The mission of the Texas Advanced Computing Center (TACC) is to power discoveries that propel science and society through the application of advanced computing technologies.

Why is this important? Science and engineering enable us to better understand the environment, the human body, the universe, and improve our quality of life. Science at this scale is incredibly complex and requires advanced computing.

TACC’s computing ecosystem includes a range of advanced technologies as well as skilled experts who help researchers use these resources effectively. Here’s a snapshot of the technologies and services TACC offers.

### POWERING DISCOVERIES

By the Numbers

177,827,112
Computing Hours Provided in 2018

>10K
Annual Users

872
Institutions Globally

Enabling Research

>150
Fields of Science

DATA SERVICES
Scientific research relies on the analysis of large amounts of data. Researchers leverage TACC’s high-speed data analysis systems to make this possible.

MACHINE LEARNING
TACC supports a number of machine and deep learning frameworks that work across TACC’s many systems. The center also offers experts to help researchers apply TACC resources to machine and deep learning problems.

CLOUD COMPUTING
TACC’s comprehensive cloud services provide access to large-scale resources in a virtualized environment, allowing modeling and simulations, visualization, data analytics, and management from any location, on-demand.

STORAGE
TACC provides high-speed, shared storage including a global file system, mass archival storage, and storage systems for data preservation and sharing.

VISUALIZATION
Visualization blends science, technology, and art to bring the complexities of our world to life. TACC’s systems allow researchers to produce insightful visualizations to help solve large-scale problems.
COMPUTE
Solving the world’s biggest problems requires the most advanced computing systems. TACC’s diverse systems enable thousands of researchers to make discoveries.

SOFTWARE
TACC maintains a vast collection of software packages and libraries to support advanced computing across multiple disciplines.

NETWORK
TACC operates and maintains multiple high-speed connections to major national research networks, as well as a high-speed infrastructure connecting TACC systems to each other.

GATEWAYS & APIs
Using supercomputers can be challenging. That’s why TACC develops Gateways and APIs that make using cyberinfrastructure systems easier.

EXPERT TACC STAFF
TACC staff design, deploy, and operate world-class advanced computing resources; provide expert assistance and consulting to researchers; develop tools to make supercomputers more useful; and train the next generation of computational scientists.

EDUCATION & OUTREACH
The next generation of researchers will have access to unprecedented computational tools. Promoting STEM education and careers to students and their teachers will ensure they know how to use them effectively.

SOFTWARE
TACC maintains a vast collection of software packages and libraries to support advanced computing across multiple disciplines.

TRAINING
TACC staff offer college courses, virtual and onsite training, and an ever-expanding series of in-depth institutes aimed at making TACC’s computing ecosystem more accessible and understandable.

VISUALIZATION
Visualization blends science, technology, and art to bring the complexities of our world to life. TACC’s systems allow researchers to produce insightful visualizations to help solve large-scale problems.

MACHINE LEARNING
TACC supports a number of machine and deep learning frameworks that work across TACC’s many systems. The center also offers experts to help researchers apply TACC resources to machine and deep learning problems.

NETWORK
TACC operates and maintains multiple high-speed connections to major national research networks, as well as a high-speed infrastructure connecting TACC systems to each other.

TRAINING
TACC staff offer college courses, virtual and onsite training, and an ever-expanding series of in-depth institutes aimed at making TACC’s computing ecosystem more accessible and understandable.

CLOUD COMPUTING
TACC’s comprehensive cloud services provide access to large-scale resources in a virtualized environment, allowing modeling and simulations, visualization, data analytics, and management from any location, on-demand.

DATA SERVICES
Scientific research relies on the analysis of large amounts of data. Researchers leverage TACC’s high-speed data analysis systems to make this possible.

STORAGE
TACC provides high-speed, shared storage including a global file system, mass archival storage, and storage systems for data preservation and sharing.

TRAINING
TACC staff offer college courses, virtual and onsite training, and an ever-expanding series of in-depth institutes aimed at making TACC’s computing ecosystem more accessible and understandable.

EDUCATION & OUTREACH
The next generation of researchers will have access to unprecedented computational tools. Promoting STEM education and careers to students and their teachers will ensure they know how to use them effectively.

SOFTWARE
TACC maintains a vast collection of software packages and libraries to support advanced computing across multiple disciplines.

NETWORK
TACC operates and maintains multiple high-speed connections to major national research networks, as well as a high-speed infrastructure connecting TACC systems to each other.

GATEWAYS & APIs
Using supercomputers can be challenging. That’s why TACC develops Gateways and APIs that make using cyberinfrastructure systems easier.

EXPERT TACC STAFF
TACC staff design, deploy, and operate world-class advanced computing resources; provide expert assistance and consulting to researchers; develop tools to make supercomputers more useful; and train the next generation of computational scientists.

SOFTWARE
TACC maintains a vast collection of software packages and libraries to support advanced computing across multiple disciplines.

NETWORK
TACC operates and maintains multiple high-speed connections to major national research networks, as well as a high-speed infrastructure connecting TACC systems to each other.

VISUALIZATION
Visualization blends science, technology, and art to bring the complexities of our world to life. TACC’s systems allow researchers to produce insightful visualizations to help solve large-scale problems.

MACHINE LEARNING
TACC supports a number of machine and deep learning frameworks that work across TACC’s many systems. The center also offers experts to help researchers apply TACC resources to machine and deep learning problems.

EXPERT TACC STAFF
TACC staff design, deploy, and operate world-class advanced computing resources; provide expert assistance and consulting to researchers; develop tools to make supercomputers more useful; and train the next generation of computational scientists.

SOFTWARE
TACC maintains a vast collection of software packages and libraries to support advanced computing across multiple disciplines.

NETWORK
TACC operates and maintains multiple high-speed connections to major national research networks, as well as a high-speed infrastructure connecting TACC systems to each other.

GATEWAYS & APIs
Using supercomputers can be challenging. That’s why TACC develops Gateways and APIs that make using cyberinfrastructure systems easier.

EXPERT TACC STAFF
TACC staff design, deploy, and operate world-class advanced computing resources; provide expert assistance and consulting to researchers; develop tools to make supercomputers more useful; and train the next generation of computational scientists.

SOFTWARE
TACC maintains a vast collection of software packages and libraries to support advanced computing across multiple disciplines.

NETWORK
TACC operates and maintains multiple high-speed connections to major national research networks, as well as a high-speed infrastructure connecting TACC systems to each other.

VISUALIZATION
Visualization blends science, technology, and art to bring the complexities of our world to life. TACC’s systems allow researchers to produce insightful visualizations to help solve large-scale problems.

MACHINE LEARNING
TACC supports a number of machine and deep learning frameworks that work across TACC’s many systems. The center also offers experts to help researchers apply TACC resources to machine and deep learning problems.

EXPERT TACC STAFF
TACC staff design, deploy, and operate world-class advanced computing resources; provide expert assistance and consulting to researchers; develop tools to make supercomputers more useful; and train the next generation of computational scientists.

SOFTWARE
TACC maintains a vast collection of software packages and libraries to support advanced computing across multiple disciplines.

NETWORK
TACC operates and maintains multiple high-speed connections to major national research networks, as well as a high-speed infrastructure connecting TACC systems to each other.

VISUALIZATION
Visualization blends science, technology, and art to bring the complexities of our world to life. TACC’s systems allow researchers to produce insightful visualizations to help solve large-scale problems.

MACHINE LEARNING
TACC supports a number of machine and deep learning frameworks that work across TACC’s many systems. The center also offers experts to help researchers apply TACC resources to machine and deep learning problems.

EXPERT TACC STAFF
TACC staff design, deploy, and operate world-class advanced computing resources; provide expert assistance and consulting to researchers; develop tools to make supercomputers more useful; and train the next generation of computational scientists.

SOFTWARE
TACC maintains a vast collection of software packages and libraries to support advanced computing across multiple disciplines.

NETWORK
TACC operates and maintains multiple high-speed connections to major national research networks, as well as a high-speed infrastructure connecting TACC systems to each other.
Welcome to Texascale, the new annual magazine of the Texas Advanced Computing Center. We have launched this magazine to bring news from our center to the many stakeholders who support us. The work we do at TACC has never been more central to research at The University of Texas at Austin, and around the state, the nation, and the world.

Simulation is one of the most important tools for advancing fundamental science and accelerating engineering progress. The ubiquitous availability of massive amounts of digital data from scientific instruments, mobile phones, and other edge computing devices has made the so-called “Big Data” problem part of the advanced computing world, with profound implications not just in traditional fields of simulation science, but in business, social sciences, and the humanities as well.

Progress turning this data into useful and actionable knowledge is being driven by the breathtaking pace of advances in artificial intelligence, machine learning, and deep learning (AI/ML/DL), which could impact nearly every job on the planet from marketing and customer service to healthcare and agriculture. We are only just scratching the surface of our AI capabilities and the role that AI might play in scientific research — and AI will in turn be driven by the frontiers of advanced computing.

Future generations of AI and advanced computing may be extended by another revolutionary technology: quantum information science, an emerging field with the potential to produce revolutionary advances in science and engineering.

Together, this set of technologies that converge around advanced computing — simulation, data, AI, quantum — will be key not only to advancing fundamental science and engineering, but also to tackling many of the core problems our nation will face in the future: cybersecurity, resilient and sustainable infrastructure, food and energy production, and the list goes on.

These many missions have created new challenges at TACC. Fortunately, we have never been better prepared to meet them. This year, TACC was selected as the recipient of the National Science Foundation’s new “Towards a Leadership-Class Computing Facility” award. The first visible manifestation of this award will be the deployment of the largest system we have ever built, Frontera, featured prominently in this magazine. Frontera also represents the next stage of TACC’s evolution in our ongoing quest for scale in systems, in science, and in challenges that match our home state of Texas, where “Everything is Bigger”… Texascale. 🌟

—Dan Stanzione, Executive Director, TACC
Associate Vice President for Research,
The University of Texas at Austin
# Table of Contents

- Powering Discoveries Ecosystem 01
- A Letter from the Director 03
- The Path to Texascale 05
- **Protecting Pregnancy with Smartphones** 07
- Artificial Intelligence v. Austin Traffic 11
- Modeling Without Models 12
- Texas Teens Turn Citizen Scientists 13
- Disaster-Proofing U.S. Infrastructure 15
- Beneath Hawaii's Volcanoes 17
- Overcoming a Battery's Fatal Flaw 18
- **Fighting Cancer with Supercomputers** 19
- Saving the Ross Ice Shelf 23
- Better Computing through Industry Collaboration 24
- **Computation for the Endless Frontier** 25
- Science on Repeat 35
- Accelerating Deep Learning 37
- Closing the Workforce Gap 39
- **Stampede2: A Powerhouse for Science** 41
- Serving the Lone Star State 45
- Stampede in Africa 47
- Turn on the Switch 48
- **Circulation in the Arctic Ocean** 49
Over the past 17 years, TACC has grown from a small, upstart research center to one of the world’s largest and most innovative hubs for advanced computing. Explore how our center has grown over the years.
LEADERSHIP-CLASS COMPUTING
TACC wins $60 million award from NSF to build and deploy FRONTERA, a supercomputer that will be twice as fast as Stampedede2 and among the most powerful in the world.

TACC FOR SCIENCE
TACC receives largest ever award from NSF to UT Austin to build STAMPEDE1, at 8.5 PetaFLOPS, the 7th fastest supercomputer in the world.

POWERHOUSE FOR RESEARCH
TACC debuts STAMPEDE2, at 18 PetaFLOPS, the fastest supercomputer at any university in the U.S. and 12th most powerful in the world.

TACC systems help confirm the first discovery of gravitational waves by detectors at the Laser Interferometer Gravitational Wave Observatory (LIGO).

On its 15th anniversary, TACC dedicates the new Advanced Computing Building, which includes space for growing staff and an experimental Visualization Lab.

TACC EXPANDS

TACC donates time to Japanese researchers to predict impacts from Fukushima nuclear power plant disaster.

NATIONAL LEADERSHIP
TACC launches LONESTAR4, a 302 TeraFLOPS system with shared memory and remote visualization capabilities.

