Special COVID-19 Report
Starting on page 6.
XSEDE helps the nation's most creative minds discover breakthroughs and solutions for some of the world's greatest scientific challenges. Through free, customized access to the National Science Foundation's advanced digital resources, expert consulting, training and mentorship opportunities, XSEDE enables researchers to Discover More.

XSEDE provides the following tools and services to help researchers make the most of their allocations:

- Extended Collaborative Support Services (ECSS) pairs XSEDE users with computational science experts to maximize their research potential.
- The XSEDE Cyberinfrastructure Integration (XCI) team provides advanced hardware and software architecture for a more integrated user experience.
- The XSEDE User Portal helps users access XSEDE resources, manage jobs, report issues, and view results.
- The Resource Allocation Service (RAS) team helps coordinate allocations of NSF's high-end resources and digital services.
- Training, education, workforce development, and campus engagement provided by the Community Engagement and Enrichment (CEE) team.
- Specialized community services, provided through the XSEDE Federation, allow for rapid innovation and experimentation in areas like gateway development, education, and training.
- A fellowship program that allows Campus Champions to work closely with XSEDE's advanced user support staff.
XSEDE wouldn’t be possible without our extensive network of collaborators. Led by the University of Illinois’ National Center for Supercomputing Applications, the XSEDE partnership includes the following institutions:

Center for Advanced Computing (CAC) at Cornell University
Pittsburgh Supercomputing Center (PSC), a joint effort of Carnegie Mellon University and the University of Pittsburgh
San Diego Supercomputer Center (SDSC) at the University of California San Diego
Texas Advanced Computing Center (TACC) at the University of Texas at Austin
Center for Education Integrating Science, Mathematics, and Computing (CEISMC) at Georgia Institute of Technology
Information Sciences Institute at the University of Southern California
National Center for Atmospheric Research (NCAR) at the University Corporation for Atmospheric Research (UCAR)
National Institute for Computational Sciences (NICS) at the University of Tennessee, Knoxville
Ohio Supercomputer Center
Pervasive Technology Institute (PTI) at Indiana University
Rosen Center for Advanced Computing at Purdue University
Shodor
Southeastern Universities Research Association (SURA)
OU Supercomputing Center for Education & Research at the University of Oklahoma
University of Chicago
Internet2
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Page</th>
<th>Section</th>
</tr>
</thead>
</table>
| 04   | DIRECTOR’S MESSAGE  
Annual letter from XSEDE Principal Investigator and Project Director, John Towns. |
| 06   | SPECIAL COVID-19 REPORT  
Five Projects XSEDE Enabled to Address a Global Pandemic. |
| 10   | TROUBLE IN THE AIR (AND BEYOND)  
XSEDE Systems Power DNA-Based Identification of Surface, Airborne Microbes Worldwide. |
| 12   | CONTINUING ARECIBO’S LEGACY  
XSEDE’s Ranch Storage System to Archive More Than 50 Years of Astronomy Data. |
| 14   | MITIGATING DISASTERS AT SEA  
Supercomputers Aid in Predicting Oil Dispersal During Spills. |
| 16   | FORGING STRONGER LINKS  
Online Supply Chain Tool Available to Authorities, Public Thanks to XSEDE Resources. |
| 18   | A SIMPLER RECIPE FOR THE SOUP  
| 20   | SAVING HEAVY METAL  
Supercomputer Models Describe Chloride’s Role in Metal Corrosion. |
| 22   | “IT’S COMPLICATED” FOR OREGON RODENT  
XSEDE Systems Power Discovery of Two-X-Chromosome Male Voles. |
| 24   | TALKING THE TALK  
Terminology Task Force Endeavors to Create a More Inclusive XSEDE Through Language. |
A MESSAGE FROM

THE DIRECTOR

A warm welcome to the ninth edition of the Extreme Science and Engineering Discovery Environment (XSEDE) highlights publication, featuring a selection from the year’s finest accomplishments on which we have been honored to collaborate and make possible. The incredible progress of our partners across the research community never ceases to amaze me and I am proud to share their successes with you.

Regardless of discipline, underlying each scholarly advancement is a fundamental passion to understand and resolve some of the greatest challenges our society faces today. XSEDE is proud to be part of those endeavors. As might be anticipated, the COVID-19 crisis has been a significant influence and inspiration of innovation over the past year. We provide some specific highlights with respect to our efforts to address the pandemic. Also, in this edition, you’ll find discoveries covering a wide range of disciplines, including biology, genetics, astronomy, environmental science, and materials science. Finally, we’ve included a piece regarding the efforts of the newly founded XSEDE Terminology Task Force to review, address, and define processes eliminating offensive terms in our materials and creating a community resource for communicating with respectful, inclusive language.

