Fall 2023

FRESH TRACKS™

Climate Resilience for the Western Grid

Sage Sensors Supporting Edge AI

NASA Climate Data Partnership

Blockchain Storage for Data Democratization

www.sci.utah.edu

Scientific Computing and Imaging Institute
at the University of Utah

From the Director

The Scientific Computing and Imaging (SCI) Institute at the University of Utah continues to advance its mission to transform science and society through translational research and innovation in computer and computational and data sciences, including machine learning and artificial intelligence. Building on its legacy of excellence, SCI has continued to grow and expand. This has been another exciting year marked by an impressive set of new hires, awards, grants, publications, software deployments, and other accomplishments, all with tremendous impact. I am delighted to share some of these accomplishments with you in the current issue of FreshTracks.

SCI is delighted to lead President Taylor's *Responsible AI Initiative*. This bold and ambitious initiative aims to responsibly advance translational AI research and its applications in ways that achieve societal good while also protecting privacy, civil rights and liberties, and promoting principles of accountability, transparency and equity. This concerted $100M effort will create transdisciplinary excellence in responsible AI by bringing together deep technological expertise, advanced cyberinfrastructure, and disciplinary expertise across the university to position the U as a national leader in translational AI. The project will begin with a focus on issues that have regional implications, such as health care and societal wellness, public services, and our natural surroundings, and can improves the lives of Utah's 3.4 million residents.

SCI continues to act on its an ambitious goal of hiring multiple new faculty members over the next three years with specialties that integrate computational science, data science, and science and engineering broadly (including social sciences). Our goal is to expand the core research expertise at SCI, further enhance its outstanding research portfolio, and continue to broaden the diversity of faculty, students, and staff. Specifically, we have several open faculty searches at the moment in computational oncology (in partnership with the Huntsman Cancer Institute), biomedical informatics (in partnership with the Department of Biomedical Informatics), visualization and imaging, and software systems for future computing. The initiative has already resulted in six outstanding faculty at SCI. And we plan to continue this growth in the coming years.

SCI also continues to advance its initiatives such as (1) the Cyberinfrastructure Professionals Cooperative (CIP-Co-Op) that aims to foster a vibrant and sustainable CIP community, along with models and structures for training, professional development, and sustainability at SCI and the U; and (2) the SCI-HUM initiative, in partnership with the College of Humanities, that brings together inspired technology-driven humanities research ideas with the technological innovation and research at SCI to catalyze, nurture, and advance transdisciplinary research partnerships.

The coming year promises to be even more exciting and eventful. SCI will celebrate its 30th anniversary, marking three decades of multidisciplinary research, innovation, partnerships, and impacts. The two-day event on April 24 and 25, 2024, will showcase groundbreaking research and will celebrate SCI's past, present and future. I hope you can join us at this event.

- Dr. Manish Parashar
  Director, Scientific Computing and Imaging Institute

In This Issue

---

30 SCI

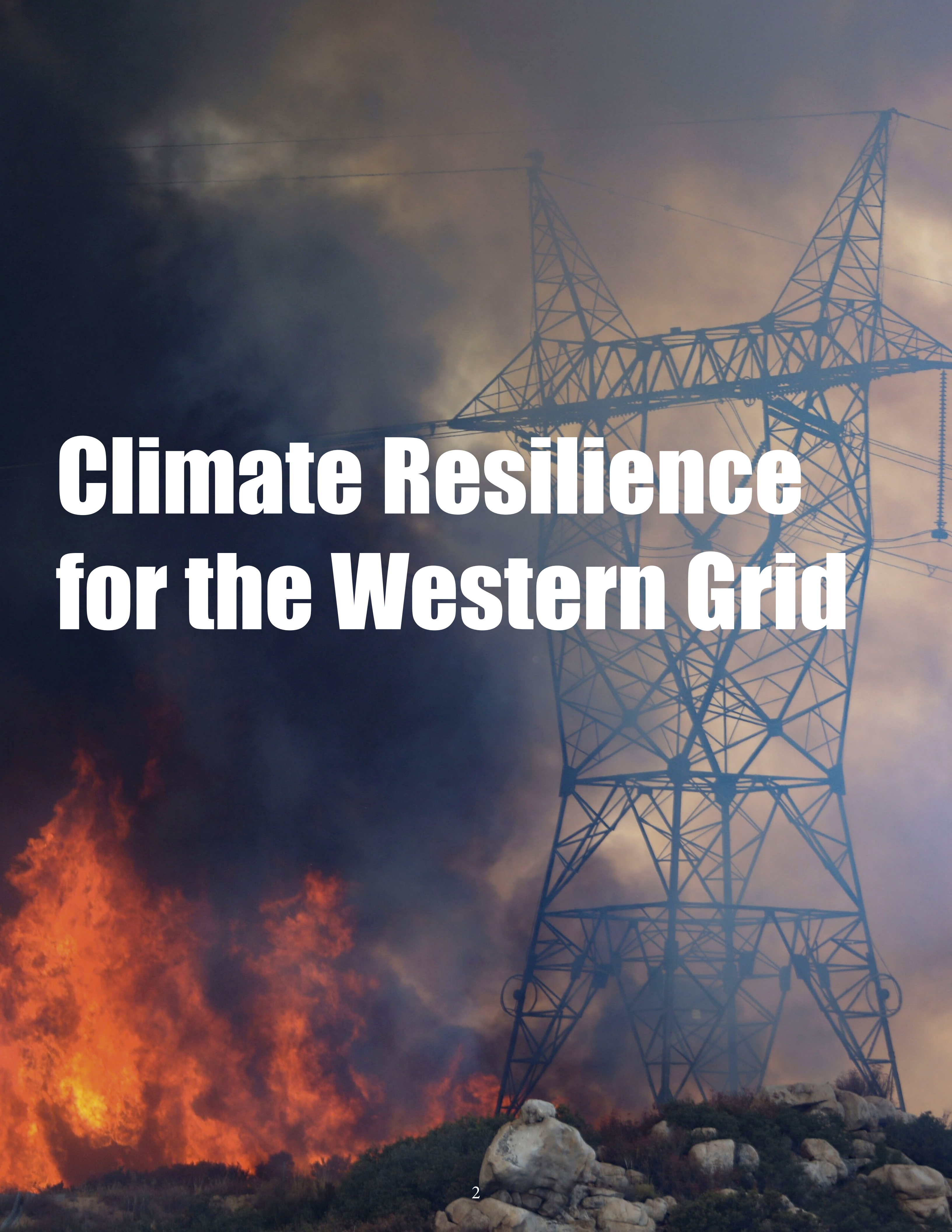
www.sci.utah.edu

The SCI Institute is Celebrating 30 years!

In 2024, the SCI Institute will mark three decades of multidisciplinary research, innovation, and impact. We hope you will join us on April 24th and 25th to help celebrate our past, present and future. We will have two days of events showcasing the ground-breaking research of SCI's faculty, staff and students and highlighting our vision for the future. We will be sending out the details soon, but for now, mark your calendars.

