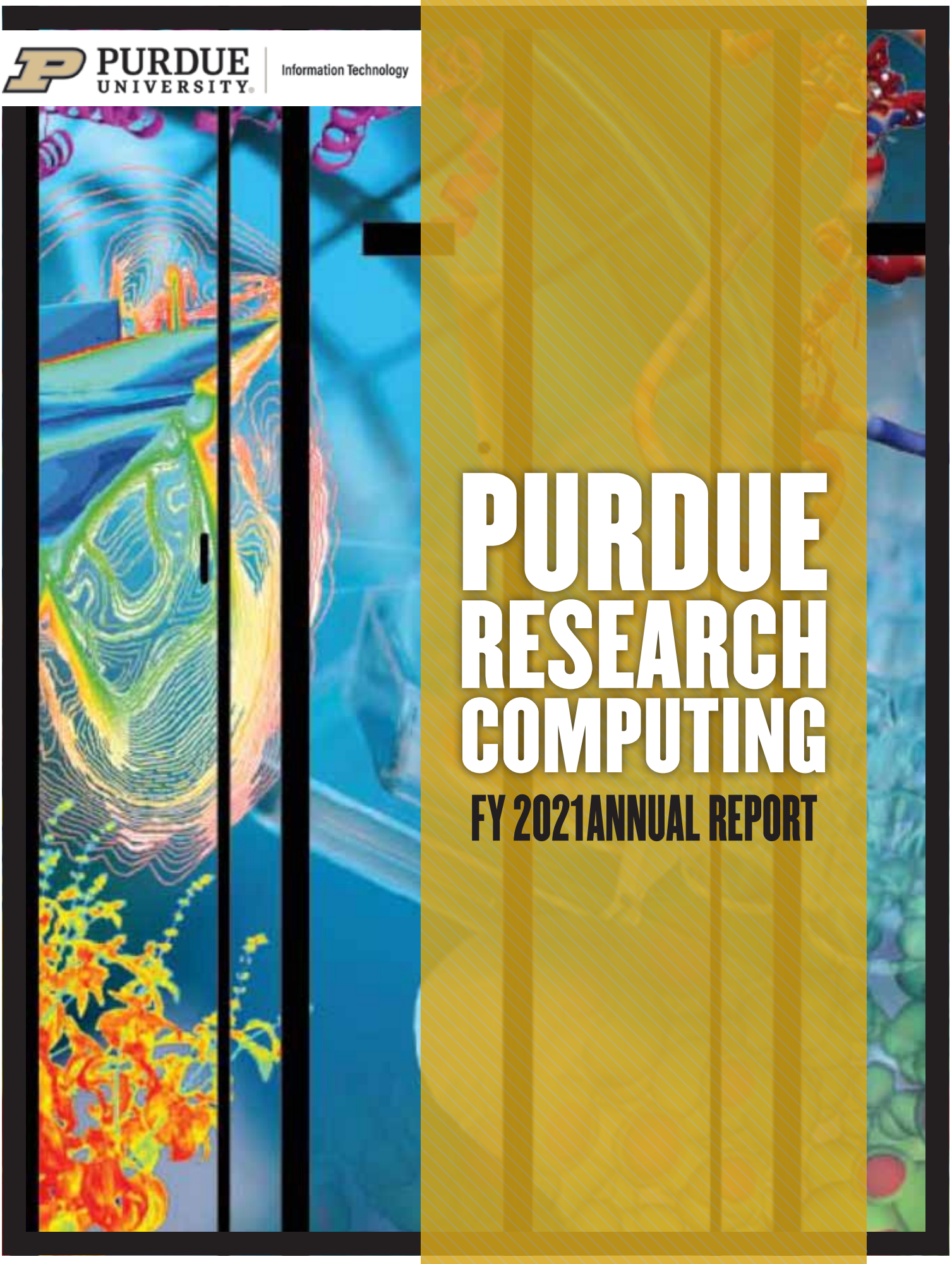
PURDUE UNIVERSITY | Information Technology



PURDUE RESEARCH COMPUTING

FY 2021ANNUAL REPORT

DIRECTOR'S WELCOME

Preston Smith,
Executive Director,
Research Computing

Dear Research Computing Partners,

I write this here in late 2021, 18 months after the COVID-19 pandemic transformed life at Purdue and around the nation. It's been since 2019 that a report of the state of research computing at Purdue has been shared with the campus community, and I'm pleased to report that during the last two years the pandemic hasn't slowed things down here around the campus cyberinfrastucture.

First, I would like to welcome to Purdue our new CIO, Ian Hyatt. Ian joins us from IGT, where he was vice president of global services, logistics and fleet since 2017 and has held several other vice president roles within IGT, including technology roles and data center operations. Prior to IGT, Ian served as the Chief Technology Officer for Lifespan, Rhode Island's first health system, a comprehensive, integrated, academic health system with The Warren Alpert Medical School of Brown University, where he was charged with support of technology and infrastructure for the 5-hospital system including both administrative and medical devices and platforms. He is also currently serving as the State Command Chief Master Sergeant in the United States Air Force and Rhode Island Air National Guard, where he has served since 1985.