TACC is selected as one of the leaders of the eXtreme Science and Engineering Discovery Environment (XSEDE) project, the foundation of the national cyberinfrastructure ecosystem.

STAMPEDE FOR SCIENCE
TACC experts largest ever award from NSF to UT Austin to build STAMPEDE1 at 8.5 PetaFLOPS, the 7th fastest supercomputer in the world.

LEADERSHIP-CLASS COMPUTING
TACC wins $60 million award from NSF to build and deploy FRONTERA, a supercomputer that will be twice as fast as Stampedede2 and among the most powerful in the world.

POWERHOUSE FOR RESEARCH
TACC debuts STAMPEDE2, at 18 PetaFLOPS, the fastest supercomputer at any university in the U.S. and 12th most powerful in the world.

TACC systems help confirm the first discovery of gravitational waves by detectors at the Laser Interferometer Gravitational Wave Observatory (LIGO).

On its 15th anniversary, TACC dedicates the new Advanced Computing Building, which includes space for growing staff and an experimental Visualization Lab.

TACC EXPANDS

TACC donates time to Japanese researchers to predict impacts from Fukushima nuclear power plant disaster.

NATIONAL LEADERSHIP
TACC launches LONESTAR4, a 302 TeraFLOPS system with shared memory and remote visualization capabilities.

TACC is selected as one of the leaders of the eXtreme Science and Engineering Discovery Environment (XSEDE) project, the foundation of the national cyberinfrastructure ecosystem.
Three leading institutions join forces to combat pregnancy-related complications and death

Women in the U.S. are more likely to die from childbirth or pregnancy-related causes than other women in the developed world. About 700 women die each year in the U.S. as a result of pregnancy or delivery complications.

Unlike other leading causes of death such as cancer or Alzheimer’s disease, birth-related deaths are largely preventable. Today, however, most adverse pregnancy outcomes are not predictable, and thus cannot be avoided, partly because of a lack of medical data.

A newly-appointed team has brought together researchers and physicians from TACC, the Dell Medical School (Dell Med) Department of Women’s Health, and the Institute for Computational Engineering and Sciences (ICES) at UT Austin. This team of experts is working to close the gap in medical evidence for complications such as maternal mortality, emergency C-sections, stillbirth, neonatal death, pre-term birth, and other important health issues.

In September 2018, the National Science Foundation awarded the team a $1.2 million Smart and Connected Health grant to support their research.

“We are very excited,” said Radek Bukowski, a co-principle investigator (PI) on the grant and associate chair of Women’s Health at Dell Med. “The project is innovative, full of promise, and unprecedented in the novelty of the approach.”

Using smart phones, the team will passively monitor the activity and behavior of 1,000 pregnant women in the Austin area. The ultimate goal is to develop digital phenotypes, or profiles, to better understand factors that influence pregnancy and can inform individualized pregnancy care.
Participants will download a smartphone application originally developed at Harvard University and modified for this project to collect physical, social, and behavioral data. The app captures activity over time such as a participant’s movement, interaction with their social network, typing speed, and screen time. Analysis of this large collection of digital data, in combination with traditional clinical data collected via participants’ medical records, informs the development of a digital phenotype of pregnancy.

"The human body is a complex system," Bukowski said. "There are a lot of elements that interact with each other, which makes it very difficult to analyze in a traditional manner. It’s difficult to draw conclusions from averages, and it’s difficult to predict adverse outcomes."

"From their first appointment to six weeks post-partum, we’ll be able to analyze data to determine the impact of their everyday lives on their medical outcomes, and determine whether we see any digital markers of significant events such as labor," said Kelly Gaither, the grant’s PI as well as director of Health Analytics at TACC and an associate professor for Women’s Health with the Dell Medical School. "We can examine, for example, whether these women have an active social structure, changes in speech patterns, and unusual sleep patterns — even the number of steps they take in a day."

With the ubiquitous use of smartphones, it is now possible to collect lived experiences or data reflecting markers of pregnancy. "And the beauty of it is it’s completely passive and unobtrusive collection," Gaither explained. "We can collect these behaviors, generically, without needing to collect sensitive data such as text messages and conversations."

All of us, first and foremost, are motivated to improve medical care and outcomes for women and babies. —Kelly Gaither, TACC

"MACHINE LEARNING HELPS PAVE THE WAY"

"We know from the computational side that there’s a lot we can do to help," Gaither said. "At TACC, we know how to bring advanced computing to bear on this issue, including using data-driven science, mathematical models, and emerging computational techniques such as more advanced forms of machine learning."

Currently, the team is using machine learning to analyze the risk of having an unplanned C-section. The hope is to improve the outcomes seen in labor and delivery using maternal characteristics recorded during labor to better quantify the risk of requiring an emergency C-section, according to Karl W. Schulz, a co-PI on the grant, a research associate professor for ICES, and an associate professor for Women’s Health with Dell Med.

"This is a problem where we leverage risk factors and patient characteristics using data at a national scale over multiple years," Schulz said. The data used for this analysis is published by the Centers for Disease Control and Prevention (CDC) and provides vital statistics for all U.S. births, approximately four million per year.

Schulz said that machine learning techniques can be used to determine which patient factors are the best at predicting certain health risks by training and evaluating a variety of models, and assessing the accuracy of their results against available test data.

Modeling parameters from the CDC data are divided into two categories: those that are known during pregnancy versus those that are known during labor and delivery. Influential factors belonging to the former include maternal age, live birth order, diabetes, hypertension, and the number of previous cesareans.

Schulz said that moving forward, TACC’s newly announced supercomputer, Frontera, is going to allow analysis on confidential HIPAA data and be “a great resource for research in computational health.”
DATA CHALLENGES IN HEALTHCARE

There are four major data challenges in healthcare — access, quality, inconsistency, and integration. All data, whether it’s from the CDC, Medicaid, electronic medical records, or private insurance, are going to have similar problems.

In terms of access, for example, researchers are unable to get data about mothers and babies from death records. A researcher might know that the mother died of complications from childbirth, but the records don’t contain data about anything leading up to the complications or about the birth and post-partum period.

To help with this challenge, the team is partnering with the School of Public Health at the UT Texas Health Science Center at Houston to create a data architecture capable of housing large collections of heterogeneous types of health data, clinical records, and medical claims, among others.

“We’re creating a large health data infrastructure with the UT School of Public Health,” said Tomislav Urban of TACC, a key member of the team. “They have outgrown their capacity and would like to move the data to TACC for larger space, but also for analysis and reporting.”

Urban has created a database, primarily from CDC data, that correctly maps data points to meta-fields and meta-values over a 10 year period from 2006 to 2016. The CDC data captures statistics from various states including vital statistics such as births, causes of death, weight gain during pregnancy, and cigarette smoking. There are about 300 different fields on these annual data sets.

“However, the agency de-identifies them, so they don’t include any personal or geographical information,” Urban said.

Mapping the fields to a common set of values, organizing the data, and cleaning the data is important because the team is looking for rare events, trying to create more accurate models and develop a more personalized look at the attributes of an individual, not just of a general population.

“All of us, first and foremost, are motivated to improve medical care and outcomes for women and babies,” Gaither concluded.

“We believe this type of truly interdisciplinary team working together for this common goal leverages the diverse set of expertise, perspectives, and experiences that we all have, and is key to advancing the state of the art in health. While we have very different skill sets, we all have one thing in common: we either have been a patient in the past or will be one in the future.”

Health informatics applies information technology and advanced computing to the field of healthcare. Other TACC-enabled health informatics projects include LungMAP, a detailed molecular atlas of developing lungs, and the Virtual Drug Discovery Portal, a graphical interface for identifying small molecules that bind to target proteins. Learn more about this growing research area: useta.cc/health
Look above the traffic light at a busy intersection in your city and you will probably see a camera. These devices monitor traffic conditions and provide visuals in the case of a collision. But can they do more?

Researchers from TACC, The University of Texas Center for Transportation Research, and the City of Austin developed a tool that uses artificial intelligence to recognize cars, bikes, and pedestrians in raw traffic camera footage and characterize how they move and interact. This information can then be analyzed by traffic engineers and officials to improve the safety and performance of the city’s transportation network.

The project earned a 2018 Smart 50 Award, recognizing it as one of the most innovative and influential smart cities projects in the world.

The team at TACC is now working on a more sophisticated algorithm for pedestrian detection and tracking to enable the City of Austin to identify potential areas of concern and ultimately improve pedestrian safety.
If you look at many of the industrial products that we build today — airplanes, microchips, steel beams — there are mathematical models of the underlying physics that determine how they will behave.

Engineering starts with those robust mathematical models which allow inventors to predict the properties of a process or a product. For example, knowledge of how a material conducts electricity or bends lets engineers use computer-aided design to improve these systems predictably. The combination of theory, experiment, and simulation has created fantastic increases in design efficiency over the years.

However, there are some scientific domains where good, predictive models are still unavailable. Biology and chemistry are two such domains. Researchers would seize the opportunity to engineer new drugs or solar technologies that work more effectively if only they had mathematical formulas to represent how the human body or the quantum world worked.

TACC and the Defense Advanced Research Projects Agency (DARPA) are working to solve this problem.

A little over a year ago, TACC and DARPA began developing data-driven ways to accelerate design and discovery in research areas where predictive models don’t yet exist as part of the Synergistic Discovery and Design (SD2) program.

The program is building computational tools that use experimental data to develop hypotheses and new designs in areas such as synthetic biology, neuro-computation, and polymer chemistry. The team is using artificial intelligence to consider a universe of hypotheses; converge on the ones that fit the data; design experiments to test those hypotheses; then evaluate outcomes and feed that back into the training process.

“Humans learn this way,” said Matt Vaughn, director of Life Sciences Computing at TACC. “But machines can consider a much wider number of variables and examples.”

SD2 employs TACC’s powerful computing and data ecosystem and takes advantage of software and services that enable collaboration, sharing, and reproducible science tailored to the evolving needs of the four-year program.

“This approach has served TACC and its partners well,” Vaughn said.

One of the SD2E partners is Bree Cummins, a mathematics professor at Montana State University, who uses TACC’s environment for code collaboration, data sharing, and data analysis on a regular basis.

“TACC’s ecosystem is backed by a helpful and responsive staff that provides timely support and feature enhancements,” Cummins said. “These resources have enabled me to participate in successful group projects with researchers scattered across the country.”

David Baker, another SD2 participant and a professor of Biochemistry at the University of Washington, works to create new proteins.

“TACC is making an important contribution to the creation of a whole new world of designed proteins to address current day challenges,” he said.

TACC supports other large research communities through projects like CyVerse, which serves life scientists, and DesignSafe, which empowers the natural hazards engineering community. The center also supports hundreds of long-term data and archival curation projects that make use of TACC resources.

“Researchers shouldn’t feel constrained by their environment,” Vaughn said. “Whether they need graphical interfaces, APIs, or batch data analytics, we provide it and make sure they can use it.”

By the end of the project, real and tangible outcomes are possible. These include complex biological circuits that can sense and respond to target molecules; therapeutically proteins that are more resilient to their surrounding environment; and more efficient, flexible solar cells.

“TACC is excited to provide DARPA with a diverse computational infrastructure, including HPC, GPU, and cloud technologies, coupled with powerful data-sharing mechanisms that democratize progress across domains of science by making it much easier for people to reproduce one another’s work and share compelling results,” Vaughn said.

“Ultimately, these capabilities accelerate the pace of turning research into useful applications for society.”

David Baker, another SD2 participant and a professor of Biochemistry at the University of Washington, works to create new proteins.

“TACC is making an important contribution to the creation of a whole new world of designed proteins to address current day challenges,” he said.

TACC supports other large research communities through projects like CyVerse, which serves life scientists, and DesignSafe, which empowers the natural hazards engineering community. The center also supports hundreds of long-term data and archival curation projects that make use of TACC resources.

“Researchers shouldn’t feel constrained by their environment,” Vaughn said. “Whether they need graphical interfaces, APIs, or batch data analytics, we provide it and make sure they can use it.”

By the end of the project, real and tangible outcomes are possible. These include complex biological circuits that can sense and respond to target molecules; therapeutically proteins that are more resilient to their surrounding environment; and more efficient, flexible solar cells.

