XSEDE

ACCELERATING
SCIENTIFIC
DISCOVERY

XSEDE

Extreme Science and Engineering
Discovery Environment

PARTNER INSTITUTIONS

Cornell University Center for Advanced Computing
Indiana University
Jülich Supercomputing Centre
Michigan State University
National Center for Atmospheric Research
National Center for Supercomputing Applications - University of Illinois at Urbana-Champaign
National Institute for Computational Sciences - University of Tennessee Knoxville/Oak Ridge National Laboratory
Ohio Supercomputer Center - The Ohio State University
Pittsburgh Supercomputing Center - Carnegie Mellon University/ University of Pittsburgh
Purdue University
Rice University
San Diego Supercomputer Center - University of California, San Diego
Shodor Education Foundation
Southeastern Universities Research Association
Texas Advanced Computing Center - The University of Texas at Austin
University of California, Berkeley
University of Chicago
University of Southern California
University of Virginia

John Towns
XSEDE PI and Project Director

WELCOME

to the annual highlights publication of the Extreme Science and Engineering Discovery Environment (XSEDE). As the fifth issue of this publication, it is exciting and gratifying to see the stories that have been selected and how they have continued to represent the larger collection of world class research and education activities supported by XSEDE. There was a tremendous portfolio of articles from which to choose. This issue represents the work of more than 8,000 researchers participating in more than 2,700 projects, producing more than 3,500 publications during the past year.

XSEDE has matured significantly over the five years it has been providing resources and services to the community. The organization has come a great distance from the early days of a freshly forged team of partners without significant experience in such close collaboration to an organization well aligned with its goals and committed to attaining a shared vision: a world of digitally-enabled scholars, researchers, and engineers participating in multidisciplinary collaborations while seamlessly accessing computing resources and sharing data to tackle society's grand challenges. More pointedly, research now requires more than just supercomputers, and XSEDE represents a step toward a more comprehensive and cohesive set of advanced digital services through our mission: to substantially enhance the productivity of a growing community of scholars, researchers, and engineers through access to advanced digital services that support open research; and to coordinate and add significant value to the leading cyberinfrastructure resources funded by the NSF and other agencies.

Five years ago, XSEDE was looking forward to accomplishing a great deal, and it has. XSEDE has supported over 17,000 publications, collected during the program period. Analysis of these publications has indicated that publications supported by XSEDE are, on average, cited twice as often as other publications in the same journal, indicting the greater impact these publications are having in the various domains of research supported by XSEDE. Thus far (July 2011-April 2016), over $2.13 billion in funded research projects have had allocations on XSEDE resources to support those research activities. The growth in size of the community is complemented by a greater diversity of domains from which those researchers come. Researchers from computational finance, machine learning, textual analysis, phylogeny, genomics, archaeology, and digital humanities have begun to make use of the national cyberinfrastructure ecosystem with the support of XSEDE.

Looking forward, XSEDE has submitted a proposal to the National Science Foundation (NSF) to continue support of XSEDE for an additional five years. That proposal suggests a number of changes induced by the lessons we have learned over the past five years in deploying and supporting an ever-advancing infrastructure within the broader cyberinfrastructure ecosystem. While currently under consideration by the NSF, it is hoped that by the time this publication is released, some very positive news can be shared with the community about the future availability of the resources and services XSEDE has provided to support the national research enterprise.

XSEDE looks forward to an opportunity to continue to deliver value and to accelerate scientific discovery!

Each story in this @XSEDEscience Highlights book comes w/ its own tweet! Enjoy the stories and share with your friends on social media!

RESCUED HISTORY

We say, "History is written by the victors." But it's probably more true to say it's written by the people who have the opportunity to write.

One example of this phenomenon is the study of Black women, their lives and their experiences. Many earlier documents don't mention them directly, though these works may offer clues. Those that do mention Black women often are historically obscure. Until recently researchers had no good way of recovering this "lost history."

Ruby Mendenhall, an associate professor of Sociology, African American Studies, Urban and Regional Planning and Social Work at the University of Illinois at Urbana-Champaign, is leading a collaboration of social scientists and digital researchers that hopes to harness the power of high-performance computing to find and understand the historical experiences of Black women. They're using XSEDE resources to analyze two massive databases of written works from the 18th- through 20th centuries, along the way creating a common toolbox that can help other digital humanities projects.

"With a Big Data approach we get a chance to make use of hundreds of thousands of texts—journals, books, periodicals," says Mendenhall. "The number is greater than what you would normally be able to look at during an entire career."

Researchers use XSEDE resources to recover Black women's "lost history" by analyzing literary works from 18th to 20th centuries.

Powering Up

To search tens or even hundreds of thousands of books, articles and letters, the researchers needed considerably more computing power than available on a typical campus "cluster" of commodity computers. They consulted with colleagues on campus who were XSEDE members, identifying the now-retired XSEDE resource *Blacklight* at the Pittsburgh Supercomputing Center (PSC) as a good fit for their project.

With help from PSC's Sergiu Sanielevici, head of XSEDE's New and Innovative Projects group in the Extended Collaborative Support Service (ECSS), they adapted their work to *Blacklight*. They have since moved the project to PSC's interim *Greenfield* system, a precursor to the new XSEDE resource at PSC, *Bridges*.

"We chose *Blacklight* specifically because the tools we're using need huge amounts of shared memory," a strength of that system, says Mark Van Moer, an XSEDE staff member at the National Center for Supercomputing Applications at Illinois who worked as the team's visualization specialist.

The researchers analyzed 20,000 documents known to contain information about Black women in the HathiTrust and JSTOR databases to create a computational model. They're now using this model on *Greenfield* to study the entire 800,000 documents in both databases.

Making Sense of Words via Numbers, Graphics

To make sense of the numbers, the investigators turned to two sets of computational techniques: topic modeling and data visualization.

Topic modeling looks at how often certain key words associate with each other. For example, a book that contains the word "negro"—at the time considered the most respectful term to describe Black men and women—the word "vote" and the word "women" might offer clues about Black women's participation in the women's suffrage movement. Mike Black, then at Illinois and currently at the University of Massachusetts, headed this topic modeling project for the team.

"We're hoping, in the next stage, to ramp up and check these topics against the larger corpus of works," Mendenhall adds. "Do the 'recovery' part."

Van Moer's data visualization work is building ways of displaying results in a way that helps humans make more intuitive sense of them. A "tree map" displays key words in boxes that correspond to each word's frequency. A "network graph" charts how often key words appear close to each other, offering insight based on context. Yet another visualization technique plots key terms in histographs, allowing users to track a given topic's prominence over time.

Making Sense of the Numbers

An initial finding confirmed that documents that referenced the post-World-War-I Black Women's Club movement also mentioned the New Negro movement. The Illinois researchers hope to begin answering questions about how these movements interacted in their current work on XSEDE systems.

"The beauty of computation and Big Data lies in how it complements the traditional close reading," says Nicole Brown, a postdoctoral fellow in Mendenhall's group who is interpreting the computational results in light of Black feminist theory. "The two methods complement each other to give you a full picture of what's going on."

"Working with the social science and humanities people has been a real eye opener," adds Van Moer. "Previously, I worked with physical scientists. Humanities and social science researchers are worried about more than what the numbers mean. They have a whole theory behind how you go about interpreting things as they relate to the larger society—that's really interesting to me."

Another goal will be to create a computational toolbox that researchers in many fields can use to search various texts for topics of interest—and how they relate.

"We're generally interested in Black women and their life experience," Mendenhall says. "But we also see this as a tool that social scientists and people in the humanities can use to study many topics."

Above: Sculptor Edmondia Lewis (1844-1907) was the first woman of African- and Native-American descent to achieve notoriety in the fine arts world. She spent most of her career in Rome. Credit: Henry Rocher – National Portrait Gallery, Smithsonian Institution, Public Domain

Left: Harriet Tubman, famous as an abolitionist, Underground Railroad leader and women's suffrage pioneer. Credit: H. B. Lindsley – National Portrait Gallery, Smithsonian Institution, Public Domain (PD-1875)

HAVE = HAD + CHANGE

Everybody has a favorite teacher from high school or professor from college, and these folks teach eager minds about every subject imaginable. But who prepares these instructors to excite, energize, and educate the future generations of American scientists, technologists, engineers, mathematicians, and scholars in all fields of study?