Save the Date - April 24-25, 2024

Climate Resilience for the Western Grid



University of Utah and University of Calgary Launch U.S.-Canada Center on Climate-Resilient Western Interconnected Grid

Through $5M funding by the U.S. National Science Foundation (NSF) and $3.75M funding by the Natural Sciences and Engineering Research Council of Canada (NSERC), the University of Utah and University of Calgary will establish and co-lead the U.S.-Canada Center on Climate-Resilient Western Interconnected Grid.

The Western Interconnected Grid, commonly known as "the Western Interconnection," is one of the two major interconnected power grids in North America, which stretches from the northern edge of British Columbia, Canada to the border of Baja, Mexico, and from the California coast to the Rockies, and serves roughly 80 million people over 1.8 million square miles across two Canadian provinces and fourteen western states in the United States. The Western Interconnection is the backbone of one of the largest regional economic engines in the world.

Masood Parvania, associate professor of Electrical and Computer Engineering at the University of Utah's John and Marcia Price College of Engineering will co-lead the center along with Hamid Zareipour, professor of Electrical and Software Engineering at the University of Calgary's Schulich School Engineering.

"Our center is being established at a critical time when the region is experiencing more frequent and severe extreme weather disturbances such as wildfires, heatwaves, drought, and flooding, the impacts of which not only pose threats to human health and the environment but also affect the ability of the western interconnection to continue powering the communities," says Parvania.

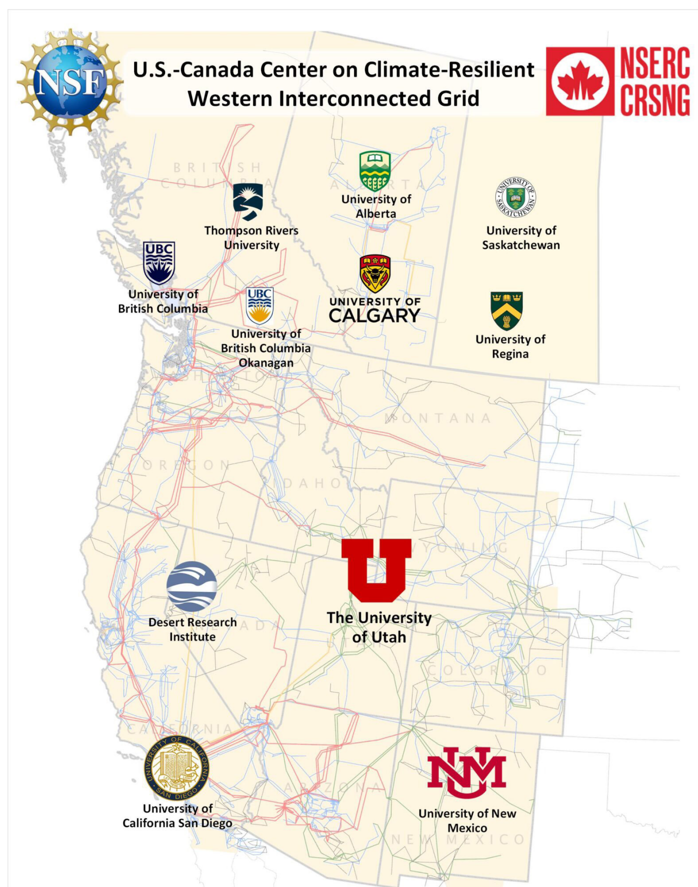
At the University of Utah, the center involves co-principal investigators Valerio Pascucci, professor at the Scientific Computing and Imaging Institute and Kahlert School of Computing, William Andregg, director of the Wilkes Center for Climate Science and Policy, and Divya Chandrasekhar, associate professor in the Department of City and Metropolitan Planning in the College of Architecture and Planning, among multiple other partners and faculty.

The U.S.-Canada Center on Climate-Resilient Western Interconnected Grid creates an interdisciplinary and international partnership that brings together leading experts in power engineering, climate, forestry, data, policy, and social sciences, as well as industry, entrepreneurs and community knowledge holders from a network of 35 partners across academia, industry, government, and communities with the mission of enhancing the power grid resilience to the rising frequency, intensity, and duration of extreme weather events, such as wildfires and heatwaves.

"The Center will work closely with the various communities that are served by the western interconnection, which include some of the most densely populated cities in the world, as well as remote and rural areas with minimal power infrastructure. Establishing a comprehensive understanding of the unique needs of these communities is necessary to develop effective climate-resilience strategies," says Zareipour.

The center will develop customized models for risk quantification and forecasting of regional extreme climate disturbances and will build a cyberinfrastructure for collecting and sharing both climate and grid data among the Western Interconnection's stakeholders.

"The center will showcase a new international innovation ecosystem for rapid transformation of use-inspired research into technologies with global applications to overcome common challenges posed by the extreme weather events. This is a mission-critical process aimed at safeguarding the future of our energy infrastructure and, by extension, communities across North America and beyond," says Mostafa Farrokhabadi, assistant professor of Electrical and Software Engineering at the University of Calgary's Schulich School Engineering, and the Chair of the Innovation Ecosystem for the U.S.-Canada Center.



Sage sensors monitor environment, support 'edge AI'

Smokey Bear is mostly right—you can help prevent wildfires—but it's not all on you. Using fire-resistant building materials, establishing vegetation-free "ignition zones," and avoiding fire-related activities when it's hot, dry and windy are actions that, ideally, we all can take.

The scientific community, too, has a preventive role to play around natural disasters, urbanization and climate change. Recent advances in this area, like attempts to predict fire behavior before it becomes unmanageable, are the result of an ambitious project to build a continent-wide network of intelligent sensors that monitor environmental changes.

The Sage project, of which the University of Utah is a partner institution, is a $9 million National Science Foundation (NSF)-funded initiative launched in 2019 and led by researchers at Northwestern-Argonne Institute of Science and Engineering (NAISE), a collaboration between Northwestern University and the U.S. Department of Energy's Argonne National Laboratory. Other Sage partners include the University of Colorado, the University of California-San Diego, Northern Illinois University, and George Mason University.

"Sage is a next-generation software infrastructure that's flexible enough to accommodate both urban and environmental monitoring," said Dan Reed, professor of computational science at the U; chair of the National Science Board, the NSF's policymaking body; and the Sage project's chief architect.

Manish Parashar, director of the U's Scientific Computing and Imaging (SCI) Institute, meanwhile, is overseeing an effort to write code for scheduling and analysis functions on the three U-maintained sensor nodes—one at the Taft-Nicholson Environmental Humanities Center in Lakeview, Montana; one atop 102 Tower in downtown Salt Lake City; and one on the roof of the Rio Tinto Center, which houses the Natural History Museum of Utah (NHMU).

When Reed presented the project to the U's Council of Academic Deans, NHMU Executive Director Jason Cryan said he "immediately put up my hand to offer

Scientists from Argonne National Laboratory conduct experiments with a Sage monitoring system near a controlled burn in the Konza prairie in Kansas. Select the image to enlarge. (Image by Rajesh Sankaran, Argonne)

NHMU as a potential first installation site for Utah."

The idea behind Sage is to move advanced machine learning algorithms to "edge computing." The traditional method involved deploying sensors and collecting data later. Previous systems might save data on hard drives retrieved from sensors a few times a year or upload only a fraction of data to a cloud server through a slow wireless connection. Edge computing, however, allows data to be analyzed and measured almost immediately, by instruments near or at the data collection site. Sage devices, Reed explained, process images, sound, vibration and other data to create measurements that can't be as easily obtained from conventional sensor networks.

