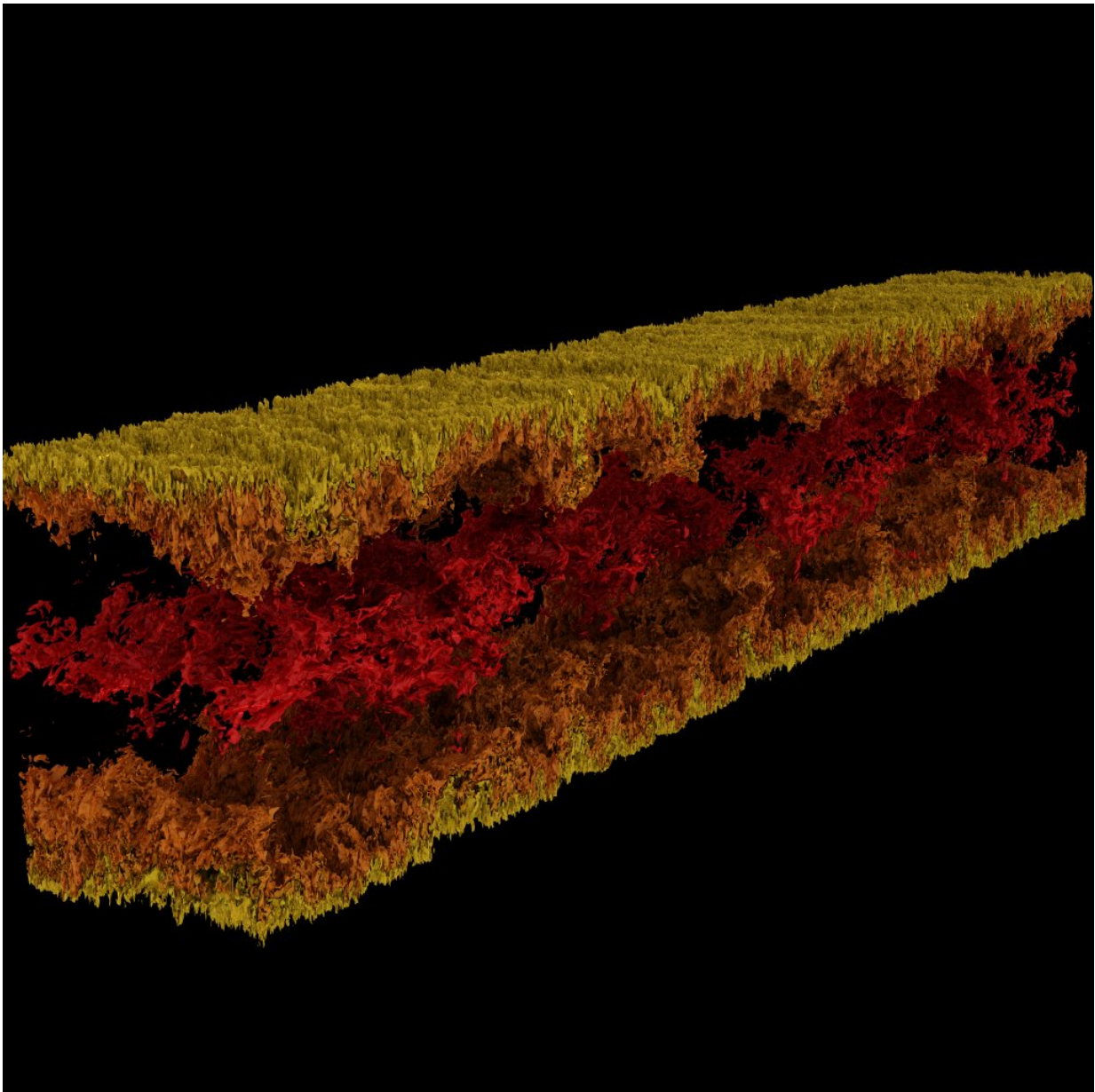


TACC builds seamless software for scientific innovation

April 27 2018, by Aaron Dubrow



Turbulent channel flow visualization produced using GraviT. Credit: Visualization: Texas Advanced Computing Center. Data: ICES, The University of Texas at Austin.

Big, impactful science requires a whole technological ecosystem to progress. This includes cutting-edge computing systems, high-capacity storage, high-speed networks, power, cooling... the list goes on and on.

Critically, it also requires state-of-the-art software: programs that work together seamlessly to allow scientists and engineers to answer tough questions, share their solutions, and conduct research with the maximum efficiency and the minimum pain.