XSEDE's education team, which includes Steve Gordon and Kate Cahill from the Ohio Supercomputer Center, and Bob Panoff and Aaron Weeden from the Shodor Education Foundation, are leading this charge across the country. They work side-by-side with faculty across the country to help them incorporate computational and data-enabled technologies, resources, and methods into the curriculum in all fields of study. They conduct this effort through a collection of campus visits, strategic planning sessions, tutorials, workshops and online courses.

"The mission of the XSEDE education program is to prepare a diverse community of researchers, scholars, educators, and practitioners in the use of data analysis and management, modeling, simulation, and visualization techniques," said Gordon.

"We welcome inquiries from educators in all fields of study to discuss opportunities for extending this impact to classrooms across the nation," said Cahill.

Panoff added, "Our workshops are reaching a national audience of faculty and K-12 teachers, who in turn teach the future practitioners, innovators and leaders within a diverse mix of institutions including two- and four-year colleges and universities, minority serving institutions, and K-12 schools."

Even though computation has become the third pathway to discovery—along with theory and experimentation—many faculty have not integrated the use of computation into their classes. "Our workshops provide explicit examples of computational models and tools that can easily be assimilated into a variety of classroom situations," said Panoff.

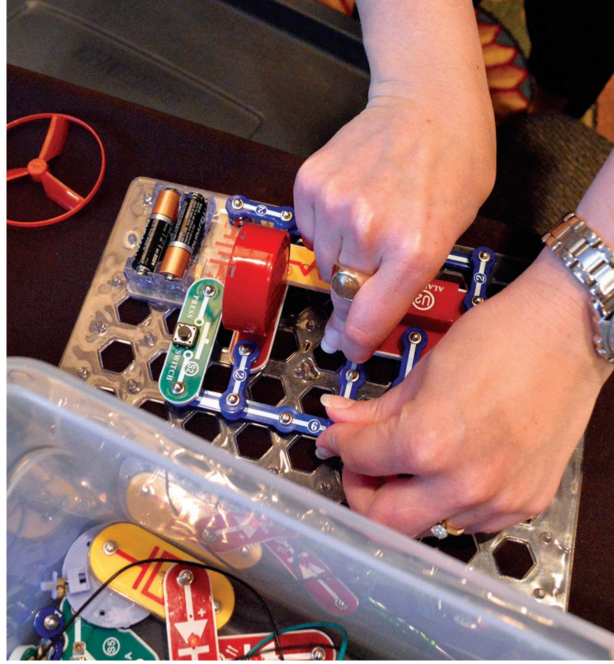
Below is a sampling of the range of education workshops hosted by XSEDE:

- *Computational Biology for Biology Educators*
- *Computational Chemistry for Chemistry Educators*
- *Computing MATTERS: Inquiry-Based Science and Mathematics Enhanced by Computational Thinking*
- *Computational Physics for Use in the Undergraduate Modern Physics Course*
- *Intermediate Parallel Computing: Introduction to Parallel Computing*
- *Introduction to Computational Thinking*
- *Kinder, Gentler Supercomputing: Using HPC Resources and Visualization Tools*
- *LittleFe Curriculum Module Buildout*

"What we've seen is educators want—and their students need—this type of material for a changing and evolving workforce. Computational science, including parallel computing and basic modeling knowledge, is essential for the future of all fields of scholarly research and applications in industry," said Gordon.

The response from the educators who have attended these courses has been overwhelmingly positive, and numerous faculty members indicate that they have used materials from the workshops in their classes.

.@XSEDEscience helps educate nation's educators: from junior high to graduate classes, in STEM subjects and even national security.



"I incorporated the information about Vensim that I learned at the seminar into my Modeling and Simulation (CSCI 323) course that I subsequently taught," said Don Allison, a faculty member at SUNY-Oneonta in the computer science department. "I also purchased one of the recommended books on computational science to serve as a starting point to develop a future course in the topic for our institution."

Stephen Angel, Washburn University: "I attended an introduction to computational chemistry course many years ago and developed a freshman lab exercise upon return. It has been well-received by students and faculty."

Melanie Correll, University of Florida: "I use materials for my ABE 4662 –Quantification of Biological Processes for projects and homework assignment ideas for more than 50 undergraduate engineering students over the past couple of years."

Eric Frost at San Diego State University even uses XSEDE materials for a graduate class on Homeland Security, which shows XSEDE's value outside of physics or chemistry courses. "Materials were used as people try to understand how modeling, simulation, and statistics are used in regard to public safety and security. Where lives are at stake for making real decisions and computational sciences can assist, XSEDE has been wonderfully helpful in being able to be a bridge from mathematical complexity to practical decision-making and confidence in approximating results."

Leonard Freidhof at Eastland Vocational School in northwest Illinois said, "I have been using information I picked up in my high school chemistry and physics classes to show virtual events in kinetics, thermodynamics, equilibrium and velocity, acceleration, pseudoderivations and more. My students create spreadsheets to show what can happen when variables change in the equations we use and the use of graphs of this data to visualize the changes as they vary. And we also have a Lego robotics class in 7th and 8th grade and this helps with computational thinking as well, at a younger age."

In summary, Gordon concluded, "The key is that faculty have used the examples from the workshops to make substantial changes in their courses and curriculum to include computational science. Those changes often take several years beyond the workshops to get put into place. However, they clearly have made significant impacts, in some cases resulting in entire new courses, for example on modeling or computational chemistry, while in others using computational science to enhance mathematics education for a wide range of students."

LEARNING MORE ABOUT ECSS

One of the most robust services offered by XSEDE is its Extended Collaborative Support Service, or ECSS. This service provides expert staff to assist research teams for either a short or an extended period using large-scale computing resources in order to advance the team's work.

Any U.S. researcher with an XSEDE allocation can request ECSS support. Researchers can receive help in areas such as code optimization, optimizing complex workflows, porting to new system architectures, and using of accelerators such as GPUs for detailed visualizations, all at no cost to their grants.

Comprising roughly one-third of XSEDE's annual budget, the ECSS staff covers a wide variety of expertise, and is sized to support about 50 ECSS projects and 50 training events per year. Nancy Wilkins-Diehr of San Diego Supercomputer Center (SDSC) and Ralph Roskies of Pittsburgh Supercomputing Center (PSC), both co-PIs of XSEDE, lead ECSS. While most ECSS staff members come from research sites that have also received NSF grants to operate large-scale resources, the group also includes staff from other institutions who bring valuable expertise in areas such as digital humanities, workflows, and science gateways.

"We find this expertise increasingly necessary as the domain areas, the programming methods, and the advanced cyberinfrastructure resources allocated by XSEDE continue to diversify," said Wilkins-Diehr.

The ECSS program, which evolved from the Advanced Support for TeraGrid Applications (ASTA) in the NSF program that preceded XSEDE, historically focused on improving code performance on supercomputers, including performance analysis, optimization, debugging, and on the use of specific computing architectures.

One recent example of ECSS's assistance involved a project led by Ramon Huerta, a research scientist with UC San Diego's BioCircuits Laboratory, to place chemical sensors – known as "e-noses" – in the homes of elderly participants in an effort to continuously and non-intrusively monitor their living environments. Huerta's lab developed software that could detect anomalous behavior that would suggest a change in a resident's health status. This software was paired with sensor data and was ported to SDSC's *Gordon* supercomputer.

"We were able to improve the performance of the code through a combination of smart compiler choices and options, replacement of calls to the generic math library with calls to Intel's highly tuned MKL (Math Kernel Library), parallelization of key loops, and restructuring of the code to greatly reduce the number of linear algebra operations," said Robert Sinkovits, who leads SDSC's scientific applications efforts. "The end result was that calculations which previously took nearly twelve hours to finish could be completed in just a few minutes, with a total speedup of 167x relative to the original code."

Today, the expertise of ECSS staff members extends to science gateway development and workflow systems, the use of R, Java, and MATLAB, database optimization, cloud computing, data mining, and more. Moreover, while ECSS was established to provide researchers the support and expertise needed to conduct computationally-based science, the service also gives them an opportunity to learn new computational techniques.