XSEDE has continued its efforts in response to the call from our country in March 2020 by the Office of Science and Technology Policy at the White House to support the efforts of a newly formed collaboration under the moniker COVID-19 HPC Consortium. This group brings together federal agencies, academic institutions, and private companies across the nation and internationally to make available an unprecedented portfolio of resources to address the COVID-19 pandemic. We also continue to manage the scientific review process for the Consortium, whose array of resources is truly impressive and represents a collaborative effort that has never before taken place.

As the pandemic has played out over the past year – dominating headlines and becoming integral to the fabric of our global society – it cannot be forgotten that XSEDE continues to provide resources, services, and support to the national research community as we have now for more than a decade. We are incredibly grateful to the National Science Foundation for their continued support, including the project supplement and extension, which will enable us to continue this meaningful work through August 31, 2022.

Finally, I want to express my deepest gratitude and appreciation to every member of the XSEDE team across our many partner institutions for their untiring effort on this project. It is only through the contributions of all that we are successful as a team and make meaningful contributions to the work of our research community collaborators. I continue to be honored and humbled by the privilege of leading this project, but even more so, deeply appreciate the relationships across the community built through our working together.

Looking forward, I am excited to continue this journey into a safe and productive future.

Sincerely yours,

John Towns
XSEDE PI and Project Director
Scientists trying to understand the virus causing COVID-19 used XSEDE resources across the nation to gain speed and clarity in their studies. Studies paramount to getting the virus – and its variants – under control as the world looks forward to putting the devastation of this modern pandemic behind us.

Using these supercomputing resources enabled research, “that could not have been completed on standard hardware,” and saved great expense in computer time. Even in using XSEDE to provide simple quantification, the resources gave researchers “high confidence in the results.”

Here are five stories illustrating XSEDE’s essential impact on COVID-19 research.
RE-ENGINEERING ANTIBODIES TO FIND RELIEF FOR COVID-19

During the pandemic, people turned to antibody tests to determine whether they had been exposed to the coronavirus that causes COVID-19.

"Antibodies are important because they prevent infection and heal patients affected by diseases," said Victor Padilla-Sanchez, a researcher at The Catholic University of America in Washington, D.C. "If we have antibodies, we are immune to disease – as long as they are in your system, you are protected."

Padilla-Sanchez uses computer models to understand the structure of viruses on the molecular level and uses this information to figure out how the virus functions. For his research, Padilla-Sanchez relies on supercomputing resources allocated through XSEDE.

A June 2020 study in Research Ideas and Outcomes described Padilla-Sanchez’s efforts to unravel this problem using computer simulation. He discovered sequence differences prevent antibodies 80R and m396 from binding to COVID-19.

"Understanding why 80R and m396 did not bind to the SARS-CoV-2 spike protein could pave the way to engineering new antibodies that are effective," he said. "Mutated versions of the 80r and m396 antibodies can be produced and administered as a therapeutic to fight the disease and prevent infection."

The XSEDE-allocated systems Stampede2 at the Texas Advanced Computing Center (TACC) at The University of Texas at Austin and Bridges at the Pittsburgh Supercomputing Center (PSC), a joint effort of Carnegie Mellon University and the University of Pittsburgh, supported docking experiments, macromolecular assemblies, and large-scale analysis and visualization.

Padilla-Sanchez’s experiments showed that amino acid substitutions should increase binding interactions between the antibodies and SARS-CoV-2, providing new antibodies to neutralize the virus.

"XSEDE resources were essential to this research," Padilla-Sanchez said.

XSEDE ALLOCATIONS ASSIST VANDERBILT UNIVERSITY HUMAN VACCINES PROJECT

A team from the San Diego Supercomputer Center (SDSC) at UC San Diego contributed to a study led by Vanderbilt Vaccine Center of Vanderbilt University Medical Center (VUMC) on T cell receptors, which play a vital role in alerting the adaptive immune system to mount an attack on invading foreign pathogens including Coronavirus SARS-CoV-2.

Using XSEDE allocations, SDSC’s Comet performed complex calculations on the receptor sequence data from sorted human T cells to allow scientists to better understand the size and diversity of receptor repertoire in healthy individuals. The team’s findings were published in Cell Reports as a follow-up study to earlier findings about B cells published in Nature.

"Our most recent study puts us one step closer to truly understanding the extreme and beneficial diversity in the immune system and identifying features of immunity that are shared by most people," said James E. Crowe, Jr., director of VUMC’s Vanderbilt Vaccine Center. "Now we continue to identify T cell receptors and antibodies that can be targets for vaccines and treatments that work more universally across populations."

