intensive modeling method.

Generating a single 3D MHD model requires more than 60,000 core hours, estimates ud-Doula. That's the equivalent of running a personal computer for several years, but he achieved it using hundreds of Anvil cores in less than two weeks.

Ud-Doula and his collaborators have taken advantage of RCAC expert staff support for Anvil, as well as the Slack channel for Anvil early users.

"When you're trying to do science, you don't want to spend a lot of time resolving technical issues," says ud-Doula. "Having someone to help us with that was a tremendous benefit."



Anvil used to train cancer researchers in big data analysis

Min Zhang, Purdue professor of statistics, used Anvil in her 2022 Big Data Training for Cancer Research ("Big Care") workshop, the latest in a series of biomedical big data analysis workshops she's organized. Since 2020, the workshops have been supported by funding from the National Cancer Institute and have specifically focused on cancer research.

Anvil's speed and processing power meant that this year the organizers were able to invite more participants than originally planned.

"There's no way we could do this for so many people without Anvil," says Zhang.

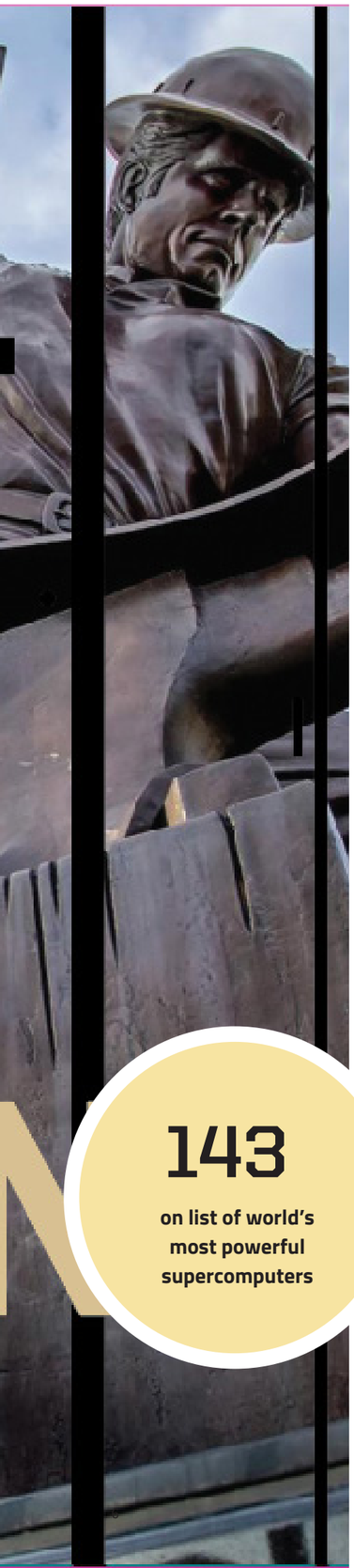
Anvil accelerates genome sequencing for studying psychiatric disorders

Purdue's Anvil supercomputer has helped one research team better understand how DNA changes relate to psychiatric disorders such as schizophrenia and post-traumatic stress disorder.

Richard Wilton, an associate research scientist at Johns Hopkins University, and his collaborators used Anvil to carry out the analysis of whole-genome DNA sequencing data from over 600 patients with these psychiatric disorders. They used a variety of bioinformatics tools on Anvil, including high-performance computing software that exploited Anvil's GPU infrastructure to obtain twice the processing speed that they had obtained in comparable work on another national supercomputer.

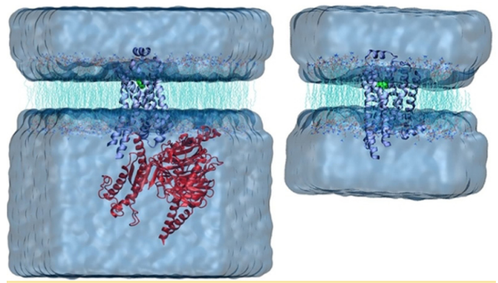
"The team that put Anvil together has realized what researchers are going to be doing over the next few years and what the need is now for high-performance computing," says Wilton.

"When we measured how fast we could get the data processed on Anvil, it's pretty much as fast as you can get it processed anywhere."



Purdue's Anvil Supercomputer

aids in drug discovery



GPU-accelerated molecular dynamics simulations of G protein activation by three different ligands binding to b1-adrenergic receptor (b1-AR), a key G-protein-coupled receptor (GPCR) targeted for treating heart disease. Image courtesy of Yinglong Miao.

Purdue's powerful new Anvil supercomputer is helping one researcher accelerate his biomolecular simulations and unlock new treatments for heart disease.

Yinglong Miao, an assistant professor at the Center for Computational Biology and Department of Molecular Biosciences at the University of Kansas, develops novel theoretical and computational methods and applies these methods to advanced biomolecular modeling and drug discovery.

As an early user of Anvil, Miao and his colleagues have harnessed Anvil's impressive GPU power to accelerate molecular dynamics simulations. In particular, Jinan Wang, an associate researcher in Miao's lab, and graduate

student Hung Do study G protein activation by the beta1-adrenergic receptor, a key membrane protein that has served as a drug target for treating heart failure.

"Using CPUs this work could take months, if not years," says Miao. "With the GPUs on Anvil, we can run these simulations much faster. Instead of needing months, we just need a couple of weeks."

Accelerated molecular dynamics simulations by Miao and his colleagues were not only performed faster on Anvil than on other supercomputers they've used, they also matched well with experimental results.