"Most ECSS staff members have advanced degrees in scientific fields as well as technology expertise, and that means they speak the language of the researchers," according to PSC's Roskies. "While our domain expertise includes areas traditionally associated with high-end computing such as astrophysics, computational chemistry, climate modeling, and molecular dynamics, we are also prepared to assist researchers in emerging disciplines – areas that haven't traditionally used advanced computation such as genomics, machine learning, or predictive analytics."

How ECSS Works

ECSS assistance is requested through the XSEDE allocations process, the same process used to request CPU, data, and other resources. Researchers are encouraged to identify areas where they might need specific assistance at the same time that they are planning their computational objectives. ECSS can be requested with small startup allocations as well as at any point in an allocation lifecycle as needs develop. ECSS collaborations can last from a number of weeks up to one year, and renewals are possible.

At the top level, ECSS is divided into two main areas: ECSS Projects, led by Roskies, where work is done primarily with individual research groups, and ECSS Communities, led by Wilkins-Diehr, where work benefits larger scientific communities. The areas work together closely, with ECSS staff often shifting back and forth between the two areas. Sub groups include:

Within ECSS Projects:

- *Extended Support for Research Teams (ESRT), led by Lonnie Crosby of National Institute for Computational Sciences (NICS): This is a collaborative effort between an XSEDE user group and one or more ECSS staff members, with the goal of enhancing the research group's capability to transform knowledge using XSEDE-allocated resources and related technologies. These projects currently comprise about 60% of all ECSS projects.*
- *Novel and Innovative Projects (NIP), led by Sergiu Sanielevici of PSC: This group seeks out projects by researchers in communities that don't typically use high-performance computing and advanced cyberinfrastructure – areas that may include bioinformatics, economics, the social sciences, and arts and humanities.*

Within ECSS Communities:

- *Extended Support for Community Codes (ESCC), led by John Cazes of Texas Advanced Computing Center (TACC): ESCC's focus is on deploying, hardening, and optimizing software systems necessary for extensive research communities to create new knowledge using XSEDE resources and related technologies. ESCC staff also evaluate software that might be appropriate for installation on XSEDE resources.*
- *Extended Support for Science Gateways (ESSGW), led by Marlon Pierce of Indiana University: This group's primary mission is to provide support to existing and new scientific communities that would like to use XSEDE resources through Science Gateways. ESSGW does not directly develop and support gateways, but supports science gateway operators via integration with XSEDE resources and choice of technologies.*
- *Extended Support for Training, Education and Outreach (ESTEO), led by Jay Alameda of NCSA: ESTEO and Training staff survey existing online training materials delivered by all service providers and identify new areas where materials might benefit the community. Online training is increasingly popular with the user community when both time and travel budgets are limited.*

More information about ECSS can be found at https://www.xsede.org/ecss.

.@XSEDEscience Extended Collab. Support Service helps researchers across diff. domains use advanced computational tools to advance their work.

CIREN TO THE RESCUE

More than 33,000 Americans die in motor vehicle crashes annually, according to the Centers for Disease Control and Prevention. Modern restraint systems save lives, but some deaths and injuries remain—and restraints themselves can cause some injuries.

"Crash-test dummies" help engineers design safer cars, but provide only limited information about forces the body experiences during impact. Computer models of vehicle crashes, on the other hand, provide more sophisticated information on how to improve restraints and other safety systems. The models also help researchers simulate the effects of thousands of variables that would be impossible to assess in physical crash tests.

"There's really limited information you can get from a crash-test dummy," says Ashley A. Weaver, an assistant professor at the Virginia Tech-Wake Forest University Center for Injury Biomechanics and a former NSF graduate research fellow. "The [virtual] human body model gives us much more, predicting injuries in organs that aren't in that dummy, such as lung contusions."

.@XSEDEscience resources power study of virtual "crash-test dummies," pointing to safer passenger restraints.

Turning on the CIREN

With Joel Stitzel and graduate students and staff from the Center for Injury Biomechanics, Weaver used as their starting point a database of real-world vehicle crashes from the National Highway Traffic Safety Administration-sponsored Crash Injury Research and Engineering Network (CIREN).

"Wake Forest is a CIREN center, enrolling four to five cases per month," Weaver says. "Engineers and physicians discuss the injuries and how the vehicles may have contributed. The result is a database that researchers can use to model injuries computationally."

CIREN's anatomical detail allowed the scientists to detect injuries to soft-tissue organs not present in mechanical dummies.

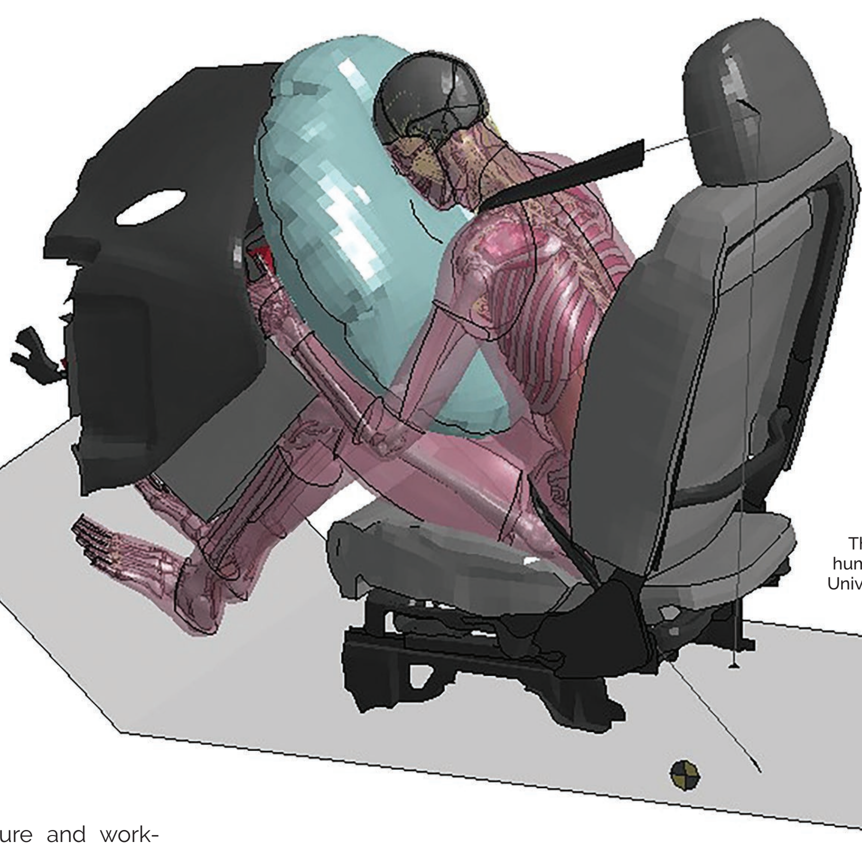
"Lung injuries are really important in respect to elderly vehicle occupants," Weaver says. "They put patients at risk for pneumonia or acute respiratory distress syndrome. In addition, the human body model gives us the ability to predict fractures with much more detail than do dummies." For example, she adds, the model helps identify which specific areas of the rib cage are experiencing forces that can break bones, instead of only calculating the overall force on the chest.

Stepping up to XSEDE power

Initially, the team used the DEAC Cluster at Wake Forest University for their work. The simulations used virtual versions of the Toyota Camry and Chevrolet Cobalt.

But the researchers found they needed more computational power than they could get from a campus computational cluster like DEAC. Assistance from XSEDE Extended Collaborative Support Service (ECSS) expert David O'Neal at Pittsburgh Supercomputing Center (PSC) was instrumental in their success. The architecture of XSEDE resource *Blacklight* at PSC, they found, would be optimal for rapid assessment of the large numbers of simulations they needed to run in parallel.

The ECSS team—staff with expertise in advanced computing as well as the scientific background needed to understand each user's research at a deep level—helped the Wake Forest team set up the cyberinfrastructure and workflows they needed to run the simulations. Using the Total Human Model for Safety, developed by the Toyota Central Research and Development Laboratory, Weaver and her team showed that simulations can reproduce real-world injury patterns and predict details crash-test dummies can't provide.



This simulated real-world crash uses a human-body model. Credit: Wake Forest University Center for Injury Biomechanics

Results: Tailored restraints?