While over the last two years, ITaP's overall focus has been around cost-effectively "keeping the lights on," Ian views ITaP as an enabling force for Purdue's strategic goals. With a newly articulated mission and vision (see right), Ian has laid out three priorities for guiding ITaP's strategic investments – in no particular order:

- ***Security Posture***: Protect Purdue at higher level
- ***Student Experience***: Easy-to use services, tools, and support
- ***Research Support:*** World-class cyberinfrastructure and facilities to support research computing

Previous research has demonstrated the value of our campus' investment in cyberinfrastructure, with CI investment demonstrating a significant impact on institutional outputs of research expenditures, publications, and earned doctorates.

Mission:

ITaP provides the technology infrastructure, services, solutions, and information security that support teaching and learning, enhance research, and enable faculty and staff to achieve their objectives while providing a positive student experience.

Vision

To empower our students, faculty and staff with the technology to make giant leaps to advance knowledge, impact

This ongoing commitment will ensure that Purdue's cyberinfrastructure remains in the top echelon compared to our peers. Operating a world-class campus cyberinfratructure requires facilities to support it, and Purdue leadership is now focused on how to ensure that Purdue's datacenter facilities are prepared for the computing systems of the future.

In these pages you will see highlights both of the accomplishments of our own center, and of the Purdue scientists that require advanced computing for discovery. 2020-2021 saw numerous accomplishments within RCAC, including:

DIRECTOR'S WELCOME

- The deployment of our 2020 community cluster "Bell," which debuted at number 429 on the November 2020 Top 500 list.
- Continued investment in GPU resources to support the growth in AI research
- Long-awaited upgrades to data platforms:
    - The Data Depot has been refreshed, now with a capacity of over 5PB
    - Acquisition of a new tape library to support Fortress and deal with long-term growth
- A $400k award from the NSF campus cyberinfrastructure program to fund "Geddes," a composable campus cloud for research
- Multiple sponsored research awards in partnership with researchers both at Purdue and nationwide:
    - AccelNet: GLASSNET: Networking Global to Local Analyses to Inform Sustainable Investments in Land and Water Resources
    - HDR Institute: Geospatial Understanding through an Integrative Discovery Environment
    - CCRI: ENS: Collaborative Research: Open Computer System Usage Repository and Analytics Engine
- The 2020 award of the $22 million NSF Anvil system, and the deployment currently underway
- New partnerships with campus centers including the Data Mine, Regenstrief Center for Healthcare Engineering, and Agriculture Data Services

As always, thank you for your continued partnership. I truly believe that our model of supporting campus cyberinfrastructure as a partnership between the faculty and the institution is both a competitive advantage to Purdue, and the ideal way to deliver a crucial capability in the most cost effective and sustainable way. To borrow a phrase – "High performance computing at the highest proven value."

Preston Smith
Executive Director, Research Computing

Anvil: Forging the Future of Computing

Purdue University will soon be the home of Anvil, a powerful new supercomputer that will provide advanced computing capabilities to support a wide range of computational and data-intensive research spanning from traditional high-performance computing to modern artificial intelligence applications.

Anvil, which is funded by a $22 million award from the National Science Foundation (NSF), will significantly increase the capacity available to the NSF's Extreme Science and Engineering Discovery Environment (XSEDE), which serves tens of thousands of researchers across the U.S., and in which Purdue has been a partner for the past nine years.

The name "Anvil" reflects the Purdue Boilermakers' strength and workmanlike focus on producing results, and the Anvil supercomputer will enable important discoveries across many different areas of science and engineering. Anvil also will serve as an experiential learning laboratory for students to gain real-world experience using computing for their science, and for student interns to work with the Anvil team for construction and operation.

At Purdue, early Anvil users are working on projects in key areas of research such as hypersonics, deep learning theory, archeological reconstruction and bone biomechanics.