The Sage project relies heavily on lessons learned from the Array of Things (AoT) project, part of the smart cities initiative announced in 2015 by former President Barack Obama. AoT leverages an open-source intelligent sensing and edge computing platform called Waggle, developed at Argonne National Laboratory.

Primarily funded by the NSF, AoT was a collaborative effort among scientists, universities, federal and local government, industry partners and communities to collect real-time data on urban environments. AoT was the brainchild of Charlie Catlett, a senior computer scientist in Argonne's Mathematics and Computer Science (MCS) Division, Sage co-principal investigator, and Reed's longtime friend and colleague.

Catlett's vision was to create a "fitness tracker for the city": a vast network of low-cost sensors placed throughout Chicago capable of measuring everything from urban heat islands to noise pollution.

"When you think about the way we've traditionally done social science, people go out and conduct surveys. When you survey people, they sometimes say what they think you want to hear," Reed said. "It's a little like your doctor advising you to eat a better-balanced diet and exercise more, which ends up being a promise most of us make and break the moment we leave the doctor's office. With AoT, the idea became, instead of surveys, why don't we just measure what really happens?"

Reed said getting the AoT project off the ground required a considerable amount of civic discussion over the acceptable use of data "because putting cameras up

The U maintains Sage sensor nodes (L-R) at the Taft-Nicholson Environmental Humanities Center (image courtesy of Dan Reed); Rio Tinto Center (image courtesy of Natural History Museum of Utah staff); and 102 Tower (image courtesy of Tom Jennings).

in the middle of the city doesn't exactly inspire faith in the community."

"One of the agreed-upon rules was that raw imagery would never leave the cameras, only derived and anonymized statistics," Reed said. "The other realization was that the volume of data is too large to push back to the central site, so being able to do edge AI lowered the bandwidth requirements while allowing greater privacy."

Sage cyber infrastructure links small, powerful computers directly to nodes, most of which feature high-resolution cameras (including a thermal camera), microphones, weather measurements and air quality sensors, and sends the information back to central servers. The distributed system allows researchers to analyze and respond to enormous amounts of data quickly, without having to transfer it all back to the laboratory.

Though Sage devices are intended to measure environmental changes locally and regionally, sensors are often sensitive enough to detect changes thousands of miles away. For example, Reed said Sage sensors detected the air pressure wave from a volcanic eruption in Tonga that produced the largest atmospheric explosion in recorded history.

In addition to U-maintained nodes, Sage technology continues to be tested in AoT environments; the University of California-San Diego's WIFIRE project, which provides real-time wildfire prevention and response data at about 80 towers across southern California; and the NSF's National Ecological Observatory Network (NEON), an array of hundreds of terrestrial and aquatic measurement sites across the United States that collects data on plants, animals, soil, water, and the atmosphere.

"NEON provides expert ecological data from sites across the continent to power the most important science being done today," Reed said. "NEON is essential to tracking and understanding how human activities impact local flora and fauna as the environment has shifted."

Reed is also excited about Sage's potential for "engaging K-12 students in real citizen science."

"The nodes have places where young people can plug in low-cost sensors of their own," Reed said. "Just imagine students interacting with this technology, too, getting excited by the notion that they can do real science, write code, build sensors and capture data in real-time, instead of just reading examples in a textbook."

A brief overview of Sage sensor nodes (source: Dan Reed)

- Each node runs on wireless or wired ethernet connections, or Starlink, the highest-bandwidth option at the Taft-Nicholson Center.
- The nodes periodically send data back to the Waggle platform and other Argonne-hosted infrastructure.
- A dashboard of real-time information displays still images recorded by a top and bottom camera, infrared views, 30-second audio snippets, and environmental conditions, such as pressure, temperature, and humidity.
- Sensors are managed by Raspberry Pi microcontrollers and an Nvidia graphics processing unit (GPU)-powered AI engine, with plug-in connectors capable of supporting additional sensors.
- Code can be scheduled to run on nodes.
- The nodes accomplish myriad environmental monitoring tasks, from detecting changes in air quality and weather patterns to recognizing different species of birds and their migratory habits.

Seal Storage Technology and the University of Utah Announce NASA Climate Data Democratization Partnership

Over 200TB of NASA climate data is now accessible through the collaboration between Seal Storage Technology and the University of Utah's Scientific Computing and Imaging Institute (SCI), enabling researchers around the world to access and visualize the data with zero egress fees.

In an effort to democratize data access, decentralized cloud storage provider Seal Storage Technology and the University of Utah's Scientific Computing and Imaging Institute (SCI) announced today that over 200 terabytes (TB) of OpenViSUS IDX NASA atmospheric wind velocity and ocean current datasets are now openly available to researchers globally. The data will be accessible via a Jupyter notebook, which is a digital document that makes it easy to access and visualize data by combining text, code, and visualizations in a single medium. This project is part of SCI's participation in The National Science Data Fabric (NSDF), a trans-disciplinary approach to integrated data delivery and access, democratizing data-driven scientific discovery through shared storage, networking, computing, and educational resources.

"The democratization of data is essential to scientific progress, and we are thrilled to partner with the University of Utah to make this possible," said Seal Storage Technology COO Alex Altman. "Seal's decentralized cloud storage platform ensures that this valuable data is accessible to researchers around the world, without barriers. We are excited to see the future discoveries and insights that will be made possible by making this important data available and by supporting the mission of NSDF."

Researchers around the globe will now be able to access the NASA OpenVisus data with zero egress fees as the data is stored on Seal's decentralized cloud storage platform, which is powered by the Filecoin network. Egress fees are typically charged by centralized cloud storage providers when users want to move or access their data. In contrast, decentralized cloud storage leverages distributed ledger technology to store data at centers across the globe, offering a novel way to store data and remove egress fees. This global distribution ensures that the NASA data is immutable, verifiable, and has a chain of custody.

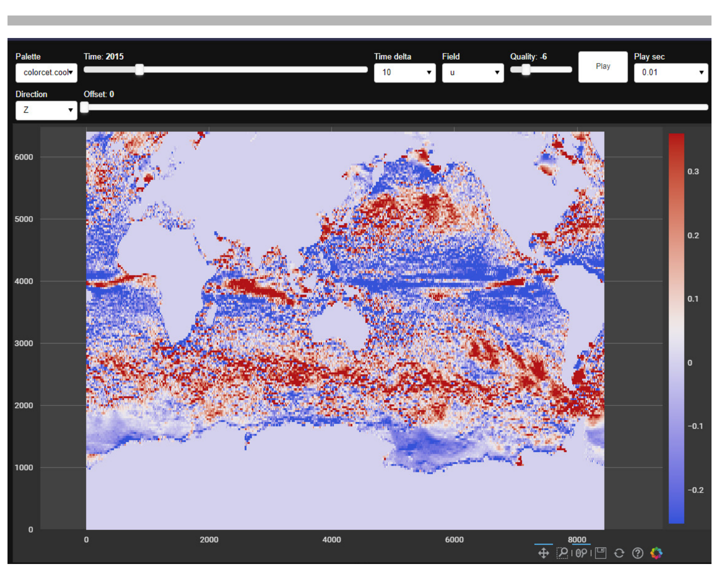
The data democratization partnership will provide researchers with new ways to access and utilize NASA's climate data, including navigation via timestamps, fields, and other dimensions. This is made possible by transferring the data from NASA supercomputers to Seal's platform, which then enables the data to be accessed via a Jupyter notebook.