Madhusudan Gujral, a senior bioinformatician at SDSC; Robert Sinkovits, XSEDE co-PI and ECSS co-director; and Cinque Soto, a Vanderbilt computational biologist and
lead author of the study, share Crowe’s enthusiasm over the implications of this research and recognize the importance of access to high-performance computing resources to make it possible.

"Being able to access Comet through an XSEDE allocation made us much more productive," said Sinkovits. "The larger memory nodes were also essential for some of the clustering calculations that could not have been completed on standard hardware."

This work was supported by a grant from the Human Vaccines Project and institutional funding from Vanderbilt University Medical Center. The authors acknowledge support from TN-CFAR grant (P30 AI110527). This work also used XSEDE, which is supported by NSF grant (ACI-1548562), and Comet supercomputer at SDSC, supported by NSF grant (ACI-1341698).

SUPERCOMPUTERS SIMULATE NEW PATHWAYS FOR POTENTIAL RNA VIRUS TREATMENT

University of New Hampshire researchers used the XSEDE-allocated Comet at SDSC and Stampede2 at TACC to identify new inhibitor binding/unbinding pathways in an RNA-based virus. The findings could be beneficial in understanding how these inhibitors react and potentially help develop a new generation of drugs to target viruses with high death rates, such as HIV-1, Zika, Ebola, and SARS-CoV2, the virus that causes COVID-19.

In their study, published in the Journal of Physical Chemistry Letters, UNH Associate Professor of Chemical Engineering Harish Vashisth and his team created molecular dynamics simulations using XSEDE allocations on Comet and Stampede2 to look specifically at an RNA fragment from the HIV-1 virus and its interaction with acetylpropazine. This small molecule is known to interfere with the virus replication process.

The scientists focused on structural elements from the HIV-1 RNA genome because they are considered a good model for studying the same processes across a wide range of RNA viruses. These simulations allowed them to discover the pathways of the inhibitor unbinding from the viral RNA in several rare events – which are often difficult to observe experimentally – that unexpectedly showed a coordinated movement in many parts of the binding pocket that are the building blocks of RNA.

The XSEDE allocations enabled researchers to run hundreds of simulations simultaneously to observe rare base-flipping events involved in the inhibitor binding/unbinding process that provided new details of the underlying mechanism of this process.

Funding for this study was provided by NSF (CBET-1554558) and NSF (OIA-1757371). Access to Comet and Stampede2 was provided by NSF XSEDE allocation (TG-MCB160183).

RAPID ID OF POTENTIAL ANTI-COVID-19 AGENTS POWERED BY XSEDE

Speed can be as important as quality in medical research. Scientists have used the XSEDE-allocated Bridges-AI system at PSC to screen about five billion chemical compounds, selecting a small number of candidates for combating COVID-19 thousands of times faster than possible with previous methods.

Simulating the interactions of candidate molecules with target proteins, such as those on the SARS-Cov-2 virus, saves time and expense by allowing scientists to lab-test only the most promising. But the quantum chemistry standard for such simulations is complex and expensive in computer time. Olexandr Isayev of Carnegie Mellon University, working with colleagues at the University of North Carolina Chapel Hill, where he began the effort, and the University of Florida, had an idea to use the speed of graphics processing units (GPUs) by unloading the complexity into computer memory. The Bridges-AI system had the perfect balance of GPU power and memory for the work.
With an allocation from XSEDE via the COVID-19 HPC Consortium, the scientists first “trained” their AI on drugs known to be effective against SARS-CoV-2. Once the AI reproduced the effects correctly, the team used it on several databases containing about five billion known antiviral or FDA-approved or investigational drugs. They made their results, consisting of 20,000 promising candidates, freely available to the research community and entered the European Union’s COVID Challenge project. The team intends to apply the AI pipeline to other medical and industrial processes unrelated to the SARS-CoV-2 virus in the future.

This work was supported by XSEDE award DMR110088, funded by NSF grant number ACI-1548562. Specifically, it used the Bridges system, which was supported by NSF award number ACI-1445606, at PSC.

**SUPERCOMPUTER SIMULATES HOW PATHOGENS TRAVEL, LAND IN THE GROCERY STORE**

A team of environmental engineers used XSEDE allocations on Comet at SDSC to simulate how COVID-19 airborne pathogens travel and land in the familiar setting of the supermarket.