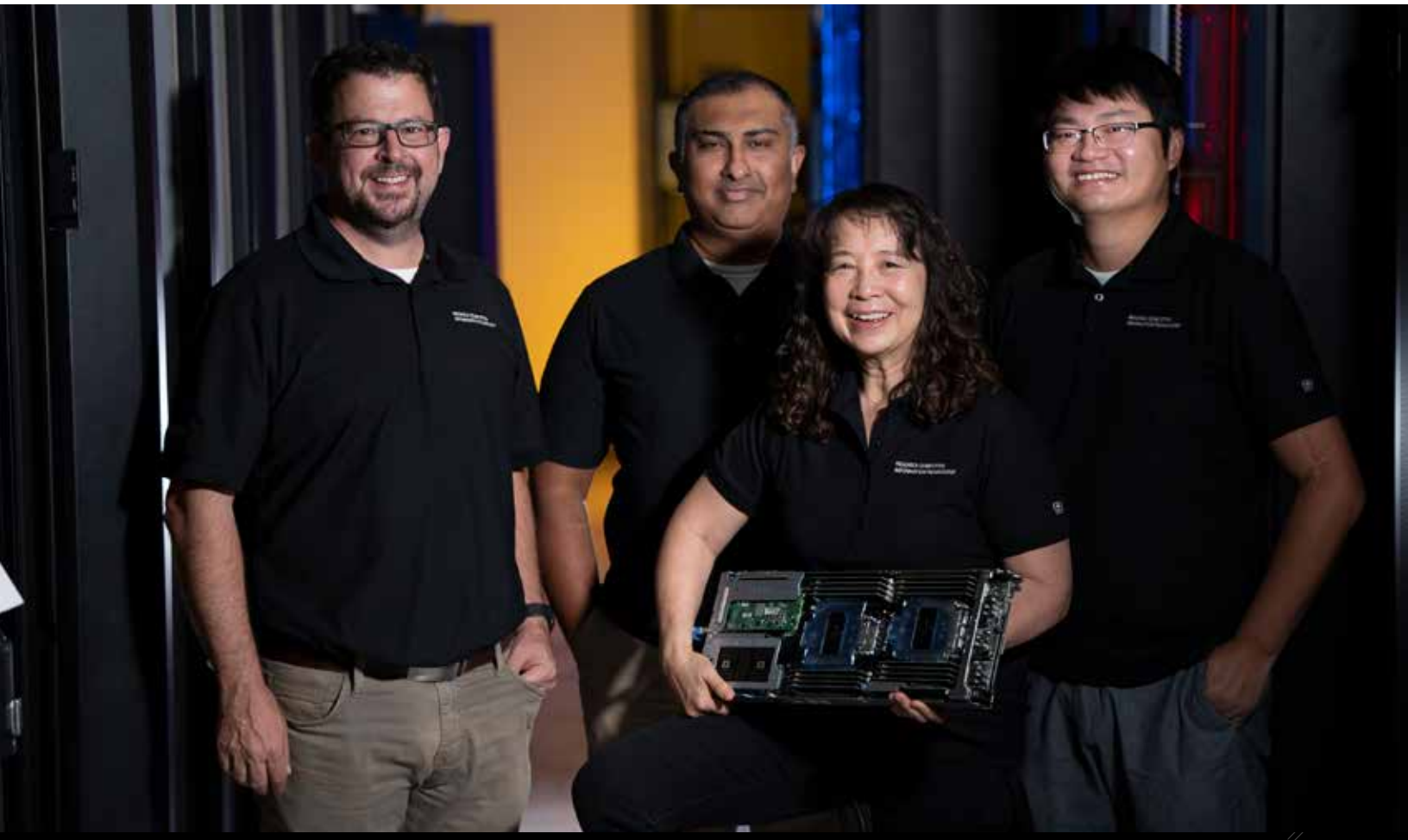
RESEARCH HIGHLIGHTS

Anvil: Forging the Future of Computing

Anvil is also available to Purdue industry partners, adding access to Purdue's world-class cyberinfrastructure to the list of available research facilities. This is intended to enable Purdue partners to use Anvil for modeling, simulation and data analysis at large scales and faster turnaround that wouldn't happen without resources like Anvil.

By building Anvil alongside its community cluster supercomputers, including 2020's Bell system, Purdue has leveraged its existing campus computing infrastructure, such as massive storage systems, high-speed networking, and its team of expert research computing staff that has already deployed 15 large supercomputers since 2008. Purdue's community clusters serve thousands of researchers and students each year.

Senior research scientist Carol Song is the principal investigator and project director. Preston Smith, executive director of Research Computing, Xiao Zhu, computational scientist and senior research scientist, and Rajesh Kalyanam, data scientist, software engineer, and research scientist, are all co-PIs on the project. The project is funded under NSF award # 2005632.



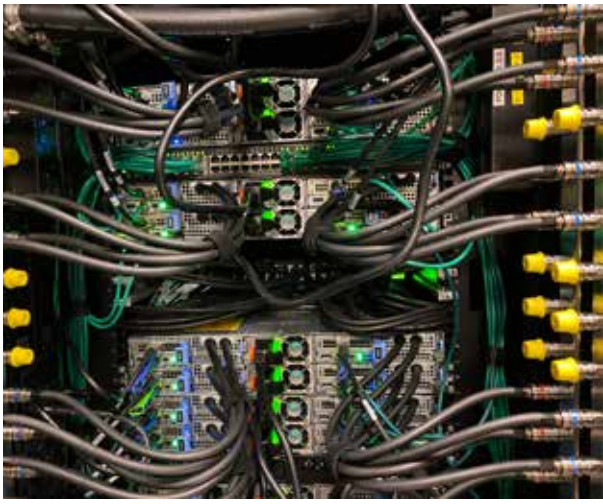
RESEARCH HIGHLIGHTS

Bell cluster makes the list of the world's most powerful supercomputers

Purdue's 2020 community cluster supercomputer debuted at number 429 on the Top500 list of the world's most powerful supercomputers. With a top processing speed of 1.624 petaFLOPs, Bell is nearly twice as fast Purdue's Conte cluster, which was number 28 on the list and the nation's fastest campus supercomputer at the time it was built just seven years before Bell. Bell is also number 44 on the Green500 ranking of the world's most energy-efficient supercomputers.

16 new AMD MI50 GPUs were added this year to support machine learning and AI applications.

Clara Bell Sessions



"Bell is the best system I've worked on. It's perfect for my lab."

- Jennifer Wisecaver, assistant professor of biochemistry

Bell's namesake is Clara Bell Sessions, a nursing professor and the director of continuing education in the School of Nursing who helped establish Purdue's Minority Student Nurses Association, now known as the Diversity in Nursing Association, and the Minority Faculty Fellows programs.

