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 you 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 & 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 which 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

04 DIRECTOR’S MESSAGE
Letter from John Towns, XSEDE’s PI and Project Director.

06 SUPERCOMPUTING DRUG SCREENING FOR DEADLY HEART ARRHYTHMIAS
Computational pipeline tests for cardiotoxicity.

08 STAR CRASH
Artificial intelligence on XSEDE-allocated systems is key to speeding simulations of neutron star mergers.

10 SUPERCOMPUTING DYNAMIC EARTHQUAKE RUPTURE MODELS
XSEDE resources support multi-fault earthquake research.

12 STUDENT WINS POSTER AWARD AT ERN CONFERENCE IN STEM
XSEDE allocation assists undergraduate student researcher obtain award-winning scientific results.

14 XSEDE AIDS INTERNATIONAL ENGINEERING TEAM’S WAVE ENERGY PROJECT
Supercomputers validate novel simulation tool.

16 LOOKING A HOMINID IN THE MOUTH
Chimpanzee “oral microbiome” shows surprising frequency of species linked to human disease.

18 XSEDE ALLOCATIONS SHED NEW LIGHT ON HOW LASER PULSES PENETRATE DENSE MATTER
Large-scale simulations used to detect super strong magnetic fields.

20 FREEDOM, THEY PRINTED
AI on XSEDE-allocated system solves mystery of who printed seminal works on liberty.

22 CAN DEEP LEARNING YIELD MORE ACCURATE EXTREME WEATHER FORECASTS?
XSEDE systems support pattern recognition-based extreme weather prediction.

24 DISCOVERING NEW MATERIALS FOR NEXT GENERATION SOLAR CELLS AND LEDS
Large-scale quantum calculations, new computational methods are key to findings.

26 SUPERCOMPUTING FUTURE WIND POWER RISE
XSEDE-allocated resources simulate promises and pitfalls of wind power expansion.

28 A GALACTIC CHOICE
AI running on XSEDE-allocated systems surpasses humans at classifying galaxies.

XSEDE 2020 HIGHLIGHTS

THE XSEDE PARTNERSHIP

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
Welcome to the 8th edition of the Extreme Science and Engineering Discovery Environment (XSEDE) highlights publication, featuring a selection from the finest accomplishments of the year that we have been honored to help make possible. I continue to be amazed at the progress of our partners across the research community and am proud to share their successes with you.

In this edition, you’ll find discoveries covering a wide range of disciplines, including molecular biology, materials research, atmospheric sciences, humanities, physics, astronomical sciences and more. Regardless of discipline, underlying each discovery is a fundamental drive to understand and resolve some of the greatest challenges our society faces today. XSEDE is proud to be part of those efforts.

In a year unlike any other, we found ourselves called to serve in a new way. In the midst of the COVID-19 global pandemic, XSEDE has not only proactively worked to put resources, services, and support in the hands of those who can help us better understand how to fight SARS-CoV-2, but we have also responded to a call from our country asking us to play a key role in coordinating efforts well beyond our own scope.

On March 19, 2020, XSEDE was asked by the Office of Science and Technology Policy at the White House to support the efforts of a newly formed collaboration under the moniker of the COVID-19 HPC Consortium, which brings together federal agencies, academic institutions, and private companies across the nation—and subsequently internationally—to make available an unprecedented portfolio of resources to bear on addressing the COVID-19 pandemic. We quickly agreed, volunteering XRAS (XSEDE Resource Allocation Service) to serve as the platform for proposal submission and review for the Consortium. 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.

While this global experience has taken the headlines in nearly all we do, it cannot be forgotten that XSEDE continues to provide resources, services, and support to the national community as we have for nearly a decade. We are incredibly grateful to the National Science Foundation for their continued support, including the recent project supplement which will enable us to continue this meaningful work for an additional 12 months through August 31, 2022.

I remain impressed by the XSEDE team in that we continue to perform and excel at pursuing our vision and executing on our mission, even in light of making unexpected adjustments during these uncertain times. It is truly an honor to lead this team. I cannot adequately express the depth of my appreciation to every one of our staff and those with whom we collaborate at our partner institutions. It is through their constant dedication that we are able to make scientific progress a reality. I am, once again, humbled.

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

Sincerely yours,

John Towns
XSEDE PI and Project Director
HOW DID XSEDE HELP?

XSEDE changed the perspective of the researchers by helping them realize that they are not limited by their own resources—that they can use the top computational resources in the world to enhance their science, which in this case led to more accurate testing which in this case led to advance their science, and feel like you belong to a large community of other scientists, and you can use these resources to advance science, and feel like you belong to this large community of other scientists, for the greater good of the U.S. and the worldwide community.

He added that access to XSEDE and other supercomputing resources changed their perspective, saying that now they are not limited by their own resources.

“We can use the top computational resources in the world to do these calculations,” Vorobyov said. “It totally changes your perspective as a scientist. You’re not limited by your own resources.”

Vorobyov’s team also received support from XSEDE staff to help fine tune the code they used, a standard code called NAMD. “We got tremendous help, and it saved us many hours to be able to use this service. The code efficiency was increased by 50 percent or more sometimes, when we worked with XSEDE staff, who helped us to fine tune the code,” he said.

Death from sudden cardiac arrest makes headlines when it strikes athletes. But it also causes the most deaths by natural causes in the United States, estimated at 325,000 per year.

According to the Cleveland Clinic, the heart’s bioelectrical system goes haywire during arrest. The malfunction can send heartbeats racing out of control, cutting off blood to the body and brain. This differs from a heart attack, which is caused by a blockage of the heart’s arteries. The leading risk factors for sudden cardiac arrest are a past attack and the presence of disease. Another risk factor is the side effects from medications, which can potentially cause deadly arrhythmias.

Using supercomputers allocated by XSEDE, scientists have developed for the first time a way to screen drugs through their chemical structures for induced arrhythmias.

“What we set out to do was to try to solve the problem by building a computer-based pipeline for screening,” said Colleen Clancy, a professor in the Department of Physiology and Membrane Biology and the Department of Pharmacology at the UC Davis School of Medicine. “It’s on the atomic scale. We have around 130,000 atoms in our system.”

The calculations involved billions of individual time steps to achieve an all-atom simulation of several microseconds, enough to get detailed information on how the drug binds to the target.

Vorobyov was awarded allocations by XSEDE on the Stampede2 system of the Texas Advanced Computing Center and Comet at the San Diego Supercomputer Center, making use of Comet’s GPU and CPU nodes.

“Stampede2 offered a large array of powerful multi-core CPU nodes, which we were able to efficiently use for dozens of molecular dynamics runs we had to do in parallel. Such efficiency and scalability rivaled and even exceeded other resources we used for those simulations including even GPU-equipped nodes,” Vorobyov added.

Clancy and team also want to move the research into a more personalized medicine approach, building models of an individual’s cellular electrophysiology that include some genetic background. One project cultures heart muscle cells from individuals to develop a model called the ‘induced pluripotent stem cell derived cardiac myocyte.’

The study, “A Computational Pipeline to Predict Cardiotoxicity: From the Atom to the Rhythm” was published February 24, 2020, in the journal Circulation Research. The study was funded by the National Institutes of Health, the American Heart Association, and other agencies.

Death from sudden cardiac arrest makes headlines when it strikes athletes. But it also causes the most deaths by natural causes in the United States, estimated at 325,000 per year.