To nurture this critical mode of scientific progress, in 2012 NSF established the Software Infrastructure for Sustained Innovation (SI2) program, with the goal of transforming innovations in research and education into sustained software resources that are an integral part of the cyberinfrastructure.

"Scientific discovery and innovation are advancing along fundamentally new pathways opened by development of increasingly sophisticated software," the National Science Foundation (NSF) wrote in the SI2 program solicitation. "Software is also directly responsible for increased scientific productivity and significant enhancement of researchers' capabilities."

With five current SI2 awards, and collaborative roles on several more, the Texas Advanced Computing Center (TACC) is among the national leaders in developing software for scientific computing. Principal investigators from TACC will present their work from April 30 to May 2 at the [2018 NSF SI2 Principal investigators Meeting](#) in Washington,

D.C.

"Part of TACC's mission is to enhance the productivity of researchers using our systems," said Bill Barth, TACC director of [high performance computing](#) and a past SI2 grant recipient. "The SI2 program has helped us do that by supporting efforts to develop new tools and extending existing tools with additional performance and usability features."

From frameworks for large-scale visualization to automatic parallelization tools and more, TACC-developed software is changing how researchers compute in the future.

Interactive parallelization tool

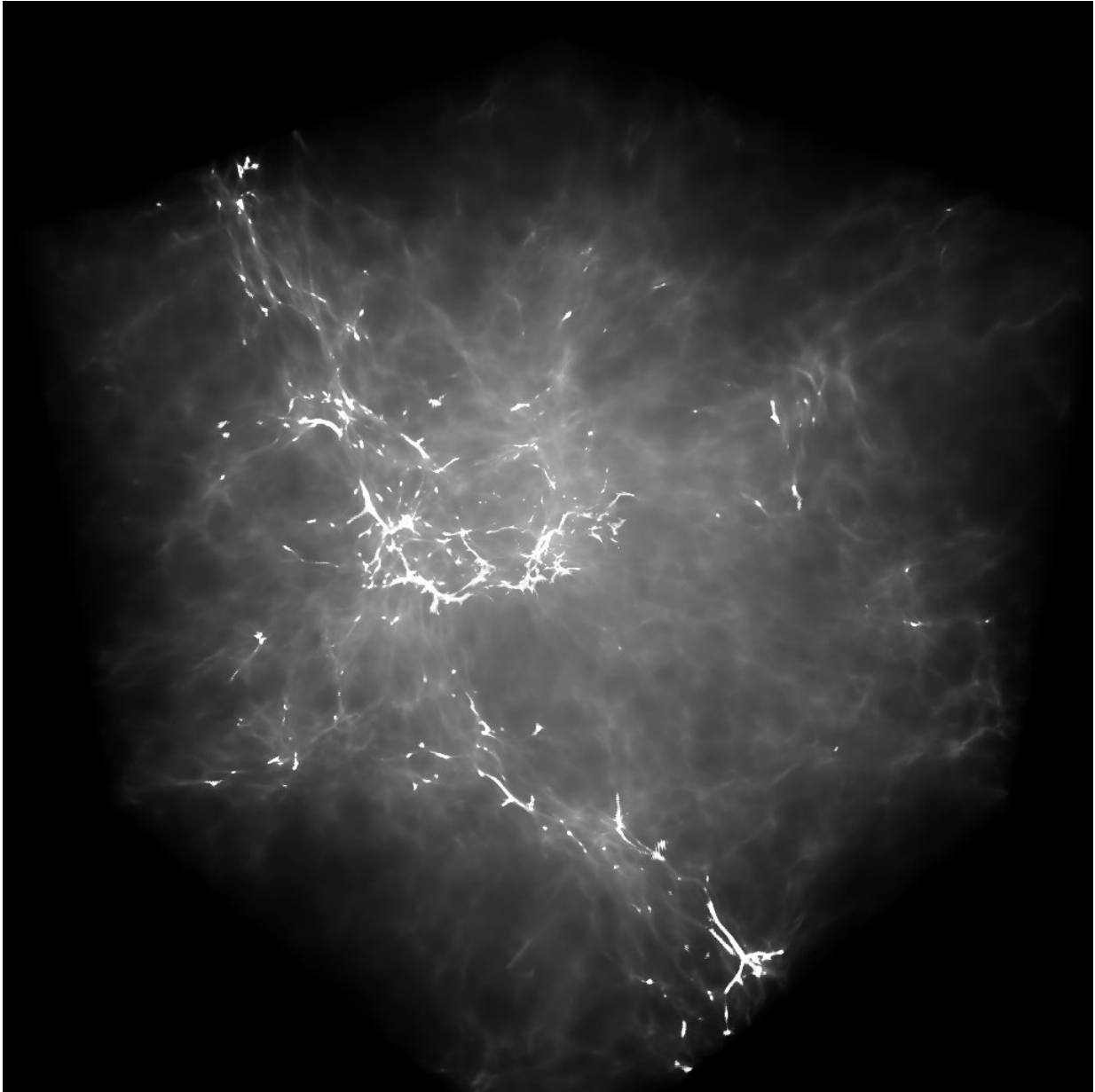
The power of supercomputers lies primarily in their ability to solve mathematical equations in parallel. Take a tough problem, divide it into its constituent parts, solve each part individually and bring the answers together again - this is parallel computing at its essence. However, the task of organizing one's problem so it can be tackled by a supercomputer is not easy, even for experienced computational scientists.