“TACC is excited to provide DARPA with a diverse computational infrastructure, including HPC, GPU, and cloud technologies, coupled with powerful data-sharing mechanisms that democratize progress across domains of science by making it much easier for people to reproduce one another’s work and share compelling results,” Vaughn said.

“Ultimately, these capabilities accelerate the pace of turning research into useful applications for society.”

David Baker, another SD2 participant and a professor of Biochemistry at the University of Washington, works to create new proteins.

“TACC is making an important contribution to the creation of a whole new world of designed proteins to address current day challenges,” he said.

TACC supports other large research communities through projects like CyVerse, which serves life scientists, and DesignSafe, which empowers the natural hazards engineering community. The center also supports hundreds of long-term data and archival curation projects that make use of TACC resources.

“Researchers shouldn’t feel constrained by their environment,” Vaughn said. “Whether they need graphical interfaces, APIs, or batch data analytics, we provide it and make sure they can use it.”

By the end of the project, real and tangible outcomes are possible. These include complex biological circuits that can sense and respond to target molecules; therapeutically proteins that are more resilient to their surrounding environment; and more efficient, flexible solar cells.

“TACC is excited to provide DARPA with a diverse computational infrastructure, including HPC, GPU, and cloud technologies, coupled with powerful data-sharing mechanisms that democratize progress across domains of science by making it much easier for people to reproduce one another’s work and share compelling results,” Vaughn said.

“Ultimately, these capabilities accelerate the pace of turning research into useful applications for society.”

David Baker, another SD2 participant and a professor of Biochemistry at the University of Washington, works to create new proteins.

“TACC is making an important contribution to the creation of a whole new world of designed proteins to address current day challenges,” he said.

TACC supports other large research communities through projects like CyVerse, which serves life scientists, and DesignSafe, which empowers the natural hazards engineering community. The center also supports hundreds of long-term data and archival curation projects that make use of TACC resources.

“Researchers shouldn’t feel constrained by their environment,” Vaughn said. “Whether they need graphical interfaces, APIs, or batch data analytics, we provide it and make sure they can use it.”

By the end of the project, real and tangible outcomes are possible. These include complex biological circuits that can sense and respond to target molecules; therapeutically proteins that are more resilient to their surrounding environment; and more efficient, flexible solar cells.

“TACC is excited to provide DARPA with a diverse computational infrastructure, including HPC, GPU, and cloud technologies, coupled with powerful data-sharing mechanisms that democratize progress across domains of science by making it much easier for people to reproduce one another’s work and share compelling results,” Vaughn said.

“Ultimately, these capabilities accelerate the pace of turning research into useful applications for society.”

David Baker, another SD2 participant and a professor of Biochemistry at the University of Washington, works to create new proteins.

“TACC is making an important contribution to the creation of a whole new world of designed proteins to address current day challenges,” he said.

TACC supports other large research communities through projects like CyVerse, which serves life scientists, and DesignSafe, which empowers the natural hazards engineering community. The center also supports hundreds of long-term data and archival curation projects that make use of TACC resources.

“Researchers shouldn’t feel constrained by their environment,” Vaughn said. “Whether they need graphical interfaces, APIs, or batch data analytics, we provide it and make sure they can use it.”

By the end of the project, real and tangible outcomes are possible. These include complex biological circuits that can sense and respond to target molecules; therapeutically proteins that are more resilient to their surrounding environment; and more efficient, flexible solar cells.

“TACC is excited to provide DARPA with a diverse computational infrastructure, including HPC, GPU, and cloud technologies, coupled with powerful data-sharing mechanisms that democratize progress across domains of science by making it much easier for people to reproduce one another’s work and share compelling results,” Vaughn said.
Students from all over Texas attended the Code@TACC Connected summer camp, which taught them how to use sensors and computers to observe, measure, and understand the world around them.
Gonzalez, who participated in the Code@TACC Connected camp, says that the camp's focus on real world problem-solving made coding more accessible. “Through this camp I’ve learned that actual data is much easier to comprehend and use rather than making assumptions. Numbers allow you to understand what is true and what is not true,” Gonzalez said.

Gonzalez said that UT Austin is one of her top school choices, and she plans to study civil engineering and computer science.

Code@TACC Connected was created thanks in part to funding from Planet Texas 2050 (PT2050), a new initiative at UT Austin that has brought together more than 100 researchers from across campus to find ways to make Texas resilient in the face of climate extremes and rapid population growth.

As part of the camp, the students visited Waller Creek, an Austin-area waterway that for many years has experienced severe flooding. Waller Creek represents an example of what happens when natural watersheds become more urbanized.

"Here at the university, Waller Creek is a primary point of interest for scientists," said Suzanne Pierce, a research scientist at TACC and an organizing committee member of PT2050. "The very same environmental sensors that we used for the summer camp can also be used for Waller Creek, and the lessons learned can be transferred to Planet Texas 2050."

"We’re very excited to have our signature Code@TACC program selected to contribute to these important efforts that the university is leading," says Rosalia Gomez, TACC Education & Outreach manager.

"Students came out of the camp learning that the landscape around them is not passive...that they can interact and observe the world more deeply, and the tools we use to do this, such as computers and coding, strengthen us," Gomez said.

In addition to summer programming, TACC provides educational tours throughout the year to increase awareness about the impact of science and technology; get students excited about STEM careers; and help students make the connection between STEM careers and their everyday lives.

*The Summer 2018 Code@TACC camps were supported by the National Science Foundation, KLE Foundation, Summer STEM Investment Hub, Planet Texas 2050, Cisco, and Dell EMC.*
The 2017 hurricane season was one of the worst on record. Seventeen named storms caused thousands of fatalities, incurred more than $280 billion in damages, and disrupted millions of lives.

We may not be able to prevent hurricanes from occurring, but we can improve our ability to predict them, move people out of harm’s way, respond quickly to their aftermath, and build our homes and infrastructure in a way that can survive major storms.

A web-based research platform called DesignSafe helps researchers manage, analyze, and understand critical information about natural hazards — from earthquakes and tornadoes to hurricanes and sinkholes. It is supported by grants from the National Science Foundation and is developed at TACC in collaboration with partners at the UT Austin Cockrell School of Engineering, Rice University, and Florida Institute of Technology.

DesignSafe helps prevent natural hazards from becoming societal disasters.

POWERING POST-STORM RECONNAISSANCE

David Roueche, an assistant professor of Civil Engineering at Auburn University, was on the front lines of the 2017 hurricane response. He participated in post-storm reconnaissance missions to coastal Texas, the Florida Keys, Puerto Rico, and several Caribbean islands.

Roueche and his collaborators inspected thousands of homes after Hurricanes Harvey, Irma, and Maria, uploading tens of thousands of photographs captured by ground-based teams and unmanned aerial vehicles.

The 2017 hurricane season put DesignSafe to the test.

Roueche and his team coordinated their deployments via a community channel established by DesignSafe. They used wind map data shared on DesignSafe to determine where they would focus their efforts. They uploaded their field data to DesignSafe and used software within DesignSafe to generate maps that synthesized their and others’ data collections. DesignSafe’s Reconnaissance Portal, which launched in 2017, provided both the computing capabilities Roueche needed for his analyses, and a place to share more than 200 gigabytes of gathered data.

The engineers found examples of houses side-by-side, built around the same time, where one was completely destroyed and the other was intact. What factors influenced survival? And how could rebuilding efforts be improved by understanding what features led some structures to withstand storms?

“We were interested in capturing data about structures before they were destroyed, torn down,
and rebuilt,” Roueche said. “Typically, it takes one to five years before products from the data are formally published. That doesn’t allow us to help communities in rebuilding.”

With DesignSafe, recommendations can be disseminated in months rather than years, accelerating what Roueche calls the ‘resilience curve’: the time it takes a community to fully recover from a disaster.

“By having a more streamlined workflow, standardizing processes, and publishing data sooner, DesignSafe allows us to affect the reconstruction process and have a greater impact,” Roueche said.

CREATING STREET-LEVEL FORECASTS OF HURRICANE IMPACTS

Working at the front-lines of Hurricane Harvey also inspired the creation of new tools to assist first responders. At the Texas State Operations Center, Dan Stanzione, TACC’s executive director and a co-principal investigator for DesignSafe, assisted decision-makers during Hurricane Harvey.

“While there, a first responder said to me: ‘I want to know a list of addresses in Houston that have been flooded above their electrical outlet height. You have a supercomputer, you can do that right?’” Stanzione recalled. “We weren’t able to provide that assistance in the moment, but it inspired us to create such a tool.”

The data needed for the analysis, such as storm surge forecasts, elevation maps, and home constructions records, were all in place. But the ability to connect these datasets and generate a list of potentially damaged homes in a reasonable amount of time was not.

So, researchers on the DesignSafe team created tools to stitch together simulations and data across scales, from the entirety of the Gulf of Mexico to a particular stretch of coastline, and from individual neighborhoods to specific homes.

Using 2008’s Hurricane Ike as a test case, the DesignSafe team showed that they could generate, in real-time, a list of potentially damaged homes. In fact, all aspects of the process could be computed within DesignSafe, using TACC’s supercomputers in the background.

“We can go from Gulf-wide, large-scale simulations to, ‘Did this house flood or not?’” Stanzione said. “It’s a long-time goal that was not technologically possible before, but that we’ve been able to achieve.”

When Hurricane Florence hit the mid-Atlantic coast in 2018, natural hazard researchers were ready to act. Multiple field reconnaissance teams relied on DesignSafe to coordinate their efforts, identify high impact targets, share data, and make decisions.

“DesignSafe is truly changing the way we plan, perform, and document our reconnaissance efforts,” noted Ellen Rathje, principal investigator for DesignSafe and the Janet S. Cockrell Chair in Engineering at UT Austin. “I have been participating in natural hazards reconnaissance for 20 years, and I’ve never seen anything like this. It’s very exciting.”

DesignSafe researchers deployed reconnaissance drones to survey the devastation after Hurricane Florence.
Hawaii's volcanoes stand as silent sentinels. They guard the secret of how they formed, thousands of miles away from the edges of tectonic plates that clash and generate magma for most other volcanoes. Conclusive evidence remains elusive of what carries lava through the Earth's mantle to Hawaii's surface.

Scientists hypothesize that mantle plumes, mushroom-shaped upwellings of hot rock from the deep Earth, caused Hawaii's volcanoes to form. The plumes are thought to originate at the base of the mantle and carry the heat from the Earth's core that generates a volcano's magma.

Scientists used the Stampede1 supercomputer at TACC to make the most detailed computational model yet of mantle plumes. They published their results in the American Geophysical Union's Journal of Geophysical Research, Solid Earth in January 2018.

The international science team showed for the first time how plumes decelerate seismic waves and, consequently, how they appear in seismic tomographic images of the Earth's mantle. Seismic tomography uses seismic waves from earthquakes to create maps of Earth's subsurface. What's more, the researchers say their work could guide future experiments with deep Earth imaging on the ocean floor and help get to the bottom of the mystery of the origin of Hawaii's volcanoes.

"We used computer modeling to find optimal imaging scenarios, so that we can recover the most detail of mantle plumes at the lowest cost," said lead author Ross Maguire, formerly a PhD student who recently graduated from the department of Earth and Environmental Sciences at the University of Michigan.

The team explored various ways of using the reflections of seismic waves to image the Earth and did extensive tests to figure out the optimal configurations for seismometers at the surface to see plumes.

The effort is particularly important for Hawaii, said study co-author Jeroen Ritsema, a professor of Earth and Environmental Science at the University of Michigan. "We believe there is a plume responsible for volcanism on the Hawaiian islands. We've determined what might be optimal offshore deployments on the seafloor that could lead to the best images of the deep mantle beneath Hawaii."

Maguire, Ritsema, and their colleagues' simulations on Stampede1 required 1.2 million computing hours — the equivalent of 136 years of computing with a single processor. Their research continues with the Stampede2 system.

"Understanding Earth dynamics is of fundamental importance because we all live here and are affected by what goes on beneath our feet," said Maguire. "The existence of mantle plumes and the role that they play in our planet is still a big question mark."
OVERCOMING A BATTERY'S FATAL FLAW

As renewable energy grows as a power source around the world, one key component has proven elusive: large-scale, stable, efficient, and affordable batteries.