The simulations allowed the researchers to quantify the sensitivity and uncertainty of the injury risk predictions based on occupant position, which is difficult to determine in real-world motor vehicle crashes. They demonstrated how injury-causing stress moves from the foot to the lower leg as a driver's head comes forward into a frontal airbag. They also discovered that more reclined seating positions can lead to a higher risk of head and chest injuries.

Weaver and her colleagues published their results in *Traffic Injury Prevention* in October 2015. The findings can help auto manufacturers design safer vehicles and restraint systems. They have since moved their work to PSC's interim *Greenfield* system, a precursor to the new XSEDE resource at PSC, *Bridges*.

"I would like to see our work used to evaluate safety changes," Weaver says. "One of the things we would be interested in is helping in the design of adaptive restraints that can be tuned to the age and size of the occupant." Such a system might shift the stress of an impact to a part of an occupant's body that is better able to sustain the shock, whether a frail elder, a child or a healthy adult.

"Of course we'd need to test new restraints computationally before moving into the crash-test phase," she adds.

TINY ZAPS, BIG RESULTS

When nanotechnology researchers make 'small talk,' it's extremely small. They are interested in physical phenomena at one-billionth of a meter – a million times shorter than the length of an ant, or 100,000 times thinner than a human hair. Despite the size of their subject, the benefits to society are huge.

In fact, according to the National Nanotechnology Initiative, more than 800 everyday commercial products rely on nanoscale materials and processes for a variety of applications in medicine, energy, information technology, and many other areas.

One method of nanotechnology research involves the use of short laser pulses at minuscule fractions of a second to produce structural changes in thin, localized surface regions of materials like gold, silver, or silicon. Leonid Zhigilei of the University of Virginia said that what attracted him to this type of research is the ability of lasers to excite and change materials in ways not possible with any other technique.

For example, short-pulse laser processing can transform a surface from very water attractant to very water repellant, reducing friction, erosion, and contamination on items that need to be kept clean, such as roof tiles and skyscraper windows.

Because laser-induced processes are complex and happen so fast, experimentation cannot provide a detailed understanding of the structural transformations triggered by the rapid laser energy deposition, Zhigilei said.

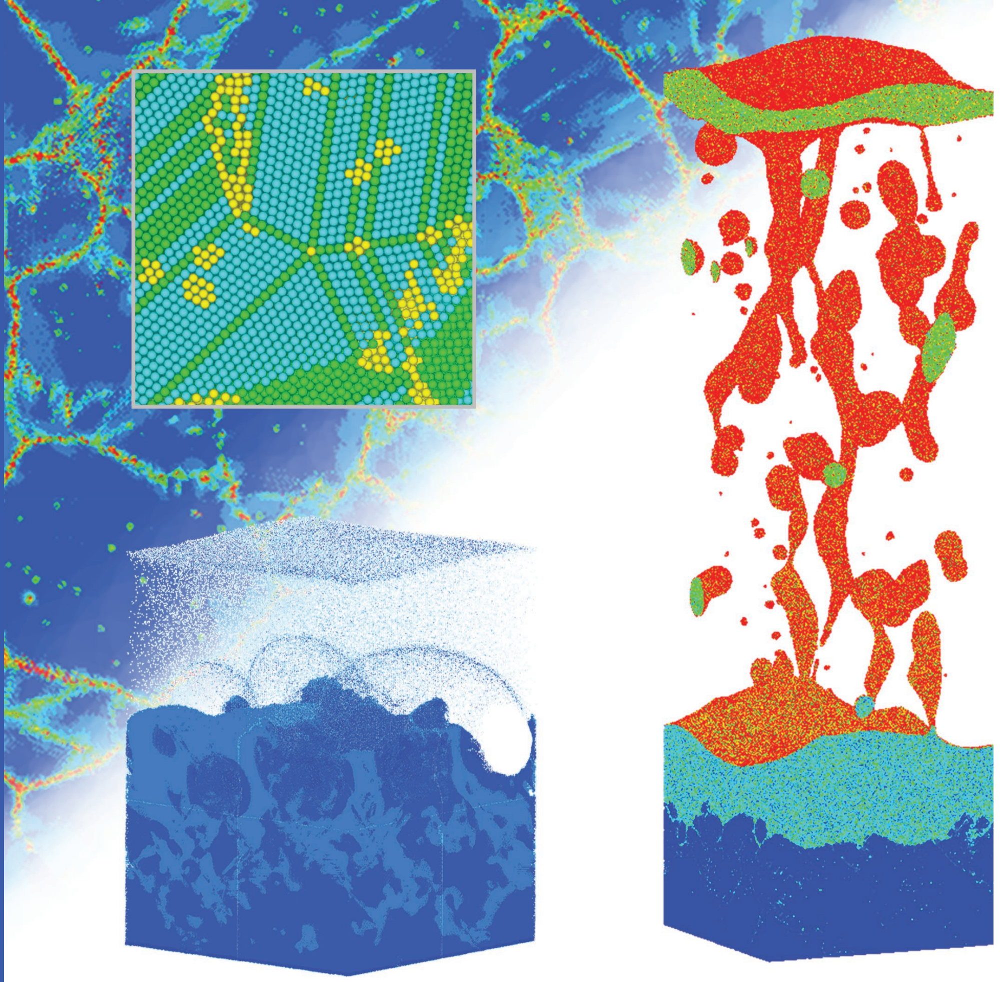
Chengping Wu, a member of Zhigilei's computational materials group, added: "Our atomistic simulations, on the other hand, provide clear visual representations, or 'atomic movies.' They are well-suited to reveal the relationships between the properties of laser-treated regions of the targets and the underlying microscopic mechanisms of laser-induced target modification."

Zhigilei's group uses the *Darter* supercomputer at the National Institute for Computational Sciences (NICS) and *Stampede* at the Texas Advanced Computing Center (TACC). High-performance computing has played a crucial role in many of their projects because the systems they simulate can consist of up to a billion atoms.

"Unlike in real experiments," Zhigilei explained, "the analysis of non-equilibrium processes in molecular dynamics can be performed with unlimited atomic-level resolution, providing complete information of the phenomena of interest."

XSEDE made Zhigilei's compute allocations on *Darter* and *Stampede* possible, as well as provided additional benefits. "When we write our allocation requests, we often ask for time on different computers, and we also take advantage of other XSEDE resources, of course," he said. "We use VisIt software for visualizations. In addition, many of my students enjoy the file transfer service, Globus Online, which is very efficient in moving large files that we are generating in our simulations."

> XSEDE supports research in nanotechnology, where complex laser processes occur at lightning speed. And although the scale is tiny, the results for society are huge.



An illustration of some of the results of large-scale atomistic computer simulations of laser-induced structural modification of silver targets irradiated by 200 femtosecond (one quadrillionth of a second) laser pulses. The simulations make predictions regarding the surface and explain the experimental observation of surface swelling. The experiments were done in the group of Peter Balling at Aarhus University, Denmark. Illustration courtesy of Leonid Zhigilei, the University of Virginia.

During 2015, Zhigilei's group published two papers in the journal *Physical Review* with different collaborators. Working with Henry Helvajian of The Aerospace Corporation, they discovered they could use surface acoustic waves (SAWs) to move tiny particles of gold on a silicon surface. In the article, they note that the use of SAWs has broad implications for applications in which heating must be avoided.

Collaboration with the experimental group of Peter Balling of Aarhus University, Denmark, led to an article titled "Generation of subsurface voids and a nanocrystalline surface layer in femtosecond laser irradiation of a single-crystal Ag target." The research showed that the volume of a metal on a surface could be increased, thereby providing new opportunities for tailoring surface properties to the needs of practical applications.

Going forward, Zhigilei's group will study how metal nanoparticles efficiently convert laser energy absorbed at a surface, a capability that is key to a growing number of imaging and therapeutic biomedical techniques, Wu said.

"For example, researchers have demonstrated that laser irradiation of gold or silver nanoparticles attached to gene markers and delivered to specific cells can be used for selective killing of cancer cells or bacteria," Wu explained. "In the area of drug delivery, doping the walls of microcapsules with metal nanoparticles opens a way for the remote release of encapsulated material into specific cells by targeting metal nanoparticles with near-infrared laser irradiation."