"The NSDF project is making significant progress towards more open access to critical cyberinfrastructure," said Manish Parashar, Director of the Scientific Computing and Imaging (SCI) Institute. "Having shared data cyberinfrastructure at scale, such as what NSDF envisions, can accelerate scientific discovery across a broad range of disciplines. It can also broaden access to the data by new groups of researchers."

This dataset is an output of a simulation that combines two models to provide high-resolution data on atmospheric and oceanic variables. The first model is a C1440 configuration of Goddard Earth Observing System (GEOS) atmospheric model, whereas the second model is a LLC2160 configuration of the MITgcm model. There are over 10000 timesteps in each output, and each of them has multiple scalar fields such as temperature, snow thickness, east-west velocity, etc. With the help of Jupyter notebooks, researchers are now able to navigate across all timesteps and dimensions of certain fields (e.g. East-west velocity 'U') of both the atmospheric dataset and LLC2160 ocean dataset.

"This is a major milestone of the National Science Data Fabric (NSDF) that gives all scientists, engineers, and regular citizens direct access to massive data of major societal importance without needing expensive hardware resources," said Valerio Pascucci, NSDF Principal Investigator, Professor at Kahlert School of Computing, and Director of the Center for Extreme Data Management and Visualization (CEDMAV) at the University of Utah. "This partnership between academia and industry shows how anyone with a laptop and an internet connection can now visualize and analyze data that until now has been the sole monopoly of a selected few with the special access and training needed to use some of the largest supercomputers in the world."

Researchers can access the NASA data on the following web dashboard http://chpc1.nationalsciencedatafabric.org:10888/nasa or visualize the data locally by downloading the Jupyter notebook here: http://bit.ly/3MkpVBa. The current 200TB of data constitutes the first installation of this ambitious multi-petabyte project demonstrating the feasibility of the NSDF-SEAL data democratization initiative.



Using blockchain-based storage to protect and democratize data

Decentralized solutions can make academic research more secure and more accessible.

Academic institutions house some of the world's most important data generated from years of research. Yet centralized data storage models are becoming a concern for many universities looking to keep critical information safe and accessible.

Danny O'Brien, a senior fellow at the Filecoin Foundation and Filecoin Foundation for the Decentralized Web (FFDW) — an independent organization that facilitates governance of the Filecoin network and funds development projects — told Cointelegraph that data stored by academic institutions is at risk of vanishing due to centralized storage models. To put this in perspective, a recent Filecoin Foundation survey found that 71% of Americans have lost information and records due to challenges like deleted hyperlinks or locked online accounts.

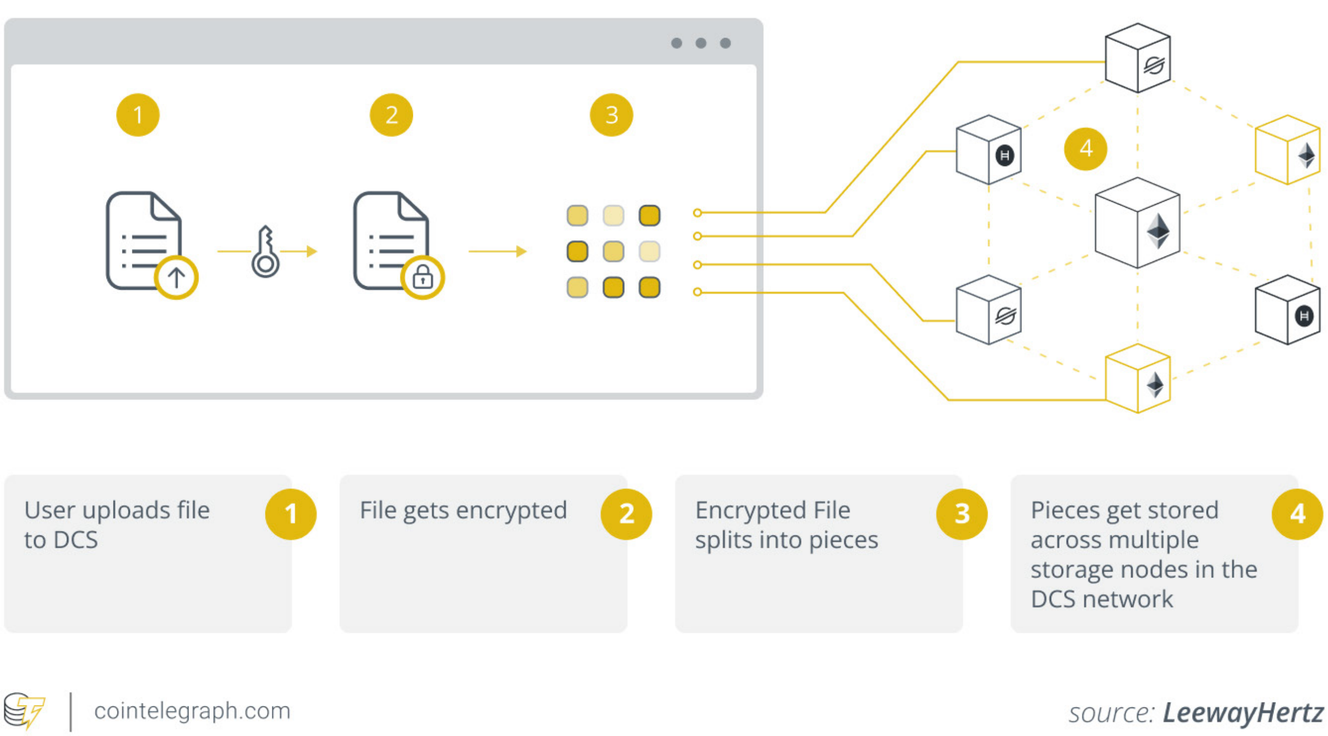
Decentralized storage helps secure and distribute data

To combat this, O'Brien explained that a handful of educational institutions have begun using decentralized data storage models to preserve data sets. "A growing number of higher education institutions, including the Massachusetts Institute of Technology (MIT), Harvard University, the University of California, Berkeley, Stanford University, the University of South Carolina, and others, are all using Filecoin to store, preserve and archive their most important data on the blockchain," he said.

For example, O'Brien pointed out that MIT is currently working on a three-year project with the FFDW to explore how decentralized technology can support its Open Learning programs. MIT's Open Learning programs include "OpenCourseWare," which is designed to provide free online materials from over 2,500 MIT courses. This will allow anyone worldwide to access MIT courses on the internet.

O'Brien explained that through the support of the FFDW, MIT's Open Learning programs will use decentralized storage to house cataloging, while preserving its OpenCourseWare materials. He added that MIT would soon host public seminars about the challenges and op-

Working of a decentralized storage system



portunities of the decentralized web. "Education's ongoing embrace of decentralized Web3 data storage offers, via cryptographic proof, a guarantee that data remains available and unchanged over time, preserving their critical data for as long as they want," he said.