Lithium-ion batteries have been successful for consumer electronics, but electric vehicles, wind turbines, and smart grids require batteries with far greater energy capacity. A leading contender is the lithium-metal battery, which differs from lithium-ion technology in that it contains lithium-metal electrodes.

"Lithium-metal batteries are basically the dream batteries since they provide an extremely high energy density," said Reza Shahbazian-Yassar, associate professor of mechanical and industrial engineering at the University of Illinois at Chicago.

However, engineers have struggled to build commercially viable lithium-metal batteries because of a fatal flaw: dendrites, sharp "needles" of lithium atoms that can cause batteries to heat up, short-circuit, and catch fire.

Recently, a team of researchers, including Shahbazian-Yassar and Perla Balbuena at Texas A&M University, have been inching closer to finding a solution. They used the Stampede1 and Lonestar5 supercomputers to understand the core chemistry and physics at work in dendrite formation and to engineer new materials that can prevent dendrite growth.

Writing in Advanced Functional Materials in February 2018, the researchers presented a new material that may solve the long-standing dendrite problem.

"The idea was to develop a coating material that can protect the lithium metal and make the ion deposition much smoother," said Balbuena, professor of Chemical Engineering at Texas A&M and co-author on the paper.

The material the researchers developed is a graphene oxide nanosheet that can be sprayed onto glass and inserted into a battery. The material allows lithium ions to pass through it, but slows down and controls how the ions combine with electrons from the surface to become neutral atoms.

They used computer models and simulations in tandem with physical experiments and microscopic imaging to reveal how and why the material works. The lithium ions, they showed, form a thin film on the nanosheet and then sift through gaps in the layers of the material before settling below the bottom layer of the graphene oxide. The material acts like the pegs in a pachinko game, slowing and directing the metal balls as they fall.

The graphene oxide-doped batteries are stable for up to 160 cycles, whereas an unmodified battery rapidly loses its efficiency after 120 cycles. The oxide can be applied simply and affordably with a spray coating gun.

"The simulations gave our collaborators ideas about the mechanism of ion transfer through the coating," Balbuena said. Some future directions involve different thicknesses or chemical compositions based on the phenomenon that they observed.

The work is supported by the Department of Energy and is aimed at creating smaller, safer, lighter, and less expensive battery packs to make electric vehicles more viable.

"The research is a combination of chemistry, physics, and engineering," said Balbuena, "all enabled by computing, this theoretical microscope that can visualize phenomena through theory."
FIGHTING CANCER WITH SUPERCOMPUTERS

Cancer kills more than 500,000 people each year. There’s a greater than 40 percent chance on average that you will be diagnosed with cancer at some point in your lifetime, and a one in five chance that it will be terminal. But the tide might be turning on this terrible disease, thanks to developments in cancer treatments, diagnostics, medical imaging, and basic knowledge.

Hundreds of cancer researchers use supercomputers at TACC to explore aspects of the disease that can’t be studied in labs or clinical trials. What follows are eight ways TACC is helping oncologists, surgeons, and computer scientists improve our fundamental understanding of cancer and the methods for its diagnosis and treatment.
CHEMOTHERAPY AND DRUG DESIGN

New drugs can cost billions of dollars to develop and take decades to reach the marketplace. Supercomputers can speed up the process by finding new uses for approved drugs.

TACC’s Lonestar5 supercomputer virtually tested more than 1,400 small molecule drugs approved by the Food and Drug Administration to see if they could be used to treat cancer. Shuxing Zhang, associate professor of experimental therapeutics at MD Anderson Cancer Center, and graduate student Zhi Tan found that the parasite-fighting drug mebendazole could effectively bind to and inhibit the activity of TNIK, an enzyme that plays a key role in cell signaling related to colon cancer.

“Such advantages render the possibility of quickly translating the discovery into a clinical setting for cancer treatment in the near future,” Zhang said.

IMMUNOTHERAPY

Immuno therapy supercharges the body’s natural defenses to fight cancer, but not every immune therapy works the same way on every patient.

Researchers from Wake Forest School of Medicine and Zhejiang University in China turned to TACC’s Stampede1 to help develop a new mathematical model that represents the interactions between prostate tumors and common immunotherapies. By doing millions of simulations to predict tumor responses to treatments, the researchers found that the depletion of T cells and the neutralization of the signaling protein Interleukin 2 can have a stronger effect when combined with androgen deprivation therapy and vaccines.

Said lead researcher Xiaobo Zhou: “TACC provides an important assistance for discovering clinically meaningful and actionable knowledge across highly heterogeneous biomedical big data sets.”

PROTON THERAPY

Proton therapy causes less damage to surrounding tissues than the commonly used X-ray radiation therapy for tumor irradiation. But the proton beam needs pinpoint accuracy and precise calibration.

Mayo Clinic researcher Wei Liu used Lonestar5 and Stampede1 to develop a model for treatment planning that is more accurate and better at sparing organs than radiation therapy.

“It’s very computationally expensive to generate a plan in a reasonable timeframe,” Liu said. “Without a supercomputer, we can do nothing.”
GENOMICS

The human genome consists of three billion base pairs, so identifying a single mutation by sight simply isn’t possible. Computers, on the other hand, are great at finding patterns in massive datasets and have been a boon to cancer researchers.

Researchers from UT Austin and the National Cancer Institute used Stampede1 to mine massive amounts of data from the Cancer Genome Atlas to identify genetic variants and patient subtypes.

They found a specific mutation in the protein FOXP1 that’s associated with an aggressive type of lymphoma.

“This knowledge can be helpful in the development of more targeted therapies that seek to eliminate cancer at its origin,” said UT Austin geneticist Vishy Iyer.

CANCER DIAGNOSTICS

Manual breast exams and mammograms are currently the most effective and widely-used techniques for early detection of breast cancer. Unfortunately, manual breast exams only provide information about the site where the force is applied, and mammograms (breast X-rays) expose patients to radiation and produce many false positives.

Researchers at the Rose-Hulman Institute of Technology developed an electro-mechanical device that gently indents breast tissue in various locations and records the tissue surface response. This data is then converted into detailed 3D maps which can be used to identify suspicious, stiffer sites for further testing.

They used Stampede1 to map the distribution of stiffness in a given tissue to find which tissue stiffness maps best match the response they see in testing.

“This system has the potential to significantly increase the early detection of breast cancer with no unnecessary radiation, essentially no risk, and with little additional cost,” said Lorraine Olson, lead researcher.

SURGERY

Surgical removal of cancer cells is risky. Removing too little of a tumor can lead to a relapse; too much — especially in a critical area like the brain — can harm the patient.

A pioneering project performed minimally invasive laser treatment on a canine tumor without a surgeon. The project team from UT Austin and MD Anderson Cancer Center used TACC’s advanced computing resources to develop an interactive system that plans, predicts, and dynamically alters the course of a laser treatment for cancer patients.

“The more data and images that can be acquired, the more confidence researchers and surgeons can have in planning surgical simulations,” said David Fuentes of the MD Anderson Cancer Center.

GENOMICS

The human genome consists of three billion base pairs, so identifying a single mutation by sight simply isn’t possible. Computers, on the other hand, are great at finding patterns in massive datasets and have been a boon to cancer researchers.

Researchers from UT Austin and the National Cancer Institute used Stampede1 to mine massive amounts of data from the Cancer Genome Atlas to identify genetic variants and patient subtypes.

They found a specific mutation in the protein FOXP1 that’s associated with an aggressive type of lymphoma.

“This knowledge can be helpful in the development of more targeted therapies that seek to eliminate cancer at its origin,” said UT Austin geneticist Vishy Iyer.
PATIENT-SPECIFIC TREATMENTS

Scientists used TACC’s Lonestar5 supercomputer to develop mathematical models of cancer that predict how the disease will progress in a specific patient.

A tumor’s response to treatments such as chemotherapy is characterized by an equation that captures its behavior. Scientists at the UT Austin Center for Computational Oncology combined these equations with data from patients in a study to simulate tumor development.

“If you have a model that can recapitulate how tumors grow and respond to therapy, then it becomes a classic engineering optimization problem. ‘I have this much drug and this much time. What’s the best way to give it to minimize the number of tumor cells for the longest amount of time?’” said Thomas Yankeelov of Dell Med at UT Austin.

Yankeelov’s group is currently conducting a clinical study in Austin, Texas to predict, after one treatment, how an individual’s cancer will progress. They will use these predictions to plan the course of treatment.

ARTIFICIAL INTELLIGENCE

An emerging way that researchers are using high performance computing for cancer research is through the application of machine and deep learning.

Researchers from Tufts University and the University of Maryland, Baltimore County used TACC’s systems to uncover the complex cellular communication networks that underlie cancer and cellular mutations, and to design methods to disrupt them.

“Getting a true understanding, given the complexity of the information, without some assistance from machine learning, is probably hopeless,” said Michael Levin, of Tufts University. “I think it’s inevitable that we use machine learning to enrich scientific and biomedical discovery.”

CONCLUSION

Whether helping scientists sift through terabytes of genomic data, plan optimal proton treatments, or understand how cancer cells communicate, TACC’s advanced computing resources are pushing the state-of-the-art in cancer research and helping turn the tide on one of the world’s deadliest diseases.
Over the past four years, researchers at the Lamont-Doherty Earth Observatory (LDEO) have been flying over the frozen waters in the polar regions and collecting field data to study the Ross Ice Shelf. The field data includes crucial information on the shelf and the underlying tectonics of the Antarctic region.

The shelf, which can be up to 10,000 feet thick, is the largest of several that hold back West Antarctica’s massive amounts of ice. If these were to collapse, global sea level could rise by up to 10 feet. In fact, NASA this year confirmed that ice loss from this critical region is happening at an increasingly fast pace.

Nick Frearson, a lead engineer on the LDEO project, and other colleagues have collected hundreds of terabytes of data. They need state-of-the-art solutions to process, analyze, and store the information.

“That’s where TACC comes in,” Frearson says. The researchers rely heavily on the Stampede2, Ranch, and Corral systems.

“The partnership we have forged with TACC shows that it’s possible to manage and disseminate this level of data in a cost-effective, user-friendly, and easily accessible manner,” Frearson said. “This data will help people in the science community who are interested in the cryosphere and the changes going on there. We hope that people across the globe will benefit from this data set.”
Through the Science & Technology Affiliates for Research (STAR) program, TACC serves as a partner to more than two dozen companies, ranging from Fortune 100 corporations to startups.

STAR began in 2007 with three members — Chevron, Aramco and Dell EMC — and has grown to include 30 members from energy and engineering to computing and aerospace. TACC helps companies test new hardware, improve software, and do innovative research and development to gain a competitive advantage.

Aramco Services Company used Ranger to run their first-ever billion-cell simulation of an oil reservoir, and then interactively visualized the data from Dhahran, Saudi Arabia.

Using 1 million hours on Ranger, TACC partnered with the American Museum of Natural History to produce “Journey to the Stars,” the most advanced planetarium show ever made.

Star partner Chevron funded the creation of a distributable version of TACC’s Scientific Computing Curriculum, so that universities and HPC centers can fill the student pipeline across all industries.

NASA scientists simulated the heat shield for the Orion capsule on Lonestar5 to better replicate the flight environment.

TACC and Dell teamed up to benchmark HPC applications and hardware, and provide a better understanding of cost-effective, standards-based supercomputer performance and scalability.

Raytheon, TACC, and the National Center for Atmospheric Research (NCAR) developed novel data management and visualization methods to improve high-resolution severe weather simulations around O’Hare International Airport.

STAR partners have supported TACC teams competing in the Student Cluster Challenge at the annual supercomputing conference, SC, for multiple years.

Engineers from Petrobras trained at TACC to apply deep learning techniques to oil field analyses.

BP consulted with TACC to make their new data-center — which hosts the world’s largest supercomputer for commercial research — more energy efficient.

Technip, an oil services company based in Houston, used TACC’s systems to improve the reliability, safety, and environmental quality of its offshore oil platforms.

Firefly Aerospace used Stampede1 to model the material properties of their rocket engine designs and to analyze their performance during flight.