According to the Cleveland Clinic, the heart’s bioelectrical system goes haywire during arrest. The malfunction can send heartbeats racing out of control, cutting off blood to the body and brain. This differs from a heart attack, which is caused by a blockage of the heart’s arteries. The leading risk factors for sudden cardiac arrest are a past attack and the presence of disease. Another risk factor is the side effects from medications, which can potentially cause deadly arrhythmias.

Using supercomputers allocated by XSEDE, scientists have developed for the first time a way to screen drugs through their chemical structures for induced arrhythmias.

“What we set out to do was to try to solve the problem by building a computer-based pipeline for screening,” said Colleen Clancy, a professor in the Department of Physiology and Membrane Biology and the Department of Pharmacology at the UC Davis School of Medicine. “It’s on the atomic scale. We have around 130,000 atoms in our system.”

The calculations involved billions of individual time steps to achieve an all-atom simulation of several microseconds, enough to get detailed information on how the drug binds to the target.

Vorobyov was awarded allocations by XSEDE on the Stampede2 system of the Texas Advanced Computing Center and Comet at the San Diego Supercomputer Center, making use of Comet’s GPU and CPU nodes.

“Stampede2 offered a large array of powerful multi-core CPU nodes, which we were able to efficiently use for dozens of molecular dynamics runs we had to do in parallel. Such efficiency and scalability rivaled and even exceeded other resources we used for those simulations including even GPU-equipped nodes,” Vorobyov added.

Clancy and team also want to move the research into a more personalized medicine approach, building models of an individual’s cellular electrophysiology that include some genetic background. One project cultures heart muscle cells from individuals to develop a model called the ‘induced pluripotent stem cell derived cardiac myocyte.’

The study, “A Computational Pipeline to Predict Cardiotoxicity: From the Atom to the Rhythm” was published February 24, 2020, in the journal Circulation Research. The study was funded by the National Institutes of Health, the American Heart Association, and other agencies.

Death from sudden cardiac arrest makes headlines when it strikes athletes. But it also causes the most deaths by natural causes in the United States, estimated at 325,000 per year.

According to the Cleveland Clinic, the heart’s bioelectrical system goes haywire during arrest. The malfunction can send heartbeats racing out of control, cutting off blood to the body and brain. This differs from a heart attack, which is caused by a blockage of the heart’s arteries. The leading risk factors for sudden cardiac arrest are a past attack and the presence of disease. Another risk factor is the side effects from medications, which can potentially cause deadly arrhythmias.

Using supercomputers allocated by XSEDE, scientists have developed for the first time a way to screen drugs through their chemical structures for induced arrhythmias.

“What we set out to do was to try to solve the problem by building a computer-based pipeline for screening,” said Colleen Clancy, a professor in the Department of Physiology and Membrane Biology and the Department of Pharmacology at the UC Davis School of Medicine. “It’s on the atomic scale. We have around 130,000 atoms in our system.”

The calculations involved billions of individual time steps to achieve an all-atom simulation of several microseconds, enough to get detailed information on how the drug binds to the target.

Vorobyov was awarded allocations by XSEDE on the Stampede2 system of the Texas Advanced Computing Center and Comet at the San Diego Supercomputer Center, making use of Comet’s GPU and CPU nodes.

“Stampede2 offered a large array of powerful multi-core CPU nodes, which we were able to efficiently use for dozens of molecular dynamics runs we had to do in parallel. Such efficiency and scalability rivaled and even exceeded other resources we used for those simulations including even GPU-equipped nodes,” Vorobyov added.

Clancy and team also want to move the research into a more personalized medicine approach, building models of an individual’s cellular electrophysiology that include some genetic background. One project cultures heart muscle cells from individuals to develop a model called the ‘induced pluripotent stem cell derived cardiac myocyte.’

The study, “A Computational Pipeline to Predict Cardiotoxicity: From the Atom to the Rhythm” was published February 24, 2020, in the journal Circulation Research. The study was funded by the National Institutes of Health, the American Heart Association, and other agencies.

Death from sudden cardiac arrest makes headlines when it strikes athletes. But it also causes the most deaths by natural causes in the United States, estimated at 325,000 per year.

According to the Cleveland Clinic, the heart’s bioelectrical system goes haywire during arrest. The malfunction can send heartbeats racing out of control, cutting off blood to the body and brain. This differs from a heart attack, which is caused by a blockage of the heart’s arteries. The leading risk factors for sudden cardiac arrest are a past attack and the presence of disease. Another risk factor is the side effects from medications, which can potentially cause deadly arrhythmias.

Using supercomputers allocated by XSEDE, scientists have developed for the first time a way to screen drugs through their chemical structures for induced arrhythmias.

“What we set out to do was to try to solve the problem by building a computer-based pipeline for screening,” said Colleen Clancy, a professor in the Department of Physiology and Membrane Biology and the Department of Pharmacology at the UC Davis School of Medicine. “It’s on the atomic scale. We have around 130,000 atoms in our system.”

The calculations involved billions of individual time steps to achieve an all-atom simulation of several microseconds, enough to get detailed information on how the drug binds to the target.

Vorobyov was awarded allocations by XSEDE on the Stampede2 system of the Texas Advanced Computing Center and Comet at the San Diego Supercomputer Center, making use of Comet’s GPU and CPU nodes.

“Stampede2 offered a large array of powerful multi-core CPU nodes, which we were able to efficiently use for dozens of molecular dynamics runs we had to do in parallel. Such efficiency and scalability rivaled and even exceeded other resources we used for those simulations including even GPU-equipped nodes,” Vorobyov added.

Clancy and team also want to move the research into a more personalized medicine approach, building models of an individual’s cellular electrophysiology that include some genetic background. One project cultures heart muscle cells from individuals to develop a model called the ‘induced pluripotent stem cell derived cardiac myocyte.’

The study, “A Computational Pipeline to Predict Cardiotoxicity: From the Atom to the Rhythm” was published February 24, 2020, in the journal Circulation Research. The study was funded by the National Institutes of Health, the American Heart Association, and other agencies.
STAR CRASH

Artificial intelligence on XSEDE-allocated systems is key to speeding simulations of neutron star mergers.

Deep learning starts with training. The AI analyzes data in which the “right answers” have been labeled. It then extracts features from the data that humans might not have recognized but which predict the correct answers. Next, the AI is tested on data without the right answers labeled.

Rosofsky designed his deep-learning AI to progress in steps. This allowed him to verify the results at each step and understand how the AI was obtaining its predictions. This is important in deep learning computation, which otherwise might not fully understand and so can’t fully trust.

Rosofsky used the XSEDE-allocated Stampede2 supercomputer at the Texas Advanced Computing Center (TACC) to produce the data to train, validate, and test his neural network models. For the training and testing phases of the project, he chose the NVIDIA Tesla P100 GPUs of the XSEDE-allocated Bridges platform at the Pittsburgh Supercomputing Center (PSC), the most advanced GPUs available at the time. He was able to obtain a correction factor for the lower resolution simulations much more accurately than with the alternatives. The ability of AI to simplify the simulation should allow the scientists to perform a large simulation in months rather than years. The NCSA team reported their results in the journal Physical Review D in April 2020. The AI computations on Bridges showed that the method would work better and faster than gradient models. They also present a roadmap for other researchers to use AI to speed other massive computations.