The University of Utah also uses decentralized storage to protect and democratize access to large data sets. Valerio Pascucci, professor of computer science at the university, told Cointelegraph that the institution's Center for Extreme Data Management Analysis and Visualization recently adopted a solution from Seal Storage — a decentralized cloud storage platform powered by Filecoin — to complement its current centralized infrastructure.

Pascucci explained that the model provided by Seal Storage allows the National Science Data Fabric (NSDF) — a pilot program working with institutions to democratize data — to further its goal of creating new mechanisms for easy access to scientific information.

"Traditionally, Minority Serving Institutions (MSIs), small colleges and other disadvantaged organizations cannot be part of scientific investigation endeavors because they cannot access the data necessary to do the work," he mentioned. The NSDF's use of decentralized storage will change that.

According to Pascucci, the NSDF-Seal Storage partnership has already demonstrated the possibility of distributing massive data collections to different communities without needing to deploy special servers or other complex processing capabilities that may be impractical for many institutions.

"For example, NASA stores on its largest supercomputer, 'Pleiades,' an open climate data set that is over 3 petabytes in size. Yet anyone who wants to use the data would need to have a special account on Pleiades and require the training needed to process the data," he explained, "NSDF has adopted an 'OpenVisus' approach that has reorganized NASA's data so that its distribution through decentralized storage allows for interactive processing and exploration virtually without any local resources."

Pascucci added that this might be the first time a data set of this size has been made available for interactive exploration directly from the cloud. Moreover, he believes that the decentralized approach has enhanced security.

Decentralized storage is beneficial, but challenges remain

Although several universities have begun leveraging decentralized storage models, challenges that may hamper adoption remain.

For example, Pascucci pointed out that to distribute NASA's open climate data set, NSDF's OpenVisus data format had to be extended from traditional file systems to meet the storage model provided by Seal Storage. Jacques Swanepoel, chief technology officer at Seal Storage, told Cointelegraph that mapping and tagging data on the blockchain is a very complicated undertaking.

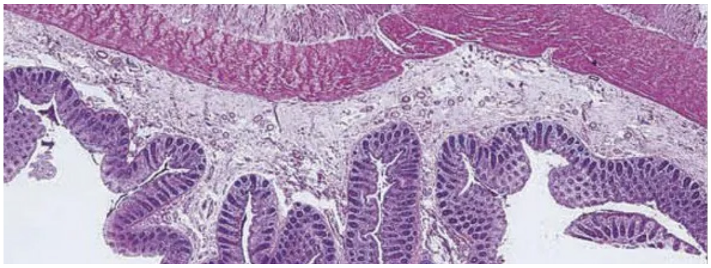
"Identifying which block on the blockchain contains specific information is key to fully utilizing the benefits of decentralized storage technology. In order to overcome these challenges, providers need to properly track where customer data is on the blockchain with creative software strategies."

Yet it remains notable that academic institutions are using decentralized storage models. "Often considered slow-moving, academia has proven to be an early adopter of blockchain-based technologies, including decentralized storage, and continues to be a leader in adopting and deploying these tools," O'Brien said.

This may very well be the case, as Pascucci shared that The University of Utah and NSDF are working on implementing additional use cases with different universities.

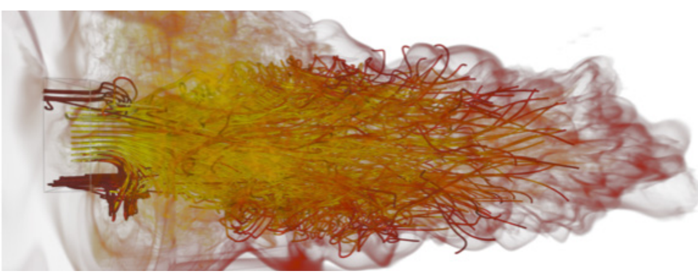
"While the NASA use case is very high profile both for size and application to the important field of global climate change, we are already working on other use cases, including the experimental facility of the Cornell High Energy Synchrotron Source. This is where thousands of scientists go every year to collect data and share it with collaborators across the nation," he said.

New Grants



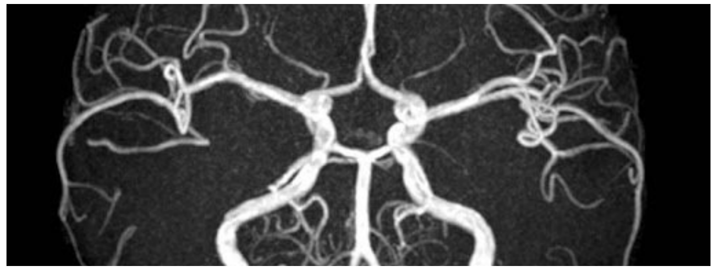
Geometry and Topology for Interpretable and Reliable Deep Learning in Medical Imaging
Bei Wang (NSF)

Abstract: The goal of this project is to develop methods for making machine learning models interpretable and reliable, and thus bridge the trust gap to make machine learning translatable to the clinic. This project achieves this goal through investigation of the mathematical foundations -- specifically the geometry and topology -- of DNNs. Based on these mathematical foundations, this project will develop computational tools that will improve the interpretability and reliability of DNNs. The methods developed in this project will be broadly applicable wherever deep learning is used, including health care, security, computer vision, natural language processing, etc.



Implicit Continuous Representations for Visualization of Complex Data
Bei Wang (DOE)

Abstract: The proposed research investigates how to accurately and reliably visualize complex data consisting of multiple nonuniform domains and/or data types. Much of the problem stems from having no uniform representation of disparate datasets. To analyze such multimodal data, users face many choices in converting one modality to another; confounding processing and visualization, even for experts. Recent work in alternative data models and representations that are continuous, high-order (nonlinear), and can be queried anywhere (i.e., implicit), suggests that such models can potentially represent multiple data sources in a consistent way. We will investigate models that are high-order, continuous, differentiable, potentially anti-differentiable (providing integrals in addition to derivatives), and can be evaluated anywhere in a continuous domain for scientific data analysis and visualization.



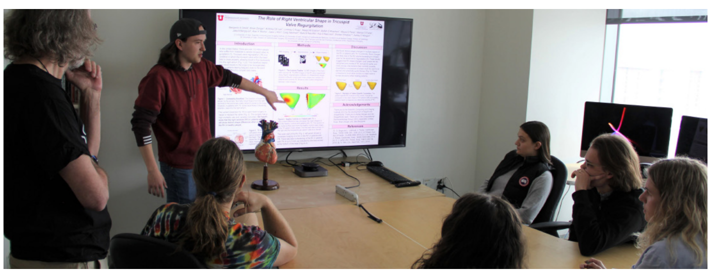
A physics-informed machine learning approach to dynamic blood flow analysis from static subtraction computed tomographic angiography imaging
Amir Arzani (NSF)

Abstract: Recent investigations have shown that interactions of blood flow with blood vessel walls plays an important role in the progression of cardiovascular diseases. Accurately quantifying blood flow or hemodynamic interactions could lead to methods for patient-specific therapies that result in better treatments and reduced mortality. In this project, the researchers will develop techniques to non-invasively inferring the complex, dynamic hemodynamic behavior using a commonly used medical imaging modality that is typically used to produce static anatomical images for analyzing blood vessel structure. In this project, the researchers propose to develop a novel physics-informed model of the blood flow using a deep-learning based processing method. This will allow the researchers infer dynamic time-resolved three-dimensional blood velocity and relative pressure field. The results will be used to accurately compute relevant hemodynamic factors. This project will train a cohort of graduate students in the latest data-driven deep learning techniques in engineering. It will engage undergraduate students in research through well-established programs at UW Milwaukee and Northern Arizona University. Outreach to high school students, particularly those belonging to under-represented communities will be accomplished through summer programs at UW Milwaukee.