Anvil ranked 143rd

on list of world's most powerful supercomputers

143

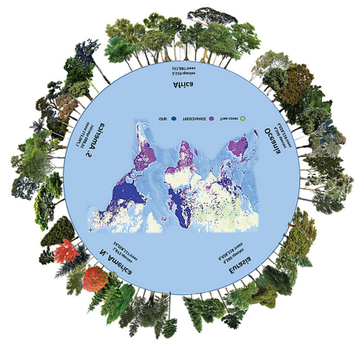
on list of world's most powerful supercomputers

With a peak processing speed of 5.1 petaFLOPs, Purdue's Anvil supercomputer debuted at number 143 on the Top500 list of the world's most powerful supercomputers.

Anvil is the tenth Purdue supercomputer to make the Top 500 list, and the seventh highest ranked academic system in the United States. Only 14 institutions in the U.S. currently have a supercomputer on the Top 500 list.

"Anvil not only provides users with the performance they need, as evidenced by its place among the world's most powerful supercomputers, it was also built with a focus on usability and easy access for those in traditionally under-served domains," says Carol Song, senior research scientist for the Rosen Center for Advanced Computing and project director for Anvil. "It's the greatest feeling when users tell us Anvil exceeded their expectations, for example, by reducing months of machine learning model training time to hours."

Clara Bell Sessions



"It's great to have these world-class HPC systems at Purdue," says Liang. "Without these systems, we would not have been able to finish this study."

Purdue professor estimates the number of tree species

with the help of RCAC supercomputers

Jingjing Liang, an assistant professor of quantitative forest ecology in the department of forestry and natural resources, recently led a team that produced the first ever estimate of the total number of tree species in the world – and they did it with the help of RCAC's Bell community cluster supercomputer.

In order to build their model, Liang and his collaborators needed to collect and compile a huge database of data from all over the world. Because this was such a large dataset – more than one million sample plots from around the world – it required the computational power of a supercomputer to analyze.

Using the high memory nodes of Bell (and the Snyder cluster before it), Liang's team arrived at their estimate of 73,000 tree species worldwide – 9,000 more species than have been documented.

The resulting paper was published in the Proceedings of the National Academy of Sciences (PNAS).

"It's great to have these world-class HPC systems at Purdue," says Liang. "Without these systems, we would not have been able to finish this study."

Liang is also an avid user of Scholar, Purdue's community cluster for classroom instruction. His graduate class "Big Data, AI and Forests" learns to use the statistical analysis software R on Scholar. He has also recently introduced Scholar into his undergraduate classes.

Liang appreciates that the Scholar cluster lets the students log into the remote desktop from anywhere and run the same type of software in a consistent matter.

"It makes my teaching methods and the classroom experience for the students a lot better," he says.

Overview of Bell

Bell is a Community Cluster optimized for communities running traditional, tightly-coupled science and engineering applications. Bell is being built through a partnership with Dell and AMD over the summer of 2020. Bell consists of Dell compute nodes with two 64-core AMD Epyc 7662 "Rome" processors (128 cores per node) and 256 GB of memory. All nodes have 100 Gbps HDR Infiniband interconnect and a 6-year warranty.

Envision Center developing 3D weather visualization technology

to assist small aircraft pilots

New technology developed by the RCAC Envision Center aims to make aviation safer by helping pilots avoid flying into adverse weather.

In collaboration with FlightProfiler, the Envision Center has created a geospatial web-based visualization that allows pilots to see meteorological conditions along their planned route and plot an optimal course.

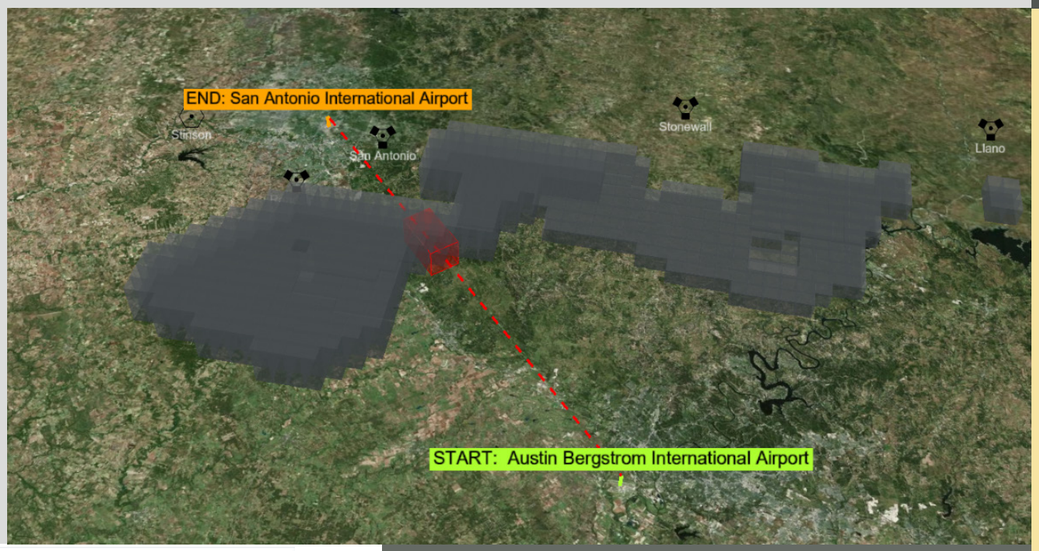
The technology was designed for use in general aviation (small aircraft) where pilots typically lack weather information and technology in the cockpit making them susceptible to adverse weather conditions.