Led by Professor Michel Boufadel and Research Associate Fangda Cui from the New Jersey Institute of Technology, the Comet simulations were the basis for analysis reported in an April 2021 *Journal of Environmental Engineering* study. The research revealed that the attachment of virus-laden particles on shelves, floors and ceilings of supermarkets reduces the maximum concentration of suspended particles in the air by as much as 50%.

According to the study, the attachment of airborne, virus-laden particles on surfaces at 25 and 100% have the same efficiency. These particles are as small as dust and don’t fall onto a surface within a short time of emission, but they significantly accumulate on surfaces after five minutes or more. Boufadel suggested that one-way lanes in stores might be a step in the right direction, but he also said display shelves in aisles help break the flight of virus particles.

"In this particular project, we used XSEDE as a tool to solve a simple problem (transport of particles) in a domain with a complicated geometry," said Boufadel. "The goal was a quantification, and the superior ability of XSEDE allowed us to have a high confidence in the results."

This study was funded by the NSF Rapid Response Research grant program (CBET 2028271). Computations on Comet were funded by XSEDE (TG-BCS190002).
COVID-19 has offered a harsh lesson on the importance of knowing what the microbial world is doing. Public health experts would like to monitor bacteria and viruses worldwide, much as weather radar scans for incoming storms.

That’s why a huge international collaboration, led by Christopher Mason of Weill Cornell Medicine, decoded the DNA of bacteria and viruses in the air in six cities and on surfaces in public transit locales in 60 cities across the globe. Using three of the most powerful XSEDE-allocated supercomputers in three different eras – PSC’s Blacklight from 2010 to 2015, Bridges from 2015 to February 2021, and Bridges-2 since – the team sequenced thousands of microbes’ DNA at once. The systems’ computational brawn enabled a many-species metagenomic map for use by scientists and public health experts.

“In a nutshell, we wanted to build a genetic, functional and geospatial map of the DNA of the world's cities,” Mason said. “Just kind of like the Google Maps of DNA of the Earth.”

The most popular DNA sequencing methods can decode a few hundred nucleotides – the A, C, T, G code of DNA – at a time. When using such a “short read,” scientists need to match millions of these overlapping fragments of DNA sequence to put a species’ genome in proper order. Mason and his collaborators wanted to do even more, sorting DNA fragments from thousands of species of bacteria and viruses at once. Blacklight, Bridges and Bridges-2 all offered large-memory nodes that made the task possible.

Graduate student David Danko at Weill Cornell Medicine and Research Associate Daniela Bezdan, at Weill Cornell Medicine and the Abdulaziz Alsaud Institute for Computational Biomedicine, worked with Mason and hundreds of scientists worldwide to collect and analyze 4,728 surface samples from mass transit locations in dozens of cities globally in 2016 and 2017. In parallel work, M. H. Y. Leung and X. Tong at the City University of Hong Kong, K. O. Bøifot at the Norwegian Defence Research Establishment and Mason performed a similar analysis of 259 airborne samples in Denver, Hong Kong, London, New York City, Oslo, and Stockholm.

“To do all of our de novo assembly at scale, [XSEDE] gave us the fastest and most expansive computational framework through which we could assemble all the sequences to find what were the real novel species and the novel genetic elements in this data set,” Mason said. “It probably literally wouldn’t have been possible in this time frame without that infrastructure.” XSEDE personnel proved critical in making it all work quickly and well, he added.

The results gave scientists their first metagenomic map of urban areas worldwide. The surface samples contained more than 15,000 species of virus, bacteria and archaea, primitive bacteria-like organisms from
which more complex plants and animals evolved. The airborne samples showed evidence of more than 450 species. Amazingly, fewer than 10% of these were known to science.

The cities had varying amounts of the 31 most common species from surfaces and 17 from the air. These “microbial fingerprints” varied over time, though the cities remained recognizably distinct. The team also detected and monitored differences in 20 known genes that make bacteria resistant to antibiotics, a finding with a large potential payoff for human health.

The collaborators reported their results in two simultaneous May 2021 papers, one in the prestigious journal *Cell* and another in *Microbiome*. The team is now expanding their research by collecting RNA data, which will open up a view of RNA viruses such as the coronavirus that causes COVID-19.

For the *Cell* paper, the Tri-I Program in Computational Biology and Medicine (CBM) was funded by the NIH grant 1T32GM083937, GitHub and XSEDE, which is supported by NSF grant number ACI-1548562. Specifically, it used the Bridges-2 system, supported by NSF award number ACI-1928147, at the PSC. The *Microbiome* paper used XSEDE, supported by National Science Foundation grant number ACI-1548562. Specifically, it used the Bridges-2 system, supported by NSF award number ACI-1928147, at PSC.