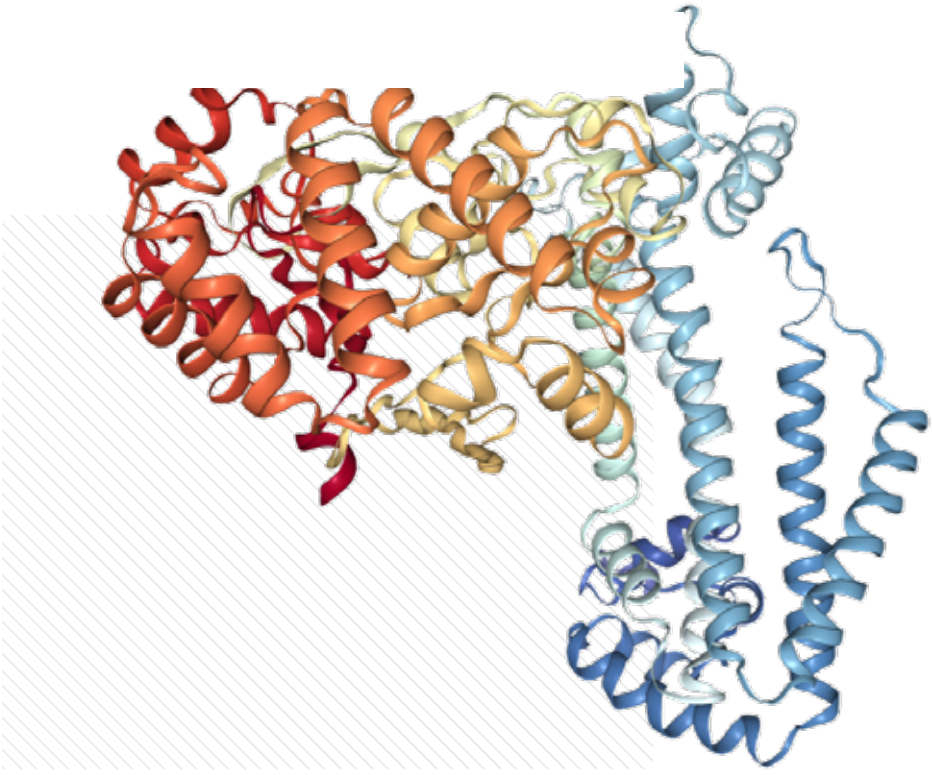
Bell delivered 51% of 2021's computational hours Purdue-wide to 163 users

RESEARCH HIGHLIGHTS

Professor models novel coronavirus proteins with Research Computing supercomputers

Daisuke Kihara, professor of biological sciences and computer science, used Purdue's community cluster supercomputers to develop computational models of the proteins in the novel coronavirus SARS-CoV-2 that may prove useful for the development of drugs to treat COVID-19.

Experimental techniques like cryo-electron microscopy can provide a more definitive picture of a protein's structure, but computational methods are much faster – something that is important for a fast-spreading virus like this where time is of the essence. Kihara used Gilbreth, Purdue's powerful GPU cluster optimized for machine learning applications, to compare the protein's amino acid sequence to sequences in a database of known proteins from other viruses, and predict which pairs of amino acids are interacting or not. He then used the Brown and Halstead clusters to run the simulation structure model and build a three-dimensional model of the protein.



Purdue research team uses community clusters to track COVID-19, malaria variants

Giovanna Carpi, assistant professor of biological sciences and a member of Purdue's Institute of Inflammation, Immunology and Infectious Disease, uses the Bell cluster to assemble the novel coronavirus genome in samples taken from individuals with COVID-19 infections and identify a constellation of mutations that characterize variants of concern.

Ultimately, the detailed COVID-19 genomic analysis from this study "will help us better understand viral spread in congregate settings and the younger population and how we can help mitigate it in the future," said Carpi.

Carpi works closely with Purdue's Animal Disease Diagnostic Laboratory, the Protect Purdue Health Center, and the Indiana State Department of Health, which has sent samples directly to her for genome sequencing instead of to the CDC, since her lab can turn results around faster.

MORE THAN 200 FACULTY PARTNERS FROM 3 PURDUE CAMPUSES FROM ALL OF PURDUE'S PRIMARY COLLEGES AND 60 DIFFERENT DEPARTMENTS

Of course the novel coronavirus is not the only pathogen that causes serious disease. In addition to working on COVID-19 genomic surveillance, Carpi is using the GPU-based Gilbreth community cluster to study malaria parasites in Southern Africa, where malaria is endemic and primarily affects children.