Ritu Arora, a research scientist at TACC, has been working to lower the bar to parallel computing by developing a tool that can turn a serial code, which can only use a single processor at a time, into a parallel code that can use tens to thousands of processors. The tool analyzes a serial application, solicits additional information from the user, applies built-in heuristics, and generates a parallel version of the input serial application.

Arora and her collaborators deployed the current version of [IPT](#) in the cloud so that researchers can conveniently use it through a web-browser. Researchers can generate parallel versions of their code semi-automatically and test the parallel code for accuracy and performance on TACC and XSEDE resources, including Stampede2, Lonestar5, and

Comet.

"The magnitude of the societal impact of IPT is a direct function of the importance of HPC in STEM and emerging non-traditional domains, and the steep challenges that domain experts and students face in climbing the learning curve for parallel programming," Arora said. "Besides reducing the time-to-development and the execution time of the applications on HPC platforms, IPT will decrease the energy usage and maximize the performance delivered by the HPC platforms through its capability to generate hybrid code."



GraviT enabled researchers to produce ray tracing visualizations using data produced from Enzo, a simulation code designed for rich, multi-physics hydrodynamic astrophysical calculations. Credit: University of Texas at Austin

As an example of IPT's capabilities, Arora points to a recent effort to parallelize a molecular dynamics (MD) application. By parallelizing the

serial application using OpenMP at a high-level of abstraction - that is, without the user knowing the low-level syntax of OpenMP—they achieved an 88% speed-up in the code.

They also quantified the impact of IPT in terms of the user-productivity by measuring the number of lines of code that a researcher has to write during the process of parallelizing an application manually versus using IPT.

"In our test cases, IPT enhanced the user productivity by more than 90%, as compared to writing the code manually, and generated the parallel code that is within 10% of the performance of the best available hand-written [parallel code](#) for those applications," said Arora. "We're very happy with its success so far."

TACC is extending IPT to support additional types of serial applications as well as applications that exhibit irregular computation and communication patterns.

(Watch [a video demonstration](#) of IPT in which TACC shows the process of parallelizing a molecular dynamics application with the OpenMP programming model.)

GraviT

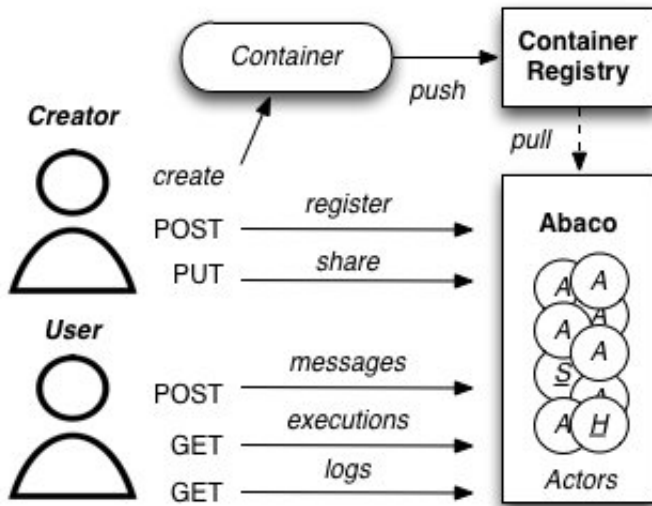
Scientific visualization—the process of transforming raw data into interpretable images—is a key aspect of research. However, it can be challenging when you're trying to visualize petabyte-scale datasets spread among many nodes of a computing cluster. Even more so when you're trying to use advanced visualizations methods like ray tracing—a technique for generating an image by tracing the path of light as pixels in an image plane and simulating the effects of its encounters with virtual objects.