"[Previously], it was sort of on the pilot to create a mental model of 'where will I be versus where will the weather be,'" explains Gary Pokodner, program manager for the Federal Aviation Administration's Weather Technology in the Cockpit program, which has been working with the Envision Center and FlightProfiler on refining the technology. "The idea is that this technology will assist pilots in performing these tasks."

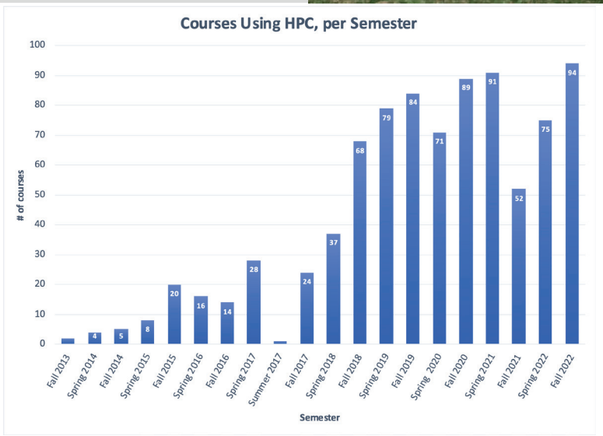
Pokodner demonstrated the visualization software at the 2022 EAA AirVenture airshow in Oshkosh, Wis., with the goal of getting feedback from the pilots in attendance about how they might use this technology and what additional features could be useful.

In the future, the technology may be used in uninhabited aircraft systems that are remotely piloted as well.

"[Previously], it was sort of on the pilot to create a mental model of 'where will I be versus where will the weather be. The idea is that this technology will assist pilots in performing these tasks."



During Fall 2021 – Spring 2022 semesters, 128 courses used HPC for instructional purposes, impacting a total of over 4400+ students. (180 courses in AY20-21)

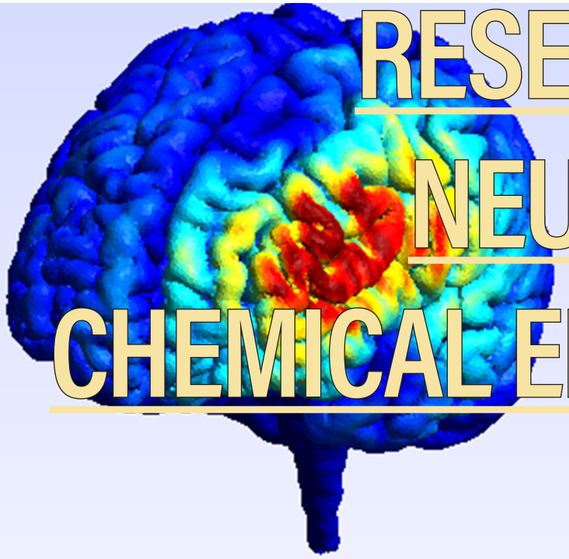


In collaboration with FlightProfiler, the Envision Center has created a geospatial web-based visualization that allows pilots to see meteorological conditions along their planned route and plot an optimal course.

Envision Center colaborates with FEMA for cybersecurity project

The Envision Center continues to build on its collaboration with faculty to create virtual labs, expanding to work with five new faculty members in the College of Science. Recent projects have included nuclear engineering, electrical and computer engineering, biology and two chemistry classes.

In collaboration with FEMA, the Envision Center is working on the Cyber Resilience Adaptive Virtual Reality Experiences (CRAVRE) project. This project, supported by $1.5 million in funding from a Department of Homeland Security grant, funded the creation of a training curriculum regarding cybersecurity resiliency during natural disasters. Purdue and Texas A&M partnered in the creation of this training project and the Envision Center has been working with Mesut Akdere, professor of human resource development, to build the training using webXR. This allowed the creation of an entire learning system that is VR-capable but can be accessed completely from anywhere via a browser both in VR and non-VR.



RESEARCHERS IN NEUROSCIENCE, CHEMICAL ENGINEERING

using new Bell cluster GPUs

Electric field distribution in Transcranial Magnetic Stimulation. Image courtesy of Nahian Hassan.

As part of a push to expand GPU computing capabilities to meet the needs of researchers working in AI and machine learning, RCAC has added eight MI50 AMD GPUs and six MI60 AMD GPUs to the Bell community cluster – and they are already making a big impact on campus.

Tianfan Jin, a graduate student working with Brett Savoie, Charles Davidson Assistant Professor of Chemical Engineering, uses the new Bell GPU nodes with machine learning tools to predict and automatically interpret chemical reactions. Since Jin's work involves using very large datasets, the computation time required could be prohibitively long, taking days to accomplish what the Bell cluster can do in hours.