RESEARCH HIGHLIGHTS

Research Computing expands GPU capacity for machine learning, AI applications

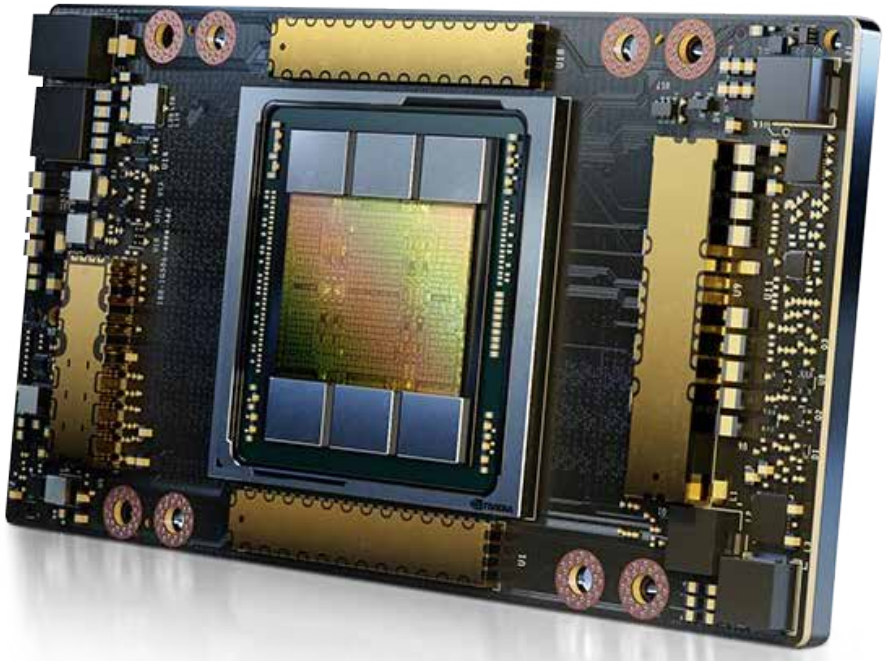
Research Computing recently added 24 new Nvidia A100 GPUs in the Gilbreth community cluster, 16 new AMD MI50 GPUs in the Bell cluster, and 8 new A100 GPUs in the Geddes composable platform. In addition, MI100 GPUs currently on loan from AMD are available to Purdue PIs for benchmarking.

This addition brings the total of centrally operated GPUs at Purdue to 158, with a total performance of 2.1 single-precision PetaFLOPs. These new GPUs represent a 45% increase in the number of GPUs, and a 70% increase in the number of single precision FLOPs to support AI and machine learning research. Additional new GPU capacity is planned for 2022.

"To make sure that our Purdue faculty have the facilities on hand to be competitive with this revolution in AI research, our next several years' plans in Research Computing reflect an increase in the annual investment in HPC resources designed for AI."

- Preston Smith, executive director of Research Computing

In FY 2021, Purdue faculty submitted 293 proposals for $281.8M that specified the need for research computing facilities



Envision Center collaborates with faculty on VR apps for research, education

With the COVID-19 pandemic moving many activities virtual, more researchers than ever before turned to Research Computing's Envision Center for help building interactive applications for use in the classroom or laboratory.

The Envision Center has collaborated with schools in the College of Engineering to design virtual labs for engineering students, such as the Aerospace Structural Analysis lab, designed to teach students about how different materials behave under stress. To expand the reach of this innovative instructional approach, the College of Engineering recently launched the Virtual Labs Faculty Fellows program. The 10 inaugural fellows received $10,000 in discretionary funding and participated in a six-week virtual lab development program this summer. Weekly workshops focus on evidence-based practices and activities to design, develop and implement virtual labs content. College experts are teaching the workshops, supported by specialists from the Envision Center, which produces the virtual labs, and the Center for Instructional Excellence (CIE). The new virtual labs are expected to be deployed by the 2022-2023 academic year.

An interdisciplinary team from the School of Nursing, the School of Industrial Engineering and the School of Biomedical Engineering worked with the Envision Center to build a mannequin that can teach nursing students to recognize non-verbal signs of a stroke, such as facial drooping. The simulation, which was built using projected augmented reality, displays an image onto a real three-dimensional object and doesn't require the user to have a headset or any special technology.

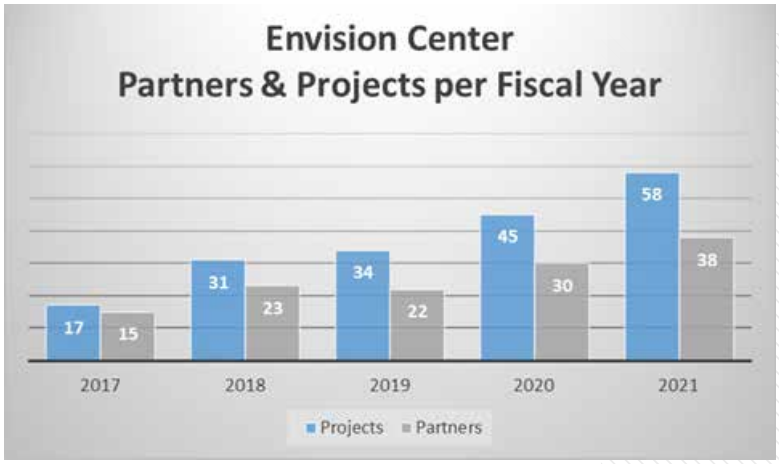
RESEARCH HIGHLIGHTS

Envision Center collaborates with faculty on VR apps for research, education

Krishna Jayant, assistant professor of biomedical engineering, and his graduate student Hammad Khan reached out to the Envision Center to design a virtual reality environment for mice that they could use to study how memories form in the brain. As the mouse runs on a physical wheel, it syncs to an optical (computer) mouse, giving the mouse the impression it's moving through a physical corridor as it navigates the virtual environment. The researchers record the electrical signals firing in the mouse's brain as it explores the virtual environment and can map how a location-specific memory forms in the hippocampus.

"This is a great resource to have at Purdue. Something that would normally cost us $50,000 to build cost much less. It was a huge benefit to us."