Yet another illustration of big implications for humanity found at the nanoscale.

CIPRES OPENS GATEWAY

An outline for a new tree of life, depicting the evolution of life on this planet that included more than 1,000 new types of bacteria and Archaea lurking in the Earth's nooks and crannies, was made possible with the help of supercomputing resources available through XSEDE, along with a phylogenetics "gateway" available on those resources.

The new tree, published in April 2016 in the new journal *Nature Microbiology* and widely publicized throughout the general press, reinforces once again that the life we see around us – plants, animals, humans and other so-called eukaryotes – represents but a tiny percentage of the world's biodiversity.

"The tree of life is one of the most important organizing principles in biology," said Jill Banfield, a University of California, Berkeley professor of earth and planetary science, policy, and management, and the study's principal investigator. "The new depiction will be of use not only to biologists who study microbial ecology, but also biochemists searching for novel genes and researchers studying evolution and earth history."

Researchers used the CIPRES (CyberInfrastructure for Phylogenetic RESearch), a web-based portal or gateway that allows researchers to explore evolutionary relationships between species.

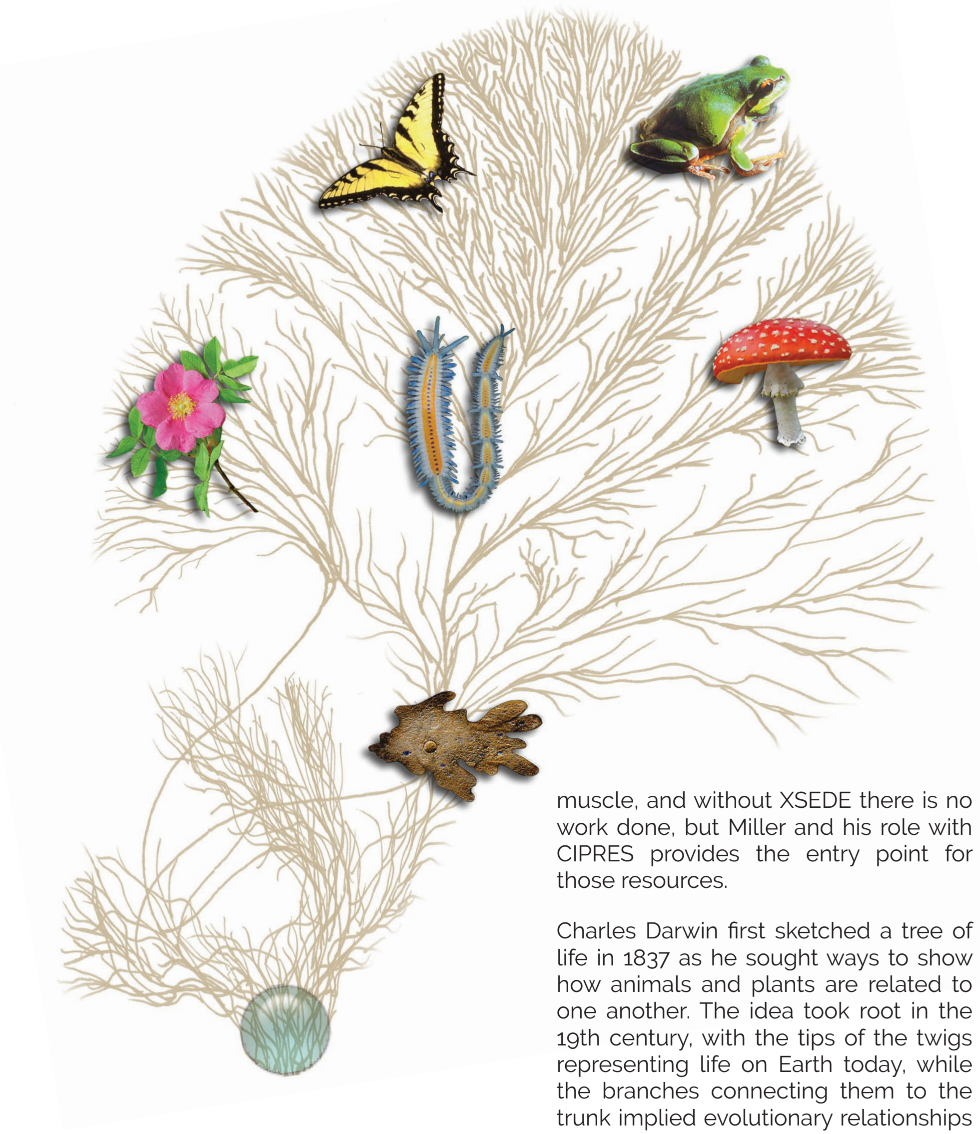
"The CIPRES Science Gateway was critical to our work," said Laura Hug, who computed the trees at the University of California, Berkeley and is now a biology faculty member at the University of Waterloo (Canada). "Previous attempts to infer the trees presented severe problems with run time, memory allocation and a lack of parallelized implementation of the RAxML (Randomized Axelerated Maximum Likelihood), a popular program for phylogenetic analysis of large datasets. No run had successfully finished prior to our introduction to CIPRES."

Access to supercomputers also was a key part of completing this study, helping researchers investigate relationships by comparing DNA sequences information between species. This type of analysis is becoming more powerful as the number of DNA sequences available is increasing rapidly, with new, larger data sets requiring higher levels of computational power.

In using the CIPRES gateway, the researchers relied on two other XSEDE resources: *Gordon*, the first high-performance supercomputer to use massive amounts of flash-based SSD (solid state drive) storage, and *Comet*, a petascale supercomputer designed to transform advanced scientific computing by expanding access and capacity among traditional as well as non-traditional research domains. The two jobs ran for a total of about five days, using 48 cores. Both *Gordon* and *Comet*, the result of two NSF grants, are housed at the San Diego Supercomputer Center (SDSC) at the University of California, San Diego.

Mark Miller, principal investigator for CIPRES, thinks of the science gateway as a broker for XSEDE: he helps bring people in to use CIPRES through one XSEDE account. XSEDE and its infrastructure provide all the computational muscle, and without XSEDE there is no work done, but Miller and his role with CIPRES provides the entry point for those resources.

> A new tree of life showing 1000 new types of bacteria developed using @XSEDEscience resources including the CIPRES phylogenetics gateway.



Charles Darwin first sketched a tree of life in 1837 as he sought ways to show how animals and plants are related to one another. The idea took root in the 19th century, with the tips of the twigs representing life on Earth today, while the branches connecting them to the trunk implied evolutionary relationships among these creatures.

Since then, researchers have built on Darwin's initial sketch, gradually adding twigs and branches to the original, including bacteria and Archaea — microbes that live in extreme environments such as hot springs and oxygen-free wetlands. Our current view of the tree of life consists of three major branches: Eukaryotes, including animals, plants, fungi and protozoans; familiar bacteria such as Escherchia coli; and the Archaea.

The revolution in DNA sequencing during the past couple of decades, however, has led to an explosion of data offering yet more complete descriptions of the genetic relationships among species. Researchers began sequencing whole communities or organisms at once and picking out the individual groups based on their genes alone. This metagenomic sequencing revealed whole new groups of bacteria and Archaea, many of them from extreme environments, such as the toxic puddles in abandoned mines, the dirt under toxic waste sites, and the human gut. Some of these had been detected before, but nothing was known about them because they wouldn't survive when isolated in a lab dish.

For the new paper, Banfield and Hug, along with more than a dozen other researchers who have sequenced new microbial species, gathered 1,011 previously unpublished genomes to add to already known genome sequences of organisms representing the major groups of life on Earth.

Their investigation, representing the total diversity among all sequenced genomes, produced a tree with branches dominated by bacteria, especially by uncultivated bacteria. A second view of the tree grouped organisms by their evolutionary distance from one another rather than current taxonomic definitions, making clear that about one-third of all biodiversity comes from cultivated bacteria, one-third from uncultivated bacteria, and one-third from Archaea and Eukaryotes.

THE SCIENCE OF JET NOISE

Daniel Bodony's love of science began with a love of airplanes. On weekends, he worked for one of his dad's colleagues who had an airplane. "I would mow his grass and he would let me fly," Bodony remembers fondly.