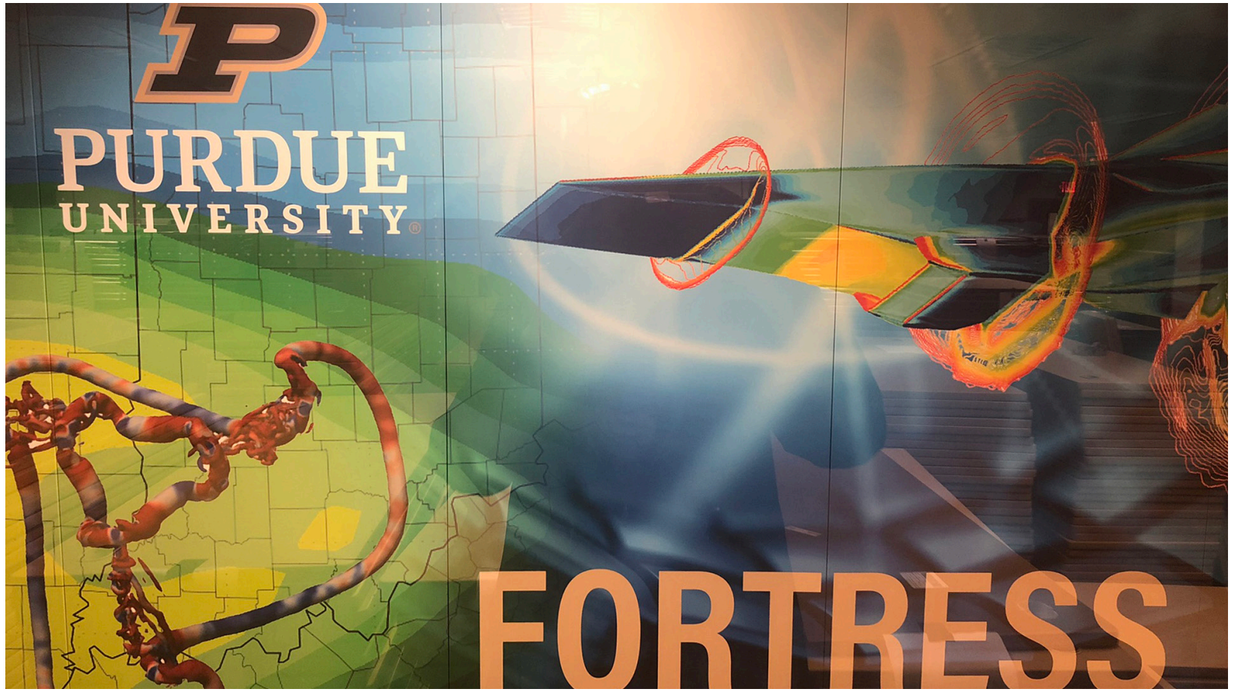
"With the help of the Bell GPUs, we can speed up this process and control the computational time to a reasonable range," explains Jin. "Bell GPU nodes are a great resource.

Nahian Hasan, a graduate student working with Luis Gomez, assistant professor of computer and electrical engineering, assembles three-dimensional head models from MRI data, which can then be used to diagnose and treat neurological diseases with non-invasive electromagnetic stimulation techniques.

Constructing these models is computationally intensive and requires a great deal of processing power and memory. Using multiple Bell GPU nodes, Hasan has assembled over 6,000 models, something that would be impossible to achieve on a personal computer where building each model takes at least 24 hours.

Scan here to learn more about Bell

www.rcac.purdue.edu/compute/bell



Fortress storage archive expands capacity

The Fortress archival storage system has a new tape library that has expanded the archive's capacity by a factor of five, from approximately 50 petabytes of data to more than 250 petabytes.

The new SpectraLogic TFinity tape library can hold 9,000 tapes, of a new generation that can hold about four times as much data as the previous generation. They are also twice as fast, meaning data can be read off the tapes at 600 megabytes per second instead of the previous 300.

The new tape library has two arms, which means it can handle more requests at the same time, significantly reducing the mean time between when a user requests data and when it's available.

"This upgrade is phase one of a multi-phase improvement to the archive system that will include more disk capacity, faster networking and newer, faster servers to handle the ever-growing demands of researchers looking for a safe place to hold their data long term," says Ramon Williamson, senior engineer for RCAC. "Tape continues to be the most cost effective way to deliver large-scale archival storage, and these improvements will help make the archiving experience faster, easier and more reliable for end-users.

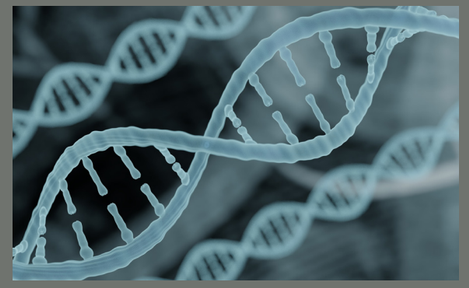
Researchers in fields such as climate modeling, agriculture, physics, chemical modeling and life sciences have long depended on Fortress for a place to store their growing datasets from increasingly more complicated models.

This has led to an exponential growth in the amount of data being ingested into the archive. From first reaching the one petabyte mark in 2013, the archive has grown to its current 60 petabytes of stored data in nine years.

Scan here to learn more about Fortress

www.rcac.purdue.edu/storage/fortress

RCAC deploys software container tools for bioinformatics research



RCAC now offers biocontainers – software containers designed for Purdue's community clusters that contain more than 400 commonly used bioinformatics tools.

Biocontainers have the advantage of keeping things organized by isolating programs inside their containers, allowing for portability and reproducibility, and making software installation and use easier for researchers.

"The packages that perform these analyses are often incredibly buggy and sometimes conflict with other Python packages that may be loaded," says Nadia Atallah Lanman, a research assistant professor of comparative pathobiology, who uses the biocontainers extensively to perform single-cell RNA velocity analyses.

"The biocontainers provide a really stellar way to run these analyses without having to deal with package conflicts."

Purdue computer scientist searching for 'buried treasure'

with cloud extension to community clusters

Sameul Wagstaff, a professor of computer science, who searches for factors of very large numbers – work that is essential to keeping everything from nuclear weapons to personal banking information secure - is using new hybrid cloud capabilities between Purdue community HPC clusters and the Microsoft Azure HPC cloud.

Wagstaff likens the search for factors of large numbers to a search for buried treasure. The longer a computer can run a job, the "deeper" a computer scientist can dig for treasure. Using Azure, Wagstaff is able to dig three to four times deeper than he could using standby queues on the community clusters.