“What we’re doing here is not just pushing the boundaries of AI,” Huerta says. “We’re providing a way for other users to optimally utilize their resources.” Future work by the group may include the even more advanced V100 GPU nodes of the XSEDE-allocated Bridges-Alt system, or the upcoming Bridges-2 platform, at PSC. Their next step will be to incorporate the AI’s correction factors into large-scale simulations of neutron-star mergers and further assess the accuracy of the AI and of the quicker simulations. Their hope is that the new simulations will demonstrate details in neutron-star mergers that can be identified in gravitational wave detectors. These could allow observatories to detect more events, as well as explain more about how these massive and strange cosmic events unfold.

Bizarre objects the size of a city but with more mass than our Sun, neutron stars spew magnetic fields a hundred thousand times stronger than an MRI medical scanner. A teaspoon of neutron-star matter weighs about a billion tons. When these cosmic bodies smash together it is dramatic.

“We don’t know the nature of matter when it is super-compressed,” says Shawn Rosofsky, a graduate student at the University of Illinois at Urbana-Champaign’s National Center for Supercomputing Applications (NCSA). “Neutron stars are ideal laboratories to get insights into this state of matter. [Other than black holes] they are the most compact objects in the Universe.”

Scientists have directly detected two neutron star mergers to date. These detections depended on two gravitational-wave detector observatories, the Laser Interferometer Gravitational-Wave Observatory (LIGO) and Virgo. LIGO consists of two detectors, one in Hanford, WA, and the other in Livingston, LA. The European Virgo detector is in Cascina, near Pisa, Italy.

Higher-quality computer simulations of neutron star mergers would allow scientists to identify what they should be looking for to better recognize and understand these events. But these simulations are slow and computationally expensive.

Bizarre objects the size of a city but with more mass than our Sun, neutron stars spew magnetic fields a hundred thousand times stronger than an MRI medical scanner. A teaspoon of neutron-star matter weighs about a billion tons. When these cosmic bodies smash together it is dramatic.

“We don’t know the nature of matter when it is super-compressed,” says Shawn Rosofsky, a graduate student at the University of Illinois at Urbana-Champaign’s National Center for Supercomputing Applications (NCSA). “Neutron stars are ideal laboratories to get insights into this state of matter. [Other than black holes] they are the most compact objects in the Universe.”

Scientists have directly detected two neutron star mergers to date. These detections depended on two gravitational-wave detector observatories, the Laser Interferometer Gravitational-Wave Observatory (LIGO) and Virgo. LIGO consists of two detectors, one in Hanford, WA, and the other in Livingston, LA. The European Virgo detector is in Cascina, near Pisa, Italy.

Higher-quality computer simulations of neutron star mergers would allow scientists to identify what they should be looking for to better recognize and understand these events. But these simulations are slow and computationally expensive.

Bizarre objects the size of a city but with more mass than our Sun, neutron stars spew magnetic fields a hundred thousand times stronger than an MRI medical scanner. A teaspoon of neutron-star matter weighs about a billion tons. When these cosmic bodies smash together it is dramatic.

“We don’t know the nature of matter when it is super-compressed,” says Shawn Rosofsky, a graduate student at the University of Illinois at Urbana-Champaign’s National Center for Supercomputing Applications (NCSA). “Neutron stars are ideal laboratories to get insights into this state of matter. [Other than black holes] they are the most compact objects in the Universe.”

Scientists have directly detected two neutron star mergers to date. These detections depended on two gravitational-wave detector observatories, the Laser Interferometer Gravitational-Wave Observatory (LIGO) and Virgo. LIGO consists of two detectors, one in Hanford, WA, and the other in Livingston, LA. The European Virgo detector is in Cascina, near Pisa, Italy.

Higher-quality computer simulations of neutron star mergers would allow scientists to identify what they should be looking for to better recognize and understand these events. But these simulations are slow and computationally expensive.

Bizarre objects the size of a city but with more mass than our Sun, neutron stars spew magnetic fields a hundred thousand times stronger than an MRI medical scanner. A teaspoon of neutron-star matter weighs about a billion tons. When these cosmic bodies smash together it is dramatic.

“We don’t know the nature of matter when it is super-compressed,” says Shawn Rosofsky, a graduate student at the University of Illinois at Urbana-Champaign’s National Center for Supercomputing Applications (NCSA). “Neutron stars are ideal laboratories to get insights into this state of matter. [Other than black holes] they are the most compact objects in the Universe.”

Scientists have directly detected two neutron star mergers to date. These detections depended on two gravitational-wave detector observatories, the Laser Interferometer Gravitational-Wave Observatory (LIGO) and Virgo. LIGO consists of two detectors, one in Hanford, WA, and the other in Livingston, LA. The European Virgo detector is in Cascina, near Pisa, Italy.

Higher-quality computer simulations of neutron star mergers would allow scientists to identify what they should be looking for to better recognize and understand these events. But these simulations are slow and computationally expensive.

Bizarre objects the size of a city but with more mass than our Sun, neutron stars spew magnetic fields a hundred thousand times stronger than an MRI medical scanner. A teaspoon of neutron-star matter weighs about a billion tons. When these cosmic bodies smash together it is dramatic.

“We don’t know the nature of matter when it is super-compressed,” says Shawn Rosofsky, a graduate student at the University of Illinois at Urbana-Champaign’s National Center for Supercomputing Applications (NCSA). “Neutron stars are ideal laboratories to get insights into this state of matter. [Other than black holes] they are the most compact objects in the Universe.”

Scientists have directly detected two neutron star mergers to date. These detections depended on two gravitational-wave detector observatories, the Laser Interferometer Gravitational-Wave Observatory (LIGO) and Virgo. LIGO consists of two detectors, one in Hanford, WA, and the other in Livingston, LA. The European Virgo detector is in Cascina, near Pisa, Italy.

Higher-quality computer simulations of neutron star mergers would allow scientists to identify what they should be looking for to better recognize and understand these events. But these simulations are slow and computationally expensive.
XSEDE resources support multi-fault earthquake rupture models

Some of the world’s most powerful earthquakes involve multiple faults, and scientists are using supercomputers to better predict their behavior. Multi-fault earthquakes can span fault systems of tens to hundreds of kilometers, with ruptures propagating from one segment to the other. During the last decade, scientists have observed several cases of this complicated type of earthquake. Major examples include the M7.5 El Mayor -Cucapah earthquake in Mexico; the M8.6 Indian Ocean Earthquake; and perhaps the most complex of all, the M7.8 2015 Kaikoura earthquake in New Zealand.

The main findings of our work concern the dynamic interactions of a postulated network of faults in the Brawley seismic zone in Southern California,” said Christodoulos Kyriakopoulos, a research geophysicist at the University of California, Riverside.

He’s the lead author of a study published in April of 2019 in the Journal of Geophysical Research: Solid Earth, published by the American Geophysical Union.

“We used physics-based dynamic models to simulate complex earthquake ruptures using supercomputers,” Kyriakopoulos said. “We ran dozens of numerical simulations, and documented a large number of interactions that we analyzed using advanced visualization software.”

Kyriakopoulos and his colleagues used XSEDE-allocated resources Stampede2 and Stampede1 at the Texas Advanced Computing Center (TACC) and Comet at the San Diego Supercomputer Center (SDSC).