Those early boyhood days fueled the fire for Bodony as he committed himself to a career as a military pilot but, at that time, a pilot who wore glasses was not allowed. Taking that in stride, Bodony decided he would instead design airplanes, only to experience another shift in his early career.

"When I got to college and started to design airplanes I realized that I liked the science behind the design more than I liked the design itself," he said.

Bodony, the Blue Waters Associate Professor in Aerospace Engineering at the University of Illinois at Urbana-Champaign, is looking into the science surrounding the aeroacoustics of jet engines and researching how to make them quieter.

A veteran user of NSF high performance computing (HPC) resources since 2008, Bodony said: "The reason we use supercomputers is because in aeroacoustics there is no simple relationship that relates an unsteady flow field to the sound it creates. So we have to resort to elaborate experiments or simulations to try to come up with the contextual underpinnings that relate cause and effect. And we still haven't done it. The fact that aircraft has gotten quieter over the years is more by accident than by design, and we're trying to change that, but it relies on bigger calculations, bigger codes, and more complex computing capabilities."

The computational challenges that Bodony and his team face invariably involve turbulence, which is an unsteady, chaotic motion of a fluid.

"It's a classical multi-scale problem," Bodony said. "Computational research is required to resolve all of those scales which requires us to use the largest computers to which we have access, XSEDE's *Stampede* being one of them."

XSEDE'S Extended Collaborative Support Service (ECSS)

Through XSEDE's Extended Collaborative Support Service (ECSS) program, researchers have access to cyberinfrastructure experts with a variety of expertise.

Bodony has used the ECSS program for a variety of projects. When asked if he would recommend the ECSS program to other researchers, his response was a quick, "Yes, wholeheartedly. The ECSS experts are able to look at the code and understand the hardware and software very quickly to make a diagnosis."

Currently, through XSEDE's ECSS program, Bodony works primarily with Luke Wilson from the Texas Advanced Computing Center (TACC). Wilson, who serves as the technical expert—with knowledge about the hardware, and how the software interacts with the hardware—is helping Bodony and his team achieve real performance results on their code.

"When we run our code we have a basic idea of what its weaknesses are, and we try to identify the biggest weakness that impacts our ability to run efficiently on XSEDE systems, including being able to utilize *Stampede's* Intel® Xeon Phi™ processors," Bodony says.

According to Wilson, "The goal has always been to get this code up and running on the Intel® Xeon Phi™, and we started out looking for some simple places we could target to improve performance, mostly through vectorization. We found that the data encoding and the original algorithm weren't well suited to the Xeon Phi… there was a lot of out-of-order memory access, which you can't vectorize very easily."

At first, Wilson executed a simple performance analysis of the code and identified the algorithmic weaknesses to find better ways to express the algorithms. Sometimes the algorithms needed new data structures, sometimes an entirely new algorithm needed to be implemented to perform the same operation, and sometimes the researchers needed to rewrite part of the code that made that algorithm no longer necessary.

It was a team effort among Bodony, Wilson and several people at Illinois, but Wilson was instrumental in taking the cumulative view of the code and speeding it up by a factor of seven. How? He figured out where the performance bottlenecks were.

"We still have a long way to go with jet noise and we're going to continue to follow jet noise for the foreseeable future," Bodony said. "We think the flow that exits the jet engines contains the information that we need to figure out how to make jet engines quieter. We just haven't probed it in the right way. We've been working on tools to extract that information. Our current hypothesis has shown that our idea has merit using small scale simulations and now we're applying these ideas to the full scale jet noise problem."

What's next for Bodony and team?

According to Bodony, future computers are going to look a lot different than *Stampede* or any of the other NSF-funded systems. Now, they're not using ECSS to focus on performance; they're working with ECSS to change how the code is programmed.

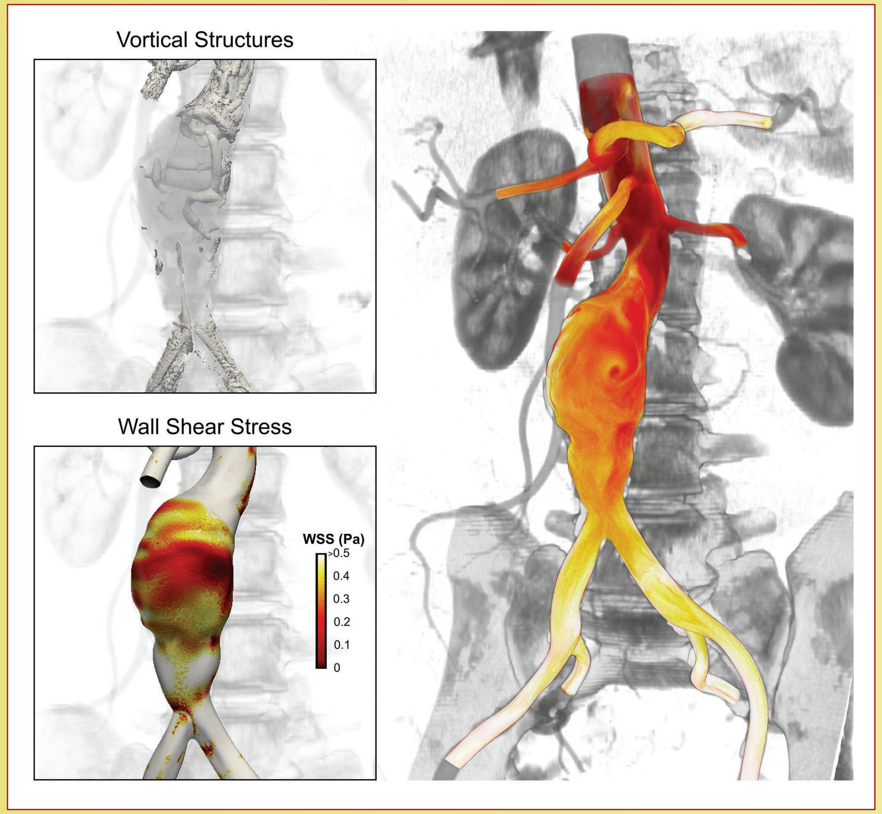
As part of the exascale applications group, Bodony and his team are focused on building scalable algorithms. "How you program on future machines is going to be very different from how we program for *Stampede*," Bodony said. "What that means is that the codes that we have now may not run on future machines. We're trying to rewrite the code in such a way that it's ready for those future machines. Luke and I are working together to figure out how to fix our current code and transform it into one that's useful at exascale."

"Most people think that Knight's Landing (a second generation Xeon Phi product using a 14 nm process) is a preview of what processors will look like going forward as we push toward exascale—lots of concurrency, many cores in a single package, the memory footprint per thread is going to be very small—so we will completely rethink the way we solve our problems. It's safe to say that every time a new processor comes out it's a completely new challenge," Wilson concluded.

Quiet the sonic boom! @XSEDEscience researchers are finding innovative ways of studying aeroacoustics in order to make jet engines quieter.

INVESTIGATING AAAs

XSEDE ECSS supports Yale researchers studying vascular disease



Computer simulations of the fluid dynamics of blood flow. Researchers integrate this information with other blood flow metrics to investigate why thrombus forms in aneurysms.

Abdominal aortic aneurysm affects thousands each year. @XSEDEscience ECSS supports Yale researchers investigating the disease.

Abdominal aortic aneurysm (AAA) is a vascular disease that affects tens of thousands of people around the world each year. The disease occurs most commonly in men over 60, and is characterized by a dilation of the abdominal aortic wall and often a persistent blood clot. When they rupture, abdominal aneurysms can be lethal.

To understand why some AAAs rupture and others do not, researchers must better understand how they form and enlarge, digging deep into both the micro and macro levels to understand the complex interactions that define the disease.

One group attempting to accomplish this task is the Continuum Biomechanics Lab in the Department of Biomedical Engineering at Yale University. To explore the hemodynamics in AAAs, the fluid dynamics of blood flows, the researchers rely on a mix of computational resources from Yale and XSEDE.

"Access to XSEDE resources is a key aspect of our research. The code we started working with was already equipped to work on supercomputers, but having access to high performance computers enables us to explore additional parallel capabilities," said Paolo Di Achille, a researcher in the group.