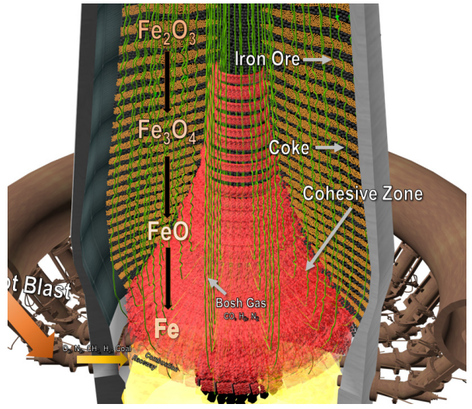
The Azure HPC environment has been set up by RCAC staff to extend the Purdue community HPC cluster environment, meaning users who are familiar with the community clusters won't have to make major adjustments to existing workflows.

"Researchers can move from an on-premises system to a cloud system without having to make any modifications to their code," says Erik Gough, lead computational scientist for Research Computing, who has led the Azure project for RCAC.

"Researchers can move from an on-premises system to a cloud system without having to make any modifications to their code"

Purdue Northwest Team Using

Bell cluster on $7 million DOE project



A Purdue Northwest team designing a multi-component tool for blast furnace operation in the steelmaking industry is using Bell to refine their models.

Tyamo Okosun, a research associate professor at Purdue Northwest who leads the team, explains that while high-performance computing can significantly accelerate the computational fluid dynamics models used to predict blast forces, even HPC isn't fast enough to allow for real-time fine-tuning by the operators at the plant.

To solve that problem and allow plant operators to quickly see how a change is going to impact the process, Okosun and the project team are developing a machine learning model. They are training the model on physics-based simulation data by first running a large database of different operating conditions on the Bell cluster.

The project, which is funded by a $7 million award from the U.S. Department of Energy, aims to reduce energy consumption in blast furnaces and downstream processes by up to 10 percent.

"RCAC's computational resources make it possible for us to work more effectively," says Okosun. "Our projects really depend on RCAC resources and in particular on the Bell cluster that we use quite a bit."

Purdue joins new quantum

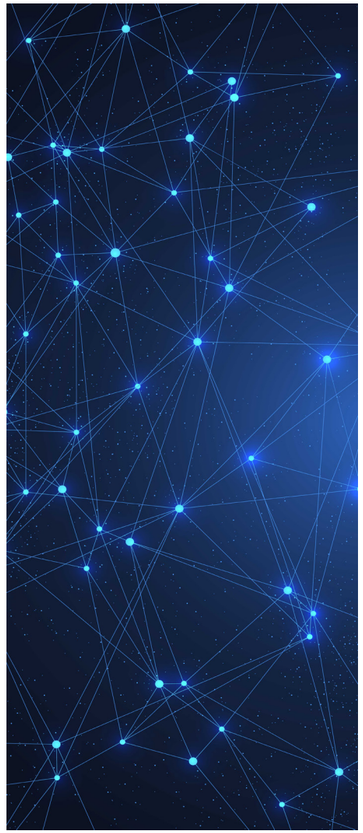
research collaborative

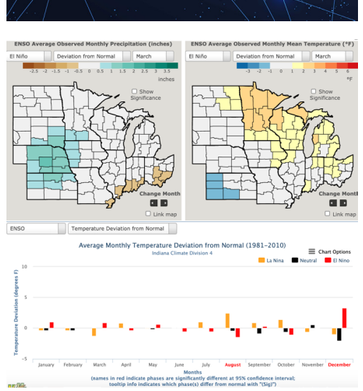
Led by RCAC, Purdue has joined as a founding partner in the new Quantum Collaborative launched by Arizona State University. The Quantum Collaborative is a major 21st century initiative poised to profoundly impact society and the American economy with new discoveries and applications in advanced quantum technology.

The Quantum Collaborative is a Hub within the IBM Quantum Network, a global community of Fortune 500 companies, academic institutions, startups and national labs with prioritized access to IBM's quantum computers, experts and resources. As a Hub in IBM's network, partners within the Quantum Collaborative can draw on IBM's quantum technology and resources to work to deliver incremental advances in the emerging technology landscape.

In addition to fundamental research and technology development, another key aim of the Quantum Collaborative is workforce and education program development, a goal that Purdue is well-positioned to achieve.

Preston Smith, Executive Director for RCAC, notes: "Joining the ASU Quantum Collaborative will provide access to new resources for Purdue's quantum initiatives that we previously did not have available to us. Our membership will allow Purdue faculty to access IBM's quantum computers, including machines ranging from 7 up to 27 qubits." RCAC plans to hire a research scientist in quantum computing to help support and drive adoption of quantum methods.





RCAC offers Research Software Engineering services

RCAC offers a wide range of research software engineering (RSE) solutions to help advance scientific discovery, enable greater research reproducibility, and generate lasting impact at Purdue and beyond.

RCAC's Scientific Solutions group, Envision Center and web developers have comprehensive experience in software design, programming languages, and deployment practices that give us the tools to help you create or modify software to take advantage of the latest technologies in advanced computation, web frameworks, data analysis, visualization, and management.

When you partner with RCAC RSE, you get a co-producer who is closely involved in defining the product, even helping to make the pitch to funding agencies, understanding at a certain depth the underlying methods, developing iteratively with frequent feedback, and providing technical support and training, as well as planning for the deployment and future maintenance before the active development phase.