Modeling realistic earthquakes on a computer isn’t easy. Kyriakopoulos and his collaborators faced three main challenges. The first was the implementation of these faults in the finite element domain in the numerical model. The second challenge was to run dozens of large computational simulations. “Having available time on Stampede2 was a game-changer for me and my colleagues because it allowed us to set the right conditions for the entire set of simulations,” he said.

The third challenge was to use optimal tools to properly visualize the 3D simulation results, which in their raw form consist simply of huge arrays of numbers. Their team also used Comet at SDSC in this research, mostly for test runs and prototyping. “The efficiency of the SDSC support team kept my optimism high and helped me think positively for the future of my project. This is important for an ongoing investigation, especially in the first stages where you are making sure that your models work properly,” Kyriakopoulos said.

The XSEDE support also helped Kyriakopoulos optimize his computational work and better organize the scheduling of computer runs. “I saved as much as 20% of personal time because of the way XSEDE is organized,” he said.

The NSF Earth Sciences Program Director Eva Zanzerkia said: “This research has provided us with a new understanding of a complex set of faults in Southern California that have the potential to impact the lives of millions of people in the United States and Mexico. Ambitious computational approaches, such as those undertaken by this research team in collaboration with XSEDE, make more realistic physics-based earthquake models possible.”

Kyriakopoulos concluded: “My participation in XSEDE gave a significant boost in my modeling activities and allowed me to explore the parameter space of my problem. I feel part of a big community that uses supercomputers and has a common goal to push forward science and produce innovation.”

The study, “Dynamic Rupture Scenarios in the Brawley Seismic Zone, Salton Trough, Southern California,” was published in April of 2019 in the Journal of Geophysical Research: Solid Earth. Funding was provided by the U.S. Geological Survey, the National Science Foundation, and the Southern California Earthquake Center.

Supercomputing Dynamic Earthquake Rupture Models

Some of the world’s most powerful earthquakes involve multiple faults, and scientists are using supercomputers to better predict their behavior. Multi-fault earthquakes can span fault systems of tens to hundreds of kilometers, with ruptures propagating from one segment to the other. During the last decade, scientists have observed several cases of this complicated type of earthquake. Major examples include the M7.5 El Mayor -Cucapah earthquake in Mexico; the M8.6 Indian Ocean Earthquake; and perhaps the most complex of all, the M7.8 2015 Kaikoura earthquake in New Zealand.

The main findings of our work concern the dynamic interactions of a postulated network of faults in the Brawley seismic zone in Southern California,” said Christodoulos Kyriakopoulos, a research geophysicist at the University of California, Riverside. He’s the lead author of a study published in April of 2019 in the Journal of Geophysical Research: Solid Earth, published by the American Geophysical Union.

“We used physics-based dynamic models to simulate complex earthquake ruptures using supercomputers,” Kyriakopoulos said. “We ran dozens of numerical simulations, and documented a large number of interactions that we analyzed using advanced visualization software.”

Kyriakopoulos and his colleagues used XSEDE-allocated resources Stampede2 and Stampede1 at the Texas Advanced Computing Center (TACC) and Comet at the San Diego Supercomputer Center (SDSC).

Modeling realistic earthquakes on a computer isn’t easy. Kyriakopoulos and his collaborators faced three main challenges. The first was the implementation of these faults in the finite element domain in the numerical model. The second challenge was to run dozens of large computational simulations. “Having available time on Stampede2 was a game-changer for me and my colleagues because it allowed us to set the right conditions for the entire set of simulations,” he said.

The third challenge was to use optimal tools to properly visualize the 3D simulation results, which in their raw form consist simply of huge arrays of numbers. Their team also used Comet at SDSC in this research, mostly for test runs and prototyping. “The efficiency of the SDSC support team kept my optimism high and helped me think positively for the future of my project. This is important for an ongoing investigation, especially in the first stages where you are making sure that your models work properly,” Kyriakopoulos said.

The XSEDE support also helped Kyriakopoulos optimize his computational work and better organize the scheduling of computer runs. “I saved as much as 20% of personal time because of the way XSEDE is organized,” he said.

NSF Earth Sciences Program Director Eva Zanzerkia said: “This research has provided us with a new understanding of a complex set of faults in Southern California that have the potential to impact the lives of millions of people in the United States and Mexico. Ambitious computational approaches, such as those undertaken by this research team in collaboration with XSEDE, make more realistic physics-based earthquake models possible.”

Kyriakopoulos concluded: “My participation in XSEDE gave a significant boost in my modeling activities and allowed me to explore the parameter space of my problem. I feel part of a big community that uses supercomputers and has a common goal to push forward science and produce innovation.”

The study, “Dynamic Rupture Scenarios in the Brawley Seismic Zone, Salton Trough, Southern California,” was published in April of 2019 in the Journal of Geophysical Research: Solid Earth. Funding was provided by the U.S. Geological Survey, the National Science Foundation, and the Southern California Earthquake Center.
XSEDE allocation assists undergraduate student researcher obtain award-winning scientific results.

Tenacious Underwood, a senior at Livingstone College in North Carolina, has already had quite a year. In early 2019, Underwood, who always had an affinity for math and computers, was encouraged by a fraternity brother he met at the National Science Foundation’s Emerging Researchers National Conference in STEM to apply for the Jetstream REU program. This program is hosted at Indiana University (IU) and is made possible, in part, through XSEDE Education Allocations.

Each summer, Jetstream—the NSF’s first production cloud computing system which provides on-demand HPC and data analysis resources for research and education co-located at IU and the Texas Advanced Computing Center (TACC)—sets students free in the cloud on projects that capitalize on IU’s leadership in fields like bioinformatics, data visualization, and advanced media. Underwood, Evan Suggs (University of Tennessee, Chattanooga), and Eliza Foran (Indiana University, Bloomington) created a workflow to record and identify animal vocalizations, easing the burden of data collection on field biologists.

Specifically, they wanted to identify amphibians local to Indiana. They gathered sample calls for four species from the Macaulay Library archive of wildlife sounds at Cornell and used 85 percent of the data to train three neural networks that could process data collected in the field. In essence, Underwood and his colleagues used machine learning and HPC to solve the problem of identifying frog migration patterns.

The team outlined a proof-of-concept workflow that makes the entire process from gathering to interpreting data, more attainable for researchers.

They simulated the data collection process by collecting animal (frog) calls using recording devices and Raspberry Pi’s (low-cost, fully functional computers the size of credit cards). The team then fed this data into a database and virtual machine hosted on TACC-allocated XSEDE resources (i.e. Jetstream and Wrangler). Processing the more than 5,000 audio files that were recorded would have taken up to 20 hours on a laptop computer, but using Jetstream, it took minutes.

Once the research was done, Underwood’s journey took off. He, Suggs, and Foran presented their work at PEARC19 in Chicago. Then, in November, Underwood presented their work at SC19 in Denver to great acclaim. Most recently, he presented Automatic Recognition of Frog Calls at the 2020 ERN conference and won first place in the Computational Computer Sciences and Information Management Division of undergraduate student posters category.

Underwood is now preparing to graduate from Livingstone College in just a few months. What’s next? Graduate study in computer science at Kentucky State University with a specialization in cybersecurity. Underwood’s bright outlook is a credit to his accomplishments.
Supercomputers validate novel simulation tool.

Similar to solar panels and wind turbines, wave energy converters harness energy from Mother Nature’s resources—specifically ocean waves—and turn them into electricity. While this process has not yet been perfected, it has the potential to reduce our reliance on fossil fuels.