Learn more about the STAR program: useta.cc/star
In an inconspicuous-looking data center on The University of Texas at Austin’s J. J. Pickle Research Campus, construction is underway on one of the world’s most powerful supercomputers.

The Frontera system (Spanish for “frontier”) will allow the nation’s academic scientists and engineers to probe questions both cosmic and commonplace — What is the universe composed of? How can we produce enough food to feed the Earth’s growing population? — that cannot be addressed in a lab or in the field; that require the number-crunching power equivalent to a small city’s worth of computers to solve; and that may be critical to the survival of our species.

The name Frontera pays homage to the “endless frontier” of science envisioned by Vannevar Bush and presented in a report to President Harry Truman calling for a national strategy for scientific progress.

The report led to the founding of the National Science Foundation (NSF) — the federal agency that funds fundamental research and education in science and engineering.

It has been basic United States policy that Government should foster the opening of new frontiers. It opened the seas to clipper ships and furnished land for pioneers. Although these frontiers have more or less disappeared, the frontier of science remains.

— Vannevar Bush, director of the Office of Scientific Research and Development, July 1945

It paved the way for investments in basic and applied research that laid the groundwork for our modern world, and inspired the vision for Frontera.

“Whenever a new technological instrument emerges that can solve previously intractable problems, it has the potential to transform science and society,” said Dan Stanzione, executive director of TACC and one of the designers behind the new machine. “We believe that Frontera will have that kind of impact.”
THE QUEST FOR COMPUTING GREATNESS

The pursuit of Frontera formally began in May 2017 when NSF issued an invitation for proposals for a new leadership-class computing facility, the top tier of high performance computing systems funded by the agency. The program would award $60 million to construct a supercomputer that could satisfy the needs of a scientific and engineering community that increasingly relies on computation.

“For over three decades, NSF has been a leader in providing the computing resources our nation’s researchers need to accelerate innovation,” explained NSF Director France Córdova. “Keeping the U.S. at the forefront of advanced computing capabilities and providing researchers across the country access to those resources are key elements in maintaining our status as a global leader in research and education.”

NSF’s invitation for proposals indicated that the initial system would only be the beginning. In addition to enabling cutting-edge computations, the supercomputer would serve as a platform for designing a future leadership-class facility to be deployed five years later that would be 10 times faster still — more powerful than anything that exists in the world today.

TACC has deployed major supercomputers several times in the past with support from NSF. Since 2006, TACC has operated three supercomputers that debuted among the Top15 most powerful in the world — Ranger (2008-2013, #4), Stampede1 (2012-2017, #7) and Stampede2 (2017-present, #12) — and three more systems that rose to the Top25. These systems established TACC, which was founded in 2001, as one of the world leaders in advanced computing.

TACC solidified its reputation when, on August 28, 2018, NSF announced that the center had won the competition to design, build, deploy, and run the most capable system they had ever commissioned.

“This award is an investment in the entire U.S. research ecosystem that will enable leap-ahead discoveries,” NSF Director Córdova said at the time.

Frontera represents a further step for TACC into the upper echelons of supercomputing — the Formula One race cars of the scientific computing world. When Frontera launches in 2019, it will be the fastest supercomputer at any U.S. university and one of the fastest in the world — a powerful, all-purpose tool for science and engineering.

“Many of the frontiers of research today can be advanced only by computing,” Stanzione said. “Frontera will be an important tool to solve Grand Challenges that will improve our nation’s health, well-being, competitiveness, and security.”

THE ARCHITECTS

When TACC proposed Frontera, it didn’t simply offer to build a fastest-in-its-class supercomputer. It put together an exceptional team of supercomputer experts and power users who together have internationally recognized expertise in designing, deploying, configuring, and operating HPC systems at the largest scale.

Experts at TACC and UT Austin will lead the science application and technology assessment efforts with partners from the California Institute of Technology, Cornell University, Princeton University, Stanford University, University of Chicago, University of Utah, University of California, Davis, Ohio State University, Georgia Institute of Technology and Texas A&M University.

“Frontera is a project that is not just about the system; our proposal is anchored by an experienced team of partners and vendors with a community-leading track record of performance.”

— Dan Stanzione, TACC Leadership
SUPERCOMPUTERS EXPAND THEIR MISSION

Supercomputers have historically had very specific uses in the world of research, performing virtual experiments and analyses of problems that can’t be easily physically experimented upon or solved with smaller computers.

Since 1945, when the ENIAC (Electronic Numerical Integrator and Computer) at the University of Pennsylvania first calculated artillery firing tables for the United States Army’s Ballistic Research Laboratory, the uses of large-scale computing have grown dramatically.

Today, every discipline has problems that require advanced computing. Whether it’s cellular modeling in biology, the design of new catalysts in chemistry, black hole simulations in astrophysics, or Internet-scale text analyses in the social sciences, the details change, but the need remains the same.

“Computation is arguably the most critical tool we possess to reach more deeply into the endless frontier of science,” Stanzione says. “While specific subfields of science need equipment like radio telescopes, MRI machines, and electron microscopes, large computers span multiple fields. Computing is the universal instrument.”

In the past decade, the uses of high performance computing have expanded further. Massive amounts of data from sensors, wireless devices, and the Internet opened up an era of big data, for which supercomputers are well suited. More recently, machine and deep learning have provided a new way of not just analyzing massive datasets, but of using them to derive new hypotheses and make predictions about the future.

As the problems that can be solved by supercomputers expanded, NSF’s vision for cyberinfrastructure — the catch-all term for the set of information technologies and people needed to provide advanced computing to the nation — evolved as well. Frontera represents the latest iteration of that vision.

Data-Driven Design

TACC’s leadership knew they had to design something innovative from the ground up to win the competition for Frontera. Taking a data-driven approach to the planning process, they investigated the usage patterns of researchers on Stampede1, as well as on Blue Waters — the previous NSF-funded leadership-class system — and in the Department of Energy (DOE)’s large-scale scientific computing program, INCITE, and analyzed the types of problems that scientists need supercomputers to solve.

They found that Stampede1 usage was dominated by 15 commonly used applications. Together these accounted for 63 percent of Stampede1’s computing hours in 2016. Some 2,285 additional applications utilized the remaining 37 percent of the compute cycles. (These trends were consistent on Blue Waters and DOE systems as well.) Digging deeper they determined that, of the top 15 applications, 97 percent of the usage solved equations that describe motions of bodies in the universe, the interactions of atoms and molecules, or electron and fluids in motion.

“We did a careful analysis to understand the questions our community was using our supercomputers to solve and the codes and equations they used to solve them,” said TACC’s director of High Performance Computing, Bill Barth. “This narrowed the pool of problems that Frontera would need to excel in solving.”

“Computation is arguably the most critical tool we possess to reach more deeply into the endless frontier of science.”

— Dan Stanzione, TACC

But past use wasn’t the only factor they considered. “It was also important to consider emerging uses of advanced computing resources for which Frontera will be critical,” Stanzione said. “Prominent among these are data-driven and data-intensive applications, as well as machine and deep learning.”

Though still small in terms of their overall use of Stampede2, and other current systems, these areas are growing quickly and offer new ways to solve enduring problems.

Whereas researchers traditionally wrote HPC codes in programming languages like C++ and Fortran, data-intensive problems often require non-traditional software or frameworks, such as R, Python, or TensorFlow.

“The coming decade will see significant efforts to integrate physics-driven and data-driven approaches to learning,” said Tommy Minyard, TACC director of Advanced Computing Systems. “We designed Frontera with the capability to address very large problems in these emerging communities of computation and serve a wide range of both simulation-based and data-driven science.”

Frontera will be the most powerful supercomputer at any U.S. university and likely top 10 in the world when it launches in 2019. It will support simulation, data analysis, and AI on the largest scales.

FRONTERA WILL HELP RESEARCHERS

Explore the origins of the universe
Forecast hurricanes and floods for first responders
Find new drug candidates for cancer, Alzheimer’s, and other diseases
Create new materials for hypersonic flight
The Large Hadron Collider (LHC) at CERN in Geneva, Switzerland, is the world's most powerful particle collider. The LHC is capable of detecting the presence of infinitesimal particles that only exist for a fraction of a nanosecond. From these traces, physicists try to understand what makes up the universe at the most fundamental level.

One of two large, general purpose detectors at the LHC, known as ATLAS (A Toroidal LHC Apparatus), was designed to observe some of the tiniest, yet most energetic, particles ever created on earth. It was one of the two LHC experiments involved in the discovery of the Higgs boson in July 2012.

ATLAS is a physical instrument, but it requires enormous amounts of computing power to interpret its results. Analyzing the signals of particles spewed out during collisions; reconstructing the reactions that led to the outburst; and developing models that explain what happened, are all major computational challenges. A world-wide network of advanced computing systems collaborates on this effort.

In the coming years, the ATLAS team plans to ramp up the number of collisions it produces by a factor of 10. With this increase in activity comes a corresponding need for significantly more computing power.

This is an area where TACC and Frontera will help. In 2017, the ATLAS team began using Stampede2 to run event generators, work that will increase on Frontera.

"We simulate the detector response to a given physics model," said Robert Gardner, a research professor in the Enrico Fermi Institute at the University of Chicago who co-leads the distributed computing facility group for the U.S. ATLAS collaboration.

"When we’re doing the analysis on the actual data, we may plot some distributions such as the particle mass, transverse momentum, or the 'missing energy' in the collision. And you get the number of candidates that we have for the raw data coming off the detector. Then we compare those to different kinds of models and see if we can match up the distributions. This provides clues to what might be actually happening during the collisions."

The Higgs Boson was the last building block predicted by the Standard Model, which explains three of the four fundamental forces in the universe, but not gravity. However it is likely not the end of the story. "Our observations only account for about five percent of the energy thought to make up the universe," Gardner explained. "There’s a huge component out there that is yet to be discovered, and some believe dark matter might provide some of this missing energy. ATLAS has the means to hunt for signs of new types of matter created in LHC collisions, but we need large scale computing and analytics to interpret the data."

Frontera will be useful not only for particle physics, says Gardner, but for all types of 'long-tail science'.

"We need systems like Frontera to answer the big questions of our time, such as the sustainability of the environment and renewable energy," he said. "We have to continue to work on frontier science and everything that comes after it, and we can’t do that without computation."
THE RIGHT CHIPS FOR THE RIGHT JOBS

Anyone following computer hardware trends in recent years has noticed the blossoming of options in terms of computer processors. Today’s landscape includes a range of chip architectures, from low energy ARM processors common in cell phones, to adaptable FPGAs (field-programmable gate arrays), to many varieties of CPU, GPUs and AI-accelerating chips.

The team considered a wide-range of system options for Frontera before concluding that a CPU-based primary system with powerful Intel Xeon x86 nodes and a fast network would be the most useful platform for most applications.

Once built, TACC expects that the main compute system will achieve 35 to 40 petaflops of peak performance. For comparison, Frontera will be twice as powerful as Stampede2 (currently the fastest university supercomputer) and 70 times as fast as Ranger, which operated at TACC until 2013.

To match what Frontera will compute in just one second, a person would have to perform one calculation every second for one billion years.

In addition to its main system, Frontera will also include a subsystem made up of graphics processing units (GPUs) that have proven particularly effective for deep learning and molecular dynamics problems.

“For certain application classes that can make effective use of GPUs, the subsystem will provide a cost-efficient path to high performance for those in the community that can fully exploit it,” Stanzione said.

DESIGNING A COMPLETE ECOSYSTEM

The effectiveness of a supercomputer depends on more than just its processors. Storage, networking, power, and cooling are all critical as well.

Frontera will include a storage subsystem from DataDirect Networks with almost 53 petabytes of capacity and nearly 2 terabytes per second of aggregate bandwidth. Of this, 50 petabytes will use disk-based, distributed storage, while 3 petabytes will employ a new type of very fast storage known as Non-volatile Memory Express storage, broadening the system’s usefulness for the data science community.

Supercomputing applications often employ many compute nodes, or devices, at once, which requires passing data and instructions from one part of the system to another. Mellanox InfiniBand interconnects will provide 100 Gigabits per second (Gbps) connectivity to each node, and 200 Gbps between the central switches.