"Our RSEs have never stopped learning as computing technologies and computational methods evolve so rapidly. Our team at RCAC has a proven track record of more than ten years in enabling and accelerating research and discovery, and deepening education and training through partnerships with faculty and students across the campus. We are increasing our capacity to support larger initiatives, more researchers, and a wider range of fields of study." - Carol Song

Examples of recent RCAC Research Software Engineering projects:

*GABBS and GeoEDF - Innovative geospatial data analysis building blocks including the latest GeoEDF geospatial data framework to help researchers wrangle complex and distributed datasets and make valuable data assets more usable and workflows more reproducible

*Cyber Training for FAIR Science - cyberinfrastructure to support the cybertraining activities, including science gateway MyGeoHub, scalable hands-on training HPC and cloud, hosting courses, supporting workshops and fellows

*GLASSNET (Global-Local-Global Analysis of Systems Sustainability) - cyberinfrastructure to support data harmonization, large scale modeling, training, teaching, etc.

*CHEESE (Cyber Human Ecosystem of Engaged Security Education) - scalable, web-based, hands-on education platform for network security, secure coding, cryptography

*AnalytiXIN – streaming data cyberinfrastructure for this large Indiana state project, collaborating with the department of computer science, Purdue Polytechnic and Indiana University and the University of Notre Dame

*Superpower – a generalized framework and flexible software environment for power analysis (NIH)

*PIOTHub – a collaborative platform for research on developing circular supply chain for more sustainable manufacturing processes

*Useful to Usable (U2U) – decision support tools translating climate data into actionable information for corn bell growers and farm advisors to respond to climate change

EDUCATION AND OUTREACH

Envision Center offers tours to student groups

The Envision Center gives tours of their virtual reality and data visualization facilities to student groups. In 2022, the center hosted Purdue's GenCyber Camp, a camp that introduces high school students to cybersecurity concepts at no cost to the student. The camp is focused on diversity and is open to 9th-12th grade female students. RCAC staff were excited to share the center's work with GenCyber Camp students and inspire the next generation of women in STEM.

Women in HPC

Purdue's Women in High-Performance Computing group, which is led by RCAC staff members, hosted Ying Zhang to discuss her experiences as a woman in HPC. Zhang, a doctoral candidate and the AI Support Team Lead in research computing at the University of Florida, is the founder of the UF Chapter of Women in HPC and organized campus wide activities to advance the participation of women in HPC and AI.

Sarah Rodenbeck, a Women in HPC member and senior research data scientist for RCAC, led a data processing workshop in April. The workshop was a beginner-level discussion about data, focusing on how data fits into machine learning and data science workflows. Topics included types of data, data collection, data processing with Pandas and considerations for exploratory data analysis.

Purdue WHPC provides travel support for students interested in attending conferences.

To learn more about Purdue Women in HPC or be added to the mailing list, contact whpcinfo@lists.purdue.edu.

Sarah Rodenbeck, a Women in HPC member and senior research data scientist for RCAC

AMD senior vice president headlines inaugural RCAC cyberinfrastructure symposium

RCAC held its inaugural cyberinfrastructure symposium "Enabling Giant Leaps in Advanced Computing" on October 4. AMD senior vice president Ivo Bolsens delivered the keynote address.

Bolsens spoke about AMD's vision for the future of computing, and AMD's collaboration with RCAC on Anvil, the $10 million NSF-funded system that is one of the most powerful campus supercomputers in the US and in the top 150 supercomputers worldwide.

After his keynote address, Bolsens toured RCAC facilities with RCAC and ITaP leaders and saw Anvil in person.

Andrew Pranger and Derik Lovejoy won the student poster award for their poster "Solving Fiducial Free Biological Structures Using High-Powered Computational Programs." Their mentor is Lauren Ann Metskas, assistant professor of biological sciences and chemistry.

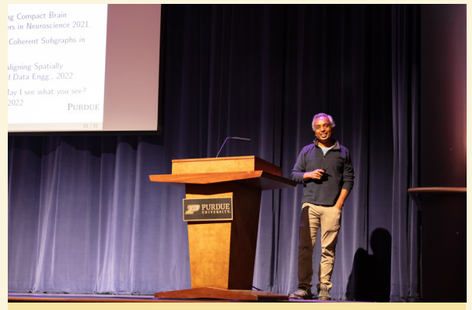
A number of Purdue faculty members also spoke at the symposium about the research they're doing using RCAC resources.

For a full list of the speakers and a list of their speeches scan here

www.rcac.purdue.edu/news/5169

PURDUE FACULTY SPEAK AT INAUGURAL

Cyberinfrastructure Symposium



Ananth Grama,
Samuel D. Conte Professor of Computer Science

The audience got to explore the power of functional neuroimages or fMRIs and their applications in "Computational Methods for Analyses of Functional Brain Connectomes." They learned about their effect across a wide range of neuroscience, psychiatry, and even in the design of human-computer interfaces like augmented reality (AR) and virtual reality (VR). The RCAC is excited to be able to support vital research like Ananth Grama's, Samuel D. Conte Professor of Computer Science.

Andrew Flachs, Ike Obi, and Dr. Lindsay Weinberg discussed the social and ethical implications of Big Data and AI in a range of contexts, including global agriculture, digital platforms, and higher education. Together, their panel helped audience members envision more just approaches to the design and development of digital tools.