Although several challenges, such as operating complex devices in harsh ocean conditions, must be overcome before wave energy becomes a realistic option for mainstream power, one step toward a solution was recently studied and validated by an international team of engineering researchers with support from XSEDE.

The team, which published findings related to their wave energy simulation tool in the December 2019 issue of *Journal of Computational Physics*, validated their results using the Comet supercomputer at the San Diego Supercomputer Center (SDSC) and Bridges at the Pittsburgh Supercomputing Center (PSC). Focused on modeling complex fluid-structure interaction problems, the team’s simulations included ones of heavy rigid structures interacting with high winds, breaking waves, and other often-severe marine characteristics.

While this multifaceted work required hundreds of processors and multiple days of runtime, the use of Comet and Bridges enabled the research team to demonstrate the parallel scalability of their simulation technique as well as validate various three-dimensional, large-scale test cases.

“We primarily used our simulation techniques to investigate inertial sea wave energy converters, which are renewable energy devices developed by our collaborators at the Polytechnic University of Turin that convert wave energy from large bodies of water into electrical energy,” said study co-author Amneet Pal Bhalla, an assistant professor of mechanical engineering at San Diego State University (SDSU). “There’s a need for these devices to perform effectively year-round using optimal control strategies, and our simulations could be an initial step in making this happen.”

Because wave energy converter engineering is still in the research and development phase, such simulations allowed the scientists to inexpensively investigate their designs before building full-scale prototypes, noted Bhalla.

The researchers encountered some large-scale modeling challenges while conducting their study. For example, they had access to SDSU College of Engineering’s cluster called Fermi but it was not suitable to evaluate the parallel scalability of the code and tune its performance for modern computer architectures. The XSEDE-allocated clusters offered a large number of homogeneous nodes with a more modern architecture, along with specialized code analytics and performance tuning software such as TAU, DDT, Valgrind, and Massif.

“The XSEDE allocations also allowed us to detect hotspots such as unnecessary Message Passing Interface (MPI) barriers, check for memory leaks, and monitor call instructions in routines,” said Bhalla. “These tools helped us to profile and optimize our code, which uses several high-performance libraries such as OpenMPI, PETSc, and HDF5. These resources saved us time, while our students also became proficient high-performance computing developers by attending various XSEDE training seminars.”

The research was also supported by the NSF’s Graduate Research Fellowship Program and the NSF’s SI2 program.
Chimpanzee “oral microbiome” shows surprising frequency of species linked to human disease.

Bacteria that cause gum and other dental disease have been linked to heart disease, pregnancy and birth complications, and pneumonia. Scientists would like to better understand how the collective bacteria in our mouths—the “oral microbiome”—changed as humans diverged from the great apes, and then from ancient human relatives like the Neanderthals. A better understanding of the ecology of the mouth could help identify which bacteria cause disease and which are just bystanders.

“Differences in bacteria in the oral cavity affect health states in different populations,” says Andrew Ozga, an assistant professor at the Halmos College of Natural Sciences and Oceanography, Nova Southeastern University. “We’re still trying to explore and understand what species are there, especially in humans. A few bacteria are known to associate with (oral) disease states; we wanted to see whether these types of bacteria are present in our closest evolutionary relatives and ancestors.”

Then at Arizona State University, Ozga and colleagues there and at the University of Minnesota and Duke University decided to study the bacterial DNA of the oral microbiome. They collected DNA from the teeth of modern humans, ancient chimpanzees, gorillas and orangutans. The scientists’ hope is to gain enough insight into the exact mix of bacteria associates with disease and the oral microbiome in the chimpanzees. Also, they’d like to include our closest evolutionary relatives and ancestors. "Differences in bacteria in the oral cavity affect health states in different populations," says Andrew Ozga, an assistant professor at the Halmos College of Natural Sciences and Oceanography, Nova Southeastern University. "We’re still trying to explore and understand what species are there, especially in humans. A few bacteria are known to associate with (oral) disease states; we wanted to see whether these types of bacteria are present in our closest evolutionary relatives and ancestors."

How Did XSEDE Help?

Scientists used data-intensive DNA sequencing on an XSEDE-allocated supercomputer to identify the bacteria of the oral microbiome in chimpanzees, humans, and Neanderthals. They discovered that recently rare bacteria associated with disease in human mouths are common in our close evolutionary relatives.

The Arizona State-led team collected samples of calculus—the hardened coating on the teeth that results from calcium reacting with dental plaque—from 19 Gombe chimpanzees that had died over five decades at Gombe National Park in Tanzania. Calculus doesn’t contain live bacteria, but the scientists were able to isolate DNA from dead bacteria in these samples. They then compared these to samples from 41 ancient and more recent humans, four Neanderthals, and one “historical” chimpanzee from outside the Gombe.

Each sample generated the DNA sequences of millions of genetic fragments. Some of the DNA belonged to the animal being studied. But much of it came from fungi, viruses and, mostly, bacteria. To sort it all out, researchers used a technique called MAL T (MEGAN ALignment Tool). MAL T requires massive computer power. So they turned to XSEDE.

"We’re creating very large datasets … we need a lot of computing power to compare the samples with all the known databases," Ozga says. "XSEDE has basically allowed me to do work that I could not have done otherwise."

The DNA assemblies offered expected results and some surprises. Chimpanzees had different oral microbiomes than humans. The Neanderthal oral microbiome was distinct from most of the humans, but was closer to ours than to chimpanzees’, overlapping to some extent with ours.

One of the two most common bacteria in the chimpanzee oral microbiome was Porphyromonas. These bacteria are associated with oral disease in humans, but are much less abundant in our mouths. It isn’t yet clear whether the bacterium causes disease in chimpanzees or is part of a healthy chimp microbiome. The Arizona and Florida team published their findings in the November 2019 issue of Nature Scientific Reports.

The next step in the research will include a deeper dive into how and whether the exact mix of bacteria associates with disease in both humans and chimps, including captive chimpanzee populations. The scientists would also like to study how age affects both dental disease and the oral microbiome in the chimpanzees. Also, they’d like to include more distant human relatives such as gorillas and orangutans. The scientists’ hope is to gain enough insight into how bacterial populations and disease states evolved. Ultimately, the goal is to identify which bacteria are causing disease, which prevent it, and which are only bystanders.
Large-scale simulations used to detect super strong magnetic fields.

While intense magnetic fields are naturally generated by neutron stars, researchers have been striving for many years to achieve similar results in their laboratories. Tao Wang, a University of California San Diego mechanical and aerospace engineering graduate student, recently demonstrated how an extremely strong magnetic field, similar to that on the surface of a neutron star, can be not only generated but detected using an x-ray laser inside a solid material.

Wang carried out his research with the help of simulations allocated by XSEDE on the Comet supercomputer at the San Diego Supercomputer Center (SDSC) and the Stampede1 and Stampede2 systems at the Texas Advanced Computing Center (TACC).

“Tao Wang’s findings were critical to our recently published study’s overall goal of developing a fundamental understanding of how multiple laser beams of extreme intensity interact with matter,” said Alexey Arefiev, a professor of mechanical and aerospace engineering at the UC San Diego Jacobs School of Engineering.

Wang, Arefiev, and their colleagues used multiple large three-dimensional simulations, remote visualization, and data post-processing to complete their study, which showed how an intense laser pulse is able to propagate into the dense material because of the laser’s relativistic intensity. Their findings were published in a January 2019 Physics of Plasmas journal article.