Looking In

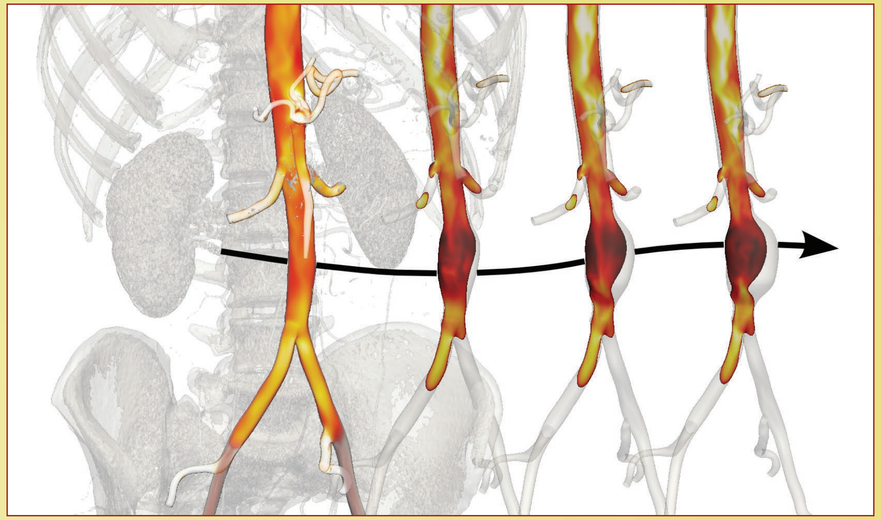
The aorta is the largest artery in the body; it allows oxygenated blood to be distributed to organs and tissues throughout the body. Aneurysms, or localized dilatations, can form in the ascending or descending portions of the aorta within the chest or within the descending portion in the abdomen. Abdominal aneurysms are more common and when they rupture, they are often lethal.

Because aneurysms are stressed by the continuous action of changing blood pressures, methods of engineering that were developed to study the flow of viscous fluids (e.g., blood) and the failure of structures (rupture of the aorta) can be used to help understand the complex history of these deadly lesions. Once an AAA ruptures, death is nearly inevitable due to extreme internal bleeding, signifying the critical need to predict which aneurysms are most vulnerable.

Using computational models in combination with biological experiments, researchers in the lab can study AAA development at both the micro and macro scales. Jay Humphrey, professor of biomedical engineering, emphasizes that, "The biological and mechanical complexity of AAAs demands a multidisciplinary team to advance our understanding." George Tellides, Yale professor of surgery and Andrew Sherman at the Yale Center for Research Computing, teamed up with Humphrey to enable the development of patient-specific computational models for multiple aspects of AAAs.

In collaboration with post-doctoral fellows and Ph.D. students, they have recently put forth new hypotheses on factors that drive intraluminal (interior areas of the artery) clotting as well as the mechanobiological stability of the diseased aortas.

To take full advantage of XSEDE's capabilities for this research, the researchers applied to XSEDE's Extended Collaboration and Support Service (ECSS), a program that pairs researchers with expert staff members in advanced cyberinfrastructure. The researchers at Yale partnered with Yifeng Cui at the San Diego Supercomputing Center (SDSC) to optimize their code and improve simulations. They found the program very helpful to adapting their new codes to a supercomputing environment.



Mechanically and hemodynamically-driven growth and remodeling models of disease progression could help improve clinical decision-making by providing indications of the future course of disease.

"I would say that's the biggest advantage of ECSS, talking with experts to get real insight into how our codes can be improved. It's the best way for us to optimize our codes, because we don't have time to read all the literature on these new architectures," said Di Achille.

In this particular case, according to Di Achille, after discussing with the ECSS expert about the limitations of a particle tracking code that they developed, the researchers introduced a different load balancing scheme that better exploited the capabilities of large computers and high performance file systems (e.g. parallel I/O libraries). With these modifications, the group brought their code to scale almost linearly for typical use cases, which in practice translated into a four-time speedup.

Working with Cui, the team also ran their code on TACC's most powerful supercomputer, *Stampede*, and used its visualization nodes to create realistic depictions of AAAs. The team also used *Trestles*, which had been at SDSC.

To advance their research, the team is using this information to develop comprehensive models to track the progression of the disease. Ultimately, they hope this research will not only increase our basic understanding of AAAs, but will also help clinicians make more informed decisions when treating patients.

"Clinical decisions are extremely complex," Di Achille noted. "If our modeling results could help in this decision process and improve outcomes that would be very satisfying."

THE TURTLE TREE OF LIFE

In all of the animal kingdom, it's hard to top the evolutionary longevity and unique features of turtles. These reptiles have existed for more than 200 million years, and their combination of shells, beaks, widely ranging body sizes, and other unusual traits has been an endless source of fascination.

From the freshwater-loving variety to those of the sea, to the bizarre ones that sport soft shells and can move at speeds of up to 15 miles an hour, turtles have adopted so many varied characteristics during their existence through the ages that their lineage has been a longstanding question in biology.

What does the turtle tree of life look like?

In an effort to add further knowledge and perspective to the existing body of scientific literature, a collaborative team composed of turtle expert James Parham, geneticist Nicholas Crawford, and other co-researchers performed an investigation to trace the changes in evolutionary history, or phylogeny, of turtles.

"In some instances, previous molecular [genetic] analyses conflicted with morphological [form, structure, feature] analyses, but it wasn't clear if this was because morphological analyses were wrong or if there simply wasn't enough molecular data," said Crawford, a postdoctoral fellow in the Department of Genetics at the University of Pennsylvania. "So we sequenced 1,000-plus individual parts from 29 turtle genomes to get a bigger picture and hopefully less-biased molecular perspective of turtle relationships."

Crawford explained that he needed to assemble numerous libraries of what are called ultraconserved elements (UCEs), across a variety of parameters. UCEs, according to a website called ultraconserved.org, can be defined as "highly conserved regions of organismal genomes shared among evolutionary distant taxa [organism populations]—for instance, birds share many UCEs with humans."

Through an allocation from XSEDE, Crawford was able to make use of the large RAM nodes on the *Gordon* system at the San Diego Supercomputer Center (SDSC) to complete the UCE work in parallel during the course of only a few days. The project also applied the *Blacklight* system (now decommissioned) at the Pittsburgh Supercomputing Center (PSC) via XSEDE.

The research team discovered the strongest genetic evidence to date for how turtles evolved, and the phylogenomic tree they developed from their findings can be used by paleontologists as a starting point to greatly advance the study of turtle fossils and evolution, according to Parham.

"Over the past 25 years or so, ideas about how different groups of turtles were related to one another—such as which group of freshwater turtles is most closely related to sea turtles, or which turtles have been evolving independently on their own branch of the tree for the longest amount of time—have been changing," Parham said. "By virtue of the huge amounts of data that we bring to the problem, we can provide strong evidence to bear on competing hypotheses."

> With XSEDE, researchers have probed the lineage of turtles and found the strongest genetic evidence yet for how these animals with eclectic features likely evolved.