These components will be integrated via servers from Dell EMC, who has partnered with TACC since 2003 on massive systems, including Stampede1 and 2.

“The new Frontera system represents the next phase in the long-term relationship between TACC and Dell EMC, focused on applying the latest technical innovation to truly enable human potential,” said Thierry Pellegrino, vice president of Dell EMC High Performance Computing.

Though a top system in its own right, Frontera won't operate as an island. Users will have access to TACC’s other supercomputers — Stampede2, Lonestar, Wrangler, and many more, each with a unique architecture — and storage resources, including Stockyard, TACC’s global file system; Corral, TACC’s data collection repository; and Ranch, a tape-based long-term archival system.

Together, they compose an ecosystem for scientific computing that is arguably unmatched in the world.
Whereas the LHC seeks to answer basic questions about the universe, another global physics experiment underway has a much more practical focus: designing a fusion reactor that can solve the world’s power needs without the drawbacks of current sources.

Fusion has been touted for decades as the holy grail of energy production — a way to produce power by merging nuclei and releasing massive amounts energy, the way the Sun does. As demonstrated in the EUROfusion Joint European Torus (JET), progress in fusion has reached the “break-even” milestone — the ability to produce as much energy as is put in.

The next step towards positive energy generation will be carried out by the International Thermonuclear Experimental Reactor (ITER), a joint effort from seven governments. Currently under construction in France, this $25 billion experimental facility is designed to produce 10 to 20 times more power than it uses. The reactor is scheduled to be operational by 2025.

An especially urgent and challenging problem facing the development of a fusion reactor is the need to reliably predict and avoid large-scale major disruptions, which can damage the machine.

After years of trying to predict disruptions using physics models and simulations, researchers still struggled to match the dynamics in a real reactor.

“If you try to use conventional theoretical methods, buttressed by high performance computing, you still aren’t going to be able to make predictions,” said William Tang, principal research physicist at the Princeton Plasma Physics Laboratory — the U.S. DOE National Lab for fusion studies. “You needed the impact of big data analytics that can deal with a lot of data that’s relevant to disruptions.”

In order to accelerate progress, the Princeton AI/Deep Learning Team led by Julian Kates-Harbeck, Alexey Svyatkovskiy, and Tang, developed the Fusion Recurrent Neural Net (FRNN) Code and successfully deployed deep learning to demonstrate exciting advances.

“We adopted a supervised machine learning approach,” he said. “This means that everything involves real physics events — with the pre-disruption classifiers determined by first-principles based physics.” They have demonstrated that their code can predict disruption events with better than 90 percent accuracy and less than 5 percent false positives and can do so more than 30 milliseconds before disruptions are triggered.

Furthermore, they showed for the first time that they could train many neural networks on signals from one reactor and make accurate predictions on a much larger device. In addition, they are now able to make predictions much more than 30 milliseconds before a disruption occurs to enable the device to have a much-expanded disruption avoidance window.

For several years, Tang has used TACC systems to develop a well-known code that simulates particle behavior in burning plasmas. He intends to extend this research on Frontera to develop an actual control system that is capable of avoiding disruptions in ITER. In particular, he is excited about Frontera’s hybrid design that can enable both HPC simulations and machine learning/deep learning, and their possible integration.

“We’ll look forward to bringing these application domains where we’ve had experience with TACC in the past to be integrated into the exciting area of AI and deep learning execution on Frontera,” he said.
NEW MODES OF ACCESS & USE
PORTALS, GATEWAYS, & APIs

Researchers traditionally interact with supercomputers through the command line—a text-only program that takes instructions and passes them on to the computer’s operating system to run.

The bulk of a supercomputer’s time (roughly 90 percent of the cycles on Stampede2) is consumed by researchers using the system in this way. But as computing becomes more complex, having a lower barrier to entry and offering an end-to-end solution to access data, software, and computing services has grown in importance.

Science gateways offer streamlined, user-friendly interfaces to cyberinfrastructure services. In recent years, TACC has become a leader in building these accessible interfaces for science.

“Visual interfaces can remove much of the complexity of traditional HPC, and lower this entry barrier,” Stanzione said. “We’ve deployed more than 20 web-based gateways, including several of the most widely used in the world. On Frontera, we’ll allow any community to build their own portals, applications, and workflows, using the system as the engine for computations.”

Though they use a minority of computing cycles, a majority of researchers actually access supercomputers through portals and gateways. To serve this group, Frontera will support high-level languages like Python, R, and Julia, and offer a set of RESTful APIs (application program interfaces) that will make the process of building community-wide tools easier.

“We’re committed to delivering the transformative power of computing to a wide variety of domains from science and engineering to the humanities,” said Maytal Dahan, TACC’s director of Advanced Computing Interfaces. “Expanding into disciplines unaccustomed to computing from the command line means providing access in a way that abstracts the complexity and technology and lets researchers focus on their scientific impact and discoveries.”

THE CLOUD

For some years, there has been a debate in the advanced computing community about whether supercomputers or “the cloud” are more useful for science. The TACC team believes it’s not about which is better, but how they might work together. By design, Frontera takes a bold step towards bridging this divide by partnering with the nation’s largest cloud providers—Microsoft, Amazon, and Google—to provide cloud services that complement TACC’s existing offerings and have unique advantages.

These include long-term storage for sharing datasets with collaborators; access to additional types of computing processors and architectures that will appear after Frontera launches; cloud-based services like image classification; and Virtual Desktop Interfaces that allow a cloud-based filesystem to look like one’s home computer.

“The modern scientific computing landscape is changing rapidly,” Stanzione said. “Frontera’s computing ecosystem will be enhanced by playing to the unique strengths of the cloud, rather than competing with them.”

SOFTWARE & CONTAINERS

When the applications that researchers rely on are not available on HPC systems, it creates a barrier to large-scale science. For that reason, Frontera will support the widest catalog of applications of any large-scale scientific computing system in history.

TACC will work with application teams to support highly-tuned versions of several dozen of the most widely used applications and libraries. Moreover, Frontera will provide support for container-based virtualization, which sidesteps the challenges of adapting tools to a new system while enabling entirely new types of computation.

With containers, user communities develop and test their programs on laptops...
Physicists have been using supercomputers since the 1940s. But biologists and life scientists constitute a newer group of computational researchers. The Human Genome Project showed how DNA sequencers and computers can combine forces to sequence genes and develop insights from their relationships. Its success spurred the widespread adoption of advanced computing in the life sciences.

Ed Buckler, a research geneticist at the U.S. Department of Agriculture and a member of the National Academy of Sciences, is among the biologists excited to use Frontera to enable important new findings.

"Global protein production needs to double in 40 years in order to feed the planet," Buckler said. "The computational modeling of crop genomes and their interactions with a wide range of environments will allow us to design a food system that can sustainably feed the planet for the coming century."

Buckler studies how corn genomes interact with their environment. The goal is to plant the right crops in the right places at the right times, and develop varieties that are heat- and drought-tolerant.

In Buckler’s experiments, rovers and drones travel over, on, and under fields gathering vast amounts of data related to the health of the crops in a given environment. Combining this approach with gene sequencing offers up new insights, but requires massive computing power.

"The goals are to make maize varieties that require fewer inputs, such as fertilizer and water, while at the same time being adapted to extreme weather," Buckler said. "A lot of our questions require the scale of computing that Frontera offers."

The methods Buckler is pioneering for maize can be applied to many other crops, as well as to the management of forests and other resources. But it’s not just Frontera’s ability to crunch data that’s important, Buckler says.

"Frontera will train the next generation of scientists on how to analyze data and what kinds of questions we can address with computing," he said. "It will bring together the scientific fields with engineering and computer science."
or in the cloud, and then transfer those same workflows to HPC systems using programs like Singularity. This facilitates the development of event-driven workflows, which automate computations in response to external events like natural disasters, or for the collection of data from large-scale instruments and experiments.

“Frontera will be a more modern supercomputer, not just in the technologies it uses, but in the way people will access it,” Stanzione said.

A FRONTIER SYSTEM TO SOLVE FRONTIER CHALLENGES

The LHC modeling effort, fusion disruption predictions, and genomic analyses represent the types of ‘frontier,’ Grand Challenge research problems Frontera will help address.

“Many phenomena that were previously too complex to model with the hardware of just a few years ago are within reach for systems with tens of petaflops,” said Stanzione.

A review committee made up of computational and domain experts will ultimately select the projects that will run on Frontera, with a small percentage of time reserved for emergencies (as in the case of hurricane forecasting), industry collaborations, or discretionary use.

It’s impossible to say what the exact impact of Frontera will be, but for comparison, Stampede1, which was one quarter as powerful as Frontera, enabled research that led to nearly 4,000 journal articles. These include confirmations of gravitational wave detections by LIGO that contributed to a Nobel Prize in Physics in 2016; discoveries of FDA approved drugs that have been successful in treating cancer; and a greater understanding of DNA interactions enabling the design of faster and cheaper gene sequencers.

From new machine learning techniques to diagnose and treat diseases to fundamental mathematical and computer science research that will be the basis for the next generation of scientists’ discoveries, Frontera will have an outsized impact on science nationwide.

PLANNING FOR PHASE 2 – LOOKING BEYOND THE FRONTIER

The NSF program that funds Frontera is titled, “Towards a Leadership-Class Computing Facility.” This phrasing is important because, as powerful as Frontera is, NSF sees it as a step toward even greater support for the nation’s scientists and engineers. In fact, the program not only funds the construction and operation of Frontera — the fastest system NSF has ever deployed — it also supports the planning, experimentation, and design required to build a system in five years that will be 10 times more capable than Frontera.

“We’ll be planning for the next generation of computational science and what that means in terms of hardware, architecture, and applications,” Stanzione said. “We’ll start with science drivers — the applications, workflows, and codes that will be used — and use those factors to determine the architecture and the balance between storage, networks, and compute needed in the future.”

Much like the data-driven design process that influenced the blueprint for Frontera, the TACC team will employ a “design — operate — evaluate” cycle on Frontera to plan Phase 2.

TACC has assembled a Frontera Science Engagement Team, consisting of a more than a dozen leading computational scientists from a range of disciplines and universities, to help determine the “the workload of the future” — the science drivers and requirements for the next generation of systems. The team will also act as liaisons to the broader community in their respective fields, presenting at major conferences, convening discussions, and recruiting colleagues to participate in the planning.

Fusion physicist William Tang joined the Frontera Science Engagement Team in part because he believed in TACC’s vision for cyberinfrastructure. “AI and deep learning are huge areas of growth. TACC definitely saw that and encouraged that a lot more. That played a significant part in the winning proposal, and I’m excited to join the activities going forward,” Tang said.

A separate technology assessment team will use a similar strategy to identify critical emerging technologies, evaluate them, and ultimately develop some as testbed systems.

TACC will upgrade and make available their FPGA testbed, which investigates new
ways of using interconnected FPGAs as computational accelerators. They also hope to add an ARM testbed and other emerging technologies.

Other testbeds will be built offsite in collaboration with partners. TACC will work with Stanford University and Atos to deploy a quantum simulator that will allow them to study quantum systems. Partnerships with the cloud providers Microsoft, Google, and Amazon, will allow TACC to track AMD (Advanced Micro Devices) solutions, neuromorphic prototypes and tensor processing units.

Finally, TACC will work closely with Argonne National Laboratory to assess the technologies that will be deployed in the Aurora21 system, which will enter production in 2021. TACC will have early access to the same compute and storage technologies that will be deployed in Aurora21, as well Argonne’s simulators, prototypes, software tools, and application porting efforts, which TACC will evaluate for the academic research community.

“The primary compute elements of Frontera represent a relatively conservative approach to scientific computing,” Minyard said. “While this may remain the best path forward through the mid-2020’s and beyond, a serious evaluation of a Phase 2 system will require not only projections and comparisons, but hands-on access to future technologies. TACC will provide the testbed systems not only for our team and Phase 2 partners, but to our full user community as well.”

Using the “design — operate — evaluate” process, TACC will develop a quantitative understanding of present and future application performance. It will build performance models for the processors, interconnects, storage, software, and modes of computing that will be relevant in the Phase 2 timeframe.