To address this problem, Paul Navratil, director of visualization at TACC, has led an effort to create GraviT, a scalable, distributed-memory ray tracing framework and software library for applications that encompass data so large that it cannot reside in the memory of a single compute node. Collaborators on the project include Hank Childs (University of Oregon), Chuck Hansen (University of Utah), Matt Turk (National Center for Supercomputing Applications) and Allen Malony (ParaTools).

GraviT works across a variety of hardware platforms, including the Intel Xeon processors and NVIDIA GPUs. It can also function in heterogeneous computing environments, for example, hybrid CPU and GPU systems. GraviT has been successfully integrated into the GLuRay OpenGL-based ray tracing interface, the VisIt visualization toolkit, the VTK visualization toolkit, and the yt visualization framework.

"High-fidelity rendering techniques like ray tracing improve visual analysis by providing the same spatial cues of light and shadow that we see in the world around us, but these are challenging to use in distributed contexts," said Navratil. "GraviT enables these techniques to be used efficiently across distributed computing resources, unlocking their potential for large scale analysis and to be used in situ, where data is not written to disk prior to analysis."

(The GraviT source code is available at the [TACC GitHub site](#)).



A diagram showing how the Abaco "Actor" model works. Credit: University of Texas at Austin

Abaco

The increased availability of data has enabled entirely new kinds of analyses to emerge, yielding answers to many important questions. However, these analyses are complex and frequently require advanced computer science expertise to run correctly.

Joe Stubbs, who leads TACC's Cloud and Interactive Computing (CIC) group, is working on a project that simplifies how researchers create analysis tools that are reliable and scalable. The project, known as Abaco, adapts the "Actor" model, whereby software systems are designed as a collection of simple functions, which can then be provided as a cloud-based capability on high performance computing environments.

"Abaco significantly simplifies the way scientific software is developed

and used," said Stubbs. "Scientific software developers will find it much easier to design and implement a system. Further, scientists and researchers that use software will be able to easily compose collections of actors with pre-determined functionality in order to get the computation and data they need."

The [Abaco API](#) (application programming interface) combines technologies and techniques from cloud computing, including Linux Containers and the "functions-as-a-service" paradigm, with the Actor model for concurrent computation. Investigators addressing grand challenge problems in synthetic biology, earthquake engineering and food safety are already using the tool to advance their work. Stubbs is working to extend Abaco's ability to do data federation and discoverability, so Abaco programs can be used to build federated datasets consisting of separate datasets from all over the internet.

"By reducing the barriers to developing and using such services, this project will boost the productivity of scientists and engineers working on the problems of today, and better prepare them to tackle the new problems of tomorrow," Stubbs said.

Expanding volunteer computing

Volunteer computing uses donated computing time on consumer devices such as home computers and smartphones to conduct scientific investigations. Early successes from this approach include the discovery of the structure of an enzyme involved in reproduction of HIV by FoldIt participants; and the detection of pulsars using Einstein@Home.

Volunteer computing can provide greater computing power, at lower cost, than conventional approaches such as organizational computing centers and commercial clouds, but participation in volunteer computing efforts is yet to reach its full potential.

TACC is partnering with the University of California at Berkeley and Purdue University to build new capabilities for [BOINC](#) (the most common software framework used for volunteer computing) to grow this promising mode of distributed computing. The project involves two complementary development efforts. First, it adds BOINC-based volunteer computing conduits to two major high-performance computing providers: TACC and nanoHUB, a web portal for nano science that provides computing capabilities. In this way, the project benefits the thousands of scientists who use these facilities and creates technologies that make it easy for other HPC providers to add their own volunteer computing capability to their systems.

Second, the team will develop a unified interface for volunteer computing, tentatively called Science United, where donors can register to participate and scientists can market their volunteer computing projects to the public.

TACC is currently setting up a BOINC server on Jetstream and using containerization technologies, such as Docker and VirtualBox, to build and package popular applications that can run in high-throughput computing mode on the devices of volunteers. Initial applications being tested include AutoDock Vina, used for drug discovery, and OpenSees, used by the natural hazards community. As a next step, TACC will develop the plumbing required for selecting and routing qualified jobs from TACC resources to the BOINC server.

"By creating a huge pool of low-cost computing power that will benefit thousands of scientists, and increasing public awareness of and interest in science, the project plans to establish volunteer computing as a central and long-term part of the U.S. scientific cyber infrastructure," said David Anderson, the lead principal investigator on the project from UC Berkeley.

Provided by University of Texas at Austin

Citation: TACC builds seamless software for scientific innovation (2018, April 27) retrieved 23 April 2024 from <https://phys.org/news/2018-04-tacc-seamless-software-scientific.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.