The goal of this project is accurate image-based hemodynamic analysis using commonly available images. Contrast concentration, three-dimensional blood velocity, and relative pressure will be modeled as deep neural nets. Training the neural nets will involve a loss function that matches actual data from time-stamped sCTA sinograms with predicted sinograms generated using line integrals computed from forward evaluation of the neural net used to model the contrast concentration. Additionally, blood flow and contrast advection-diffusion physics will be used as constraints in the solution process. System noise will be handled through a Bayesian formulation of the deep learning algorithm. The neural net formulation will allow high resolution sampling of the blood velocity and relative pressure fields and accurate computation of velocity-derive hemodynamic parameters using automatic differentiation. The methods will be validated using numerical and in vitro flow experiments using particle image velocimetry. By enabling the estimation of hemodynamic data from what, until now, has been considered to be static data, the proposed research maximizes inference that can be derived from sCTA imaging data without the need for additional computed tomography hardware or new scan protocols.

National Data Platform Pilot: Services for Equitable Open Access to Data.
Ivan Rodero (NSF)

Abstract: The National Data Platform Pilot (NDP) project aims to address the challenges faced by the scientific community in accessing vast data efficiently and equitably. The project intends to create a unified data ecosystem leveraging Cyberinfrastructure (CI) capabilities. It seeks to standardize data processes, fill existing gaps in data CI, and offer data services that promote broad access and collaboration. Furthermore, the NDP emphasizes building an interconnected data hub network to boost scientific understanding, policy formation, and societal impact, with a strong commitment to equity in AI research. The University of Utah's role involves enhancing the current data CI, integrating diverse data sources, democratizing data access, building on the foundation laid by the NSF Virtual Data Collaboratory (VDC) project, and integrating relevant science driver applications. This endeavor aims to make technology more accessible to a wide audience and amplify the value of facilities' data.



Using Behavioral Nudges in Peer Review to Improve Critical Analysis in STEM Courses
Paul Rosen (NSF)

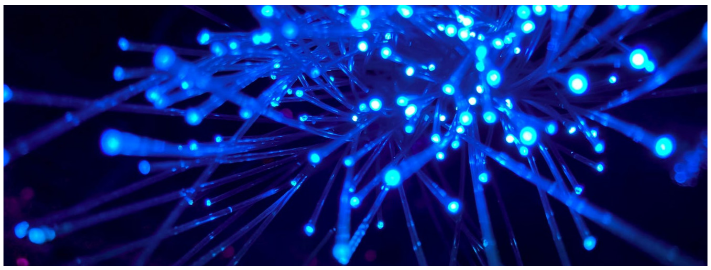
Abstract: This project aims to serve the national interest by increasing the quality of peer reviews given by students. Peer reviews, in which students have the opportunity to analyze and evaluate projects made by their classroom peers, are a widely acknowledged pedagogical method for engaging students and have become a standard practice in undergraduate education. Peer review is most often used in classes with large number of students to provide timely feedback on student assignments. However, peer review has benefits far beyond scalability. Peer review gathers diverse feedback, raises students' comfort level with having their work evaluated in a professional setting, and most importantly, the action of giving a peer review is often more valuable than receiving a peer review. The software platforms in current use that support peer review have seen limited innovation in recent decades, and potential improvements to enhance student outcomes have not been thoroughly evaluated. This project plans to develop and study an innovative peer review system that uses behavioral nudges, a method of subtlety reinforcing positive habits, to improve the evaluation skills of students and the quality of the feedback they provide in peer reviews.

This project proposes to 1) create a software system supporting peer review that includes behavioral nudges for guiding the peer review process to hone students' evaluation and critical analysis skills and 2) study how the use of behavioral nudges improves student engagement. The approach will be evaluated using four different visualization courses that serve approximately 800 students per year. Two groups of students will be studied. The control group will utilize peer review with additional feedback modalities but no nudges, while the test group will use peer review with the same feedback modalities but also include nudges. Common statistical tests such as, ANOVA, t-tests, and correlations, will be used to validate the results. The resulting system could provide a significant and measurable improvement in outcomes in courses that utilize peer review across different STEM disciplines. The resulting technologies will be disseminated as open source to enable widespread adoption. The NSF IUSE: EHR Program supports research and development projects to improve the effectiveness of STEM education for all students. Through the Engaged Student Learning track, the program supports the creation, exploration, and implementation of promising practices and tools.

EAGER: Exploring intelligent services for managing uncertainty under constraints across the Computing Continuum: A case study using the SAGE platform
Daniel Balouek-Thomert (NSF)

Abstract: Advanced cyberinfrastructures seek to make streaming data a modality for the scientific community by offering rich sensing capabilities across the edge-to-Cloud Computing Continuum. For societal applications based on multiple geo-distributed data sources and sophisticated data analytics, such as wildfire detection or personal safety, these platforms can be incredibly transformative. However, it is challenging to match user constraints (response time, quality, energy) with what is feasible in a heterogeneous and dynamic computation environment coupled with uncertainty in the availability of data. Identifying events and accelerating responses critically depends on the capacity to select resources and configure services under uncertain operating conditions. This project develops software abstractions that react at runtime to unforeseen events, and adaptation of the resources and computing paths between the edge and the cloud. This research is driven by the unique capabilities of the SAGE cyberinfrastructure as a case study to explore data-driven reactive behaviors at the network's edge on a national scale.

The project proposes intelligent services to drive computing optimization from urgent scenarios on large cyberinfrastructures. The project focuses on resource management and programming support around three interrelated tasks: (1) developing models for capturing the dynamics of edge resources; (2) researching software abstractions for reactive analytics; and (3) addressing cost/benefit tradeoffs to drive the autonomous reconfiguration of applications and resources under constraints. The outcomes are intended to directly extend the SAGE platform and provide artifacts for creating urgent analytics, which will benefit SAGE users and the scientific community.