“It’s a push/pull process,” Stanzione said. “Users must have an environment in which they can be productive today, but that also incentivizes them to continuously modernize their applications to take advantage of emerging computational technologies.”

The deployment of two to three small scale systems at TACC will allow the assessment team to evaluate the performance of the system against their model and gather specific feedback from the NSF science user community on usability. From this process, the design of the Phase 2 leadership class system will emerge.

WITH GREAT POWER COMES GREAT RESPONSIBILITY

The design process will culminate some years in the future. Meanwhile, in the coming months, Frontera’s server racks will begin to roll into TACC’s data center. From January to March 2019, TACC will integrate the system with hundreds of miles of networking cables and install the software stack. In the spring, TACC will host an early user period where experienced researchers will test the system and work out any bugs. Full production will begin in the summer of 2019.

“We want it to be one of the most useable and accessible systems in the world,” Stanzione said. “Our design is not uniquely brilliant by us. It’s the logical next step — smart engineering choices by experienced operators.”

It won’t be TACC’s first rodeo. Over 17 years, the team has developed and deployed more than two dozen HPC systems totaling more than $150 million in federal investment. The center has grown to nearly 150 professionals, including more than 75 PhD computational scientists and engineers, and earned a stellar reputation for providing reliable resources and superb user service. Frontera will provide a unique resource for science and engineering, capable of scaling to the very largest capability jobs, running the widest array of jobs, and supporting science in all forms.

The project represents the achievement of TACC’s mission of “Powering Discoveries to Change the World.”

“Computation is a key element to scientific progress, to engineering new products, to improving human health, and to our economic competitiveness. This system will be the NSF’s largest investment in computing in the next several years. For that reason, we have an enormous responsibility to our colleagues all around the U.S. to deliver a system that will enable them to be successful,” Stanzione said. “And if we succeed, we can change the world.”
SCIENCE ON REPEAT

TACC enhances computational reproducibility
Trust, but verify. The well-known proverb speaks to the heart of the scientific method, which builds on the results of others but requires that data be collected in a way that can be repeated with the same results. Scientific reproducibility extends beyond just recreating the conditions of a physical experiment. The computational analysis of data also factors into the reproducibility equation.

"Computational reproducibility is a subset of the broader and even harder topic of scientific reproducibility," said Dan Stanzione, TACC's executive director.

Computational reproducibility can be difficult to achieve. Even working with the same data, one analysis might yield small differences from another. Stanzione explained that a computer’s hardware and software systems can change a lot over time, partially because of software upgrades such as security patches. Changes in scientific software libraries, operating system components, and computer hardware upgrades can slightly alter results.

"If we can’t get the exact same answer bit-for-bit, then what’s close enough? What’s a scientifically valid way to represent that? That’s what we’re after," Stanzione said.

"Reproducibility means many things to many people, because it is in fact many things, and it has many aspects," said Doug James, former deputy director for High Performance Computing at TACC. James cited a definition for reproducibility by Lorena Barba, associate professor of mechanical and aerospace engineering at George Washington University. "She describes reproducibility as conducting your research as if someone might want to do it again. That means traceability, automation, and transparency. It means the ability to survive inspection by one’s peers, to give them the confidence that if they needed or wanted to do this again, they could," James said.

Researchers can control the software configuration of their workstations, but not on the systems at TACC and other supercomputing centers. But James explained that TACC has developed tools for researchers to control their software environments. One example is the Lmod module system, written and maintained by TACC’s manager of HPC Software Tools, Robert McLay.

"The module system has commands that offer insight into what you can load and have loaded. It allows you to save, preserve, and quickly recover your favorite collections of software so that you can come to the table tomorrow with the same software that you had today. And in particular, it’s designed to make managing and controlling the software environment easy and repeatable for the individual user. That’s the kind of thing that TACC does in the supercomputing environment to promote and enhance reproducibility," James said.

Other tools developed at TACC to enhance reproducibility include XALT, which keeps track of the software packages, libraries, and versions used to execute essentially any job, workflow or command on the system; and TACC Stats, which leverages XALT metadata to gauge how efficiently the software uses its resources.

TACC also enhances reproducibility through the expanded use of containers, added Stanzione. "Containers are a very lightweight technology for virtualization. We can store not only the code that you used, but also the environment around it, the operating system, and the libraries. That way we can go back and get the same software environment and store the whole thing," he said. Docker, an open source container platform, is one of the main tools used for this purpose.

Another opportunity for TACC to enhance reproducibility comes from the increasing use of science gateways and web services, which provide a portal for TACC systems without users having to build and compile their own code. Software versions, workflows, and other metadata can be stored for later use. "If we can preserve the data and then build a container to preserve the software environment, we have a lot of pieces that help make science more reproducible," Stanzione said. "Our strategy is to keep pushing those technologies forward and expose our users to the best practices for enhancing computational reproducibility."
Scientific computing applications traditionally start from “first principles,” mathematical formulas representing the physics of a system. They then transform those formulas into forms that can be solved by distributing the calculations to many processors.

By contrast, machine and deep learning rapidly scan very large datasets to find subtle connections in information. From those connections, they can rapidly generate, test, and optimize solutions. These capabilities let scientists derive the governing models (or workable analogs) for complex systems that cannot be modeled from first principles.

Machine learning uses algorithms that “learn” from data and improve performance based on real-world experience. Deep learning, a branch of machine learning, relies on large data sets to iteratively “train” many-layered neural networks inspired by the human brain. These trained neural networks then use a process known as inference to apply what was learned to new data.

Training can be a complex and time-consuming activity. But once a model has been trained, it is fast and easy to interpret each new piece of data to recognize, for example, cancerous versus healthy brain tissue, or enable a self-driving vehicle to identify a pedestrian crossing a street.

IN SEARCH OF DEEP LEARNING TRAINERS: HEAVY COMPUTATION REQUIRED

Just like traditional HPC, training a neural network or running a machine learning algorithm requires extremely large numbers of computations — quintillions! — making them a good fit for supercomputers and their tens of thousands of parallel processors.

In 2018, researchers from TACC, the University of California, Berkeley and the University of California, Davis, set a record for the fastest training of a neural network to classify images. The team used large batch sizes — essentially big chunks of a dataset — and the newly developed Layer-wise Adaptive Rate Scaling (LARS) algorithm, to speed up deep neural networks training on Stampede2.

They tested the effectiveness of the method on two neural networks, AlexNet and ResNet-50, trained with the ImageNet-1k dataset — a common benchmark for deep neural network training. Using 2,048 Intel Xeon processors, they reduced the 100-epoch AlexNet training time from hours to 11 minutes. They also
reduced the 90-epoch ResNet-50 training time from an hour to 20 minutes. (An epoch refers to one full pass through the entire training dataset.) Their approach showed higher test accuracy than previous efforts by Facebook researchers on batch sizes larger than 16,000.

The research won the Best Paper award at the 2018 International Conference on Parallel Processing.

High-speed, high-accuracy image classification can be used to characterize satellite imagery for environmental monitoring or to label nanoscience images obtained by scanning electron microscopes.

"These results show the potential of using advanced computing resources, like those at TACC, along with large mini-batch enabling algorithms, to train deep neural networks interactively and in a distributed way," said Zhao Zhang, a research associate at TACC. "Given our large user base and huge capacity, this will have a major impact on science."

TACC has, in recent years, created a complete ecosystem for machine learning and deep learning-based research. This includes systems like Maverick2 that are tailor-made for training deep neural networks, and support for a number of popular deep learning frameworks, including Caffe, MXNet, and TensorFlow.

These investments will impact the speed of science, as well as the kinds of projects that researchers can explore with these new methods.

Engineers from UT Austin and the University of Pennsylvania used Stampede2 to train a brain tumor identification and classification system that can recognize tumors with greater than 90 percent accuracy, roughly equivalent to an experienced radiologist. The team will soon begin using the system for clinical studies of brain tumors.

Researchers from Tufts University and the University of Maryland, Baltimore County, used machine learning on Stampede1 to model the cellular control network that determines how tadpole pigmentation develops and to reverse-engineer tadpoles with a form of mixed pigmentation never before seen in nature.

Atmospheric scientists from the University of Oklahoma used Stampede2 to develop machine learning-derived hail forecast algorithms for severe weather forecasting that will one day be able to warn the public about hailstorms hours in advance.
WHY IS TACC’S TRAINING IMPORTANT FOR TODAY’S WORKFORCE?

More and more science is dependent on advanced computing resources. But compared to the growing need, a smaller and smaller percentage of scientists are capable of integrating advanced computing resources into their research. So there’s a discrepancy in the percentage of the workforce needing advanced computing skills, and the actual percentage that have these skills. And with technologies changing so rapidly, without training and knowledge sharing, this discrepancy will only grow.

The point of our institutes is to reduce that discrepancy and broaden the number of people who can use advanced computing in their science, whether it’s for research, industry, or as a facilitator.
WHAT IS THE TACC INSTITUTES SERIES?
The TACC Institute Series is a series of one-week immersive training workshops held here at TACC. We’ve identified different topics that we specialize in — HPC leadership; visualization; advanced computing; designing and administering large-scale systems; machine learning; and computational techniques for data science. The institutes are taught by the experts at TACC using TACC systems. We want to essentially immerse participants in our environment, so they can learn the skills they need and see what our people do on a day-to-day basis.

WHAT MAKES THE INSTITUTES SPECIAL?
What makes the institutes really stand out is that they each focus on a different type of technology or approach. But they’re all taught by a combination of people from different groups. At any given institute, we’ll have someone from our large-scale systems group give a talk introducing the different systems at TACC. We might have someone from the visualization group come and talk about how one visualizes the data that one generates. Or someone from our HPC group might give an introduction to parallel programming. It’s the combination of these people that makes the TACC Institute Series special.

TACC is more than just a 15,000-square-foot data center with over 20 HPC and advanced computing resources in it. TACC is really the people behind the scenes who put those resources together; apply their knowledge to build the software tools and packages that make those machines useable; build the portals through which thousands of scientists can access HPC; develop our cloud environments; and help scientists and engineers make ground-breaking discoveries.

WHO ARE THE INSTITUTES FOR?
There’s been a vast array of attendees from every field of science — a broad spectrum of students, researchers, post-docs, and even directors of other centers who have decided they want to come learn about what makes TACC great.

Some participants have been industry researchers. For instance, a developer from Intel came to our visualization institute to learn more about how to apply visualization to the work that Intel is doing.

Other participants have come from natural hazards research centers, trying to understand how to implement parallel programing and machine learning to solve their problems. A lot of students came this summer, because we introduced a student discount to try and get more students to learn different techniques here at TACC. So there’s a range of different participants, and we strive to make sure the institutes are useful to all of them.

WHAT’S THE NEXT STEP FOR THE INSTITUTES?
I’d like to expand the role of the institutes and encourage more cross-pollination of concepts from our other departments. I think every institute can benefit from scientific visualization and a stronger data-intensive computing component. These are technologies and strategies that every domain science needs.

Also, I’d like to see a one-week institute focused mainly on undergraduate and graduate students. I want to make sure that the students aren’t feeling overwhelmed by the introduction of unfamiliar concepts.

Being an instructor and lecturer at UT Austin, I’ve seen that it’s not just the hard sciences that can benefit from advanced computing techniques. History majors can benefit from advanced computing, as can paleontologists, psychologists, and even artists.

There are so many ways the social sciences and liberal arts can benefit from our systems. What I’d love to do is create an introduction to advanced computing and have it applied beyond the hard sciences, beyond the typical advanced computing and HPC user. If we can get those students to understand how to use advanced computing techniques in their realms, I think we’ve done our job.

2019 INSTITUTES INCLUDE:

ADVANCED COMPUTING FOUNDATIONS
Introductory level programming for advanced architectures. (Also available in web cast)

ADVANCED COMPUTING ESSENTIALS
Intermediate level course exploring all aspects of advanced computing. Get hands on with MPI, OpenMP, machine learning, and much more. (Student rates available)

DESIGNING AND ADMINISTERING LARGE-SCALE SYSTEMS
Best practices for managing, maintaining, and building large-scale Linux clusters.