In other words, as the velocity of the electrons in the laser approaches the speed of light, their mass becomes so heavy that the target becomes transparent, according to the researchers. Because of this transparency, the laser pulse pushes the electrons to form a strong magnetic field. This strength is comparable to that on a neutron star’s surface—which is at least 100 million times stronger than the Earth’s magnetic field.

“Now that we have completed this study, we are working on ways to detect this type of magnetic field at a one-of-a-kind facility called the European X-Ray Free Electron Laser (XFEL), which encompasses a 3.4 kilometer (2.2 miles)-long accelerator that generates extremely intense x-ray flashes to be used by researchers like our team,” explained Arefiev.

Located in Schenefeld, Germany, the European XFEL is home to Toma Toncian, where he leads the project group construction and commissioning of the Helmholtz International Beamline for Extreme Fields at the High Energy Density instrument. He is also a co-author on the recently published study.

“The very fruitful collaboration between UC San Diego and Helmholtz-Zentrum Dresden-Rossendorf is paving the road to future high-impact experiments,” said Toncian. “As we pass from construction to commissioning and first experiments, the theoretical predictions by Tao Wang are timely and show us how to further develop and fully exploit the capabilities of our instrument.”

“Because the resulting data sets of our experiments using XFEL are very large, our research would not have been possible on a regular desktop, and we could not have completed this study without the use of XSEDE supercomputers,” said Arefiev.

The research was supported by the Air Force Office of Scientific Research and the National Science Foundation. Particle-in-cell simulations were performed using EPOCH and developed under the UK’s Engineering and Physical Sciences Research Council. Mingsheng Wei, a senior scientist at the University of Rochester’s Laboratory for Laser Energetics, is also a co-author of the published study.

> HOW DID XSEDE HELP?

XSEDE provided the research team with allocations on supercomputing systems that could handle the large-scale data sets as used by the European X-Ray Free Electron Laser (XFEL) facility in Germany, helping to blaze the trail for high-impact physics experiments.
AI on XSEDE-allocated system solves mystery of who printed seminal works on liberty.

Prior to the 18th century in England expressing your ideas on politics, religion, even divorce—anything the country’s leaders found threatening—could get you arrested or possibly killed.

The 17th century English Civil War proved a boon to free speech. A total wave of forbidden publications flooded England. It was still dangerous to write and print such books. The vital printers—without whom circulation of the books would have been far more limited—were mostly anonymous. Historians don’t know the printers of an astonishing 25 percent of English-language books in the 17th century.

One of the most important of these books was John Milton's Areopagitica, which was ideal for their approach but had not been written for supercomputers.

Their automated approach took advantage of the decidedly non-automated nature of 17th-century printing technology. To produce a book, printers would set lead type—one piece for each letter—in a wooden rack, backward. They smeared the type with ink, pressed a piece of paper onto it and then assembled the pages into a book. The ink transferred right-way letters onto the paper. But type pieces were imperfectly cast and suffered damage from use. Some developed tiny irregularities that would show up every time the piece type was used.

What the team wanted to do was to create a more sophisticated artificial intelligence at a massive scale to analyze the print of 17th century publications.

To solve the mystery, the team wanted to expand an approach previously done by the human eye. Their artificial-intelligence (AI) approach would require an ability to store and move data fluidly between many compute nodes. Having worked with XSEDE previously under a number of allocations, researchers were able to use image recognition software artificial intelligence at a massive scale to analyze the print of 17th century publications.

The solution to the mystery was worthy of an Agatha Christie novel—specifically, Murder on the Orient Express. Historians had long suspected that printer Matthew Simmons had been involved in publishing Areopagitica. He was known to print forbidden books and had printed Milton’s non-forbidden publications. Type ML analysis showed that type in Areopagitica matched books known to have been Simmons’. More surprising, the type also matched works by Simmons’ former partner, Thomas Paine. The researchers don’t know what this means; historians had thought the two had broken their partnership by late 1644 and had never worked together after.

The type also matched works printed previously by Gregory Dexter, who had been shut down in a government raid early in 1644. Soon after, Dexter left England for the colony that was to become the U.S. State of Rhode Island. The researchers don’t know how Dexter’s type pieces appear in Areopagitica. But, they wondered whether they could use Ocular to compare the type in the first edition of Areopagitica and several other forbidden books of the period with the type in about 100 books whose printers were known.

They used a machine learning (ML) approach to recognize text in old printed documents. The type is irregular, and the ML model is trained to recognize characters across books by the same publishers. Their specific method was a custom generative probability model of the printing press, in which the AI simultaneously made inferences about the parameters of the printing process and the actual text that was printed.

The researchers applied this model to type from both known and unknown volumes. They'd also like to make the analysis no longer need enhancement with human expertise. This will involve 10,000 books covering anonymous books and every known printer in a decade in the 1600s. Such work will involve the deep-learning-specialized Bridges-AI and the future Bridges-2 system.

The researchers are investigating using deep learning, in which multiple layers of inference are used to create a more sophisticated artificial intelligence, to tackle this problem. They’d also like to make the analysis no longer need enhancement with human expertise. This will involve 10,000 books covering anonymous books and every known printer in a decade in the 1600s. Such work will involve the deep-learning-specialized Bridges-AI and the future Bridges-2 system, each of which contains many coupled AI-optimized graphics processing units (GPUs) for large-scale deep learning.

In addition to XSEDE, research for this publication was supported by an A. W. Mellon Digital Humanities seed grant from Carnegie Mellon University, a resource allocation from the Pittsburgh Supercomputing Center, and a grant from the National Science Foundation.

When Dexter’s business was liquidated. As Warren and his coauthors write in the Spring 2020 edition of the journal Milton Studies, their analysis "raises nearly as many questions as it answers."

The next step will be to expand the analysis. The Milton Studies paper identified Simmons and Paine as the printer of eight other books on civil liberties, including, by Roger Williams, the founder of Rhode Island. But hundreds still lack an identified publisher. The team is investigating using deep learning, in which multiple layers of inference are used to create a more sophisticated artificial intelligence, to tackle this problem. They’d also like to make the analysis no longer need enhancement with human expertise. This will involve 10,000 books covering anonymous books and every known printer in a decade in the 1600s. Such work will involve the deep-learning-specialized Bridges-AI and the future Bridges-2 system, each of which contains many coupled AI-optimized graphics processing units (GPUs) for large-scale deep learning.

In addition to XSEDE, research for this publication was supported by an A. W. Mellon Digital Humanities seed grant from Carnegie Mellon University, a resource allocation from the Pittsburgh Supercomputing Center, and a grant from the National Science Foundation.

"Having worked with PSC and XSEDE before on a prior digital humanities project, I knew that the infrastructure and support were well-suited for a project like this one. Working with XSEDE on [our prior project] Six Degrees of Francis Bacon, I experienced XSEDE’s strong commitment to supporting projects coming from researchers like historians and literary scholars who are relatively new to supercomputing. XSEDE GPU resources, storage, and Extended Collaborative Support Services give us the capacity to scale up our project to try to identify printers for roughly 8,000 anonymously printed books from the late 17th-century era of John Locke and Isaac Newton."