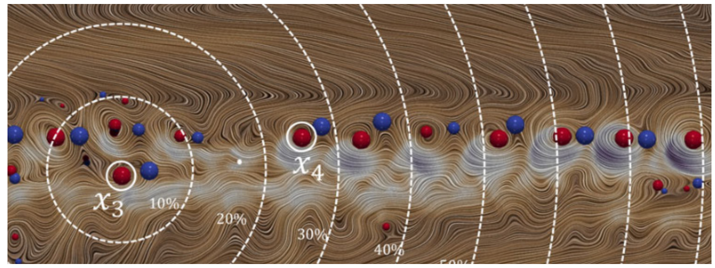


Algorithms, Theory, and Validation of Deep Graph Learning with Limited Supervision: A Continuous Perspective

Bao Wang (NSF)

Abstract: Graph-structured data is ubiquitous in scientific and artificial intelligence applications, for instance, particle physics, computational chemistry, drug discovery, neural science, recommender systems, robotics, social networks, and knowledge graphs. Graph neural networks (GNNs) have achieved tremendous success in a broad class of graph learning tasks, including graph node classification, graph edge prediction, and graph generation. Nevertheless, there are several bottlenecks of GNNs: 1) In contrast to many deep networks such as convolutional neural networks, it has been noticed that increasing the depth of GNNs results in a severe accuracy degradation, which has been interpreted as over-smoothing in the machine learning community. 2) The performance of GNNs relies heavily on a sufficient number of labeled graph nodes; the prediction of GNNs will become significantly less reliable when less labeled data is available. This research aims to address these challenges by developing new mathematical understanding of GNNs and theoretically-principled algorithms for graph deep learning with less training data. The project will train graduate students and postdoctoral associates through involvement in the research. The project will also integrate the research into teaching to advance data science education.

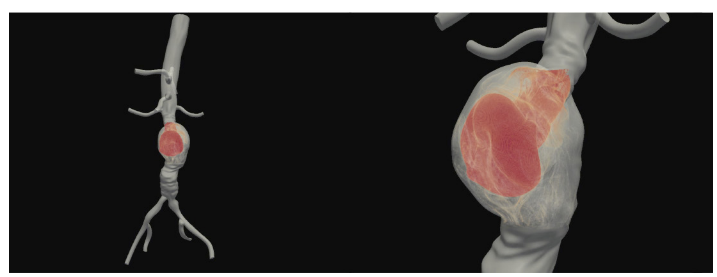
This project aims to develop next-generation continuous-depth GNNs leveraging computational mathematics tools and insights and to advance data-driven scientific simulation using the new GNNs. This project has three interconnected thrusts that revolve around pushing the envelope of theory and practice in graph deep learning with limited supervision using PDE and harmonic analysis tools: 1) developing a new generation of diffusion-based GNNs that are certifiable to learning with deep architectures and less training data; 2) developing a new efficient attention-based approach for learning graph structures from the underlying data accompanied by uncertainty quantification; and 3) application validation in learning-assisted scientific simulation and multi-modal learning and software development.



Topology-Aware Data Compression for Scientific Analysis and Visualization.

Bei Wang Phillips (NSF)

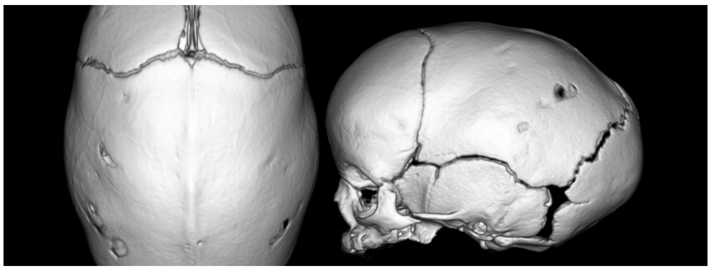
Abstract: Today's large-scale simulations are producing vast amounts of data that are revolutionizing scientific thinking and practices. As the disparity between data generation rates and available I/O bandwidths continues to grow, data storage and movement are becoming significant bottlenecks for extreme-scale scientific simulations in terms of in situ and post hoc analysis and visualization. Such a disparity necessitates data compression, where data produced by simulations are compressed in situ and decompressed in situ and post hoc for analysis and exploration. Meanwhile, topological data analysis plays an important role in extracting insights from scientific data regarding feature definition, extraction, and evaluation. However, most of today's lossy compressors are topology- agnostic, i.e., they do not guarantee the preservation of topological features essential to scientific discoveries. This project aims to research and develop advanced lossy compression techniques and softwares that preserve topological features in data for in situ and post hoc analysis and visualization at extreme scales. The data of interest are scalar fields and vector fields that arise from scientific simulations, with driving applications in cosmology, climate, and fusion simulations. This project has three research thrusts that focus on deriving topological constraints from scalar fields (I) and vector fields (II), and integrating these constraints to develop topology-aware error-controlled and neural compressors (III).



A hybrid computational-experimental framework for targeted embolization in vascular disease

Amir Arzani (NIH)

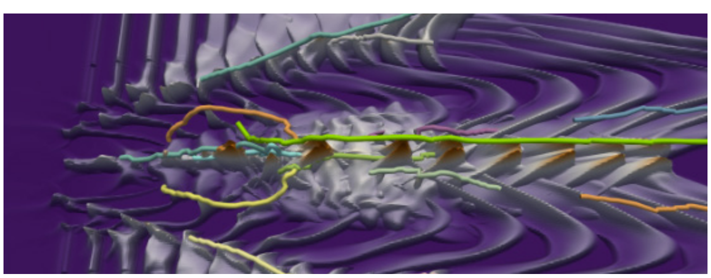
Abstract: Minimally invasive transcatheter embolization is a common nonsurgical procedure in interventional radiology used for the deliberate occlusion of blood vessels for the treatment of diseased or injured vasculature. One of the most commonly used embolic agents for clinical practice is microsphere. No systemic platform has been developed to investigate the correlation between microsphere properties and embolic outcomes. More importantly, clinicians have no technology for estimating the trajectory of emboli, and as such significant uncertainty exists in embolization treatment. In this proposal, we will develop, for the first time, a two-way interactive biomaterial-computational platform that will 1) offer rational design of multifunctional microspheres, 2) accurately guide the transcatheter location for microsphere deployment, and 3) predict microsphere in vivo trajectory and their aggregation in the vasculature to maximize embolic success for personalized therapies. We envision that this innovative technology can be applied to liquid embolic agents, and also be widely disseminated to the treatment of diverse vascular conditions, such as prostate hyperplasia, liver tumors, and fibroids, for translation to patient-specific therapy.



CranioRate: An imaging-based, deep-phenotyping analysis toolset, repository, and online clinician interface for craniosynostosis

Ross Whitaker, Shireen Elhabian (NIH)

Abstract: The purpose of this research grant application is to build on the advanced machine learning (ML) tool developed as part of a pilot study (R21EB026061) that objectively quantifies cranial dysmorphology, or deep phenotypes, in patients with metopic craniosynostosis (MC). Abnormal cranial suture fusion (craniosynostosis) occurs in one of every 2500 infants born in the US, resulting in disrupted regional skull growth and an increased risk of elevated intracranial pressure, neurocognitive impairment and visual disturbances including blindness. Impaired skull growth along the fused suture and subsequent growth compensation in other areas of the skull lead to predictable head shape patterns in patients with craniosynostosis; surgery is recommended early in childhood to restore normal head shape and prevent neurocognitive sequelae.