The Tech Justice Lab: Justice in Big Data and AI
Andrew Flachs, Associate Professor of Anthropology, Purdue University Ikechukwu Obi, Ph.D. student, Polytechnic Institute, Purdue University
Lindsay Weinberg, Clinical Assistant Professor, John Martinson Honors College, Purdue University

STAFF NEWS

Carol Song now Chief Scientist

of Research Computing

Carol Song, Chief Scientisit

Carol leads RCAC's research collaboration efforts with the campus and nationwide. She has more than 20 years of experience in high-performance and distributed computing, and software engineering to connect domain scientists to advanced cyberinfrastructure technologies.

Since 2005, she has been PI or CoPI on more than 30 sponsored research projects, representing more than $80 million in research funding. Carol is the PI of Anvil, a new large-capacity national computational system funded by NSF in 2020, continuing her leadership in HPC including the TeraGrid, XSEDE 1 & 2 projects since 2006.

She has led and collaborated in many data infrastructure projects, including NSF-funded Data Interoperability, CI-TEAM, SDCI, SI2, DIBBs, Cybertraining projects, and most recently, a $5M NSF CSSI grant to develop a reusable, plug-n-play data framework, GeoEDF, to make large geospatial scientific datasets readily usable by domain scientists.

As director of the Scientific Solutions Group in RCAC, she leads the research software engineering (RSE) group to develop innovative advanced computing and data cyberinfrastructure solutions to support scientific discovery and learning, including the MyGeoHub.org geospatial science gateway that supports data-driven collaborative research and learning since 2014.

Carol has funded and mentored more than 60 graduate, undergraduate, and high school students in research projects. She is a mentor, advisor, and advocate in the Women-In-HPC, PEARC, and Supercomputing programs to help promote and guide the professional growth of early career professionals and broader participation of the underrepresented groups.

She earned her bachelor's degree in computer science and engineering from Tsinghua University and her doctorate in computer science from the University of Illinois at Urbana-Champaign.

RCAC senior scientist awarded $25,000

Better Scientific Software fellowship

Amiya Maji, Senior Computational Scientist

Amiya Maji, a senior computational scientist for RCAC, has been named a 2022 Fellow by the Better Scientific Software (BSSw) Fellowship Program, which aims to foster and promote practices, processes, and tools to improve developer productivity and software sustainability of scientific codes.

Each 2022 Fellow will receive up to $25,000 for an activity that promotes better scientific software, such as organizing a workshop, preparing a tutorial, or creating content to engage the scientific software community.

Maji, who joined RCAC in 2015 after earning his doctorate in computer engineering from Purdue, will use his funding to document and present best practices across various HPC centers for managing Python, and develop software that will simplify Python package installation and environment management on Purdue's community clusters. His work will prove especially useful for faculty teaching courses who need hundreds of students to install Python for classroom exercises, and broadly to improve scientific productivity for a growing community of Python users in HPC centers.

"I'm very honored to be selected as one of the BSSw fellows this year," says Maji. "The fellowship provides a unique opportunity to network with experts in high-performance computing and learn from their experiences."

Betsy Hillery, Director High Performance Computing pictured second from the right

RCAC wins best paper, poster awards

at research computing conference

Several RCAC staff members received awards for their posters and papers at the 2022 Practice and Experience in Advanced Research Computing (PEARC) conference.

The paper "Understanding Factors that Influence Research Computing and Data Careers," co-authored by Betsy Hillery, director of high-performance computing, won best full paper in the Workforce Development track, as well as the Phil Andrews Award for the conference manuscript deemed to be the most impactful to the practice of research computing.

The poster "HyperShell v2: Distributed Task Execution for HPC" by lead research data scientist Geoffrey Lentner and senior computational scientist Lev Gorenstein won the conference's best poster award.

At the PEARC conference, RCAC staff also lead a tutorial about interactive computing on Anvil's composable platform.

Geoffrey Lentner, Lead Research Data Scientist

Lev Gorenstein, Senior Computational Scientist

> "Faculty using RCAC resources have produced 10,180 peer-reviewed publications since 2015.

EXTERNAL FUNDING

RCAC team develops new tools

for $15 million NSF Institute

Rajesh Kalyanam, Research Scientist and Carol Song, Chief Scientist

Chief scientist Carol Song is the co-PI on the $15 million Institute for Geospatial Understanding through an Integrative Discovery Environment (I-GUIDE), awarded last year as part of the NSF's Harnessing the Data Revolution (HDR) Big Idea, which is a national effort to enable new modes of data-driven discovery that will address fundamental questions at the frontiers of science and engineering.

As a large partner in the I-GUIDE Institute, the Purdue team consists of eight researchers and their students, including three research scientists from RCAC. The institute just finished the first year with a number of "early wins" demonstrated during its first in-person all-hands meeting in Chicago in September.

The RCAC team led the push to integrate GeoEDF, CyberGIS Compute, Globus, Singularity, JupyterHub and Anvil to form a seamless, scalable geospatial CI, resulting in a beta release of the platform. Two complex workflows were demonstrated as notebooks for (1) vulnerability analysis of potential aging dam failures and (2) scalable modeling of hydroclimatic extremes using the National Water Model (NWM) developed by NCAR. Both studies would be very difficult to carry out without this platform due to the variety and "messiness" of the data and the lack of an integrated environment of tools and large scale computational resources.

Song introduced the I-GUIDE platform during the October NSF HDR2 PI meeting via a poster and discussed the team's approach and best practice in developing the platform through integration and interoperation, leveraging the strengths of past investments at its data and cyberinfrastructure panel.