The next step will be to expand the analysis. The Milton Studies paper identified Simmons and Paine as the printer of eight other books on civil liberties, including, by Roger Williams, the founder of Rhode Island. But hundreds still lack an identified publisher. The team is investigating using deep learning, in which multiple layers of inference are used to create a more sophisticated artificial intelligence, to tackle this problem. They’d also like to make the analysis no longer need enhancement with human expertise. This will involve 10,000 books covering anonymous books and every known printer in a decade in the 1600s. Such work will involve the deep-learning-specialized Bridges-AI and the future Bridges-2 system, each of which contains many coupled AI-optimized graphics processing units (GPUs) for large-scale deep learning.

In addition to XSEDE, research for this publication was supported by an A. W. Mellon Digital Humanities seed grant from Carnegie Mellon University, a resource allocation from the Pittsburgh Supercomputing Center, and a grant from the National Science Foundation.
CAN DEEP LEARNING YIELD MORE ACCURATE EXTREME WEATHER FORECASTS?

XSEDE systems support pattern recognition-based extreme weather prediction.

**HOW DID XSEDE HELP?**

Using XSEDE allocations, researchers from Rice University were able to combine vast amounts of atmospheric data to train more accurate weather forecast models on local computational grids.

Forecasting weather patterns that cause extreme weather is still challenging despite decades of advances in numerical weather prediction (NWP). Modern forecasts use mathematical models of the atmosphere and oceans to predict the weather based on current weather conditions. Even with the increased power of today’s supercomputers, the forecasting skill of numerical weather models extends only about six days, varying on location, season, and weather pattern.

Weather patterns that are often the drivers of extreme events are particularly hard to forecast. Improving the forecast of such events using NWP requires running more simulations starting from similar weather conditions and using higher-resolution models, which demand enormous computational resources. The former is needed to tackle the chaotic nature of the atmosphere, i.e., the butterfly effect.

Pedram Hassanzadeh, an assistant professor in Mechanical Engineering and Earth, Environmental and Planetary Sciences at Rice University, and his Ph.D. students Ashesh Chattopadhyay and Ebraheem Nabizadeh, recently introduced a data-driven framework that formulates extreme weather prediction as a pattern recognition problem and employs state-of-the-art deep learning techniques.

The researchers analyzed large data sets and employed machine learning codes on supercomputers at the Texas Advanced Computing Center (TACC) and the Pittsburgh Supercomputing Center (PSC) with allocations through XSEDE. In addition, data that had already been produced by supercomputers at the National Center for Atmospheric Research (NCAR) was used as input for the deep learning models.

“Our work would not have been possible without XSEDE’s computing resources,” Hassanzadeh said. “Stampede2 is the main supercomputing resource that my group uses, and Bridges enables us to efficiently work with very large datasets.”

Deep learning is a form of artificial intelligence, in which computers are trained to make humanlike decisions without being explicitly programmed for them. The mainstream of deep learning, the convolutional neural network, excels at pattern recognition and is the key technology for self-driving cars, facial recognition, speech transcription, and dozens of other advances.

Once trained on observational and/or high-resolution numerical model data, data-driven framework can provide relatively accurate predictions at little computational cost, which can guide other NWP efforts by providing early warnings.

“Generally, the numerical weather models do a good job predicting weather, but they still have some difficulties with extreme weather,” Hassanzadeh said. “We’re trying to do extreme weather prediction in a very different way.”

As a proof-of-concept demonstration, Hassanzadeh and team predicted heat waves and cold spells over North America using limited information about the atmospheric circulation at an altitude of around five kilometers, and in some cases, the surface temperature a few days earlier. The results of their demonstration suggest that extreme weather prediction can be done as a pattern recognition problem as the researchers found that more advanced deep learning methods outperformed simpler techniques, suggesting potential benefits in developing deep learning methods tailored for climate and weather data.

“We found that because the relative position of weather patterns plays a key role in their evolution, using a more advanced deep learning method that tracks the relative position of features improves the accuracy and is also more robust when we don’t have a large amount of data for training.” Hassanzadeh said.

During the Second World War, people did weather prediction by looking through catalogs of weather patterns and pattern matching—this is called analog forecasting. Integrating an equation into the weather system, the first step in a mathematical model, was not possible. After computers became more widely available, this approach was abandoned.

The analog technique requires a forecaster to remember previous weather events that are expected to mimic upcoming events. Its difficulty comes from the fact that there’s rarely a perfect analog for a future event. Still, it remains a useful method of observing rainfall over oceans, as well as forecasting precipitation amounts and distributions.

According to Hassanzadeh, a growing number of people in the weather and climate community are interested in how deep learning can help improve climate and weather modeling.

“The next step for my group is to see if deep learning can be more accurate than the operational numerical weather models used for day-to-day weather forecasts,” Hassanzadeh said. “We’re going to focus on predictions with longer lead times, where the numerical models perform poorly. If it works, it will be a huge advance in weather prediction.”


This study was funded by NASA and an Early-Career Research Fellowship from the Gulf Research Program of the National Academies of Sciences, Engineering, and Medicine. Computing resources were made available through XSEDE and Rice’s Center for Research Computing in partnership with the Ken Kennedy Institute.
Research engineers at the University of California San Diego have developed a high-throughput computational method using XSEDE–allocated resources to design new materials for the next generation of solar cells and LEDs. According to their calculations, these materials, called hybrid halide semiconductors, would be stable and exhibit excellent optoelectronic properties. Hybrid halide semiconductors are materials that consist of an inorganic framework housing organic cations, or ions which have a positive electrical charge. They show unique material properties not found in organic or inorganic materials alone. The team published their findings in the July 2019 edition of Energy & Environmental Science.

A subclass of these materials, called hybrid halide perovskites, have attracted a lot of attention as promising materials for future solar cells and LED devices because of their exceptional optoelectronic properties and inexpensive fabrication costs. However, hybrid perovskites are not very stable and contain lead, making them unsuitable for commercial devices.

Seeking alternatives to perovskites, a team of researchers led by Kesong Yang, a nano-engineering professor at the UC San Diego Jacobs School of Engineering, used computational resources, data mining, and data screening techniques to discover new hybrid halide materials that are stable and lead-free. “We are looking past perovskite structures to find a new space to design hybrid semiconductor materials for optoelectronics,” he said.

Yang’s team began by going through the two largest quantum materials databases, AFLOW and The Materials Project, and analyzing all compounds that were similar in chemical composition to lead halide perovskites. They then extracted 24 prototype structures to use as templates for generating hybrid organic-inorganic materials. Next, they performed high-throughput quantum mechanics Density Functional Theory (DFT) calculations on the prototype structures to build a comprehensive quantum materials repository containing 4,507 hypothetical hybrid halide compounds.

Using efficient data mining and data screening algorithms, Yang’s team rapidly identified 13 candidates for solar cell materials and 23 candidates for LEDs out of all the hypothetical compounds. “A high-throughput study of organic-inorganic hybrid materials is not trivial,” Yang noted, adding that it took several years to develop a complete software framework equipped with data generation, data mining, and data screening algorithms for hybrid halide materials.

“Compared to other computational design approaches, we explored a significantly large structural and chemical space to identify novel halide semiconductor materials,” explained Yuheng Li, a nano-engineering Ph.D. candidate in Yang’s group and the first author of the study. This work could also inspire a new wave of experimental efforts to validate computationally predicted materials, he noted.