CONTINUING ARECIBO’S LEGACY

XSEDE’s Ranch Storage System to Archive More Than 50 Years of Astronomy Data.

HOW XSEDE HELPED

After Arecibo collapsed, the partners began moving more than 50 years of astronomy data to the Ranch storage system, an allocated resource of the Extreme Science and Engineering Discovery Environment (XSEDE) funded by the National Science Foundation (NSF). Arecibo plans to provide researchers access to the data and the tools necessary to continue the observatory’s legacy of discovery and innovation.

Millions of people have seen footage of the famed Arecibo radio telescope’s collapse in December 2020. What they would not have seen from those videos was Arecibo’s data center, located outside the danger zone. It stores the ‘golden copy’ of the telescope’s data – the original tapes, hard drives, and disk drives of sky scans since the 1960s.

Within weeks of the observatory’s collapse, TACC, an XSEDE service provider, entered into an agreement with the University of Central Florida (UCF), the Engagement and Performance Operations Center, the Cyberinfrastructure Center of Excellence Pilot and Globus at the University of Chicago. Together, they’re moving the radio telescope data to the Ranch storage system, an XSEDE-allocated resource at TACC, funded by the National Science Foundation (NSF). Plans include expanding access to more than 50 years of astronomy data from the observatory, which up until 2016 had been the world’s largest radio telescope. The data transfer process started in mid-January 2021 and will continue for about one year.

“Arecibo data has led to hundreds of discoveries over the last 50 years,” said Francisco Cordova, director of the Arecibo Observatory. “Preserving it, and most importantly, making it available to researchers and students worldwide will undoubtedly help continue the legacy of the facility for decades to come.”

Since 2018, UCF has led the consortium that manages the observatory, which is owned and funded by NSF. “NSF is committed to supporting the Arecibo Observatory as a vital scientific, educational,
and cultural center, and part of that will be making sure that the vast amounts of data collected by the telescope continue to drive discovery,” said NSF Program Officer Alison B. Peck. “We’re gratified to see that this partnership will not only safely store copies of the data but also provide enhanced levels of access for current and future generations of astronomers.”

Arecibo’s data landscape consists of three main sources: data in hard drives, data in tape library and data offsite. The archive holds over one petabyte of data in hard drives and over two petabytes in tapes. This data includes information from thousands of observing sessions, equivalent to watching 120 years of HD video.

Past achievements made with Arecibo include the discovery of the first-ever binary pulsar, a find that tested Einstein’s General Theory of Relativity and earned its discoverers a Nobel Prize in 1993; the first radar maps of the Venusian surface and polar ice on Mercury; and the first planet found outside our solar system.
According to the National Oceanic and Atmospheric Administration (NOAA), thousands of oil spills occur yearly in the United States. Although most incidents involve less than one barrel, the spills have wreaked economic and environmental devastation for decades. To better understand the fate of oil droplets for effective countermeasures, researchers created simulations using XSEDE-allocated supercomputers, including Comet at SDSC and Bridges at PSC. "We used the supercomputers to create high-fidelity, large-eddy simulations of underwater oil blowout in water crossflow conditions," Cosan Daskiran, a postdoctoral researcher and senior engineer at New Jersey Institute of Technology (NJIT), explained. "The main goal was to understand the fluid dynamics and estimate the trajectory of different-sized oil droplets, which is important for the counter measures following oil spill incidents."

The systems allocations were funded by XSEDE and provided the perfect resources for Daskiran, to model studies on how oil dilutes under specific conditions.

Daskiran's XSEDE-allocated supercomputer simulations showed that large oil droplets rose faster and separated from the oil plume without
spreading spatially much within the plume due to their higher individual buoyancy. Meanwhile, small droplets were trapped in a counter-rotating vortex pair, which is considered a signature of the jets in crossflow.

Daskiran worked with Michel Boufadel, a professor at the Civil and Environmental Engineering Department at NJIT who has spent much of his career examining oil dispersal after a spill. The research team compared Daskiran’s simulations with actual oil dispersal experiments before publishing their findings in the *International Journal of Heat and Fluid Flow*.

"We used Ohmsett, short for the Oil and Hazardous Materials Simulated Environmental Test Tank, here in New Jersey to create a life-like oil spill in a controlled environment," said Boufadel. "Ohmsett is operated by the U.S. Navy and provided us an environmentally safe place to conduct tests for this project."

Specifically, the researchers conducted experiments that capture the main features of an oil jet by towing a pipe horizontally in the Ohmsett wave tank and then created simulations on *Comet* and *Bridges* based on this study.