- Krishna Jayant, Assistant Professor of Biomedical Engineering



RESEARCH HIGHLIGHTS

Community cluster used to model enterovirus antibodies

A Purdue research team has used the community clusters to model antibodies that neutralize enterovirus D68, a respiratory virus that sometimes causes a polio-like illness in children.

Collaborators at Vanderbilt University had identified antibodies that appeared to neutralize the virus in laboratory tests. The Purdue team then modeled the antibodies and virus and concluded that the two identified antibodies bound very well to the virus.

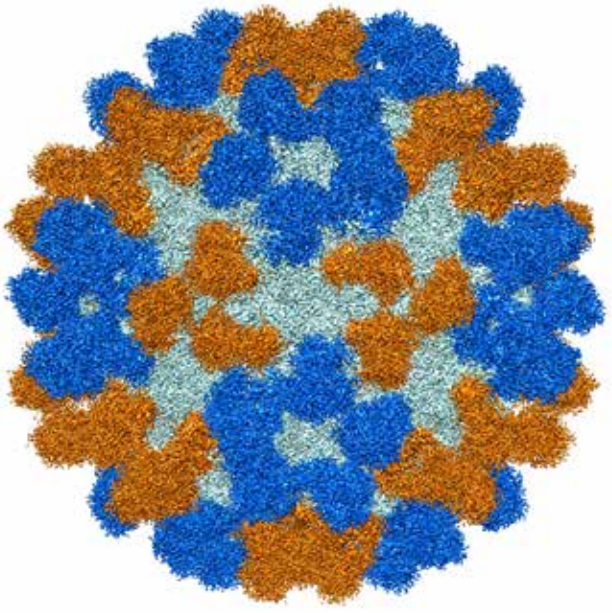
Thomas Klose, technical director of Purdue's Cryo-EM facility, and his Purdue collaborators used a technique called cryo-electron microscopy to create many different images of the virus and antibody structure, and transferred the information to Research Computing's Snyder community cluster (since retired) to run a reconstruction program that combined the two-dimensional images into a three-dimensional structure model. This work was both computationally-intense as well as memory-intense, and could not have been accomplished without the power of a supercomputer.

For the tens of terabytes of data generated by their work, Klose's team relies on the Fortress archive. Fortress preserves their data long-term, while allowing the team access to it if they ever need to review something.

61% OF PURDUE'S FY2021 RESEARCH EXPENDITURES

($346M)

WERE BY RESEARCH COMPUTING PARTNERS

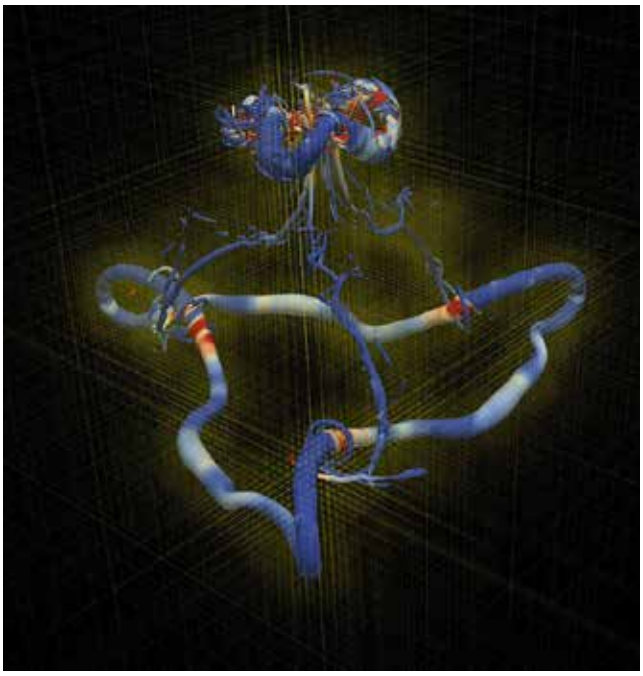


RESEARCH HIGHLIGHTS

Research Computing playing important role in Purdue's Next Moves Initative

Research Computing's computational resources are playing a critical role in Purdue's new Next Moves Initiative, five distinct strategic initiatives designed to advance the university's competitive advantage in its continuing quest for leadership among the world's top research and teaching institutions.

Senior research scientist Carol Song has collaborated with faculty projects that tie into the Plant Sciences 2.0 initiative, which will leverage and expand Purdue Agriculture's expertise in digital agriculture, phenotyping and agricultural economics to add value to plants and forest products by optimizing productivity, nutrition and sustainability traits and coupling these to consumer preferences. Song has collaborated with Jian Jin, assistant professor of agricultural and biological engineering, on developing software for LeafSpec, a portable hyperspectral leaf imaging device. The digital readout from the device can help guide when a farmer should water, fertilize a field, or address specific diseases or other plant stressors.



Research Computing's cyberinfrastructure is used by many researchers working in hypersonics and energetic materials and systems, two strategic areas of focus in the National Security and Technology Next Moves initiative. The U.S. Army has partnered with a Purdue team led by Jeffrey Rhoads, professor of mechanical engineering, to advance technology related to energetic materials - explosives, propellants and pyrotechnics. The team will be using the Bell community cluster. Hypersonics researchers like Jonathan Poggie, professor of aeronautics and astronautics, and Carlo Scalo, associate professor of mechanical engineering, have also made extensive use of the community clusters for modeling fluid mechanics. Research Computing also supports the National Security and Technology initiative through the Weber cluster, which is dedicated to data, applications, and research which are covered by export control regulations such as EAR, ITAR, or requiring compliance with the NIST SP 800-171.