NSF CSSI Data Framework: Extensible Geospatial Data framework for FAIR (Findable, Accessible, Interoperable, Reusable) Science, $400K supplemental funding

RCAC awarded $500,000 grant

for composable and HPC workflow storage

Erik Gough, Lead Computational Scientist

Lead computational scientist Erik Gough is the PI on a $500,000 NSF award that establishes a model for Software Defined Storage as a Service (STaaS) at the campus level, offering unified access to and on-demand provisioning of block, object and filesystem storage across the Geddes Composable Platform and Community Cluster resources.

The storage system will be used for both research and instruction via residential learning programs on Purdue's campus. To support workforce development, undergraduate students will play a key role in the deployment and integration of the new storage infrastructure.

"This new storage offering will impact the breadth of our user base, providing new data access methods through an S3 endpoint as well as user defined provisioning of storage for both cloud and HPC workflows," says Gough.

Chief scientist Carol Song co-PI

on $850,000 NSF grant for cyber training curriculum

RCAC is a partner in an $850,000 grant awarded to Venkatesh Merwade, Professor of Civil Engineering at Purdue, to create a cybertraining curriculum for climate, water and environment (CWE) sustainability.

Chief scientist Carol Song and research scientist Lan Zhao are Co-PI and senior personnel respectively on the award, which will address a gap between training opportunities on integrating and working with data from different CWE sources by creating a cyber training curriculum that will incorporate rapidly developing technology and community best practices.

It will provide cyber training to students, trainers and working professionals using publicly available data, simulation tools, and computing resources.

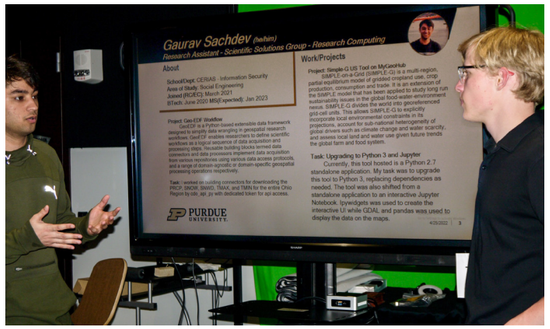
The new curriculum will offer training at foundation, expert, and developer levels using a flexible modular approach. Students can access it through an open, scalable online platform which will disseminate not only the curriculum itself but also access to tools and resources. Students from five universities, three community colleges and one high school will participate.

STUDENT PROGRAM

> “THE PROJECTS THAT THEY HAD US WORKING ON WERE AS REAL AS IT GETS, THIS IS STUFF WE DON’T LEARN IN CLASS

RCAC summer interns gain

real world high-performance computing experience



Three undergraduate students spent their summer break getting hands-on experience with high-performance computing, thanks to a Research Experiences for Undergraduates program run by the RCAC.

The students worked on a variety of projects for RCAC including benchmarking systems such as Purdue’s new $10 million NSF-funded Anvil supercomputer, monitoring cluster usage and testing to see how well scientific applications were running on the clusters.

“The projects that they had us working on were as real as it gets,” says Lucas Snyder, a junior studying intelligent systems engineering at Indiana University. “This is stuff we don’t learn in class.”

Larkin Nickle, a senior in computer information and technology at Purdue, worked on setting up a power draw monitoring system RCAC needed to analyze which systems were drawing the most power and when.

Nickle, who is continuing to work with RCAC on a part-time basis this year, says he highly recommends the internship program to fellow students.

Erik Gough, lead computational scientist for RCAC, served as a mentor to the students, although they had the opportunity to work with a number of RCAC staff members on various projects.

“I appreciated how much trust Erik put in us,” says Snyder. “He gave us a lot of freedom to get the projects done.”

Nrushad Joshi, a junior in intelligent systems engineering at IU, also participated in the program.

Students gain experience in HPC, VR/AR and Data Visualization, while learning from Research Software Engineers, and System Engineers

Practice Experience

Our purpose is to provide work opportunities and real workplace experience that enhance the student’s education through the development of professional skills, responsibilities, habits, attitudes, self-confidence and self-development.

Our PRACTICE Program has accomplished many things this last year. During the academic year, we had four new system-facing students. During this time, these students were introduced to the data center to learn beyond the racking of the systems. Three of our system-facing students will attend the Supercomputing 22 conference and compete in the Student Cluster Challenge with students from Indiana University. We recruited eighteen research-facing students embedded in the organization. The research-facing students built tools to support the research community and worked with researchers on data visualization with the latest VR and other technologies. In this program, students not only learn and improve their technical skills, they also learn valuable lessons in relationship management skills through their research-facing projects.

This year we incorporated assessment into the PRACTICE program. Surveys were conducted to evaluate the overall program and the mentors assigned to the student. With mentoring at the heart of the program, we are working to produce a mentor handbook and will share it broadly with the research computing community. Within our mentoring program, we meet bi-monthly to assist in coaching the mentors to ensure they and the students are successful.

Poster Session at the RCAC Cyberinfrastructure Symposium

After a morning filled with great talks, we took a break to visit our poster session, “Solving Science with Computation.” Each entry in the poster session showcased how departments at Purdue are instituting Research Computing, and our team of judges assessed each presentation to determine a winner. Congratulations to Andrew Pranger and Derik Lovejoy from the Department of Biological Sciences for their presentation, “Solving Fiducial-Free Biological Structures Using High-Powered Computational Programs.”

PURDUE UNIVERSITY®

Rosen Center for Advanced Computing

CONTACTS

Preston Smith
Executive Director,
Rosen Center for Advanced Computing
psmith@purdue.edu
765-494-9729

Carol Song
Chief Scientist
carolxsong@purdue.edu
765-494-7467

George Takahashi
Envision Center Technical Director
gtakahas@purdue.edu
765-494-7888

www.rcac.purdue.edu