HPC LEADERSHIP
Tailored for managers and decision makers to enhance the use of advanced computing within their organizations.

ADDITIONAL WEEK-LONG DEEP DIVE INSTITUTES

- Machine Learning Foundations
- Workflows & Reproducibility in Scientific Computing
- Applied Parallel Programming
- Visualizing & Interacting with Data
- Computational Science in the Cloud

For a full listing and details about TACC institutes, visit:
useta.cc/institutes
The Stampede2 supercomputer picked up where its predecessor, Stampede1, left off as a computing powerhouse for thousands of researchers across the U.S. Funded by NSF, Stampede2 entered production in November 2017, offering users 18 petaflops of peak performance — roughly equivalent to the processing power of 100,000 desktop computers, or one for every seat in UT Austin’s Darrell Royal Stadium. Over 3,000 direct users and 10,000 web service and gateway users have successfully completed over 1.5 million jobs on Stampede2. “The reason we build and deploy these systems is for them to be instruments of science,” said Dan Stanzione, TACC Executive Director. Stampede2 has helped scientists design new DNA readers; explain what holds protons together; understand how black holes impact star formation, and more. Here are some highlights from Stampede2’s first year of operations.
Molecular dynamics simulation of DNA capture and translocation through a graphene nanopore. (Credit: Aleksei Aksimentiev)

DNA CAPTURE BY NANOPORES

Surprising things happen when an electric field is applied to water near narrow pores in very thin materials, according to research on Stampede2 by Aleksei Aksimentiev, a computational biophysicist from the University of Illinois at Urbana–Champaign.

His team found that an electric field can compress water locally, preventing molecules from being transported through small pores. This effect, he believes, can be harnessed to create a new type of DNA sequencer. Aksimentiev and his team published their findings in the *Physical Review Letters* in June 2018.

The simulations took into account the motion of more than 100,000 atoms, which was critically important for the discovery of the phenomenon.

“It’s been amazing how fast and how accurate the Stampede2 machine works,” Aksimentiev said.

James Wilson, a post-doc working with Aksimentiev, added that “by running the simulations on Stampede2, we were able to finish twenty simulations in a couple of days, cutting down our time to solution immensely.”

SEDIMENT TRANSPORT

Stampede2 has allowed researchers to perform grain-resolving simulations of cohesive sediment dynamics involving thousands of discrete particles, in work co-authored by Eckart Meiburg, director of the Center for Interdisciplinary Research in Fluids at the University of California at Santa Barbara.

“These simulations provide insight into a wide range of situations of great importance to the environment, such as the transport of mud, silt, clay, and nutrients by flows in rivers, lakes and the oceans,” Meiburg said.

The simulations were carried out in parallel with experiments conducted on the International Space Station. The researchers expect this combination of simulations and experiments to lead to the development of novel models for cohesive sediment transport that will enhance our predictive modeling capabilities for rivers, lakes, and estuaries.

“Stampede2 has enabled us to perform computational simulations for problems that had been intractable to date, and to address parameter regimes in our simulations that have so far been inaccessible,” Meiburg said.
Magnetized winds and turbulence drive the growth, or accretion, of disks in space, be they around black holes or protoplanets. Zhaohuan Zhu of the University of Nevada, Las Vegas, modeled the relative importance of turbulence and disk wind in accretion using numerical simulations on Stampede2. He co-authored work that appeared in the Astrophysical Journal in April 2018.

"In the field of astronomy, to do cutting edge research and advance our knowledge, we need the best tools. For observers, they need big telescopes. For theorists, we need large computers to understand physical processes in more detail," Zhu said.

For research on protoplanetary disks and planet formation, recent observations provide so much detail that previous low-resolution simulations are not sufficient to explain them.

"Stampede2 is absolutely essential for both observational and theoretical astronomy," Zhu said.

**ASTROPHYSICAL ACCRETION DISCS**

Stampede2 simulations are helping Diego Donzis, an aerospace engineer at Texas A&M University, study shock turbulence interactions, a critical phenomenon that must be accounted for when designing supersonic and hypersonic aircraft.

"Stampede2 is allowing us to run simulations, some of them at unprecedented levels of realism, needed to study processes that depend both on the large and the small scales of turbulent flows," Donzis said.

Stampede2 lets Donzis and his team run a wide range of simulations and conditions for a range of problems.

"Some of these simulations ended up forming some of the largest databases of their kind," he added. These in turn will help engineers design the high-speed aircraft of the future.

**TURBULENT FLOW**

When turbulent flows are in thermal non-equilibrium, the molecular energy modes in the fluid become important to the dynamics of the flow. (Credit: Diego Donzis)
Scientists will take what they’ve learned from particle physics simulations on Stampede2 and test their results on the planned Electron-Ion Collider at Brookhaven National Laboratory.

“Stampede2 enabled me and my collaborators to calculate new aspects of the gluon structure of the proton for the first time,” said Phiala Shanahan, assistant professor of Physics at the Massachusetts Institute of Technology. (Gluons are the force-carrying particle of the strong force that bind the proton together.)

“The quantities we calculated will be able to be measured for the first time at the planned Electron-Ion Collider, a new particle collider that’s currently in development. Being able to set theory benchmarks ahead of first experimental measurements is extremely exciting,” Shanahan said.

Resources such as Stampede2 are exactly what are needed to do groundbreaking scientific calculations efficiently. “They’re essential tools,” Shanahan added.

---

**INSIDE THE PROTON**

Scientists will take what they’ve learned from particle physics simulations on Stampede2 and test their results on the planned Electron-Ion Collider at Brookhaven National Laboratory.

“Stampede2 enabled me and my collaborators to calculate new aspects of the gluon structure of the proton for the first time,” said Phiala Shanahan, assistant professor of Physics at the Massachusetts Institute of Technology. (Gluons are the force-carrying particle of the strong force that bind the proton together.)

“The quantities we calculated will be able to be measured for the first time at the planned Electron-Ion Collider, a new particle collider that’s currently in development. Being able to set theory benchmarks ahead of first experimental measurements is extremely exciting,” Shanahan said.

Resources such as Stampede2 are exactly what are needed to do groundbreaking scientific calculations efficiently. “They’re essential tools,” Shanahan added.

---

**HIV-1 TARGETS**

Stampede2 helped model a key building block in the HIV-1 protective capsid, which could lead to strategies for potential therapeutic intervention in its replication. Scientists found the naturally-occurring compound inositol hexakisphosphate (IP6) promotes both the assembly and the maturation of HIV-1.

“We discovered, in collaboration with other researchers, that HIV-1 uses this small molecule to complete its function,” said Juan R. Perilla, Department of Chemistry and Biochemistry, University of Delaware. “HIV-1 has evolved to make use of these small molecules present in our cells to essentially be infectious.” Perilla co-authored the study in the journal *Nature* in August 2018.
Though TACC’s reach is nationwide, its heart bleeds burnt orange. Since its inception, the center has aimed to help scholars throughout The University of Texas System’s 14 institutions.

The University of Texas Research Cyberinfrastructure (UTRC) initiative provides advanced computing systems, large data storage, and high bandwidth data access between the institutions, enabling researchers to collaborate and compete at the forefront of science and discovery. The efforts have led to impressive discoveries at every UT System institution.

To learn more about how TACC is aiding Texas researchers, check out our special report: useta.cc/serving-the-state
The Square Kilometre Array (SKA) will be among the first scientific instruments requiring exascale computing. And TACC supercomputers are playing a role in preparing researchers to use it.

The SKA — whose early science observations are expected to start in 2020 — will be the world’s largest radio telescope, 50 times more sensitive than any other, with thousands of small dishes that combined total one square kilometer spread out mainly over South Africa, with some in Australia.

Processing the daily data, estimated at an exabyte, will require exascale supercomputers. But that’s not all.

“There is going to be a huge demand for high performance computing expertise in South Africa through projects like the Square Kilometre Array,” said Happy Sithole, director for the Centre for High Performance Computing (CHPC) in South Africa and a champion of supercomputing in Africa.

In 2013, the CHPC received 20 racks of the decommissioned Ranger supercomputer from TACC, which the CHPC distributed to Botswana, Ghana, Kenya, Madagascar, Mauritius, Mozambique, Namibia, and Zambia.

Sithole takes stewardship of Ranger and Stampede1 seriously. “It is the responsibility of South Africa to make sure that, when we roll out the SKA, there are sufficient skills on high performance computing that can be utilized on the Square Kilometre Array.”

What’s more, African researchers will use their advanced computing knowledge to reach beyond just the starry African skies.

Sithole added, “It doesn’t just stop at the SKA, as the processing capabilities and expertise that will be generated by the SKA can be used for other...”
domains, such as health, energy, mineral processing, and climate modeling.

The partnerships TACC forged in Africa earned it a 2018 Readers’ Choice Award from HPCwire in its Workforce Diversity Leadership category. TACC, Cambridge University, CHPC, Dell, and the Department of Science & Technology (South Africa) were recognized in the award for their efforts with the HPC Ecosystems Project.

"The HPC Ecosystems Project is an initiative of the CHPC and is responsible for the distribution and re-deployment of decommissioned HPC hardware as mid-tier systems to research institutions both locally within South Africa and regionally across Africa," said Bryan Johnston, senior HPC technologist in the Advanced Computer Engineering Lab at CHPC and also the project lead for HPC Ecosystems.

HPC Ecosystems focuses on HPC adoption and skills development to prepare for projects such as the Square Kilometre Array. "TACC continues to provide invaluable assistance through ongoing collaboration, a regular pipeline of HPC resources for re-purposed deployment in Africa, and continuing opportunities for the development of human capital," Johnston said.

The Computer History Museum, home to the largest collection of computers and related materials in the world, recently accepted a piece of TACC’s history into their permanent collection — sealing its place as a milestone in computing.

"We’re always searching around the world for new, interesting, and important computing objects," said Dag Spicer, senior curator at the museum, "and TACC’s Sun Microsystems 2007 Magnum switch was a critical part of high performance computing at that time in history. The TACC switch was the largest of its class and is an example of InfiniBand technology, of which we had few examples."

The switch was part of TACC’s Ranger supercomputer and connected tens of thousands of Ranger’s processors together into a blazingly fast high-speed network. Ranger debuted as the fifth most powerful computer in the world on the June 2008 Top 500 list, and for a time was the most powerful supercomputer for open science research — up to 50,000 times more powerful than a PC at that point in history.

Said Gordon Bell, a pioneer in high performance computing and co-founder of the museum, "Behind nearly every artifact, exhibit, and pioneering effort is a story that the museum is dedicated to understand and tell."
CIRCULATION IN THE ARCTIC OCEAN

Increases in atmospheric temperatures have coincided with the shrinking of the Greenland ice sheet and Arctic glaciers, and a decrease in Arctic summertime sea ice coverage. An improved understanding of how heat and freshwater are transported by ocean currents is essential to assessing the impact of these changes on the Arctic climate system, on Greenland's contribution to global sea level rise, and on the global climate system.

Using the Stampede2 and Maverick supercomputers, and ParaView visualization software, TACC visualization expert Greg Foss illustrates daily-averaged subsurface seawater temperatures in the Arctic Ocean.

The differing salinity and temperature of the various waters entering the Arctic Ocean, as well as Earth's rotation and the steep bottom topography, all impact how these currents interact with each other, flow through, and impact the region.

Capturing detailed features is key for faithfully representing the complex and evolving Arctic atmosphere, ocean, and cryosphere systems, and is impossible without high performance computing resources.

Research:
An Nguyen, Arash Bigdeli, Victor Ocaña, and Patrick Heimbach
Institute for Computational Engineering and Sciences, Jackson School of Geosciences, UT Institute for Geophysics, UT Austin

Visualization:
Greg Foss
Texas Advanced Computing Center, UT Austin

Arctic map:
Saddle Up Partners!

TACC is headed to the far frontiers of science and we want you by our side. Whether you're a researcher, student, educator, CEO, software designer, or science aficionado, we invite you to join us as an employee, a user of our systems, an industrial affiliate, or a friend of the center. Visit us online or in person to learn about the many ways you can partner with TACC.

TACC.UTEXAS.EDU

TACC.UTEXAS