Purdue selected as Microsoft Azure HPC and AI Collaboration Center

In 2021, Purdue was selected as one of five inaugural Microsoft Azure HPC and AI Collaboration Centers.

The Purdue center will demonstrate best practices for using Microsoft's Azure HPC cloud computing platform to augment both on-campus supercomputers and Anvil.

Purdue plans to use Microsoft Azure for three major purposes. It will allow the Research Computing team that is currently building Anvil to benchmark and test applications on the new 3rd Gen AMD EPYC processors before they're physically delivered to Purdue. Azure will also allow for surge capacity on Anvil for high-throughput work. Finally, the collaboration center will partner with researchers to use Azure for cloud-native workflows in areas such as data analytics and bioinformatics.

"Purdue is very excited to work with Microsoft on this new initiative. Investing in these cloud resources will allow us to offer both Purdue researchers and users of Anvil from across the country the flexibility to complement their traditional HPC workflows and make their workflows more portable."

- Preston Smith, Executive Director for Research Computing at Purdue, and PI for the Purdue Collaboration Center

Beyond supporting Anvil, the center will advance the Azure cloud as a complement to Purdue's award-winning community cluster program. The community cluster program serves researchers from every college and more than 60 departments. In addition to Anvil, Purdue recently deployed the Bell cluster, which is optimized for traditional, tightly-coupled science and engineering applications and named after Clara Bell Sessions, a Purdue nursing professor and diversity advocate.

Purdue-led workshop series produces report on best practices for regulated research data

A Purdue-led workshop series about how higher education institutions manage regulated research data, such as Department of Defense work or health sciences research, has led to a paper published on the EDUCAUSE Cybersecurity Resources page. The workshop report is available online at https://www.rcac.purdue.edu/cuiworkshop/community_report.

Over the course of six virtual workshops, co-organized by Purdue in collaboration with Duke University, University of Florida and Indiana University, 155 participants from 84 research institutions came together to discuss how to improve the support of individual program's efforts to secure regulated data.

Workshop participants included research computing directors, information security officers, compliance professionals, research administration officers and personnel that support and train researchers.

Regulated Research at Purdue

Research Computing's Weber cluster is designed for research subject to export control regulations, such as EAR or ITAR, or that require compliance with NIST SP 800-171 standards. Purdue does a great deal of contract-based research with industry, the Department of Defense and other sponsors that fall under those controls. Weber is named for Astronaut Mary Ellen Weber, a Purdue alumna, engineer, scientist, and skydiver, who flew on two Space Shuttle missions including one to the International Space Station.

The resulting Community Report, which identifies challenges, shares best practices and provides recommendations to the community on how to handle regulated research data on campus, was co-authored by Carolyn Ellis, former program manager for Research Computing and an expert on managing regulated research data.

The workshops and resulting paper were funded by a $600,000 award from the National Science Foundation (NSF award #1840043). Ellis and Research Computing executive director Preston Smith are co-PIs on the grant and Baijian Yang, associate professor of computer and information technology, is the PI.

EDUCATION AND OUTREACH

Research Computing training the next generation of cyberinfrastructure professionals

Research Computing hosted a workshop last year that brought together 100 attendees to discuss issues related to building and enhancing the cyberinfrastructure professional workforce that develops software and operates and supports computing facilities for research. The goal of the workshop was to identify problem areas and discuss and identify potential solutions, and participants produced a written report (available online at https://www.rcac.purdue.edu/files/ciworkforce2020/report.pdf) with recommendations for the cyberinfrastructure community, higher education institutions and the NSF. This project was funded by NSF Award No. 2036534. The principal investigator is Thomas Hacker, professor of computer and information technology, and Research Computing Executive Director Preston Smith is a co-PI.

This past summer, Research Computing hosted a Research Experience for Undergraduates (REU) program with two senior computer science majors from the University of Washington Bothell. The students worked on two projects for the new Geddes private-cloud resource, with one focused on user experience and one focused on security. The students learned how to deploy virtual machines in Research Computing infrastructure, work with configuration management tools, deploy Kubernetes and work with GPUs and machine learning toolkits from NVIDIA and software used by Research Computing researchers. The REU was supported by funding from NSF Award No. 2018926, on which Smith is the PI and Hacker and Research Computing senior scientific applications analyst Erik Gough are co-PIs.

Taylor Johnson

Victor Shaw

In November 2021, the Envision Center hosted the sixth annual Campus Alliance for Advanced Visualization (CAAV) conference. The conference focused on collaborative technologies and applications developed to aid in people's ability to innovate, collaborate and educate from a distance. Roy C. Anthony, the global head of research at the visual effects and animation studio DNEG, and Mark Subbarao, who leads NASA's Scientific Visualization Studio, delivered keynote addresses, and Envision Center lead visualization scientist George Takahashi spoke on a panel about diversity and inclusion and also presented his work with Dan Milisavljevic, assistant professor of physics and astronomy, regarding supernova remnant VR visualizations and how the technology was used to engage remote students during the pandemic. Envision Center assistant director of center operations and visualization Laura Theademan was the conference chair.