Funding for this project include an allocation from XSEDE (TGBCS190002). The researchers also acknowledge support by the Department of Fisheries and Oceans Canada through Multi-Partner Research Initiative (MECTS-39390783-v1-OFSCP), and funding from the Center for Offshore Oil and Gas Environmental Response, or COOGER.
Online Supply Chain Tool Available to Authorities, Public Thanks to XSEDE Resources.

FORGING STRONGER LINKS

Supply chains of goods like food, fuel, and water have created wealth and greatly reduced human hunger. But they also bring with them vulnerabilities, when supplies are disrupted by natural disasters, human conflicts, or sudden demand. It isn’t always easy to see in a timely way how a sudden bottleneck can affect our ability to bring what we need to where we are.

“[Supply chains] are the flows that underlie the economy, and they connect us to people all over the world,” said Benjamin Ruddell of Northern Arizona University (NAU). “Those connections create efficiencies; they are a foundation of our wealth and prosperity and resilience. But those connections also carry risks and [environmental] impact.”

That’s why Ruddell and Richard Rushforth of NAU developed a set of National Science Foundation (NSF)-funded online tools called FEWSION. The first of its kind to integrate data for many national supply chains in one package, this freely available collection of software and databases can help authorities, non-governmental organizations, and the general public identify and plan for disruptions of critical supplies.

The first version of FEWSION, a snapshot of supply chains in 2012, was a success. Still, as the collaborators began tracking supply chains over time and their data approached the petascale (quadrillions – 1 followed by 15 zeros) of data elements, they turned to XSEDE. With help from the XSEDE Extended Collaborative Support Services (ECSS) and the flexibility of accessing numerous XSEDE-allocated systems – including Bridges at PSC and Wrangler, Ranch, and Stampede2 at TACC – they were able to select Bridges as the best fit to handle the massive datasets they needed. Using Bridges, they deployed the FEWSION for Community Resilience Network™ (F4R™), a citizen-led effort to collect supply-chain data for public use. This
unique tool will allow individual citizens, nonprofit organizations, companies and governments to understand the local “last-mile” supply chains of their specific interest. Another important tool is the team’s mesoscale FEW-View™ tool, a mapping interface for FEWSION that displays supply chain analysis for entire cities and communities.

FEWSION showed the collaborators how, for example, the city of Flagstaff, where NAU is based, depends on food from outside its region. Local food producers are specialized, small, or seasonal – they can’t feed the area alone. Getting other food products depends on exporting them. The FEWSION analysis showed where the weak links are and helped identify alternative sources.

The NAU scientists are preparing a series of papers on their Flagstaff analysis. To date, they’ve used FEWSION for studies on the fresh-water supply throughout the U.S. and how beef-production water use affects wild fish populations. They reported these results in a series of peer-reviewed publications.

Through F4R, they’re also recruiting new communities, adding new research collaborators, and offering consultation to communities and local and regional governments on how to use FEWSION.

FEWSION was founded in 2016 by grant number ACI-1639529 from the INFEWS program sponsored by NSF and the U.S. Department of Agriculture (USDA). Funding includes IARPA 2021-21010800001 and other sources. This work also used XSEDE supported by NSF grant number ACI-1548562. Specifically, it used the Bridges system, supported by NSF award number ACI-1445606, at PSC.
MuST, a new, open-source supercomputing code, radically increases the ease of simulating complex, industrially interesting materials. Developed in part with expertise from XSEDE’s Extended Collaborative Support Services (ECSS), the software promises quicker predictions of the properties of samples large enough to be relevant to the real world.

Named for the Multiple Scattering Theory on which it’s based, MuST uses density function theory (DFT) for ab initio investigation of disordered materials. In other words, it predicts materials’ properties from first principles.

“It’s a many-body problem, so it is impossible to solve,” said Yang Wang, senior computational scientist at PSC, ECSS staff member, and a leader in the research. “[Instead] we need to solve the distributed-equation-effective potential – it’s called ‘the soup.’ By solving this equation, [we] solve the electron density, which [answers] the question of the energy of the original many-body problem.”

MuST takes advantage of locally self-consistent multiple scattering theory (LSMS) to simplify calculating the environment of a given atom surrounded by a complex and often disordered mixture of other atoms. Instead of simulating every atom in a material, it simulates an atom’s environment as a kind of average “soup” made up of the different atoms around it. MuST accomplishes this with a combination of the Korringa-Kohn-Rostoker and coherent potential approximation methods.

“The soup reproduces the total behavior of these other atoms in their proportions in that alloy instead of accounting for each atom individually,” Wang said. “Each domain has its own potential, which you then add together to get an effective potential for the whole space. You can treat it as single-site scattering potential.”