Multiparameter Topological Data Analysis

Bei Wang Phillips (NSF)

Abstract: Although TDA involving a single parameter has been well researched and developed, the same is not true for the multiparameter case. At its current nascent stage, multiparameter TDA is yet to develop tools to practically handle complex, diverse, and high-dimensional data. To meet this challenge, this project will make both mathematical and algorithmic advances for multiparameter TDA. To scope effectively, focus will be mainly on three research thrusts to: (I) explore multiparameter persistence for generalized features and develop algorithms to compute them; (II) exploit the connections of zigzag persistence to multiparameter settings to support dynamic data analysis, and (III) generalize topological descriptors such as merge trees, Reeb spaces, and mapper. The overarching goal of lacing all three thrust areas remains that of developing actionable and practicable tools in applications including Cytometry, Materials Science, Climate Simulations and Ecology.

Understanding complex wind-driven wildfire propagation patterns with a dynamical systems approach

Amir Arzani (NSF)

Abstract: Intense and long wildfire seasons have unfortunately become a normal routine in certain parts of the US. Dangerous wildfires are often driven by intense winds. The chaotic nature of wind patterns makes prediction and fundamental understanding of wildfire growth a challenging task. In this study, dynamical systems theory will be employed to define coherent structures customized to the transport problems used to model wind-driven wildfire growth. A set of benchmark problems motivated by the field of dynamical systems and chaotic advection together with more complex realistic wind patterns will be leveraged to study the role of coherent structures in wildfire growth. Specifically, the hypothesis that generalized Lagrangian coherent structures could be defined to provide a template for wildfire growth under certain scenarios will be explored. This study will provide a new theory that not only simplifies our understanding of wildfire growth under complex wind patterns but also guides wildfire management and mitigation.

News and Notes

Manish Parashar Receives Achievement Award in High Performance Distributed Computing

Congratulations to Manish Parashar for receiving the 2023 Achievement Award in High Performance Distributed Computing for pioneering strategic and technical contributions, and international leadership in high performance parallel and distributed computing, computational science, and cyberinfrastructure.

Jake Bergquist Receives HRS Fellowship Award

Congratulations to Jake Bergquist who received the Heart Rhythm Society fellowship award. This is a one year, $50,000 fellowship to support research in cardiac electrophysiology.

Manish Parashar Receives HPDC'23 Achievement Award

We are happy to announce that Manish Parashar is the recipient of the 2023 Achievement Award in High Performance Distributed Computing for pioneering contributions in high performance parallel and distributed computational methods, data management, in-situ computing, and international leadership in cyberinfrastructure and translational computer science.

Valerio Pascucci Receives IEEE VGTC Technical Achievement Award

Valerio has pioneered the effective use of topological analysis for visualization and analytics for many different types of data, including scalar fields, vector fields, and high dimensional data with applications ranging from clean energy and climate modeling to cosmology and nuclear engineering. Valerio's topology research started in 1998 with "*Visualization of scalar topology for structural enhancement*" presented at the IEEE VIS conference where he has continued to publish every single year for over 25 years. His body of work has been recognized with seven Best Paper Awards and the 15-year Test of Time Award for his 2002 SciVis paper on "*Efficient computation of the topology of level sets*." Valerio published 6 edited books and over 250 refereed papers with a breadth of work including the theoretical foundations of topological data analysis, its robust implementation in software packages, and its direct use in science and engineering. In addition to topology, Valerio has focused a major portion of his research on the development of techniques to manage and visualize massive scientific data. The results have been implemented in the OpenVisus library and deployed via the National Science Data Fabric initiative.

Penny Atkins Named Associate Director of One Utah Data Science Hub

Penny Atkins, PhD has been named associate director of the One Utah Data Science Hub, which connects the Data Exploration and Learning for Precision Health Intelligence (DELPHI) Initiative, the Data Science and Ethics of Technology (DATASET) Initiative, and the Utah Data Science Center to expand data science research, education, outreach, infrastructure, and datasets at the University of Utah.

Valerio Pascucci Receives 2023 Distinguished Research Award

The Office of the Vice President for Research (VPR) has selected SCI faculty member Valerio Pascucci as a 2022-2023 Distinguished Research Award (DRA) recipient. The DRA is designed to shine a spotlight on the outstanding achievements of University of Utah research faculty.

Manish Parashar Receives the 2023 Sidney Fernbach Memorial Award

Manish's academic career has focused on translational computer science with a specific emphasis on computational and data-enabled science and engineering, and has addressed key conceptual, technological, and educational challenges.

Manish recently completed an IPA appointment at the US National Science Foundation (NSF), serving as Office Director of the NSF Office of Advanced Cyberinfrastructure. At NSF, he oversaw strategy and investments in national cyberinfrastructure and led the development of NSF's strategic vision for a National Cyberinfrastructure Ecosystem and blueprints for key cyberinfrastructure investments. He also served as Co-Chair of the National Science and Technology Council's Subcommittee on the Future Advanced Computing Ecosystem (FACE) and the National Artificial Intelligence Research Resource (NAIRR) Task Force. In 2002, Manish served as Assistant Director for Strategic Computing at the Whitehouse Office of Science and Technology Policy, where he led strategic planning for the Nation's Future Advanced Computing Ecosystem, and the formulation of the National Strategic Computing Reserve (NSCR) concept.

Manish is the founding chair of the IEEE Technical Community on High Performance Computing (TCHPC), and is Fellow of AAAS, ACM, and IEEE.

Three New Paper Awards

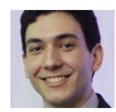
Jadie Adams, Nawazish Khan, Alan Morris, Shireen Y. Elhabian. *Spatiotemporal Cardiac Statistical Shape Modeling: A Data-Driven Approach*. The 13th International Workshop on The Statistical Atlases and Computational Modeling of the Heart (STACOM) - MICCAI, 2022, won the "Best Oral Presentation Award."

Daniel Klötzl, Tim Krake, Youjia Zhou, Ingrid Hotz, Bei Wang, Daniel Weiskopf. *Local Bilinear Computation of Jacobi Sets*. Computer Graphics International (CGI), 2022, won the "Second Best Paper Award."

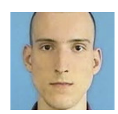
Surojit Saha, Ouk Choi, Ross Whitaker. *Few-Shot Segmentation of Microscopy Images Using Gaussian Process*. "Best Paper" at 1st International Workshop on Medical Optical Imaging and Virtual Microscopy Image Analysis.

Recent SCI Ph.D.s

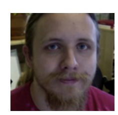
Jen Rogers - Traceability in Design-Oriented Visualization Research



Ricardo Bigolin Lanfredi - Rich visual information exchange for deep chest x-ray analysis: interpretability and gaze supervision



Pavol Klacansky - Scalable Topological Data Analysis by Loose Coupling of Topological Descriptors and Data Structures Through Formal Queries



Jake Bergquist - Advances in Electrocardiographic Imaging: Technical Improvements and Novel Applications

Siddhant Ranade - Inferring Shape and Appearance of 3D Scenes - Advances and Applications

Youjia Zhou - Topology-Based Visualization of Graphs and Hypergraphs

Duong Hoang - Compact and Progressive Encodings for Task-Adaptive Queries of Scientific Data

Haihan Lin - Data Hunches: Expressing Personal Knowledge in Data Visualizations

Scientific Computing and Imaging Institute
72 S. Central Campus Drive, Room 3750
Salt Lake City UT 84112
www.sci.utah.edu