EXTERNAL FUNDING

Geddes composable cloud platform advancing science through national computing network

Purdue's new cloud-based composable platform, Geddes, is one of the top resources used by researchers on the Open Science Grid's (OSG) Open Science Pool, a collection of computing resources available to researchers across the country.

Geddes is a composable platform that allows users to use container technology to customize resources to their needs, and complements Purdue's community clusters, which are designed for a broad variety of traditional batch science and engineering applications. The system is named after LaNelle Geddes, a former Head of Purdue's School of Nursing.

The Open Science Pool recently delivered 1.1 million core hours in a single day for the first time, and that's in no small part thanks to Geddes, which has provided over 20 million core hours to Open Science Pool in the last six months.

Through the Open Science Pool, Geddes has been used by researchers working on projects such as:

- Simulating antimatter particle collisions at different cosmic-ray energies
- The Event Horizon Telescope, an international collaboration that created a virtual Earth-sized telescope to capture the first ever image of a black hole.
- A high-intensity frontier experiment searching for physics beyond the Standard Model

Geddes is supported by a $400,000 award from the National Science Foundation (NSF award number #2018926). Preston Smith, executive director of Research Computing, is the PI on the project and Erik Gough, a senior computational scientist for Research Computing, is a co-PI.

179 of Purdue's 808 earned doctorates in 2020 (22%) were HPC users during their Purdue career

In 2020, Research Computing faculty were authors of 38% of Purdue's 34 articles in Science, Nature, and Proceedings of the National Academy of Sciences of the United States of America.

EXTERNAL FUNDING

Research Computing executive director presents to federal agencies

Preston Smith, executive director of Research Computing, recently spoke at the Networking and Information Technology Research and Development (NITRD) Program's working group on high-end computing about the campus perspective on the health of the HPC vendor ecosystem.

NITRD is the nation's primary source of federally funded R&D in advanced information technologies. The 23 NITRD member agencies, including the National Science Foundation, invest approximately $6.5 billion annually in advanced networking and IT capabilities.

Smith gave the agency representatives an overview of Purdue's Community Cluster program, which delivered 378 million computational hours to more than 200 principal investigators last year, and explained how the Community Cluster Program's widely emulated "condo" computing model allows researchers to affordably access powerful computational resources through economies of scale, while Research Computing maintains the clusters and provides expert staff support.

He outlined to the working group how Purdue evaluates and chooses vendors for processors, filesystems, archival storage, workload managers, interconnects, accelerators and cooling solutions.

"Our success in the community cluster program has been made possible by competitive markets in processors, networks, and accelerators, allowing us to provide high performance computing at the highest proven value," said Smith. "I hope that this remains a viable option in the face of increasing consolidation in the HPC vendor space."

HPC investment from early-career faculty has grown from 5% in Carter (2012) to 52% of Bell's (2020) faculty partners.

In FY 2021, Research Computing faculty earned 66 times Purdue's cyberinfrastructure investment in new grant awards.

EXTERNAL FUNDING

Research Computing senior research scientist awarded tens of millions in sponsored research funding

Carol Song

Carol Song, senior research scientist for Research Computing, is a co-PI on a $15 million NSF institute for geospatial data-driven research based at the University of Illinois Urbana-Champaign. The Institute for Geospatial Understanding through an Integrative Discovery Environment (known as I-GUIDE) is part of the NSF's Harnessing the Data Revolution initiative and will help researchers better estimate and predict risk and anticipate impacts from natural disasters or climate change.

Song is co-leading the project's cyberinfrastructure core and her team's role will be to develop and integrate geospatial cyberinfrastructure to enable the research agenda of the institute and engage with the broader community.

"Purdue Research Computing's research capacity and recent development in advanced cyberinfrastructure and data frameworks, such as the NSF Anvil and GeoEDF projects, has positioned us well for the CI leadership role in this institute," says Song.

Song is also a co-PI on the Global to Local Analysis of Systems Sustainability (GLASSNET) project, a collaborative effort to address sustainability issues related to land and water use that is led by Thomas Hertel, distinguished professor of agricultural economics and funded by a $2 million award from the NSF.

Song is co-chairing the project's working group on data and computational infrastructure, and key GLASSNET activities will be facilitated by MyGeoHub, a science gateway developed by Song's team, which will be used for easy access to supercomputers, publishing tools and sharing datasets. Among the existing MyGeoHub tools that the project will make use of are the SIMPLE-G and GGCMI crop modeling tools and the AgMIP tool connecting crop scientists with economists and policy analysts.

MyGeoHub will also host tutorials and other interactive learning materials to assist the project in achieving one of its main goals, developing the next generation of sustainability researchers.

Song is also the PI on Purdue's Anvil system, which is supported by a $22 million award from the NSF.

CONTACTS

Preston Smith
Executive Director, Research Computing
psmith@purdue.edu
765-494-9729

Carol Song
Director of Scientific Solutions
carolxsong@purdue.edu
765-496-7467

George Takahashi
Envision Center Technical Director
gtakahas@purdue.edu
765-496-7888

www.rcac.purdue.edu

PURDUE UNIVERSITY | Information Technology

Purdue University is an EA/EOU