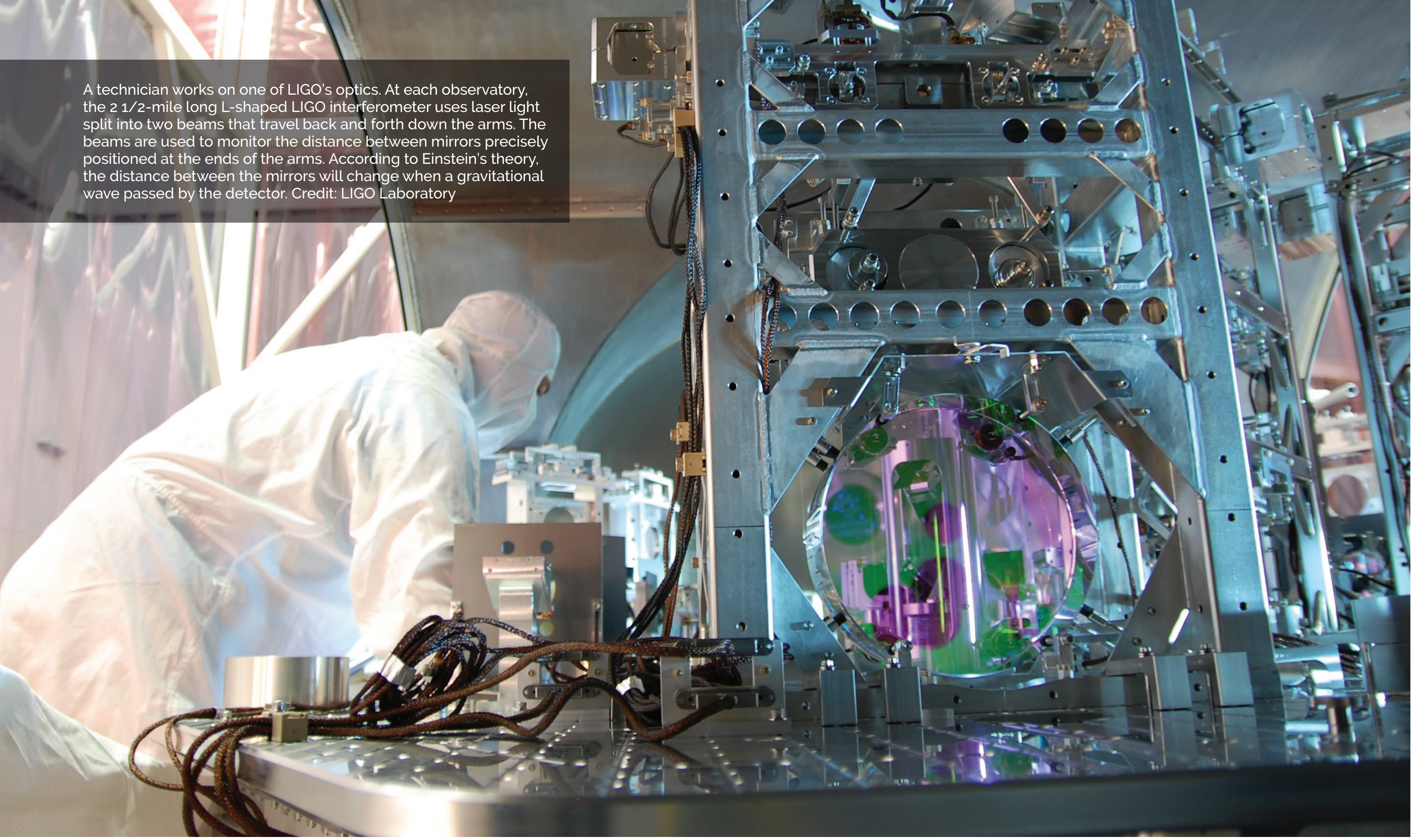
The researchers were, for example, able to show that softshell turtles have been evolving on their own branch of the tree for a very long time. The findings also illustrate that sea turtles are closely related to particular groups of North American freshwater turtles (mud and snapping turtles) and, therefore, likely originated in North America as well.

"We compared our tree of life to the timing and geography of the fossil record to provide a very complete picture of turtles around the world through time," Parham said.

The project also took the initiative of assigning a name to the grouping of what previous research has suggested to be turtles' nearest relatives, crocodiles and dinosaurs (including birds). The name they coined is Archelosauria, a combination of archosaurs (crocodiles and dinosaurs) and chelonians (turtles).

"I think that naming the Archelosauria grouping will solidify the idea of this clade (a group of organisms believed to have evolved from a common ancestor) in people's minds," Parham said.

Details of the research are contained in "A phylogenomic analysis of turtles," in the journal *Molecular Phylogenetics and Evolution*.



A technician works on one of LIGO's optics. At each observatory, the 2 1/2-mile long L-shaped LIGO interferometer uses laser light split into two beams that travel back and forth down the arms. The beams are used to monitor the distance between mirrors precisely positioned at the ends of the arms. According to Einstein's theory, the distance between the mirrors will change when a gravitational wave passed by the detector. Credit: LIGO Laboratory

XSEDE HELPS CONFIRM LIGO DISCOVERY

Scientists for the first time detected gravitational waves—ripples in the fabric of space-time hypothesized by Albert Einstein a century ago—in a landmark discovery announced in early 2016 that opened a new era in astrophysics while relying on XSEDE and Open Science Grid (OSG) resources to verify the findings.

The discovery of gravitational waves was made by researchers at the NSF-funded twin Laser Interferometer Gravitational Wave Observatory (LIGO) detectors in September 2015, but scientists wanted to be sure of what they detected on the two detectors, which are about 1,900 miles apart in Louisiana and Washington State.

LIGO researchers used the OSG as a gateway to XSEDE's *Stampede* and *Comet* supercomputers. "We also leveraged resources at other institutions in the OSG, but *Stampede* and *Comet* were among the most used," said Peter Couvares, the Advanced LIGO Data Analysis Computing Manager. "For LIGO especially, OSG and XSEDE are highly complementary and becoming critical to our research efforts."

Staff at the Texas Advanced Computing Center (TACC) at the University of Texas at Austin, including those from XSEDE's Extended Collaborative Support Service (ECSS) program, were able to recreate the computing environment that LIGO researchers were accustomed to on *Stampede*, which was valued for its large number of cores. The bulk of data analysis requires large scale high-throughput computing with parallel workflows at the scale of tens of thousands of cores for long periods of time. In addition, TACC staff were able to improve the performance of the code by a factor of four, and with the techniques learned from TACC, LIGO staff were able to improve that by another factor of two. The result was that LIGO's original estimate for their computing needs going forward were reduced by a factor of eight, which obviated an investment of tens of millions of dollars in computer hardware.

"The staff at TACC was instrumental in helping us guide where we looked for optimization; ultimately, we ended up with a team of almost five full-time LIGO people doing code optimization; the directions we chose were formed by our interactions with TACC staff and we wouldn't have been as successful without them," Couvares said.

LIGO used *Comet's* new Virtual Cluster interface for the analysis of their present data. The total data set was five terabytes (TB), stored at the Holland Computing Center at University of Nebraska–Lincoln (UNL). Distributing the 5TB dataset from UNL to the many clusters on OSG, including *Stampede* and *Comet*, required a total transfer of one petabyte of data, which was accomplished using GridFTP. *Comet* is housed at the San Diego Supercomputer Center (SDSC) at UC San Diego.

Couvares said the discovery has opened a new window on astronomy and astrophysics. "We're now seeing things in the universe that we've never seen before, and we have a medium for information that we've never had before. In addition to telescopes that provide electromagnetic observations, we now have gravitational wave observations and that's very powerful."

Since LIGO scientists don't know what the waves look like before they detect them, they have to search over a large space of possible signals. "The gravitational waves depend sensitively on the black hole's masses and spins, making the search computationally challenging," said Duncan Brown, the Charles Brightman Professor of Physics at Syracuse University and a gravitational-wave astrophysicist who studies the mergers of binary black holes.

LIGO's discovery of gravitational waves from the binary black hole required large-scale data analysis to validate the discovery claim. "This includes measuring how significant the signal is compared to noise in the detector, and re-analyzing the data with simulated signals to ensure that the search is working properly," said Brown. "Once we have made detections, extracting the physics from these waves requires careful modeling of the black holes. The waves that LIGO detected were produced by very strong gravitational fields. By comparing calculations of the detected waves with our observations, we can explore the nature of black holes and gravity itself." The factor of 8x speedup in code performance allowed LIGO to publish the result much earlier than would have been possible with their original code.

Couvares noted that LIGO's computing demands are very much in flux. "If nature is kind and we see a lot more events than we expected, we'd need a lot more computing power than anticipated. We wanted to have capabilities that went beyond our in-house computing resources, so XSEDE and OSG were an excellent option," he said.

Moreover, he believes that XSEDE is at the forefront of defining what large scale computing and these national facilities will look like in the future.

"LIGO is going to move into an era of regular detections where we're recording more than one event," he said. "At the same time, we'll be pushing the limits of physics, astrophysics, and engineering."

Scientists used XSEDE supercomputing resources to verify the discovery of gravitational waves theorized by Albert Einstein 100 years ago.

XSEDE is supported by the National Science Foundation.

XSEDE

Extreme Science and Engineering Discovery Environment

xsede.org | info@xsede.org | /XSEDEscience | @XSEDEscience