Previous methods typically scaled to the third power. In other words, twice as many atoms takes eight
times as much computational power to simulate. MuST scales to the number of atoms, so twice as many atoms only takes twice as much computation. This scaling makes MuST unique among such methods.

Initial work with the software has produced results as good as those of gold-standard methods that require much more computing power. MuST also goes beyond the reach of those methods when large numbers of atoms (thousands or more) are involved.

Two goals for the future are incorporating the LSMS method with typical medium embedding to capture the metal-insulator transition phenomena driven by disorder in quantum materials and integrating the Kubo-Greenwood formula into the package. This will enable the investigation of electronic transport – the sequential movement of electrons through a material – in disordered structures. The phenomenon underlies electrical flow and resistance, which are important in designing new electronic components. It also underlies important natural processes like photosynthesis, by which plants use sunlight to turn carbon dioxide into sugars.

The product of an international collaboration of PSC with the Oak Ridge National Laboratory, Universität Augsburg, University of the Chinese Academy of Sciences, Louisiana State University, and Middle Tennessee State University, MuST was funded by the NSF’s Cyberinfrastructure for Sustained Scientific Innovation program and developed with the help of XSEDE’s ECSS. Scientists can download the program – and join the team to develop MuST – at:

https://github.com/mstsuite/MuST.

Ridge National Laboratory, Universität Augsburg, University of the Chinese Academy of Sciences, Louisiana State University, and Middle Tennessee State University, MuST was funded by the NSF’s Cyberinfrastructure for Sustained Scientific Innovation program and developed with the help of XSEDE’s ECSS. Scientists can download the program – and join the team to develop MuST – at:

https://github.com/mstsuite/MuST.

Using the KKR-CPA method, the MuST software converts the complex surroundings of an atom in a random alloy (brass, an alloy of copper and zinc, in this case) to an “effective medium” that averages the properties of the surrounding atoms. Credit: Yang Wang
Supercomputer Models Describe Chloride's Role in Metal Corrosion.

Researchers have been studying chloride's corrosive effects on various materials for decades. Thanks to XSEDE allocations on high-performance computers at SDSC and TACC, detailed models have now been simulated to provide new insight on how chloride leads to corrosion.

Conducted by a team from Oregon State University's (OSU) College of Engineering, a study discussing this newfound information was published in Materials Degradation, a Nature partner journal.

"Steels are the most widely used structural metals in the world, and their corrosion has severe economic, environmental, and social implications," said study co-author Burkan Isgor, an OSU civil and construction engineering professor. "Understanding the process of how protective passive films break down helps us custom design effective alloys and corrosion inhibitors that can increase the service life of structures that are exposed to chloride attacks."

Isgor worked closely with OSU School of Engineering colleague Líney Árnadóttir as well as graduate students Hossein DorMohammadi and Qin Pang on conducting the study. As a chemical engineering associate professor, Árnadóttir said her work often uses computational methods to study chemical processes on surfaces with applications in materials degradation.

This work was supported by the NSF, CMMI (1435417). Part of the calculations used the XSEDE (TG-ENG170002, TG-DMR160093), which is supported by NSF (ACI-1053575).

Surface structural changes to iron passive films caused by the adsorption of OH and/or Cl. (a) Fe(OH)3, (b) Fe(OH)2Cl, (c) Fe(OH)Cl2, d FeCl3. The location of the edge Fe atom prior to the adsorption is shown with an orange dashed line. Atoms types indicated by white (H), royal blue (Cl), large pink (Fe) and small red (O) spheres.

Credit: Oregon State University, College of Engineering.
In nature, sex can be really complicated. Earthworms are both sexes at once. When a queen bee mates, her offspring will all be female. If she doesn’t mate, they’ll all be male. And some female lizards can produce female offspring without mating.

Over 150 million years ago, when mammals arose, they “settled” on the familiar X and Y sex-chromosome arrangement. A mammalian egg with a single X chromosome merges with a sperm. If the sperm has a Y chromosome, the offspring will be genetically male; if it’s an X, female.

Roughly 1.2 million years ago, the creeping vole, also known as the Oregon meadow mouse, turned this regularity on its head. In 1963, the famous geneticist Susumo Ohno proposed that the rodent’s females were “XO,” having just a single X chromosome, while males were XY. But Ohno died before he could untangle what was really going on.

“This is a system that [Ohno] originally, really thought was incredibly interesting,” said Brian Couger of Brigham and Women’s Hospital and Harvard Medical School. “But he and his contemporaries were not able to finish [it].”