Looking ahead, Yang and his team are using their high-throughput approach to discover new solar cell and LED materials from other types of crystal structures. They are also developing new data mining modules to discover other types of functional materials for energy conversion, optoelectronic, and spintronic applications.

Yang attributes much of his project’s success to the use of the Comet supercomputer at UC San Diego’s San Diego Supercomputer Center (SDSC), a National Science Foundation-funded resource allocated via XSEDE. “Our large-scale quantum mechanics calculations required a large number of computational resources,” he said. “Since 2016 we have been awarded some 3.46 million core-hours on Comet which made the project possible.”

Having access to Comet and XSEDE not only saved valuable research time, according to Yang. “The value of these awarded computing resources is about $115,000, which also saved our project a great deal of money!”

The study was supported by the Global Research Outreach Program of Samsung Advanced Institute of Technology and the National Science Foundation. The ab-initio molecular dynamics calculations used computational resources supplied by the Department of Defense High Performance Computing Modernization Program.
SUPERCOMPUTING FUTURE WIND POWER RISE

XSEDE-allocated resources simulate promises and pitfalls of wind power expansion.

**HOW DID XSEDE HELP?**

XSEDE-allocated resources simulate promises and pitfalls of wind power expansion.

Wind power surged worldwide in 2019, but will it sustain? More than 340,000 wind turbines generated over 591 gigawatts globally. In the United States, wind powered the equivalent of 32 million homes and sustained 509 U.S. factories. What’s more, in 2019 wind power grew by 19 percent, thanks to both booming offshore and onshore projects in the U.S. and China.

A study by Cornell University researchers used supercomputers allocated by XSEDE to look into the future of how to make an even bigger jump in wind power capacity in the U.S. The study focused on a potential pitfall of whether adding more turbines in a given area might decrease their output or even disrupt the local climate, a phenomenon caused by what’s referred to as ‘wind turbine wakes.’ Like the wake behind a motorboat, wind turbines create a wake of slower, choppy air that eventually spreads and recovers its momentum.

“This effect has been subject to extensive modeling by the industry for many years, and it is still a highly complex dynamic to model,” Pryor said. These simulations are massively computationally demanding.

“All our simulations were performed on the Department of Energy’s National Energy Research Scientific Computing Center (NERSC) computational environment or on Cori. Simulations presented in our paper consumed over 500,000 CPU hours on Cori and took over a calendar year to complete on the NERSC Cray. That resource is designed for massively parallel computing but not for analysis of the resulting simulation output,” Pryor said.

“Thus, all of our analyses were performed on the XSEDE Jetstream resource using parallel processing and big data analytics in MATLAB,” Pryor added. The Jetstream cloud environment is supported by Indiana University, the University of Arizona, and the Texas Advanced Computing Center (TACC).

“Our work is unprecedented in the detail of wind turbine descriptions, the use of self-consistent projections for doubled installed capacity, study domain size, and the duration of the simulations,” Pryor said. However, she acknowledged uncertainty is the best way to parameterize the action of the wind turbines on the atmosphere and specifically the downstream recovery of wakes.

The authors chose two sets of simulations, the use of self-consistent projections for doubled installed capacity and quadrupled installed capacity, which represents the capacity necessary to achieve the 20 percent of electricity supply from wind turbines in 2030.

“Using these three scenarios we can assess how much power would be generated from each situation and thus if the electrical power production is linearly proportional to the installed capacity or if at very high penetration levels the loss of production due to wakes starts to decrease efficiency,” Pryor said.

The team is currently working on how to design, test, develop, and improve wind farm parameterizations for use in WRF. The Cornell team recently had a publication on this matter in the *Journal of Applied Meteorology and Climatology*. The study, “20% of US electricity from wind will have limited impacts on system efficiency and regional climate,” was published in the journal *Nature Scientific Reports* in January 2020. The U.S. Department of Energy and Cornell University’s Atkinson Center for a Sustainable Future funded this research.

Left: Sara C. Pryor, Professor, Department of Earth and Atmospheric Studies, Cornell University. Bottom: Simulation domain and the location and installed capacity of wind turbines. (a) WT deployed in each 4 × 4 km grid cell in the three scenarios studied. (b) Location of grid cells containing WT (magenta) in 2014 within the inner domain and the background grid cells (cyan) used in the statistical testing for regional effects. Red outline denotes the area covered by the outer domain (d01) and the shading denotes the terrain elevation.
HOW DID XSEDE HELP?

AI running on XSEDE-allocated systems surpasses humans at classifying galaxies.

Astronomers estimate there are at least 100 billion galaxies in the observable universe. Scientists would like to get a better handle on these huge collections of stars for a number of reasons. For one, most of the mass of the universe seems to be invisible. One way we “see” the presence of this dark matter is through its effects on galaxies. Also, the motions of galaxies tell us that the expansion of the universe is accelerating. The reason for this may be that most of the energy of the universe is in an unknown form called dark energy. Astrophysical Surveys, such as the recent Dark Energy Survey (DES) and the upcoming Legacy Survey of Space and Time (LSST), are collecting data to study these fundamental questions.

As a first step, scientists are studying the shapes of galaxies. The shape of a galaxy tends to be strongly intertwined with the history of its evolution. Shape also sheds light on a galaxy’s star-formation rate, past mergers and interactions with other galaxies as well as other properties. The logical starting point for astronomers in modern surveys is to classify and sort the vast number of galaxies observed. The main classification is whether a galaxy has a spiral shape, with curving arms like the Milky Way, or elliptical, which looks like a uniform ball of stars. This simple task is enormous owing to the tremendous number of galaxies. For the data set, the scientists used a multi-level artificial neural network. They employed Comet in the early phases of the work, transitioning to Bridges to take advantage of the most advanced processors available for deep learning at the time—NVIDIA Tesla P100 GPUs. Today, both Bridges and Comet contain P100 nodes.

“XSEDE was pretty helpful for quickly running these huge data sets,” says Khan. The team reported their results in the journal Physics Letters B in August 2019. They presented their visualization the following November at the annual SC19 supercomputing conference.

The training phase required roughly 36,000-galaxy training data set, and a 1,000-galaxy validation data set, and a 12,500-galaxy testing data set. They chose the latter two data sets so that the galaxies in them lie in parts of the sky that both the SDSS and the DES had surveyed, taking advantage of the lessons learned by the earlier study. To generate and process all of the data sets that they used for training and testing, they used the Blue Waters supercomputer at NCSA, an XSEDE resource at that time.

In the testing phase, the AI matched the human classifications 85 percent of the time. But when they adjusted for the known error rate in these classifications, they found their AI was over 99 percent accurate—better than the humans. As a last step, the scientists applied their AI to predict galaxy types in a set of about 10,000 not-yet-labelled galaxies. In addition, they had built their AI so that its processes for classifying the galaxies could be examined by humans. This step, which explained how the work was done, was important for convincing astronomers that the AI’s methods can be trusted.

The team reported their results in the journal Physics Letters B in August 2019. They presented their visualization the following November at the annual SC19 supercomputing conference.

Future work will be to apply the method to larger groups of unidentified galaxies, automating galaxy identification to keep pace with the hundreds of millions expected to be discovered in the near future. The team has also begun using XSEDE-allocated Bridges-Al, whose NVIDIA Tesla V100 GPUs are currently the most advanced GPUs for deep learning. The platform’s NVIDIA DGX-2 enterprise AI research system enables high-performance deep learning across 16 V100s.