With collaborators at Brigham and Women’s, San Francisco State University and the University of California, Couger decided to complete Ohno’s work. He used the XSEDE-allocated Bridges and Bridges-2 advanced research computing platforms to accomplish this. What he found was even more surprising than Ohno had guessed – surprising enough to earn the team a cover article in the prestigious journal *Science*.

Many of the most popular laboratory DNA sequencing methods can only sequence roughly 75 to 400 DNA bases – the A, C, T, G letters of the genetic language – at a time. So sequencing a *genome* – a species’ entire DNA sequence – produces...
a huge mess of jumbled DNA snippets. A supercomputer, though, can put these millions of puzzle pieces together the right way in a reasonable amount of time. Particularly important in this kind of work is massive memory – the same as RAM on a personal computer, but by thousands of times more. That was a particular strength of the now-retired *Bridges*.

“This paper really required a lot of assemblies, which are really RAM-hungry for computational purposes,” Couger said. “So I did a lot of the work on *Bridges* … If we didn’t have this wonderful public resource to run it on, the work wouldn’t have been possible.”

The *Bridges* assembly showed that Ohno was correct that females were XO. But instead of an X and a Y chromosome in males, there were two strange types of X chromosome. The single X<sup>M</sup> chromosome the team found in both females and males looked mostly like a traditional X chromosome – except that it contains 10 genes only found in the Y chromosomes of other mammals. In males, the X<sup>M</sup> is paired with an X<sup>p</sup> chromosome that’s even stranger. Ninety percent of it looks like a “standard” mammalian X chromosome. The other 10% is more similar to a Y.

“Some features of this genome were so atypical that they were near-unbelievable,” Couger said. “However, other hallmarks of sex chromosomes in mammals [were] conserved even in this remarkable system.”

The collaborators followed up this startling discovery by confirming the result with a more definitive but harder-to-generate “long-read” sequence, using lab-derived snippets of DNA sequence of more than 15,000 bases. The new *Bridges*-2 platform’s 488 regular memory nodes were perfect for the job. The team’s work appeared on the cover of the May 6, 2021, issue of *Science*.

This work used XSEDE, supported by NSF grant number ACI-1548562. Specifically, it used the *Bridges*-2 system, supported by NSF award number ACI-1928147, at PSC. This work was also funded by NSF IOS 1558109 and start-up funds from the University of California, Riverside (P.C.), NSF MCB 1616878 (S.W.R.), and National Institutes of Health R44GM134994 (K.J.L.).
Over the past year, we’ve watched as the world has changed around us. For many in XSEDE, it was clear that, as an organization, we could do more to act on our diversity and inclusion goals, pillars of our organizational foundation. As a response, a diverse group of XSEDE staff members created, with support of project leadership, the XSEDE Terminology Task Force which seeks to reaffirm the organization’s commitment to building and sustaining a more diverse, inclusive and equitable research community, and also to serve as a resource for researchers and beyond.

XSEDE has always kept inclusion at the forefront of its mission, and will continue to do so, but in order to turn this from an idea into action, it became clear that a group was needed for XSEDE to actively establish and maintain inclusive standards, most notably via our own professional communications.

As a result, staff members from across the project have formed the XSEDE Terminology Task Force, a project-wide collaboration meant to address non-inclusive language and phrasing in XSEDE materials and recommend a suitable alternative replacement, for example, making sure to use non-gendered pronouns in project materials. Additionally, this list can be used as a community resource and leveraged by other advanced computing and research organizations.

“The social justice activity in the spring of 2020 brought to the fore that an organization like XSEDE, which touches a broad cross section of our community, was in a position to do something more – to act as an agent of change as opposed to just following those leading change,” said PI and Project Director John Towns.

This task force commits to reviewing the terminology used throughout XSEDE materials, both digital and physical, and assures that the language used by XSEDE meets our inclusivity standards and aligns with our Code of Conduct. This collaboration will require diligent cataloging and discussion of terminology use-cases, ongoing internal meetings, external discussions at conferences like PEARC and SC, and includes members from nearly every area of XSEDE.

While XSEDE has no control over external third-party documentation, it’s taking steps to effect change by contacting any relevant organizations; we hope this will be addressed by all third parties soon.

Through this thorough examination and continued vigilance, XSEDE hopes to foster a more inclusive and aware XSEDE and broader high-performance computing research community.

“Our inclusivity is our strength, and the Terminology Task Force has made, and will continue to make, concerted efforts to foster diversity throughout our collaboration and respective communities,” concluded Towns.

In the event that XSEDE has included inappropriate materials, verbal or written, please let the Terminology Task Force know at terminology@xsede.org.

You can read through the full Terminology Task Force database at: http://xsede.org/terminology